

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

LNCS 6363

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010

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

3 Part III



MICCAI

 Springer

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

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 III

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 3-642-15710-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15710-3 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

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, Djamel
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
Farak, 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
Işgum, 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
 Olabariaga, 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
 Penneç, 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, Tomaz
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 III

Segmentation and Modeling

Combining Morphological Information in a Manifold Learning Framework: Application to Neonatal MRI	1
<i>P. Aljabar, R. Wolz, L. Srinivasan, S. Counsell, J.P. Boardman, M. Murgasova, V. Doria, M.A. Rutherford, A.D. Edwards, J.V. Hajnal, and D. Rueckert</i>	
Fast Random Walker with Priors Using Precomputation for Interactive Medical Image Segmentation	9
<i>Shawn Andrews, Ghassan Hamarneh, and Ahmed Saad</i>	
Extraction of the Plane of Minimal Cross-Sectional Area of the Corpus Callosum Using Template-Driven Segmentation	17
<i>Neda Changizi, Ghassan Hamarneh, Omer Ishaq, Aaron Ward, and Roger Tam</i>	
Incorporating Priors on Expert Performance Parameters for Segmentation Validation and Label Fusion: A Maximum a Posteriori STAPLE	25
<i>Olivier Commowick and Simon K. Warfield</i>	
Automated Segmentation of 3-D Spectral OCT Retinal Blood Vessels by Neural Canal Opening False Positive Suppression	33
<i>Zhihong Hu, Meindert Niemeijer, Michael D. Abramoff, Kyungmoo Lee, and Mona K. Garvin</i>	
Detection of Gad-Enhancing Lesions in Multiple Sclerosis Using Conditional Random Fields	41
<i>Zahra Karimaghloo, Mohak Shah, Simon J. Francis, Douglas L. Arnold, D. Louis Collins, and Tal Arbel</i>	
Automated Sulci Identification via Intrinsic Modeling of Cortical Anatomy	49
<i>Yonggang Shi, Bo Sun, Rongjie Lai, Ivo Dinov, and Arthur W. Toga</i>	
In Vivo MRI Assessment of Knee Cartilage in the Medial Meniscal Tear Model of Osteoarthritis in Rats	57
<i>Zhiyong Xie, Serguei Liachenko, Ping-Chun Chiao, Santos Carvajal-Gonzalez, Susan Bove, and Thomas Bocan</i>	

Construction of Neuroanatomical Shape Complex Atlas from 3D Brain MRI	65
<i>Ting Chen, Anand Rangarajan, Stephan J. Eisenschenk, and Baba C. Vemuri</i>	
Non-parametric Iterative Model Constraint Graph Min-Cut for Automatic Kidney Segmentation	73
<i>M. Freiman, A. Kronman, S.J. Esses, L. Joskowicz, and J. Sosna</i>	
Synthetic MRI Signal Standardization: Application to Multi-atlas Analysis	81
<i>Juan Eugenio Iglesias, Ivo Dinov, Jaskaran Singh, Gregory Tong, and Zhuowen Tu</i>	
Multi-organ Segmentation from Multi-phase Abdominal CT via 4D Graphs Using Enhancement, Shape and Location Optimization	89
<i>Marius George Linguraru, John A. Pura, Ananda S. Chowdhury, and Ronald M. Summers</i>	
A Semi-automatic Method for Segmentation of the Carotid Bifurcation and Bifurcation Angle Quantification on Black Blood MRA	97
<i>Hui Tang, Robbert S. van Onkelen, Theo van Walsum, Reinhard Hameeteman, Michiel Schaap, Fufa. L. Tori, Quirijn J.A. van den Bouwhuijsen, Jacqueline C.M. Witteman, Aad van der Lugt, Lucas J. van Vliet, and Wiro J. Niessen</i>	
<i>Standing on the Shoulders of Giants: Improving Medical Image Segmentation via Bias Correction</i>	105
<i>Hongzhi Wang, Sandhitsu Das, John Pluta, Caryne Craige, Murat Altinay, Brian Avants, Michael Weiner, Susanne Mueller, and Paul Yushkevich</i>	
Layout Consistent Segmentation of 3-D Meshes via Conditional Random Fields and Spatial Ordering Constraints	113
<i>Alexander Zouhar, Sajjad Baloch, Yanghai Tsin, Tong Fang, and Siegfried Fuchs</i>	
Cross-Visit Tumor Sub-segmentation and Registration with Outlier Rejection for Dynamic Contrast-Enhanced MRI Time Series Data	121
<i>G.A. Buonaccorsi, C.J. Rose, J.P.B. O'Connor, C. Roberts, Y. Watson, A. Jackson, G.C. Jayson, and G.J.M. Parker</i>	
Nonlocal Patch-Based Label Fusion for Hippocampus Segmentation	129
<i>Pierrick Coupé, José V. Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles, and D. Louis Collins</i>	
Cellular Automata Segmentation of Brain Tumors on Post Contrast MR Images	137
<i>Andac Hamamci, Gozde Unal, Nadir Kucuk, and Kayihan Engin</i>	

Agreement-Based Semi-supervised Learning for Skull Stripping	147
<i>Juan Eugenio Iglesias, Cheng-Yi Liu, Paul Thompson, and Zhuowen Tu</i>	
Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces	155
<i>Liliane Ramus, Olivier Commowick, and Grégoire Malandain</i>	
Automatic Lung Lobe Segmentation Using Particles, Thin Plate Splines, and Maximum a Posteriori Estimation	163
<i>James C. Ross, Raúl San José Estépar, Gordon Kindlmann, Alejandro Díaz, Carl-Fredrik Westin, Edwin K. Silverman, and George R. Washko</i>	
Graph Search with Appearance and Shape Information for 3-D Prostate and Bladder Segmentation	172
<i>Qi Song, Yinxiao Liu, Yunlong Liu, Punam K. Saha, Milan Sonka, and Xiaodong Wu</i>	
Segmentation of Cortical MS Lesions on MRI Using Automated Lamina Profile Shape Analysis	181
<i>Christine L. Tardif, D. Louis Collins, Simon F. Eskildsen, John B. Richardson, and G. Bruce Pike</i>	
3D Knowledge-Based Segmentation Using Pose-Invariant Higher-Order Graphs	189
<i>Chaohui Wang, Olivier Teboul, Fabrice Michel, Salma Essafi, and Nikos Paragios</i>	
Markov Random Field Driven Region-Based Active Contour Model (MaRACel): Application to Medical Image Segmentation	197
<i>Jun Xu, James P. Monaco, and Anant Madabhushi</i>	
Robotics, Motion Modeling and Computer-Assisted Interventions	
Predicting Target Vessel Location for Improved Planning of Robot-Assisted CABG Procedures	205
<i>Daniel S. Cho, Cristian A. Linte, Elvis Chen, Chris Wedlake, John Moore, John Barron, Rajni Patel, and Terry M. Peters</i>	
Quantification of Prostate Deformation due to Needle Insertion during TRUS-guided Biopsy	213
<i>Tharindu De Silva, Aaron Fenster, Jagath Samarabandu, and Aaron D. Ward</i>	
Optimized Anisotropic Rotational Invariant Diffusion Scheme on Cone-Beam CT	221
<i>Dirk-Jan Kroon, Cornelis H. Slump, and Thomas J.J. Maal</i>	

Control of Articulated Snake Robot under Dynamic Active Constraints	229
<i>Ka-Wai Kwok, Valentina Vitiello, and Guang-Zhong Yang</i>	
Estimating Radiation Exposure in Interventional Environments	237
<i>Alexander Ladikos, Cedric Cagniard, Reza Ghotbi, Maximilian Reiser, and Nassir Navab</i>	
Force Adaptive Multi-spectral Imaging with an Articulated Robotic Endoscope	245
<i>David P. Noonan, Christopher J. Payne, Jianzhong Shang, Vincent Sauvage, Richard Newton, Daniel Elson, Ara Darzi, and Guang-Zhong Yang</i>	
Motion Tracking in Narrow Spaces: A Structured Light Approach	253
<i>Oline Vinter Olesen, Rasmus R. Paulsen, Liselotte Højgaard, Bjarne Roed, and Rasmus Larsen</i>	
Tracking of Irregular Graphical Structures for Tissue Deformation Recovery in Minimally Invasive Surgery	261
<i>Marco Visentini-Scarzanella, Robert Merrifield, Danail Stoyanov, and Guang-Zhong Yang</i>	
Graph Based Interactive Detection of Curve Structures in 2D Fluoroscopy	269
<i>Peng Wang, Wei-shing Liao, Terrence Chen, Shaohua K. Zhou, and Dorin Comaniciu</i>	
Automated Digital Dental Articulation	278
<i>James J. Xia, Yu-Bing Chang, Jaime Gateno, Zixiang Xiong, and Xiaobo Zhou</i>	
Image-Based Respiratory Motion Compensation for Fluoroscopic Coronary Roadmapping	287
<i>Ying Zhu, Yanghai Tsing, Hari Sundar, and Frank Sauer</i>	
Surgical Task and Skill Classification from Eye Tracking and Tool Motion in Minimally Invasive Surgery	295
<i>Narges Ahmadi, Gregory D. Hager, Lisa Ishii, Gabor Fichtinger, Gary L. Gallia, and Masaru Ishii</i>	
Micro-force Sensing in Robot Assisted Membrane Peeling for Vitreoretinal Surgery	303
<i>Marcin Balicki, Ali Uneri, Iulian Iordachita, James Handa, Peter Gehlbach, and Russell Taylor</i>	
C-arm Pose Estimation in Prostate Brachytherapy by Registration to Ultrasound	311
<i>Pascal Fallavollita, Clif Burdette, Danny Song, Purang Abolmaesumi, and Gabor Fichtinger</i>	

Cognitive Burden Estimation for Visuomotor Learning with fNIRS <i>David R.C. James, Felipe Orihuela-Espina, Daniel R. Leff, George P. Mylonas, Ka-Wai Kwok, Ara W. Darzi, and Guang-Zhong Yang</i>	319
Prediction Framework for Statistical Respiratory Motion Modeling <i>Tobias Klinder, Cristian Lorenz, and Jörn Ostermann</i>	327
Image Estimation from Marker Locations for Dose Calculation in Prostate Radiation Therapy <i>Huai-Ping Lee, Mark Foskey, Josh Levy, Rohit Saboo, and Ed Chaney</i>	335
A Machine Learning Approach for Deformable Guide-Wire Tracking in Fluoroscopic Sequences <i>Olivier Pauly, Hauke Heibel, and Nassir Navab</i>	343
Collaborative Tracking for MRI-Guided Robotic Intervention on the Beating Heart <i>Y. Zhou, E. Yeniaras, P. Tsiamyrtzis, N. Tsekos, and I. Pavlidis</i>	351
Calibration and Use of Intraoperative Cone-Beam Computed Tomography: An In-Vitro Study for Wrist Fracture <i>Erin Janine Smith, Anton Oentoro, Hisham Al-Sanawi, Braden Gammon, Paul St. John, David R. Pichora, and Randy E. Ellis</i>	359
A Strain Energy Filter for 3D Vessel Enhancement <i>Changyan Xiao, Marius Staring, Denis Shamonin, Johan H.C. Reiber, Jan Stolk, and Berend C. Stoel</i>	367
Virtual Stent Grafting in Personalized Surgical Planning for Treatment of Aortic Aneurysms Using Image-Based Computational Fluid Dynamics <i>Guanglei Xiong and Charles A. Taylor</i>	375
MRI-Guided Robotic Prostate Biopsy: A Clinical Accuracy Validation <i>Helen Xu, Andras Lasso, Siddharth Vikal, Peter Guion, Azel Krieger, Aradhana Kaushal, Louis L. Whitcomb, and Gabor Fichtinger</i>	383
Online 4-D CT Estimation for Patient-Specific Respiratory Motion Based on Real-Time Breathing Signals <i>Tiancheng He, Zhong Xue, Weixin Xie, and Stephen T.C. Wong</i>	392
Modeling and Segmentation of Surgical Workflow from Laparoscopic Video <i>Tobias Blum, Hubertus Feußner, and Nassir Navab</i>	400

Fused Video and Ultrasound Images for Minimally Invasive Partial Nephrectomy: A Phantom Study	408
<i>Carling L. Cheung, Chris Wedlake, John Moore, Stephen E. Pautler, and Terry M. Peters</i>	
Probabilistic 4D Blood Flow Mapping	416
<i>Ola Friman, Anja Hennemuth, Andreas Harloff, Jelena Bock, Michael Markl, and Heinz-Otto Peitgen</i>	
Rotational Encoding of C-arm Fluoroscope with Tilt Sensing Accelerometer	424
<i>Victor Grzeda and Gabor Fichtinger</i>	
Robotic Hand-Held Surgical Device: Evaluation of End-Effector's Kinematics and Development of Proof-of-Concept Prototypes	432
<i>Ali Hassan Zahraee, Jérôme Szewczyk, Jamie Kyujin Paik, and Guillaume Morel</i>	
Guide-Wire Extraction through Perceptual Organization of Local Segments in Fluoroscopic Images	440
<i>Nicolas Honnorat, Régis Vaillant, and Nikos Paragios</i>	
Single-Projection Based Volumetric Image Reconstruction and 3D Tumor Localization in Real Time for Lung Cancer Radiotherapy	449
<i>Ruijiang Li, Xun Jia, John H. Lewis, Xuejun Gu, Michael Folkerts, Chunhua Men, and Steve B. Jiang</i>	
A Method for Planning Safe Trajectories in Image-Guided Keyhole Neurosurgery	457
<i>Reuben R. Shamir, Idit Tamir, Elad Dabool, Leo Joskowicz, and Yigal Shoshan</i>	
Adaptive Multispectral Illumination for Retinal Microsurgery	465
<i>Raphael Sznitman, Diego Rother, Jim Handa, Peter Gehlbach, Gregory D. Hager, and Russell Taylor</i>	
Motion Artifact Correction of Multi-Photon Imaging of Awake Mice Models Using Speed Embedded HMM	473
<i>Taoyi Chen, Zhong Xue, Changhong Wang, Zhenshen Qu, Kelvin K. Wong, and Stephen T.C. Wong</i>	
Image Reconstruction, Enhancement and Representation	
Diagnostic Radiograph Based 3D Bone Reconstruction Framework: Application to Osteotomy Surgical Planning	481
<i>Pavan Gamage, Sheng Quan Xie, Patrice Delmas, and Wei Liang Xu</i>	

Comparative Analysis of Quasi-Conformal Deformations in Shape Space	489
<i>Vahid Taimouri, Huiquan He, and Jing Hua</i>	
Establishing Spatial Correspondence between the Inner Colon Surfaces from Prone and Supine CT Colonography	497
<i>Holger Roth, Jamie McClelland, Marc Modat, Darren Boone, Mingxing Hu, Sebastien Ourselin, Greg Slabaugh, Steve Halligan, and David Hawkes</i>	
Heat Kernel Smoothing Using Laplace-Beltrami Eigenfunctions	505
<i>Seongho Seo, Moo K. Chung, and Hourri K. Vorperian</i>	
Under-Determined Non-cartesian MR Reconstruction with Non-convex Sparsity Promoting Analysis Prior	513
<i>Angshul Majumdar and Rabab K. Ward</i>	
A Statistical Approach for Achievable Dose Querying in IMRT Planning	521
<i>Patricio Simari, Binbin Wu, Robert Jacques, Alex King, Todd McNutt, Russell Taylor, and Michael Kazhdan</i>	
Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures	529
<i>Nikhil Singh, P. Thomas Fletcher, J. Samuel Preston, Linh Ha, Richard King, J. Stephen Marron, Michael Wiener, and Sarang Joshi</i>	
Shape Analysis of Vestibular Systems in Adolescent Idiopathic Scoliosis Using Geodesic Spectra	538
<i>Wei Zeng, Lok Ming Lui, Lin Shi, Defeng Wang, Winnie C.W. Chu, Jack C.Y. Cheng, Jing Hua, Shing-Tung Yau, and Xianfeng Gu</i>	
Value-Based Noise Reduction for Low-Dose Dual-Energy Computed Tomography	547
<i>Michael Balda, Björn Heismann, and Joachim Hornegger</i>	
Automatic Detection of Anatomical Features on 3D Ear Impressions for Canonical Representation	555
<i>Sajjad Baloch, Rupen Melkisetoglu, Simon Flöry, Sergei Azernikov, Greg Slabaugh, Alexander Zouhar, and Tong Fang</i>	
Probabilistic Multi-Shape Representation Using an Isometric Log-Ratio Mapping	563
<i>Neda Changizi and Ghassan Hamarneh</i>	
Efficient Robust Reconstruction of Dynamic PET Activity Maps with Radioisotope Decay Constraints	571
<i>Fei Gao, Huafeng Liu, and Pengcheng Shi</i>	

Nonlinear Embedding towards Articulated Spine Shape Inference Using Higher-Order MRFs	579
<i>Samuel Kadoury and Nikos Paragios</i>	
Improved Method for Point-Based Tracking	587
<i>Andrei Danilchenko, Andrew D. Wiles, Ramya Balachandran, and J. Michael Fitzpatrick</i>	
Computer Aided Diagnosis	
A Texton-Based Approach for the Classification of Lung Parenchyma in CT Images	595
<i>Mehrdad J. Gangeh, Lauge Sørensen, Saher B. Shaker, Mohamed S. Kamel, Marleen de Bruijne, and Marco Loog</i>	
Active Learning for an Efficient Training Strategy of Computer-Aided Diagnosis Systems: Application to Diabetic Retinopathy Screening	603
<i>C.I. Sánchez, M. Niemeijer, M.D. Abràmoff, and B. van Ginneken</i>	
Sparse Bayesian Learning for Identifying Imaging Biomarkers in AD Prediction	611
<i>Li Shen, Yuan Qi, Sungeun Kim, Kwangsik Nho, Jing Wan, Shannon L. Risacher, Andrew J. Saykin, and ADNI</i>	
Computer-Aided Detection of Pulmonary Pathology in Pediatric Chest Radiographs	619
<i>André Mouton, Richard D. Pitcher, and Tania S. Douglas</i>	
Toward Precise Pulmonary Nodule Descriptors for Nodule Type Classification	626
<i>Amal Farag, Shireen Elhabian, James Graham, Aly Farag, and Robert Falk</i>	
Morphology-Guided Graph Search for Untangling Objects: <i>C. elegans</i> Analysis	634
<i>T. Riklin Raviv, V. Ljosa, A.L. Conery, F.M. Ausubel, A.E. Carpenter, P. Golland, and C. Wählby</i>	
Automatic Cephalometric Evaluation of Patients Suffering from Sleep-Disordered Breathing	642
<i>Lior Wolf, Tamir Yedidya, Roy Ganor, Michael Chertok, Ariela Nachmani, and Yehuda Finkelstein</i>	
Fusion of Local and Global Detection Systems to Detect Tuberculosis in Chest Radiographs	650
<i>Laurens Hogeweg, Christian Mol, Pim A. de Jong, Rodney Dawson, Helen Ayles, and Bram van Ginneken</i>	

Novel Morphometric Based Classification via Diffeomorphic Based Shape Representation Using Manifold Learning	658
<i>Rachel Sparks and Anant Madabhushi</i>	
Semi Supervised Multi Kernel (SeSMiK) Graph Embedding: Identifying Aggressive Prostate Cancer via Magnetic Resonance Imaging and Spectroscopy	666
<i>Pallavi Tiwari, John Kurhanewicz, Mark Rosen, and Anant Madabhushi</i>	
Author Index	675

Combining Morphological Information in a Manifold Learning Framework: Application to Neonatal MRI

P. Aljabar^{1,*}, R. Wolz¹, L. Srinivasan², S. Counsell³, J.P. Boardman³,
M. Murgasova¹, V. Doria², M.A. Rutherford³, A.D. Edwards²,
J.V. Hajnal³, and D. Rueckert¹

¹ Department of Computing, Imperial College London, UK
Paul.Aljabar@Imperial.ac.uk

² Division of Neonatology, Imperial College Healthcare NHS Trust, London, UK

³ Institute of Clinical Sciences, Imperial College London and MRC Clinical Sciences
Centre, Hammersmith Hospital, London, UK

Abstract. MR image data can provide many features or measures although any single measure is unlikely to comprehensively characterize the underlying morphology. We present a framework in which multiple measures are used in manifold learning steps to generate coordinate embeddings which are then combined to give an improved single representation of the population. An application to neonatal brain MRI data shows that the use of shape and appearance measures in particular leads to biologically plausible and consistent representations correlating well with clinical data. Orthogonality among the correlations suggests the embedding components relate to comparatively independent morphological features. The rapid changes that occur in brain shape and in MR image appearance during neonatal brain development justify the use of shape measures (obtained from a deformation metric) and appearance measures (obtained from image similarity). The benefit of combining separate embeddings is demonstrated by improved correlations with clinical data and we illustrate the potential of the proposed framework in characterizing trajectories of brain development.

1 Introduction

A number of machine learning techniques have been developed for converting data from a high to a low dimensional representation more suitable for further processing steps such as clustering or regression. This class of methods, described under the terms ‘manifold learning’ or ‘dimensionality reduction’, have recently begun to be applied to the field of medical image analysis. Medical images, or derived features, are natural candidates as raw data for manifold learning where a typical structural magnetic resonance (MR) image, for example, can contain around a million voxels.

The intuition underlying the application of manifold learning to medical images is that, while each image may be viewed as a single point in a very high-dimensional space, a set of such points for a population of images may be well

* Corresponding author.

represented by sub-manifold of the space that is likely to be non-linear and of a significantly lower dimension. Although it may be possible to learn a manifold directly from the image data, it is more typical to use measures that relate pairs of images in a dataset.

As an example, pairwise similarity (based on model description length) was used in [1] in an application of Laplacian eigenmaps [2] to identify coherent sets of landmarks in images that can be clustered. The ‘shape maps’ produced have clear applications in medical imaging data where landmark based analysis is common. The Laplacian eigenmap approach was also applied in [3] where similarities between images were derived from structural segmentation overlaps and the resulting embedding was used to discriminate clinical groups of elderly patients with Alzheimer’s Disease (AD) and controls. Another manifold learning algorithm, Isomap [4], was used to estimate the manifold structure of brain MRI acquired from AD patients and controls [5]. In that study, distance measures between image pairs are derived from the non-rigid deformations aligning them. Pairwise distances derived from the deformations were also used to empirically construct a manifold for images in [6] using a k-NN graph building approach. By ‘navigating’ the resulting sparse graph, in which edges join pairs of similar images represented by the nodes, it becomes possible to estimate transformations between pairs of very different images by concatenating successive transformations between pairs of similar images. Navigation of a low-dimensional coordinate embedding was also used in [7] where it was shown that accurate atlas-based segmentations may be obtained by propagating manually labelled structures from a small group of healthy brain images to a large set of images acquired from subjects with a range of pathology.

In this work, we contribute a framework in which separate manifold learning steps, based on different pairwise measures, can be applied to MR images. The resulting coordinate embeddings are then fused to produce a single combined embedding. The use of multiple pairwise measures allows the decomposition of morphology into different aspects. We give a motivating example using neonatal brain MR images which show both shape changes associated with growth and complementary signal changes due to microscopic structural processes such as myelination. The two pairwise measures used were therefore chosen to represent shape and appearance in the data. We compare the embeddings with non-imaging data for the subjects and the separate shape- and appearance-based embeddings correlate in a consistent way with these clinical data. The results also indicate the benefit of fusing embeddings as this leads to improved correlation and provides a useful representation for characterizing trajectories of change during the important period of neonatal brain development.

2 Methods

Our approach begins with registrations to obtain measures reflecting shape and appearance relations for image pairs in the data set. Each measure produces a coordinate embedding via a manifold learning technique appropriate for the measure. The separate embeddings are then fused by scaling and concatenation

and, after a final dimensionality reduction step, a combined embedding characterizes the manifold in which the image data reside. A schematic illustration of the proposed approach (as applied to neonatal MRI) is provided in Figure [1](#)

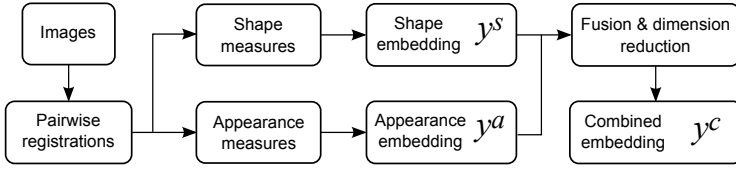


Fig. 1. An overview of the proposed framework as it is applied to neonatal brain MRI

2.1 Pairwise Measures between Images

Given images $\mathcal{I}_A, \mathcal{I}_B : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ and a non-rigid deformation $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ between them, we derive an appearance measure for the pair from their intensity-based similarity and a shape measure from ϕ . These measures are chosen for application to neonatal data but other measures possible in different applications.

The intensity similarity of the images is given by a function on the pairings $\{(\mathcal{I}_A(x), \mathcal{I}_B(\phi(x))) : x \in \Omega\}$ and we define the pairwise appearance measure to be the normalised cross correlation

$$\mathcal{M}^a(\mathcal{I}_A, \mathcal{I}_B; \phi) = \frac{\sum(\mathcal{I}_A(x) - \mu_A)(\mathcal{I}_B(\phi(x)) - \mu_B)}{\sqrt{\sum(\mathcal{I}_A(x) - \mu_A)^2 \sum(\mathcal{I}_B(\phi(x)) - \mu_B)^2}}$$

where μ_A and μ_B are the average intensities.

The pairwise shape measure is based on a metric induced by ϕ . In a diffeomorphic setting, such as the one given in [8](#), $\phi(x)$ is defined as the point at $t = 1$ of an integral curve $\phi(x, t)$ of a time-varying velocity field $v(x, t) : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$. The metric induced on an image pair is then defined as the minimal value of the energy integral $\int_0^1 \|v(x, t)\|_L^2 dt$ such that ϕ aligns the images and L denotes a choice of differential operator.

Local structure is important when learning the low-dimensional manifold and distances between images in local neighbourhoods play a greater role in characterizing manifold geometry than longer range distances. In our context, this means that transformations between similar images carry more weight. In this case, transformations can be approximated in a small deformation setting: $\phi(x) = x + u(x)$ for a displacement field u . We follow the work of [5](#) and [6](#) in adopting this approximation and define a pairwise shape measure between images as

$$\mathcal{M}^s(\mathcal{I}_A, \mathcal{I}_B; \phi) = \|u(x)\|_L^2$$

where $u(x)$ is obtained from registering \mathcal{I}_A and \mathcal{I}_B and $L = Id + \alpha \nabla$, with Id and ∇ denoting the identity matrix and divergence operator and α is a weighting parameter.

Both the shape and appearance measures depend on the resolution of the transformation that aligns the image pair, i.e. the scale at which local structure

in the images is aligned. A coarse transformation only accounts for large-scale shape differences between the images; small-scale variations remain to influence the appearance measure, \mathcal{M}^a , for the images. We found that differences in appearance are more diffuse under a fine transformation and the way image pairs relate is encoded more in the deformation $u(x)$ between them. In this case, the shape measure becomes more informative. These considerations are similar to those arising in voxel-based morphometry (VBM) studies [9] which seek local group differences in aligned image sets. The resolution of the aligning transformations must allow such differences to remain detectable. In contrast, when images are finely aligned, group differences become encoded in the aligning transformations and deformation-based morphometry (DBM) becomes possible [10].

2.2 Manifold Learning and Fusion of Embeddings

The pairwise measures for shape (\mathcal{M}^s) and appearance (\mathcal{M}^a) are used as input data for separate manifold learning steps, with each giving a coordinate embedding for the image data. Since \mathcal{M}^a represents a similarity measure and \mathcal{M}^s a distance measure, manifold learning techniques specific to each type of measure are applied: A Laplacian eigenmap [2] is used for \mathcal{M}^a and Isomap [4] is used for \mathcal{M}^s . In each case, the image data can be viewed as nodes in a graph where edges are assigned weights derived from \mathcal{M}^a or \mathcal{M}^s .

For N images $\mathcal{I}_1, \dots, \mathcal{I}_N$, the Laplacian eigenmap approach assigns a weight W_{ij} to each edge measuring the similarity of images \mathcal{I}_i and \mathcal{I}_j . We set $W_{ij} = \mathcal{M}^a(\mathcal{I}_i, \mathcal{I}_j)$. The method then seeks a coordinate embedding $y_i, i = 1, \dots, N$ for the data that minimises the cost function:

$$\sum_{i,j} W_{ij} \|y_i - y_j\|^2$$

where $\|\cdot\|$ is the L_2 norm. The local structure of the data can be modelled by selecting a neighbourhood size K such that W_{ij} is set to zero when \mathcal{I}_j is outside the K -nearest neighbourhood of \mathcal{I}_i .

The Isomap algorithm assigns a distance \mathcal{D}_{ij} to edges reflecting the distance between end nodes. We set $\mathcal{D}_{ij} = \mathcal{M}^s(\mathcal{I}_i, \mathcal{I}_j)$. Local structure is modelled by constructing a sparse graph in which edges are restricted to be those of the K -nearest neighbourhood for each node. The manifold structure is then obtained using geodesic distances D_{ij} between all pairs of nodes estimated by summing the values of $\mathcal{D}_{\alpha\beta}$ along the shortest path between each node pair in the graph. A coordinate embedding $y_i, i = 1, \dots, N$ is then generated to minimise the cost function

$$\sum_{i,j} (D_{ij} - \|y_i - y_j\|)^2$$

Isomap and Laplacian eigenmaps are examples of spectral methods since they produce embedding coordinates from eigendecompositions of matrices derived from the edge weights of the graph. Although the use of a single algorithm is possible by converting between distance and similarity measures (for example

with heat kernel functions [2]), we elected to avoid this and apply manifold learning techniques specifically developed for each type of measure. This also means that additional parameter choices during conversion are avoided. We distinguish shape embedding coordinates y_i^s from the appearance embedding y_i^a . In each case a target dimension, d^α , $\alpha \in \{a, s\}$, needs to be chosen and the coordinates may be explicitly written as $y_i^\alpha = (y_{i,1}^\alpha, \dots, y_{i,d^\alpha}^\alpha)$.

The embedding coordinates y_i^s and y_i^a were then fused by uniformly scaling each set so that the variance of the first coordinate, $y_{i,1}^\alpha$, $\alpha \in \{a, s\}$, becomes one and subsequently concatenating. This provides $(d^a + d^s)$ -dimensional coordinates $(f_a y_i^a, f_s y_i^s)$ for the appropriate scale factors f_a and f_s . Overlap in the descriptions of the data provided by the two input embeddings was then removed by applying a further dimensionality reduction step to the concatenated coordinates. This was achieved by finding all pairwise L_2 distances for the concatenated coordinates. A final Isomap step applied to the distances provides a combined coordinate embedding y_i^c with dimension d^c .

3 Data and Results

3.1 Image Data and Pre-processing

The images studied were 140 T2W scans of neonatal subjects. Fast-spin echo scans were acquired on a 3 Tesla Philips Intera scanner (TR = 1712 ms, TE = 160 ms, FA = 90°) with resolution $0.86 \times 0.86 \times 1$ mm. Mean gestational age (GA) at birth was 29.4 ± 2.9 weeks (range 23.4–34.9) and mean age at scan was 36.8 ± 4.8 weeks (range 28.6–47.7). A clinical expert selected a reference scan displaying an intermediate level of maturation for the group (GA at birth/scan 32.0/36.3 weeks). All scans were aligned to the reference using affine transformations and resliced to remove global size and orientation differences. Each image pair was non-rigidly registered using free-form deformations [11] (FFDs). The resolution of the transformations was varied using a coarse-to-fine optimisation giving three FFDs, with control point spacings of 20, 10 and 5mm, for each pair of images.

3.2 Results

The target dimension, d^α , $\alpha \in \{a, s, c\}$, for each embedding was empirically selected by varying it and calculating the correlation between the input edge

	$y_{i,1}^c$	$y_{i,1}^s$	$y_{i,1}^a$	$y_{i,2}^c$	$y_{i,2}^s$	$y_{i,2}^a$
GA (S)	0.93*	0.88*	0.92*	0.01	0.07	0.03
W (S)	0.91*	0.85*	0.89*	0.01	0.08	0.04
HC (S)	0.92*	0.88*	0.88*	0.10	0.18	0.07
GA (B)	0.18	0.21	0.07	0.68*	0.61*	0.49*
W (B)	0.24	0.24	0.17	0.48*	0.45*	0.35†
HC (B)	0.24	0.26	0.16	0.54*	0.56*	0.35†

Table 1. Correlation of clinical data and embedding components for each method. Gestational age (GA), weight (W) and head circumference (HC) data were obtained at scan (S) or at birth (B). [* $p < 10^{-4}$, † $p < 10^{-3}$]

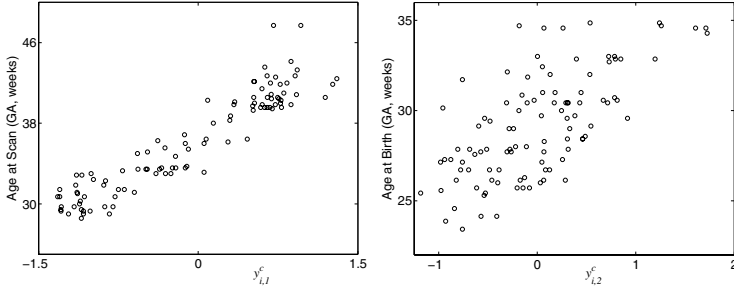


Fig. 2. Scatter graphs of combined embedding components with different age data. Left: $y_{i,1}^c$ against gestational age (GA) at scan. Right: $y_{i,2}^c$ against GA at birth.

weights and the embedding distances. For the Laplacian eigenmap method, the best correlation was obtained for $d^a = 5$ and reduced thereafter. Using Isomap, the correlation increases asymptotically as the dimension increases and the minimal dimension for which the correlation exceeded 0.7 was selected to avoid over-fitting. This led to choices of $d^s = 4$ and $d^c = 5$. For the Isomap steps, the residual variance was 0.30 for y_i^s and a much lower value of 0.05 when deriving y_i^c , reflecting already near-linear structure of the input data. The best control point spacing for the transformations was empirically chosen as above and matched expectations regarding the transformation resolutions for each type of measure with an intermediate spacing of 10mm used to derive y_i^a and a finer spacing of 5mm for y_i^s .

The explanatory power of each set of embedding coordinates was assessed by its correlations with clinical data for the group which were recorded at birth and at scan. These were gestational age, head circumference and weight. The first component in each embedding, $y_{i,1}^s$ and $y_{i,1}^a$, correlates strongly with the measures obtained at scan (see Table 1) and there are weaker but still significant correlations for the second components $y_{i,2}^s$ and $y_{i,2}^a$ with data obtained at birth.

Viewing the entries in Table 1 as a 2×2 block matrix with 3×3 elements, we note that the off-diagonal values are much lower than those on the diagonal which suggests orthogonality among the correlations. Within each diagonal block, the correlations for the combined embedding are, with one exception, greater than the corresponding correlations for the separate embeddings demonstrating the benefit of combining them.

Scatter graphs for the combined embedding components against ages at scan and at birth are shown in Figure 2. Visualisations are given in Figure 3 for the manifold structure of the image data provided by the first two components of the combined embedding. A continuous trajectory $y^c(t)$ through the combined embedding coordinates, parametrised by the age at scan t , was defined using kernel regression as $y^c(t) = \frac{1}{\nu} \sum_i K(t_i - t) y_i^c$ where t_i is the age at scan for image \mathcal{I}_i , K represents a Gaussian kernel and ν is a normalising constant. The trajectory (projected onto two dimensions) for our data is shown in Figure 3 as a solid line. Figure 4 illustrates reconstructions of images at five equally spaced

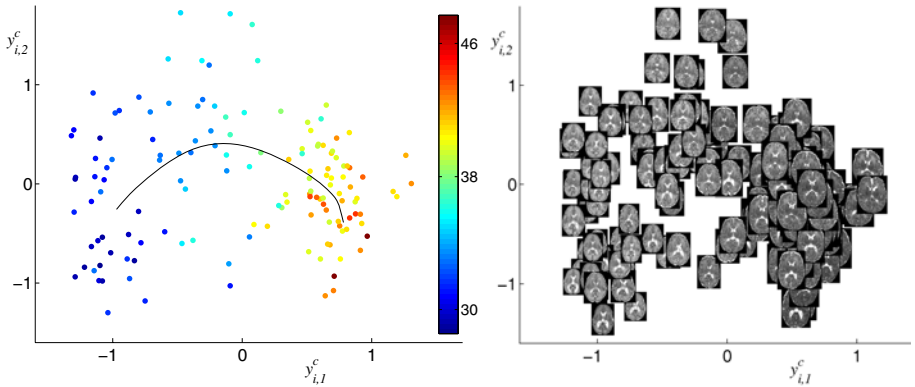


Fig. 3. Left: Scatter graph of combined embedding components $y_{i,1}^c$ and $y_{i,2}^c$ with colour coding for age at scan. The solid line shows a regressed trajectory for age at scan (see text). Right: The same scatter graph illustrated with image data.

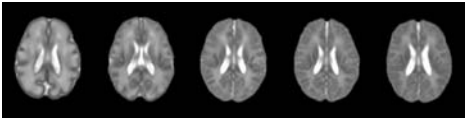


Fig. 4. Reconstructions of images along the trajectory for age at scan in Figure 3 (N.B. Original image data were normalised for global size).

points on the trajectory estimated using the method described in [5]. It should be noted that the reconstructions retain the global size normalisation of the pre-processed image data.

4 Discussion

We have presented a framework for applying manifold learning steps to imaging data using different measures of morphology. We have utilised shape and appearance measures although other applications may require different measures specific to their clinical context. Shape and appearance are important aspects of neonatal brain development and we derive their measures from image similarities and deformations. The measures provide coordinate embeddings after applying different manifold learning techniques appropriate to each type of measure used. Correlations with clinical data suggest overlap in the descriptions that the separate embeddings provide despite their separate derivations. This level of consistency across embeddings and the strong correlations with clinical data indicate that the manifold learning steps provide biologically relevant representations. The proposed method also combines separately obtained embeddings through scaling, concatenation and dimensionality reduction. While the separate embeddings may overlap in the information they provide, the improved correlations after combination also indicates some independence in the descriptions.

Considering the patterns of correlation between embeddings and clinical data, strong correlations between the first feature of all embeddings and age at scan is

perhaps unsurprising as it is the dominant factor affecting brain shape and MR appearance. It is also expected that correlations with weight and head circumference should parallel correlations with age (because size measures are themselves correlated with age) but this at least provides further proof of principle for the results. The second components, however, correlate with gestational age at *birth* which suggests that, after scan age, variation in brain morphology is mainly determined by age at birth. A possible explanation for this is that gestational age at birth measures the degree of prematurity for an infant which in turn may affect morphology. The apparent orthogonality among the observed correlations (Table I) suggests an independence to the morphological features described by the first and second components in each embedding.

Finally we have shown in a preliminary example how trajectories may be described in the embedding coordinates using kernel regression and, following the work in [5], how image reconstructions of such a trajectory may be obtained. The trajectory parameter was age at scan but clearly other clinical covariates can be used. This ability to use the proposed approach to characterize change in cohorts, with respect to chosen covariates, has clear clinical potential, for example in identifying pathology or in tracking growth.

References

1. Langs, G., Paragios, N.: Modeling the structure of multivariate manifolds: Shape maps. In: Proc. CVPR 2008. IEEE, Los Alamitos (2008)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* 15(6), 1373–1396 (2003)
3. Aljabar, P., Rueckert, D., Crum, W.: Automated morphological analysis of magnetic resonance brain imaging using spectral analysis. *NeuroImage* 43(2), 225–235 (2008)
4. Tenenbaum, J., de Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500), 2319–2323 (2000)
5. Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R.: On the manifold structure of the space of brain images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 305–312. Springer, Heidelberg (2009)
6. 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)
7. Wolz, R., Aljabar, P., Hajnal, J., Hammers, A., Rueckert, D.: The Alzheimer’s Disease Neuroimaging Initiative: LEAP: Learning embeddings for atlas propagation. *NeuroImage* 49(2), 1316–1325 (2010)
8. Beg, M., Miller, M., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61(2), 139–157 (2005)
9. Ashburner, J., Friston, K.: Voxel-based morphometry – the methods. *NeuroImage* 11(6), 805–821 (2000)
10. Gaser, C., Nenadoc, I., Buchsbaum, B., Hazlett, E., Buchsbaum, M.: Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI. *NeuroImage* 13, 1140–1145 (2001)
11. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Non-rigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Medical Imaging* 18(8), 712–721 (1999)

Fast Random Walker with Priors Using Precomputation for Interactive Medical Image Segmentation

Shawn Andrews, Ghassan Hamarneh, and Ahmed Saad

Medical Image Analysis Lab, Simon Fraser University, Canada
{sda56, hamarneh, aasaad}@sfu.ca

Abstract. Updating segmentation results in real-time based on repeated user input is a reliable way to guarantee accuracy, paramount in medical imaging applications, while making efficient use of an expert's time. The random walker algorithm with priors is a robust method able to find a globally optimal probabilistic segmentation with an intuitive method for user input. However, like many other segmentation algorithms, it can be too slow for real-time user interaction. We propose a speedup to this popular algorithm based on offline precomputation, taking advantage of the time images are stored on servers prior to an analysis session. Our results demonstrate the benefits of our approach. For example, the segmentations found by the original random walker and by our new precomputation method for a given 3D image have a Dice's similarity coefficient of 0.975, yet our method runs in $1/25^{th}$ of the time.

1 Introduction

Segmentation is a crucial task in medical imaging. Manual segmentation by an expert is accurate, but is also very time consuming, while fully automatic and accurate segmentation techniques are not yet a reality, thus semi-automatic techniques become a necessity. While many semi-automatic techniques assume only user initialization, repeated user interaction is necessary to guarantee the accuracy required for medical imaging. Therefore, it is critical to speed up these techniques, especially in 3D, in order to minimize the time spent waiting between a user inputting information and seeing the results [1,2].

A full survey of semi-automatic algorithms is beyond the scope of this work [3,4]. At a high level, semi-automatic algorithms can be divided into several classes. One class involves the specification of an approximate boundary, which evolves towards the correct segmentation by minimizing a cost function derived from shape priors and image information [5,6]. Another class of algorithms requires the user to specify sequential points on or near the boundary, and then the boundary is filled in between these points using a minimal path approach [7,8]. A third class of algorithms asks the user to provide seeds, or pixels within specific regions, and then uses these seeds as a basis for the segmentation [9,10].

An example of the last class of algorithms is the seeded random walker (RW_SD) [10], which is a graph-based approach to image segmentation that, along with its extensions, has garnered hundreds of citations in only a few years. It boasts many advantages, including weak boundary detection, robustness to noise, trivial generalization

to simultaneous multi-region and 3D segmentations, a globally optimal solution ensuring repeatability, a probabilistic segmentation that can be very useful in directing a user to areas of uncertainty, and the straightforward user input method of providing seed pixels - all important features in medical imaging. On the downside, once seeds are given, RW_SD computes the segmentation by solving a large system of equations, which can be slow. In [11], this problem is alleviated by introducing RW with precomputation (RW_PREC). Since medical images usually exist “offline” on servers for some time before they are segmented, some precomputation can be done before user input that allows a fast approximation to the segmentation once seeds are given, or “online”. The speedup RW_PREC provides allows for user interaction with RW_SD in real-time.

Unfortunately, RW_SD has some limitations, specifically the segmentation is calculated based only on localized image data and disconnected regions must be seeded individually. These problems are addressed in [12], where regional intensity priors are introduced into the formulation. The priors result in more accurate segmentations and the ability to segment disjoint regions easily. However, in RW with priors (RW_PR), the image graph is not completely known offline, since priors are usually derived from the seeds and precomputation must be performed before seeds are given. This obsoletes the methods introduced in [11], which require the graph to be known. An algorithm with the robustness of RW_PR and the online speedup of RW_PREC would be a very useful interactive segmentation tool.

In this paper, we make the following contributions. First, starting with the random walker equations from [11], we derive an offline precomputation and an online approximation that allows for a significant online speedup that can be used in conjunction with priors. Secondly, we derive some additional precomputations that are performed offline to further speed up the online segmentation. Combining the robustness of RW_PR and the online speedup of RW_PREC, we create a useful interactive segmentation tool applicable to a more general class of problems than RW_PREC. Code demonstrating our method is available from <http://mial.cs.sfu.ca>.

2 Methods: Random Walker Improvements

We begin by giving a brief review of existing RW algorithms for later reference. In the following derivations, we consider binary segmentations, but we note that our methods extend trivially to multiple labels, just as all previous RW algorithms discussed do. RW_SD [10] constructs a graph and then defines L as the graph’s Laplacian matrix and x as a vector of the probabilities that each node belongs to the object being segmented. We define $N, S, U = N - S \gg S$ as the numbers of nodes in the graph, seeded nodes, and unseeded nodes, respectively. In RW_PR [11], λ is introduced as a vector of the prior probabilities for each node, weighted by a scalar γ . In RW_PREC [11], the first K eigenvalue/eigenvector pairs of L are found and stored in matrices A and Q respectively, so $L \approx QAQ^T$. These are used to construct a pseudo-inverse for L , $E = QA^{-1}Q^T$. For all three algorithms, all of the variables involved are expressed in terms of their components corresponding to seeded (S) and unseeded (U) nodes:

$$L = \begin{bmatrix} L_S & B \\ B^T & L_U \end{bmatrix}, \quad x = \begin{bmatrix} x_S \\ x_U \end{bmatrix}, \quad E = \begin{bmatrix} E_S & R \\ R^T & E_U \end{bmatrix}, \quad Lx = f = \begin{bmatrix} f_S \\ f_U \end{bmatrix}, \quad g = \begin{bmatrix} g_S \\ g_U \end{bmatrix}, \quad (1)$$

where g is a constant eigenvector corresponding to the zero eigenvalue of L .

In RW_PR, x_U , the vector of RW probabilities for the unlabeled pixels, is obtained by solving a $U \times U$ system of equations (with I defined as the identity matrix):

$$(L_U + \gamma I_U)x_U = -B^T x_S + \gamma \lambda, \quad (2)$$

where $\gamma = 0$ gives RW_SD. In RW_PREC, several equations of size $S \times S$ are given whose solutions can be used to approximate x_U in the case that $\gamma = 0$. This method [11] is reviewed in [12]. When priors are added to the RW formulation ($\gamma > 0$), the graph is no longer known offline, and this precomputation method is rendered useless.

2.1 Precomputation with Priors

Following [11], our goal is to derive a method for using offline precomputation to speed up the online computation for the more general RW_PR, where the image graph is altered after seeds are given. We define L and x as in RW_PR [11]. Following preliminary steps similar to RW_PREC, we derive

$$(I_U + \gamma E_U)x_U - g_U \alpha = R^T f_S + \gamma E_U \lambda. \quad (3)$$

The details of the derivation are available in [12]. To proceed in our derivation, we will assume for the moment that we can calculate $J_U^{-1} = (I_U + \gamma E_U)^{-1}$ and define $\hat{B} = B J_U^{-1}$. Now, replacing B with \hat{B} in the derivation and taking $f_S = \hat{f}_S + \bar{f}_S \alpha$ and $\hat{P} = (I_S - \hat{B} R^T)$, we derive

$$\hat{P} \hat{f}_S = L_S x_S + \gamma \hat{B} E_U \lambda, \quad \hat{P} \bar{f}_S = \hat{B} g_U, \quad \alpha = \frac{g_S^T \hat{f}_S + \gamma g_U^T \lambda}{\gamma - g_S^T \bar{f}_S}. \quad (4)$$

The details are again in [12]. Now that we have f_S we plug it back into (3) to get

$$x_U = J_U^{-1} (g_U \alpha + R^T f_S + \gamma E_U \lambda). \quad (5)$$

The issue still exists of how to compute J_U^{-1} efficiently, as defining J_U requires the seeds and inverting a matrix of size U is too expensive of an operation to perform during the online phase. We will define J to be the extension of J_U to size N given by

$$J = (I + \gamma E) \approx (Q Q^T + \gamma Q \Lambda^{-1} Q^T) \approx Q (I_K + \gamma \Lambda^{-1}) Q^T \quad (6)$$

$$\Rightarrow J^{-1} \approx Q (I_K + \gamma \Lambda^{-1})^{-1} Q^T \quad (7)$$

since $Q Q^T \approx I_N$, which implies that $Q_U Q_U^T \approx I_U$. Thus

$$Q_U (I_K + \gamma \Lambda^{-1})^{-1} Q_U^T J_U \approx I_U \quad \Rightarrow \quad J_U^{-1} \approx Q_U (I_K + \gamma \Lambda^{-1})^{-1} Q_U^T. \quad (8)$$

Since $(I_K + \gamma \Lambda^{-1})$ is a diagonal matrix with all positive entries, it is easily invertible, and can be inverted during the offline phase. The simple multiplication of Q_U to both sides during the online phase produces an adequate approximation to J_U^{-1} , fulfilling our goal of a RW formulation combining priors and precomputation (RW_PR_PREC).

2.2 Increased Speed Using Extended Precomputation

By using the proposed RW_PR_PREC, we can reduce the online solving of (2) from RW_PR, an equation of size $O(U)$, to solving (4), two equations of size $O(S)$. However, due to the low connectivity of the graph, L_U is very sparse, and (2) can be solved in $O(U)$ time. The online phase of RW_PR_PREC must take $O(U)$ time also, since it returns a probability vector of size U . We want to perform as few $O(U)$ cost computations as possible by minimizing the number of matrix multiplications between matrices of size $O(U)$. Analysis and optimizations of these asymptotic run times were not considered previously in [1], which we do here by analysis of matrix operations.

Currently, in the offline phase, we compute Q and Λ^{-1} , but now we will precompute additional matrices to be used to speed up the online phase. The speedup will come from being able to retrieve the components of these matrices corresponding to the seeded nodes in $O(S)$ time. We note that Q is an $N \times K$ matrix and for storage space considerations we do not want our precomputed matrices to be larger than that. The details of the additional precomputations are in Algorithm 1, denoted OPT_RW_PR_PREC. By precomputing these 4 matrices, we save $(SK + K + K^2)U$ scalar multiplications, with the detailed calculations in [12]. We note that when not using priors, with RW_PREC, A_1 can still be used for additional speedup.

Algorithm 1. OPT_RW_PR_PREC:

Offline:

- 1: Calculate Q and Λ^{-1} from L
- 2: $A_1 = LQ$
- 3: $A_2 = Q^T Q$
- 4: $A_3 = (I_K + \gamma \Lambda^{-1})^{-1}$
- 5: $A_4 = Q^T g$

Online:

- 6: $\hat{P} = (I_S - (A_{1S} - L_S Q_S) A_3 (A_2 - Q_S^T Q_S) \Lambda^{-1} Q_S^T)$
 - 7: $\hat{P} \hat{f}_S = L_S x_S + \gamma (A_{1S} - L_S Q_S) A_3 (A_2 - Q_S^T Q_S) \Lambda^{-1} (Q_U^T \lambda)$
 - 8: $\hat{P} \bar{f}_S = (A_{1S} - L_S Q_S) A_3 (A_4 - Q_S^T g_S)$
 - 9: $\alpha = \frac{g_S^T \hat{f}_S + \gamma g_U^T \lambda}{\gamma - g_S^T \bar{f}_S}$
 - 10: $f_S = \hat{f}_S + \bar{f}_S \alpha$
 - 11: $x_U = Q_U A_3 ((A_4 - Q_S^T g_S) \alpha + (A_2 - Q_S^T Q_S) \Lambda^{-1} (Q_S^T f_S + \gamma (Q_U^T \lambda)))$
-

3 Results

We would like our results to show that using our precomputed data we make the RW online phase fast enough for interactive segmentation without compromising much accuracy. Therefore, we present results showing our high speed gains on real 2D and 3D data while maintaining negligible (and controlled) reduction in accuracy. We note that the speed increase allows much quicker seed editing and thus will translate to much improved accuracy per time spent by user.

The experiments here were performed using unoptimized MATLAB code run on an Intel Core 2 Duo (2.4GHz) with 4GB of RAM. The algorithms were implemented by the authors, utilizing Grady’s MATLAB Graph Toolbox [<http://www.cns.bu.edu/~lgrady>]. A negative exponential function was used for the edge weights, $w_{ij} = \exp(-\beta(|i_a - i_b|))$, where i_a is the intensity of pixel a . All experiment data was collected over 100 trials, and all parameters were chosen empirically and fixed across all compared methods. The only parameter effecting the speed of our method is K , the number of retained eigenvectors.

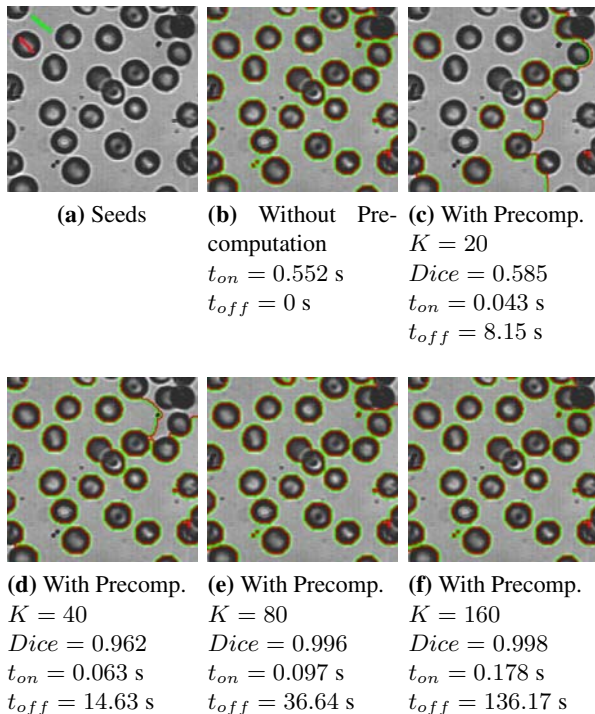


Fig. 1. (color figure) Comparison of results with and without precomputation for segmentation using priors on an image of size $N \approx 72,000$ pixels. For K , the number of eigenvectors used, we report $Dice$, the Dice similarity coefficient between RW_PR’s segmentation and OPT_RW_PR.PREC’s segmentation, t_{on} , the online time taken, and t_{off} , the offline time taken. We note that we are only concerned with t_{on} , and with $K = 80$, our method achieves excellent results in less than a fifth of the time taken when not using precomputation. Red and green correspond to different region boundaries.

The accuracy of the segmentations generated by our algorithms are evaluated by their similarity to the segmentations generated by RW_PR; the accuracy of RW_PR is well justified in other works [11]. We note that the speed and accuracy of our algorithms depend on the image only through K , and while we leave analytical methods for finding optimal K as future work, it was reported in [1] that $K = 40$ – 80 is often enough for 2D images, as our results in Section 3.1 corroborate, and we needed no more than $K = 350$ for larger 3D images as is seen in Section 3.2.

The resolution and noisiness of an image affect how large of a K is needed, and different noiseless images at the same resolution can require different values of K to be accurately segmented using precomputation. As seen in Fig. 1, even larger values of K provide online runtimes much faster than can be achieved without precomputation, so choosing K large enough to guarantee accuracy is our prime concern. As different

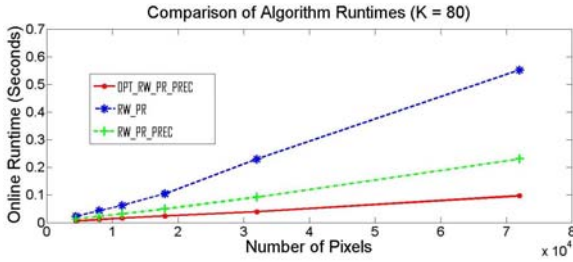


Fig. 2. Comparison of the runtimes of the original RW_PR (blue) and our proposed methods RW_PR_PREC (green) and OPT_RW_PR_PREC (red) for different resolutions of the image in Fig. 1. Note the standard deviation for the online runtimes are all under 0.05 seconds.

images affect the speed and accuracy of our algorithms only through how large K needs to be, and since we do not yet have an image dependent way to choose K (except based on resolution and noise), results for a variety of images would be redundant. Thus we focus our results on single 2D and 3D images at varying resolutions and with varying levels of noise. We note that offline runtime increases with K , but does not affect the application to interactive segmentation.

3.1 2D Results

Tests were performed on the 2D image in Fig. 1 of size $N = 265 \times 272 \approx 72,000$ pixels with an 8-connected image graph, $\beta = 30$, $\gamma = 0.001$, and two regions, where one region was divided into multiple disconnected sections and seeds were only put in one of these sections. These segmentation times do not include calculating the priors, an efficient step which was performed online, and is similar in all cases. The priors were calculated using a non-parametric density estimation with a Gaussian kernel [11]. Fig. 1 shows the Dice similarity coefficient and the average runtimes in seconds for both the online and offline phases of OPT_RW_PR_PREC for different values of K and compares the results to RW_PR, showing excellent speedup and minimal accuracy lost. Fig. 2 compares the runtimes of the different methods for different sized resolutions of the image in Fig. 1, again showing our precomputation gives excellent speedup.

3.2 3D Results

Tests were performed on a 3D CT image of the knee in Fig. 3 of size $N = 55 \times 55 \times 36 \approx 109,000$ voxels, a 26-connected image graph, and two regions, bone and non-bone. The bone region consists of 3 disconnected subregions (the femur, tibia, and patella). We tested the algorithms using priors by segmenting all the bones but placing seeds only in the tibia. We used RW_PR and OPT_RW_PR_PREC with $\beta = 100$ and $\gamma = 0.01$ and compared their average runtimes and the Dice similarity coefficient of their resulting segmentations. The average runtime of RW_PR was about 40.5 seconds, and when $K = 350$ eigenvectors are used, the average runtime of OPT_RW_PR_PREC was about 1.56 seconds. The Dice similarity coefficient between RW_PR's segmentation and OPT_RW_PR_PREC's segmentation was 0.975. Thus our method achieved a speedup of 25 times over RW_PR while maintaining excellent accuracy. The standard deviation of the runtimes of RW_PR was less than 1.0 s, and the standard deviation of the runtimes of OPT_RW_PR_PREC was less than 0.1 s.

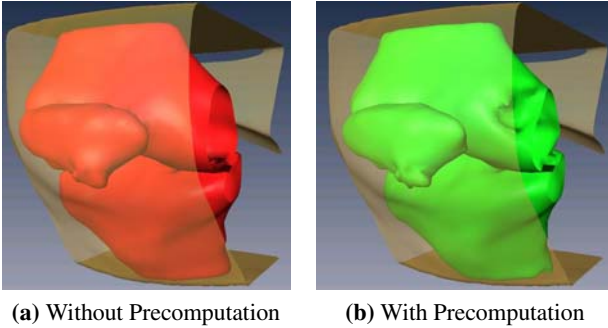


Fig. 3. The bones of a knee segmented with $S = 100$ seeds in one of the bones. (a) was found using RW_PR in 40.5 seconds and (b) was found using OPT_RW_PR_PREC with $K = 350$ in 1.56 seconds. Dice’s similarity coefficient between the two is 0.975.

3.3 Robustness to Noise

Here, we test the robustness to noise of OPT_RW_PR_PREC. We measured the similarity of the segmentations provided by the exact and approximate algorithms using the Dice similarity coefficient. The pixel intensities in our test images range from 0 to 1 and various levels of Gaussian noise with standard deviations $\sigma \in [0, 1]$ were added to the 2D image in Fig. 1 of size $N = 265 \times 272 \approx 72,000$ pixels with an 8-connected image graph, $\beta = 30$, and $\gamma = 0.001$. From Fig. 4a we see that OPT_RW_PR_PREC still provides good segmentations for small amounts of noise up to $\sigma = 0.2$ (with Dice’s similarity coefficient > 0.95) if a large enough K is used. As the noise increases to $\sigma = 0.7$, Dice decreases. We can see the same trend in Fig. 4b, where $K = 200$ eigenvectors are used in the precomputation and the noise ranges from $\sigma = 0$ to 1.

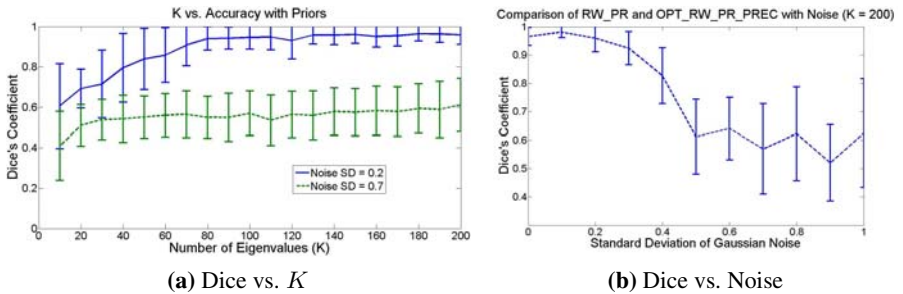


Fig. 4. Effect of K and noise on segmentation accuracy. (a) compares the Dice similarity coefficient between the segmentations found using RW_PR and OPT_RW_PR_PREC. Results are shown for two levels of noise and for multiple numbers of eigenvectors. (b) shows the Dice similarity coefficient between the segmentations at varying levels of noise with $K = 200$ eigenvectors and 20 trials for each level of noise. We see that large enough K lets us account for reasonable amounts of noise.

4 Discussion and Conclusions

The above tests give some strong results. We see from Figs. 1 and 2 that the additional precomputation of OPT_RW_PR_PREC greatly outperforms RW_PR in 2D, achieving

a segmentation in about one fifth of the time and with over 99% similarity. From Fig. 3 we see the results are more pronounced in 3D, with OPT_RW_PR_PREC achieving speedups of a factor of 25 over RW_PR while still finding an almost identical segmentation with over 97% similarity. Furthermore, Fig. 2 shows that all the algorithms appear to increase linearly in runtime with the number of pixels, as predicted in Section 2.2.

Overall, we have derived a way of combining both precomputation and the use of priors into the popular RW algorithm. This allows RW to perform much faster segmentations when seeds and priors are either given or changed. Additionally, we've shown that some precomputations can be performed in addition to finding Q and Λ^{-1} that can greatly speed up the online phase of the algorithm. These improvements in speed provide a feasible way to enable the real-time editing of a wider variety of 2D and 3D images than was previously possible by allowing updating of both seeds and priors. This allows the user to ensure the accuracy of complex segmentations with minimal effort. Thus our contributions increase the usability and effectiveness of RW algorithms.

Future work will relate to using information from the image to automatically determine K , which needs to be set high enough to maintain accuracy. However, since the effect of a larger K is seen mostly in the offline phase which doesn't effect interactivity, we currently simply err on the side of caution when selecting K .

References

1. Grady, L., Kemal Sinop, L.: Fast approximation random walker segmentation using eigenvector precomputation. *IEEE Trans. PAMI* (2008)
2. Armstrong, C.J., Price, B.L., Barrett, W.A.: Interactive segmentation of image volumes with live surface. *Computers & Graphics* 31(2), 212–229 (2007)
3. Olabarriaga, S.D., Smeulders, A.W.M.: Interaction in the segmentation of medical images: A survey. *Medical Image Analysis* 5(2), 127–142 (2001)
4. Kang, Y., Engelke, K., Kalender, W.A.: Interactive 3D editing tools for image segmentation. *Medical Image Analysis* 8(1), 35–46 (2004)
5. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *Int'l. J. on Computer Vision* 1(4), 321–331 (1987)
6. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. on Image Processing* 10(2), 266–277 (2001)
7. Cohen, L., Kimmel, R.: Global minimum for active contour models: A minimal path approach. *International Journal of Computer Vision* 24, 57–78 (1997)
8. Barrett, W., Mortensen, E.: Interactive live-wire boundary extraction. *Medical Image Analysis* 1, 331–341 (1997)
9. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-D images. In: *Proc. ICCV*, pp. 105–112 (2001)
10. Grady, L.: Random walks for image segmentation. *IEEE TPAMI* 28(11), 1768–1783 (2006)
11. Grady, L.: Multilabel random walker image segmentation using prior models. In: *IEEE Conf. CVPR*, vol. 1, pp. 763–770 (June 2005)
12. Andrews, S., Hamarneh, G., Saad, A.: Fast random walker with priors using precomputation. Technical Report TR 2010-07, Simon Fraser University (June 2010)

Extraction of the Plane of Minimal Cross-Sectional Area of the Corpus Callosum Using Template-Driven Segmentation

Neda Changizi¹, Ghassan Hamarneh¹, Omer Ishaq^{1,2},
Aaron Ward^{1,3}, and Roger Tam⁴

¹ Medical Image Analysis Lab, Simon Fraser University, Canada

² Department of Computer Sciences, Air University, Pakistan

³ Robarts Research Institute, The University of Western Ontario, Canada

⁴ MS/MRI Research Group, University of British Columbia, Canada

Abstract. Changes in corpus callosum (CC) size are typically quantified in clinical studies by measuring the CC cross-sectional area on a midsagittal plane. We propose an alternative measurement plane based on the role of the CC as a bottleneck structure in determining the rate of interhemispheric neural transmission. We designate this plane as the Minimum Corpus Callosum Area Plane (MCCAP), which captures the cross section of the CC that best represents an upper bound on interhemispheric transmission. Our MCCAP extraction method uses a nested optimization framework, segmenting the CC as it appears on each candidate plane, using registration-based segmentation. We demonstrate the robust convergence and high accuracy of our method for magnetic resonance images and present preliminary clinical results showing higher sensitivity to disease-induced atrophy.

1 Introduction

The corpus callosum (CC) is an anatomic structure that acts as a communication bridge connecting the two brain hemispheres [1]. Certain neurological diseases are known to affect the shape and size of the CC. In particular, there have been numerous studies correlating CC measurements to multiple sclerosis [2], schizophrenia [3], autism [4], and many other mental and physical ailments. The accurate measurement of CC area changes is dependent on the repeatable identification of the cross-sectional plane of interest in all studied images. In previous studies, changes in CC size have been quantified by measuring its sagittal cross-sectional area on a midsagittal plane (MSP).

MSP identification approaches can be classified as either symmetry- or feature-based. Symmetry-based approaches assume bilaterally symmetric hemispheres, with the MSP chosen to maximize this symmetry. Published approaches suggest different symmetry criteria; e.g., intensity ratios [5], cross-correlation [6], or edge-based [7]. In feature-based approaches, the MSP is defined as the plane best matching the cerebral interhemispheric fissure. The Hough transform [8],

orthogonal regression of B-spline snakes [9], and minimization of white matter area (including the cerebellum) [10], are examples of the feature-based approach.

It has been shown that even small errors in the MSP selection confound the interpretation of actual CC area changes due to pathology [11,12]. Therefore, it may be clinically more meaningful to define the plane for CC atrophy measurement according to the characteristics of the CC itself rather than more global attributes such as brain symmetry or the interhemispheric fissure.

It is known that the cross-sectional area of the CC is proportional to the number of nerve fibers passing through it [1], and correlates with neural transmission [13,14,15,16]. Therefore, in this paper, we propose adopting the minimum CC cross-sectional area as the primary factor in defining the *bottleneck* (upper bound) of interhemispheric communication and hypothesize that the minimum CC cross-sectional area is a potentially more representative measure of CC degeneration as compared to areas of other cross sections.

We emphasize that we are proposing a new plane for the measurement of CC atrophy, and not a new criterion for MSP extraction. We refer to this plane as the Minimum CC Area Plane (MCCAP). We note that multiple planes in a given brain may have the same minimum CC area. However, since we postulate that all of these planes restrict the neural transmission similarly, the identification of any one of these planes results in satisfying our objectives for measuring the CC’s structural health. We also note that the rate of CC degeneration may not be uniform for the whole CC and hence it is possible that the parameters defining the MCCAP may vary over time for the same patient.

2 Method

2.1 Overview and Definitions

We extract the MCCAP from a brain magnetic resonance imaging (MRI) volume by first defining an objective function that measures the cross-sectional area A_{cc} of the CC in a given extracted plane. Then, using an iterative optimization procedure, A_{cc} is minimized with respect to the set of plane parameters P_{ext} (Fig. 1). Using an ASL coordinate system, the parameters of P_{ext} are the angles R_x and R_y (in degrees) between the plane normal vector and the grid axes in the anterior and superior directions, respectively, and the distance T_z (mm) of the plane from the origin. The CC is segmented on every candidate plane by deformably registering a pre-segmented 2D template image from the same MRI volume to the extracted slice on the candidate plane. An alternative approach that one might suggest is to pre-segment the whole CC-bridge in 3D and find the plane with minimum CC area within the segmented volume. However, by using registration-based segmentation we avoid the complications of a full 3D segmentation, which is difficult to validate, even for anatomical experts.

2.2 Template Preparation

To create the template for the deformable registration step (Sec. 2.3), we extract the central sagittal slice from the MRI volume and provide an expert-segmented

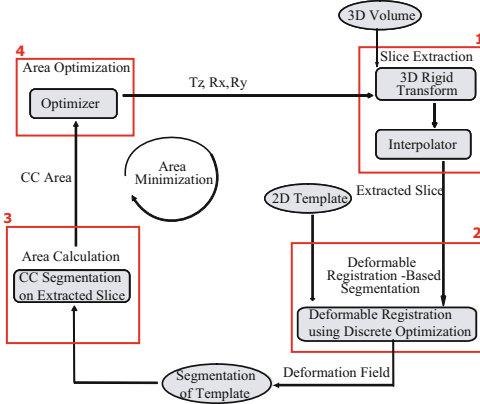


Fig. 1. Flow diagram of our method for finding the MCCAP. There are 4 primary steps. 1) A 2D slice is extracted (upper right) from the volume using P_{ext} . 2) The pre-segmented 2D template is registered (lower right) to the extracted slice through deformable registration. 3) The displacement field from the registration step is applied to the segmentation of the 2D template to segment the CC on the extracted slice. 4) The CC area and its derivatives with respect to P_{ext} are used (upper left) for updating the values of P_{ext} for the next iteration. Steps 1–4 are repeated until convergence.

CC on this slice. This central slice is positioned, either during acquisition or with post-processing, to align with the MSP. We hypothesize that most MSPs are spatially not very far from the MCCAP because the CC narrows from both sides toward the center of the brain and hence only a small spatial transformation is needed to reach the MCCAP from the MSP. In addition, we employ a multi-level optimization (Sec. 2.5) to further reduce the chances of converging onto a local minimum between the MSP and the MCCAP.

2.3 Deformable Registration-Based Segmentation

We nonlinearly register the presegmented 2D template to the extracted slice and apply the deformation field to the CC segmentation on the template to segment the CC on the extracted slice [17]. The area of the deformed CC on the template is taken to be the area of CC on the extracted slice. We use deformable registration using the discrete optimization (DROP) method [18] because it is initialization-free, does not require gradient calculations, and the Markov random field energy it adopts is optimized quickly using an efficient primal dual approach [19]. We use the sum of squared intensity differences as the similarity metric because the template and target come from the same scan.

2.4 Area Optimization

We minimize A_{cc} with respect to P_{ext} . During each iteration of the optimizer, $\frac{\partial A_{cc}}{\partial T_z}$, $\frac{\partial A_{cc}}{\partial R_x}$ and $\frac{\partial A_{cc}}{\partial R_y}$ are calculated using central differencing. A sequential quadratic programming method is employed for optimization. The optimizer updates an estimate of the Hessian of the Lagrangian at each iteration using the BFGS formula [20, 21]. We set the lower and upper search bounds empirically at -2.0° and 2.0° for R_x and R_y , and at -2.0 mm and 2.0 mm for T_z . The

bounds are in place largely for computational tractability, but they also ensure that the optimizer does not converge to a degenerate case in which a plane does not intersect with the CC bridge and has therefore zero CC area.

2.5 Multi-level Optimization

In addition to initializing with the MSP, to further decrease the possibility of converging to a local minimum of the MCCAP, we use a multi-level optimization scheme, where, in each step, the level of image smoothness is decreased and the optimization algorithm is initialized at the MCCAP parameters from the previous step. Smoothing is realized via 3D isotropic Gaussian filters with standard deviations: 10, 5, 3, 2, 1.8, 1.6, 1.4, \dots , 0.6, 0.4, and 0.2 mm (set empirically based on 3D visualizations of the variations in CC area over the entire search space, computed using brute force search in several sample MRI volumes). The size of each filter in each dimension is five times the standard deviation.

3 Results

We performed three experiments. In the first two experiments (Sec. 3.1 and Sec. 3.2), we used 3D T1 MRI brain scans of 20 normals from the internet brain segmentation repository (IBSR) (<http://www.cma.mgh.harvard.edu/ibsr/>). Each volume consists of 60 to 65 coronal slices, each of dimensions 256×256 , and was positionally normalized using the midpoints of the decussations of the anterior and posterior commissures and the midsagittal plane at the level of posterior commissure as points of reference for rigid transformation [22]. The repositioned scans were then re-sliced into 3 mm thick coronal slices with 1×1 mm pixels. In the third experiment (Sec. 3.3), we used 3D T1 MRI axial brain volumes of 10 secondary progressive MS patients and 5 normals. Each dataset dimension is $256 \times 256 \times 160$, with voxel size $1.17 \times 1.17 \times 1$ mm. The MSP was extracted by an MRI technologist trained on the Philips Achieva 3.0 Tesla scanner by initiating a 3-plane localizer in the coronal, axial, and sagittal planes, using the interhemispheric fissure as the internal landmark, and correcting for ear to shoulder tilt on the coronal plane and right to left rotation on the axial plane. We use an MSP obtained by careful manual placement because it gives us greater confidence in accuracy than automatic MSP methods that would still have to be visually inspected.

3.1 Achievement of Global Minima

We show that the MCCAP area found by the proposed optimization method is generally very close to the ground truth. We performed a brute force (BF) search in the space of T_z , R_x and R_y , with search resolution 0.1 mm, 0.1° , and 0.1° , respectively, and calculated the CC area on each slice to find the minimal CC area, up to the resolution of the search parameters. The upper and lower limits on P_{ext} were set as in Sec. 2.4. We compared the CC areas on the MCCAP using our

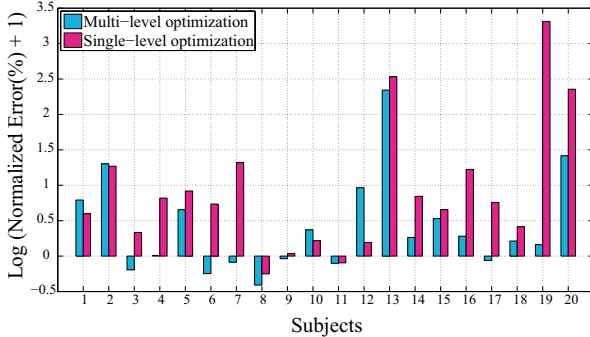


Fig. 2. Normalized error (%) in MCCAP areas (shown in logarithmic scale) is calculated as the difference between the MCCAP area computed by either optimization method and the minimum CC area computed by brute force, normalized by the minimum brute force area. The results from the multi-level optimization method is in general closer to the ground truth (global optimum). $\text{Log}(x+1)$ plot is used to take care of zero and negative values.

optimization algorithm with single- (Sec. 2.4) and multi-level optimization (Sec. 2.5) with BF results. With BF considered as ground truth, the error is computed as: $\text{Normalized Error (\%)} = (\text{Area}_{\text{MCCAP}} - \text{Area}_{\text{BF}}) / \text{Area}_{\text{BF}} \times 100\%$ (Fig. 2). In all cases except one, the minimum area from multi-level smoothing optimization is within 3.12% of the corresponding brute force result. In 7 cases, the minimum area from multi-level method is even smaller than brute force result due to the limited resolution of brute force search. The multi-level method produced a lower MCCAP area in comparison to the single-level in 85% of the cases (17 out of 20 subjects) by up to 26%. In the remaining 3 cases, the CC areas on the MCCAP found using the single-level optimization are slightly smaller (within 1.4%) than the multi-level counterpart, which may be due to the parameters of the multi-level smoothing resulting in over fitting of the optimizer at lower resolutions. This supports the assertion that the multi-level method yields results closer to the ground truth.

To investigate the uniqueness of the MCCAP, we compared the MCCAP parameters from the multi-level and brute force methods, and found that the parameters of the extracted planes can be substantially different (differences in $T_z = 0.5 \pm 1.2$ mm, $R_x = 0.51 \pm 0.66^\circ$, and $R_y = 0.65 \pm 0.82^\circ$), even when their computed minimal areas are very close. This supports the intuition that the MCCAP is not necessarily unique for a given brain MRI volume.

3.2 Quantitative Results for Method Robustness

To evaluate the robustness of our method, we compared the CC areas on the MCCAP in all 20 subjects for different initializations with respect to a default initialization of $T_z = R_x = R_y = 0$. The different initializations simulate minor variations in head position or slice angle, assuming some effort has been made

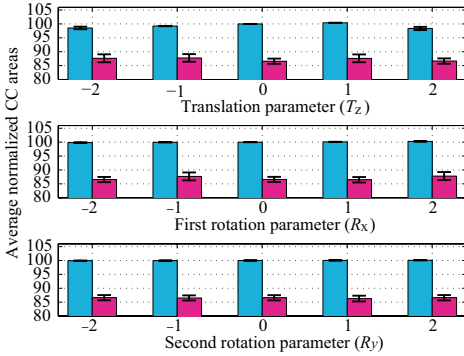


Fig. 3. Averaged normalized MCCAP areas for different initializations (while varying each parameter, the other two are set to zero) are shown in pink bars, while blue bars show the averaged normalized CC area on the initial slice. Normalization is done to the CC area on the slice with default initialization. Convergence of our method to almost the same normalized MCCAP area for different initializations demonstrates the robustness of the proposed method.

to center the volume. Fig. 3 shows that our method is robust to a wide range of initializations and is likely to converge very closely to a global minimum. The variance in MCCAP areas in all initializations is less than 1.4%. Also note that comparing the mean error between the MCCAP and brute force (from section 3.1) which is 1.08% with the 9–10% reduction in Fig. 3 shows that we recover approximately 90% of area difference between MSP and brute force.

3.3 MCCAP vs. MSP in Multiple Sclerosis Patients

We compared the discriminatory value of the CC areas calculated using an MSP vs. the MCCAP in distinguishing between multiple sclerosis (MS) patients and healthy controls. The means and standard deviations of the CC areas computed using the MSP and two MCCAP optimization methods are shown in Fig. 4. The multi-level smoothing optimization seems to be the best method at distinguishing the two groups. Note the larger separation of areas between the MS and normal groups using MCCAP multi-level optimization, as compared to the MSP. The area reduction for the normals ($\sim 12\%$) is much lower than for the patients ($\sim 50\%$), which we attribute to the true 3D shape differences between the groups. We quantified the advantage by computing the standardized mean differences between the normal and MS groups for MSP and MCCAP values using $(\overline{\text{Normal}} - \overline{\text{MS}}) / \sqrt{\sigma_{\text{Normal}}^2/n_{\text{Normal}} + \sigma_{\text{MS}}^2/n_{\text{MS}}}$, where σ is the

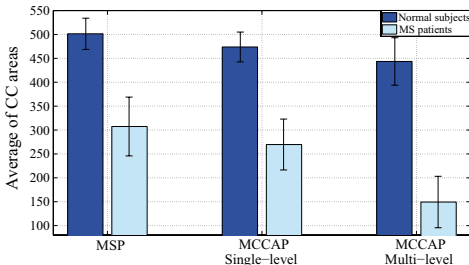


Fig. 4. Average of CC areas using different methods in MS patients and normal subjects. The larger difference in MCCAP areas found by multi-level smoothing optimization method shows that the two groups can be better distinguished than when MSP or single-level MCCAP areas are used.

standard deviation of each group and n is the sample size. If a method yields a bigger standardized mean difference, this suggests that it is better at distinguishing the two groups. The standardized mean difference for multi-level optimization is 10.51 which, compared to 7.97 for the MSP, is a 24% increase. Further, the p-values, associated with rejecting the hypothesis that the two populations come from the same distribution, show statistically significant differences between the normal and the MS subjects for all of the tested methods: $p_{\text{MSP}} = 1.94\text{e-}5$, $p_{\text{MCCAP-single}} = 2.82\text{e-}6$, $p_{\text{MCCAP-multi}} = 1.4\text{e-}7$. These initial results are supportive of our hypothesis that the interhemispheric communication bottleneck is an important bio-marker for MS, and that the MCCAP is better motivated than MSP to measure this bottleneck.

4 Conclusion

The plane with minimum CC area is proposed for the first time as the plane for studying CC atrophy. This choice is based on the hypothesis that the MCCAP is the cross section that is most representative of the bottleneck of interhemispheric transmission. We have developed an accurate and largely automatic method that employs an optimization process for extraction of the MCCAP along with the simultaneous segmentation of the embedded CC. We have shown that our optimizer converges closely to the global optimum area and is robust with respect to initialization. Preliminary results demonstrate better class separation between MS patients and normal subjects when the areas from the MCCAP are compared to those from an MSP. Future work includes longitudinal studies with larger data sets and more rigorous statistical tests (e.g. not necessarily assuming a normal distribution) to assess the usefulness of the MCCAP in monitoring disease progression and response to therapy.

Acknowledgements

Ghassan Hamarneh, Neda Changizi, and Omer Ishaq were partially funded by the Natural Sciences and Engineering Research Council of Canada. Advice on statistical methods was provided by Yinshan Zhao from Faculty of Medicine (Neurology), University of British Columbia.

References

1. Aboitiz, F., Scheibel, A.B., Fisher, R.S., Zaidel, E.: Fiber composition of the human corpus callosum. *Brain Res.* 598(1-2), 143–153 (1992)
2. Simon, J.H., et al.: A longitudinal study of brain atrophy in relapsing multiple sclerosis. *Neurology* 53(1), 139–148 (1999)
3. Woodruff, P.W., McManus, I.C., David, A.S.: Meta-analysis of corpus callosum size in schizophrenia. *J. Neurol. Neurosurg. Psychiatry* 58(4), 457–461 (1995)
4. Hardan, A.Y., Minshew, N.J., Keshavan, M.S.: Corpus callosum size in autism. *Neurology* 55(9), 1033–1036 (2000)
5. Smith, S., Jenkinson, M.: Accurate robust symmetry estimation. In: Taylor, C., Colchester, A. (eds.) *MICCAI 1999*. LNCS, vol. 1679, pp. 308–317. Springer, Heidelberg (1999)

6. Ardekani, B.A., Kershaw, J., Braun, M., Kanuo, I.: Automatic detection of the mid-sagittal plane in 3-D brain images. *IEEE TMI* 16(6), 947–952 (1997)
7. Liu, Y., Collins, R., Rothfus, W.: Robust midsagittal plane extraction from normal and pathological 3D neuroradiology images. *IEEE TMI* 20(3), 175–192 (2001)
8. Brummer, M.E.: Hough transform detection of the longitudinal fissure in tomographic head images. *IEEE TMI* 10(1), 74–81 (1991)
9. Marais, P., Guillemaud, R., Sakuma, M., Zisserman, A., Brady, M.: Visualising cerebral asymmetry. In: Höhne, K.H., Kikinis, R. (eds.) *VBC 1996*. LNCS, vol. 1131, pp. 411–416. Springer, Heidelberg (1996)
10. Anbazhagan, P., Carass, A., Bazin, P.L., Prince, J.L.: Automatic estimation of midsagittal plane and AC-PC alignment on nonrigid registration. In: *IEEE ISBI*, pp. 828–831 (2006)
11. Ishaq, O., Hamarneh, G., Tam, R., Trabouise, A.: Effects of Mid Sagittal Plane Perturbation and Image Interpolation on Corpus Callosum Area Calculation. In: *IEEE ISSPIT*, pp. 197–202 (2006)
12. Skoglund, K.V., Stegmann, M.B., Ryberg, C., Ólafsdóttir, H., Rostrup, E.: Estimation and Perturbation of the Mid-Sagittal Plane and its Effects on Corpus Callosum Morphometry. In: *International Society of Magnetic Resonance in Medicine* (2005)
13. Barkhof, F.J., Elton, M., Lindeboom, J., Tas, M.W., Schmidt, W.F., Hommes, O.R., Polman, C.H., Kok, A., Valk, J.: Functional correlates of callosal atrophy in relapsing-remitting multiple sclerosis patients. A preliminary MRI study. *J. Neurol.* 245(3), 153–158 (1998)
14. Bonzano, L., Tacchino, A., Roccatagliata, L., Abbruzzese, G., Luigi Mancardi, G., Bove, M.: Callosal contributions to simultaneous bimanual finger movements. *J. Neurosci.* 28(12), 3227–3233 (2008)
15. Pozzilli, C., Bastianello, S., Padovani, A., Passafiume, D., Millefiorini, E., Bozzao, L., Fieschi, C.: Anterior corpus callosum atrophy and verbal fluency in multiple sclerosis. *Cortex* 27(3), 441–445 (1991)
16. Pelletier, J., Suchet, L., Witjas, T., Habib, M., Guttmann, C.R., Salamon, G., Lyon-Caen, O., Chérif, A.A.: A longitudinal study of callosal atrophy and interhemispheric dysfunction in relapsing-remitting multiple sclerosis. *Arch. Neurol.* 58(1), 105–111 (2001)
17. Collins, D.L., Holmes, C.J., Peters, T.M., Evans, A.C.: Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping* 3(3), 190–208 (1995)
18. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense Image Registration through MRFs and Efficient Linear Programming. *Med. Image Anal.* 12(6), 731–741 (2008)
19. Komodakis, N., Tziritas, G., Paragios, N.: Fast, Approximately Optimal Solutions for Single and Dynamic MRFs. In: *IEEE CVPR*, pp. 1–8 (2007)
20. Fletcher, R., Powell, M.J.D.: A Rapidly Convergent Descent Method for Minimization. *Computer Journal* 6, 163–168 (1963)
21. Goldfarb, D.: A Family of Variable Metric Updates Derived by Variational Means. *Mathematics of Computing* 24, 23–26 (1970)
22. Talairach, J., Tournoux, P.: *Co-planar stereotaxic atlas of the human brain*. Thieme, New York (1988)

Incorporating Priors on Expert Performance Parameters for Segmentation Validation and Label Fusion: A Maximum a Posteriori STAPLE

Olivier Commowick^{1,2} and Simon K. Warfield²

¹ INRIA Rennes - VISAGES Team, France

Olivier.Commowick@irisa.fr

² CRL, Children's Hospital - Harvard Medical School, Boston, USA

Abstract. In order to evaluate the quality of segmentations of an image and assess intra- and inter-expert variability in segmentation performance, an Expectation Maximization (EM) algorithm for Simultaneous Truth And Performance Level Estimation (STAPLE) was recently developed. This algorithm, originally presented for segmentation validation, has since been used for many applications, such as atlas construction and decision fusion. However, the manual delineation of structures of interest is a very time consuming and burdensome task. Further, as the time required and burden of manual delineation increase, the accuracy of the delineation is decreased. Therefore, it may be desirable to ask the experts to delineate only a reduced number of structures or the segmentation of all structures by all experts may simply not be achieved. Fusion from data with some structures not segmented by each expert should be carried out in a manner that accounts for the missing information. In other applications, locally inconsistent segmentations may drive the STAPLE algorithm into an undesirable local optimum, leading to misclassifications or misleading experts performance parameters.

We present a new algorithm that allows fusion with partial delineation and which can avoid convergence to undesirable local optima in the presence of strongly inconsistent segmentations. The algorithm extends STAPLE by incorporating prior probabilities for the expert performance parameters. This is achieved through a Maximum A Posteriori formulation, where the prior probabilities for the performance parameters are modeled by a beta distribution. We demonstrate that this new algorithm enables dramatically improved fusion from data with partial delineation by each expert in comparison to fusion with STAPLE.

1 Introduction

Among numerous tools for the evaluation of automatic segmentation algorithms with respect to manual delineations [1,2,3,4], an algorithm named STAPLE (for Simultaneous Truth And Performance Level Estimation) [5] was proposed by Warfield et al. as a novel way to compute simultaneously a reference segmentation and performance parameters from a set of segmentations. This algorithm

is very versatile as it allows the evaluation of intra- and inter-rater variability as well as the comparison of segmentation algorithms with respect to multiple manual segmentations. It has therefore been used for many applications. Among them, it has been either embedded in new atlas construction methods [6], utilized to fuse segmentation decisions in multiple atlas-based segmentation [7], or to compute atlas segmentations from registered manual delineations [8].

Manual delineation is a very time consuming and burdensome task, even more when several structures have to be segmented in each image. Applications of manual segmentation, such as delineation of brain structures for neuroscience research, may be accelerated, and the quality of each segmentation improved, by having more experts who each delineate fewer structures. Some structures may then be missing in each rater segmentation. Performance estimation however requires observations of segmentation decisions of each structure by each rater. This can cause STAPLE to fail to provide accurate estimates of the reference standard and expert performance parameters. It would therefore be extremely valuable to take into account the missing structures to get accurate estimates of the reference and performance parameters. This would also help for existing datasets delineated in clinical conditions where structures are missing.

With this objective, Landman et al. [9] proposed an ad-hoc solution by fixing the parameters for missing structures and ignoring background voxels. This approach cannot be extended easily to take into account any prior on the expert parameters. This would however be valuable as the estimation of the parameters and reference segmentation may be incorrect when strong inconsistencies exist between the input segmentations. Inconsistent delineations may indeed lead the algorithm to an undesired local maximum where the performance parameter estimates converge to values incompatible with our prior information about rater performance. We can introduce an explicit prior model for rater performance parameters to drive the estimation algorithm to a better local optimum.

We propose a new algorithm that incorporates a prior probability for the performance parameters estimated through STAPLE. This is performed by extending the expression of the expected value of the complete data log-likelihood to a Maximum A Posteriori formulation incorporating prior probabilities as a beta distribution on each performance parameter. We applied our algorithm to label fusion with missing structures, and demonstrate its efficiency for improving label fusion and reducing manual rater delineation burden.

2 Method

2.1 Summary of STAPLE

We first summarize the principle of STAPLE [5]. It takes as an input a set of segmentations from J experts (either manual or automatic segmentations). These segmentations may be binary or multi-category segmentations, i.e. several structures are delineated with each structure represented by one specific label. The labeling of each voxel, in an image of I voxels, provided by the segmentation generators is referred to as segmentation decisions d_{ij} , indicating the

label given by each expert j for voxel i , $i \in [1 \dots I]$. The goal of STAPLE is then to estimate both a reference standard segmentation T , and parameters $\theta = \{\theta_1, \dots, \theta_j, \dots, \theta_J\}$ describing the agreement between each expert and the reference standard. Each θ_j is represented by an $L \times L$ matrix, where L is the number of labels, and $\theta_{j_s's}$ is the probability that the expert j gave the label s' to a voxel i when the reference standard label is s , i.e. $\theta_{j_s's} = P(d_{ij} = s' | T_i = s)$.

If the reference standard was known, then estimating the performance parameters for each expert would be straightforward. However, as this reference standard is unknown, an Expectation-Maximization approach [10,11] is used to estimate T and the expert performance parameters through the maximization of the expected value of the complete data log-likelihood $Q(\theta|\theta^{(k)})$:

$$Q(\theta|\theta^{(k)}) = \sum_i \sum_j \sum_s W_{si} \log(\theta_{j d_{ij} s}) \quad (1)$$

where W_{si} denotes the posterior probability of T for label s : $P(T_i = s | D, \theta^{(k)})$. The EM algorithm, which is guaranteed to converge to a local maximum, proceeds to identify the optimal estimate $\hat{\theta}$ by iterating two steps:

- E-Step: Compute $Q(\theta|\theta^{(k)})$, the expected value of the complete data log-likelihood given the current estimates of the expert parameters at iteration k : $\theta^{(k)}$. This requires computing $P(T|D, \theta^{(k)})$, i.e. the W_{si} values [5].
- M-Step: Estimate new performance parameters $\theta^{(k+1)}$, maximizing $Q(\theta|\theta^{(k)})$.

2.2 Introducing Priors: A Maximum a Posteriori Formulation

We consider the possibility of utilizing a prior probability for the performance parameters to modify the local maximum to which the estimator converges. This can be done by utilizing Maximum A Posteriori (MAP) estimation rather than Maximum Likelihood. MAP estimation is equivalent to augmenting the expected value of the complete data log-likelihood $Q(\theta|\theta^{(k)})$ with a term $\log(P(\theta))$ corresponding to the prior probability of the parameters:

$$Q_{MAP}(\theta|\theta^{(k)}) = Q(\theta|\theta^{(k)}) + \log(P(\theta)) \quad (2)$$

As the performance parameters for each label are independent, $P(\theta)$ can be expressed as a product of the independent probabilities $P(\theta_{j_s's})$. The appropriate form for the prior probability density function for each parameter $\theta_{j_s's}$ must be chosen. Several properties are desirable for this prior distribution:

- $\theta_{j_s's}$ is a probability and therefore must take its values in $[0, 1]$,
- it must be able to model any prior on the parameters (close to 1 e.g. diagonal parameters, close to 0 e.g. non-diagonal parameters, or uniform prior),
- a function for which the logarithm is easily obtained as well as its derivatives.

The beta distribution, $B_{\alpha,\beta}$, is particularly well suited to these requirements. Its support ranges between 0.0 and 1.0 and it allows, based on two parameters α and β , to consider a broad range of prior distributions for the parameters for

each expert (particularly the specific combination $\alpha = \beta = 1$ corresponds to the uniform prior used in the regular STAPLE). Further, the relative weight of each prior in Q can be modified by modeling the prior distribution as:

$$(B_{\alpha,\beta}(x))^\gamma = \left(\frac{1}{Z} x^{\alpha-1} (1-x)^{\beta-1} \right)^\gamma \quad (3)$$

with $\gamma \geq 0.0$ a scaling parameter. Z is the normalizing constant of the beta distribution. Moreover, the logarithm and the derivatives of $B_{\alpha,\beta}$ are easily computed.

2.3 Solving the MAP Formulation in the Multi-category Case

We associate each parameter $\theta_{js's}$ with a prior defined as a γ -weighted beta distribution $(B_{\alpha,\beta}(x))^\gamma$. The new expected value of the complete data log-likelihood function for the expert j is then expressed as:

$$Q'_{MAP}(\theta_j | \theta^{(k)}) = \gamma \sum_{s'} \sum_s \left((\alpha_{js's} - 1) \log(\theta_{js's}) + (\beta_{js's} - 1) \log(1 - \theta_{js's}) \right) + \sum_i \sum_s W_{si} \log(\theta_{jd_{ij}s}) \quad (4)$$

The computation of the posterior probability of the reference standard segmentation $P(T|D, \theta^{(k)})$ remains the same as in [5]. It indeed only depends on the current estimates $\theta^{(k)}$ and not on the prior on these parameters. However, the M-Step is modified by the prior distribution on the parameters.

M-Step: A Fixed Point Iterative Solution. The new estimates of the expert performance parameters are computed by differentiating Q'_{MAP} with respect to each $\theta_{js's}$ and equating the derivatives to 0 under the constraint that $\sum_{s'} \theta_{js's} = 1$. This leads to the following system for the parameters of each expert j :

$$\theta_{js's} = \frac{\left(\sum_{i:d_{ij}=s'} W_{si} \right) + \gamma(\alpha_{js's} + \beta_{js's} - 2) + \gamma \frac{\beta_{js's} - 1}{\theta_{js's} - 1}}{\sum_{n'} \left[\left(\sum_{i:d_{ij}=n'} W_{si} \right) + \gamma(\alpha_{jn's} + \beta_{jn's} - 2) + \gamma \frac{\beta_{jn's} - 1}{\theta_{jn's} - 1} \right]} \quad (5)$$

In this form, we can readily see that, for a particular label s and the set of decisions s' , the expression in the numerator is calculated once for each s' and the denominator is simply the sum of the numerators. When using a uniform prior on parameters ($\alpha_{js's} = \beta_{js's} = 1$) this system further simplifies to the regular STAPLE M-Step [5]. It also admits a closed form in two specific cases: first in the binary case, where the non-diagonal parameters are entirely determined by the values of the diagonal parameters, and also when all prior parameters $\beta_{js's} = 1$.

In the general multi-category case, with $\beta_{js's} \neq 1$, this system of equations does not admit any closed form solution. We therefore propose a simple iterative method to solve this system of equations. Equation (5) is a continuous mapping of the form $\theta_j = f(\theta_j)$, with $f :]0, 1[^N \rightarrow]0, 1[^N$ (where N is the number of parameters to compute for expert j). The iterative approach consists of applying

the f mapping recursively to the current estimate, i.e. computing the sequence $\{x_n\}_{n \geq 1}$ where $x_{n+1} = f(x_n)$ until convergence to the fixed point. Because of the configuration of the mapping f , Schaefer’s fixed point theorem applies, which guarantees that a fixed point solution to this system ($\theta_j = f(\theta_j)$) exists.

The $\{x_n\}$ sequence can be initialized from the previous parameters estimates $\theta_j^{(k)}$ or from the regular STAPLE parameters estimates. These initializations ensure that the sequence rapidly converges to the fixed point $\theta_j^{(k+1)} = f(\theta_j^{(k+1)})$.

3 Results

We have applied our algorithm to multi-label segmentations fusion with missing data. The manual segmentation of all structures in the entire brain is very long and costly. Repeated segmentations of the same images are necessary to estimate intra- and inter-rater variability, but this further increases the burden on each rater. It would be much more practical if each rater could focus on only a subset of structures, therefore lowering the segmentation burden of each rater. However this leads to segmentations in which some structures are missing and in which different error rates, associated with different raters, are present.

To simulate this situation, we have used a database of 15 adult images (T1 images, size $256 \times 256 \times 175$, 1 mm^3) where all structures (CSF, subcortical, cortical and cerebellar grey matter, white matter, cerebellar white matter) were delineated over the whole brain (see images (a,c,e,g) on Fig. 1). For each image, we then removed randomly 4 structures out of 6 (by replacing their labels

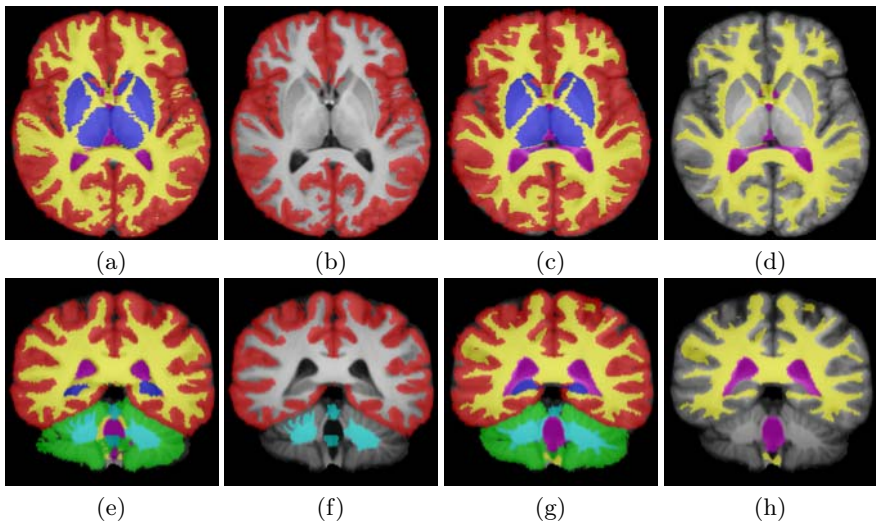


Fig. 1. Database of Segmentations. Individual manual segmentations registered on an average image. (a,c,e,g): original segmentations, (b,d,f,h): segmentations with 4 missing structures. Legend: red, blue, green: cortical, sub-cortical and cerebellar grey matter, yellow: white matter, pink: CSF, cyan: cerebellar white matter and brainstem.

with background label 0, see images (b,d,f,h) on Fig. 1) in such a way that all structures are segmented an equal number of times overall subjects.

We have aligned these images in a common template using Guimond et al.’s method [12], and run STAPLE first without taking into account missing structures (regular STAPLE algorithm as proposed in [5]). Then, we have run MAP STAPLE with a weight $\gamma = 10$, assuming a prior distribution close to 1 ($\alpha = 5$, $\beta = 1.5$) on diagonal elements for the delineated structures, on the background for missing structures, and a prior close to 0 ($\alpha = 1.5$, $\beta = 5$) on other parameters. These results as well as a regular STAPLE on the dataset without missing structures are presented in Fig. 2. Another option to account for missing structures would be to consider the case where raters were asked to delineate structures on an image where voxels are initially given an illegal label (e.g. -1) and ignore in STAPLE those voxels with the illegal label. We implemented this option with and without priors on parameters and the conclusions were similar.

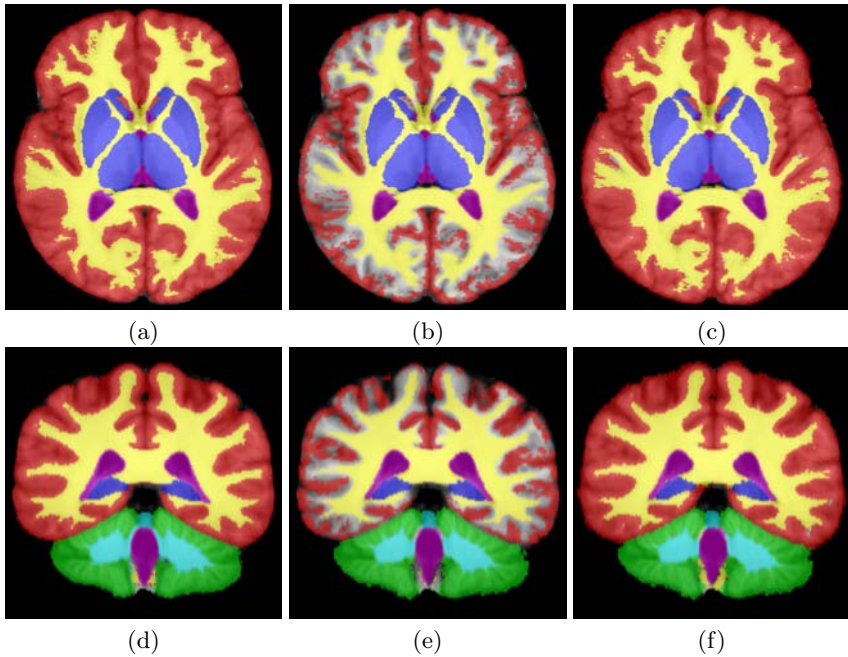


Fig. 2. Results on Label Fusion with Missing Data. (a,d): reference label fusion (all structures used), (b,e): label fusion with a third of the segmentations, (c,f): label fusion with a third of the segmentations with prior information.

Not taking into account the missing structures in the STAPLE algorithm leads to erroneous label fusion. We can indeed see on images (b) and (e) in Fig. 2 that the interface between cortical grey matter and white matter gets segmented as the background. Because missing structures are not taken into account, experts who segmented the structures obtain poor performance scores

and the background becomes the most typical structure in this region. On the contrary, when taking into account the missing structures (images (c) and (f)) by introducing an appropriate prior for the performance parameters values, the label fusion is much closer to what would be expected and also very close to regular STAPLE with all structures.

Table 1. Quantitative Evaluation of MAP STAPLE. Dice scores between the STAPLE reference estimated from all segmentations (images (a,d) on Fig. 2) and from the dataset with missing delineations using the regular STAPLE or MAP STAPLE. Legend: CGM, CeGM, SCGM: cortical, cerebellar and sub-cortical grey matter, WM, CeWM: brain and cerebellar white matter, CSF: cerebrospinal fluid.

Structure	CGM	CeGM	SCGM	WM	CeWM	CSF
Regular STAPLE	0.678	0.959	0.957	0.866	0.940	0.939
MAP STAPLE	0.939	0.960	0.958	0.947	0.939	0.939

This qualitative evaluation is confirmed by the Dice scores (shown in Table 1) between the results from the two methods and the reference segmentation obtained from all structures. The MAP STAPLE formulation therefore facilitates the accurate estimation of the reference segmentation and performance parameters by enabling accurate label fusion when expert raters are each asked to delineate only some of the brain structures.

4 Conclusion

We have presented a new algorithm to incorporate in STAPLE prior information for each of the expert performance parameters. This is obtained by utilizing a Maximum A Posteriori formulation for the expected value of the complete data log-likelihood and modeling the prior probability for each expert performance parameter with a beta distribution, whose parameters α and β allow for any prior distribution. We have derived a simple fixed point iterative solution for the performance parameters estimates for the most general multi-category case. Further, we identified specific cases where closed forms can be derived.

The MAP formulation we have presented may have many applications in validation studies and label fusion. We have illustrated our algorithm on a database with missing delineations (e.g. some structures are not segmented and assigned the background level), showing how MAP STAPLE allows to deal with these images and produce meaningful results. This experiment is particularly interesting as it will allow in the future for the design of validation experiments with multiple experts and multiple structures while minimizing the delineation burden for the experts. Apart from this application, this algorithm may be used to drive the STAPLE algorithm out of undesirable local maxima and obtain realistic tissue classifications even in the presence of strongly inconsistent input segmentations. This could be of great interest in the future to take into account registration errors or inconsistencies among manual segmentations.

In the future, we will use this algorithm to define new validation protocols with a lower delineation burden on the experts. This could be achieved for multiple structures as proposed here, or, for large structures, by asking the experts to delineate different slices and fuse them using our multi-category MAP algorithm, assigning each slice with a different label. Finally, the parameters α, β for each $\theta_{j_s'}$ and the weight γ may have an important effect on fusion results. We will perform a cross-validation study on these parameters and determine if γ can be optimized automatically to get the best trade-off between prior and data.

Acknowledgments. This investigation was supported in part by a research grant from CIMIT, grants RG 3478A2/2 and RG 4032A1/1 from NMSS, and by NIH grants R03 EB008680, R01 RR021885, R01 GM074068 and R01 EB008015.

References

1. Huttenlocher, D., Klanderman, D., Rucklidge, A.: Comparing images using the Hausdorff distance. *IEEE TPAMI* 15(9), 850–863 (1993)
2. Dice, L.: Measures of the amount of ecologic association between species. *Ecology* 26(3), 297–302 (1945)
3. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index. *Acad. Radiol.* 11(2), 178–189 (2004)
4. Gerig, G., Jomier, M., Chakos, M.: VALMET: A new validation tool for assessing and improving 3D object segmentation. In: Niessen, W.J., Viergever, M.A. (eds.) *MICCAI 2001*. LNCS, vol. 2208, pp. 516–523. Springer, Heidelberg (2001)
5. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE TMI* 23(7), 903–921 (2004)
6. Weisenfeld, N.I., Warfield, S.K.: Automatic segmentation of newborn brain MRI. *Neuroimage* 47(2), 564–572 (2009)
7. Rohlfing, T., Brandt, R., Menzel, R., Maurer, C.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage* 21(4), 1428–1442 (2004)
8. Commowick, O., Grégoire, V., Malandain, G.: Atlas-based delineation of lymph node levels in head and neck computed tomography images. *Rad. Oncol.* 87(2), 281–289 (2008)
9. Landman, B.A., Bogovic, J.A., Prince, J.L.: Efficient anatomical labeling by statistical recombination of partially label datasets. In: *Proc. of ISMRM*, p. 269 (2009)
10. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* 39(Series B) (1977)
11. McLachlan, G., Krishnan, T.: *The EM Algorithm and Extensions*. John Wiley and Sons, Chichester (1997)
12. Guimond, A., Meunier, J., Thirion, J.: Average brain models: A convergence study. *Computer Vision and Image Understanding* 77(2), 192–210 (2000)

Automated Segmentation of 3-D Spectral OCT Retinal Blood Vessels by Neural Canal Opening False Positive Suppression

Zhihong Hu¹, Meindert Niemeijer^{1,2}, Michael D. Abramoff^{2,1,3},
Kyungmoo Lee¹, and Mona K. Garvin¹

¹ Departments of Electrical and Computer Engineering, and

² Ophthalmology and Visual Sciences, The University of Iowa, Iowa City, IA, USA

³ Iowa City VA Medical Center, Iowa City, IA, USA

{zhihong-hu,mona-garvin}@uiowa.edu

Abstract. We present a method for automatically segmenting the blood vessels in optic nerve head (ONH) centered spectral-domain optical coherence tomography (SD-OCT) volumes, with a focus on the ability to segment the vessels in the region near the neural canal opening (NCO). The algorithm first pre-segments the NCO using a graph-theoretic approach. Oriented Gabor wavelets rotated around the center of the NCO are applied to extract features in a 2-D vessel-aimed projection image. Corresponding oriented NCO-based templates are utilized to help suppress the false positive tendency near the NCO boundary. The vessels are identified in a vessel-aimed projection image using a pixel classification algorithm. Based on the 2-D vessel profiles, 3-D vessel segmentation is performed by a triangular-mesh-based graph search approach in the SD-OCT volume. The segmentation method is trained on 5 and is tested on 10 randomly chosen independent ONH-centered SD-OCT volumes from 15 subjects with glaucoma. Using ROC analysis, for the 2-D vessel segmentation, we demonstrate an improvement over the closest previous work with an area under the curve (AUC) of 0.81 (0.72 for previously reported approach) for the region around the NCO and 0.84 for the region outside the NCO (0.81 for previously reported approach).

1 Introduction

Spectral-domain optical coherence tomography (SD-OCT) is a noncontact, non-invasive imaging technique used to obtain high resolution images of the retina or optic nerve head (ONH). It is a powerful modality to qualitatively assess retinal features and pathologies or to make quantitative measurements of retinal or ONH morphology.

Due to the fact that the retinal blood vessels absorb the wavelengths of light used in SD-OCT, the vessels in these volumes are mostly not directly visible. However, this causes vessel shadows to appear below the position of the vessels. As reported by Wehbe *et al.* [1], the retinal vessels are located right above the vessel shadows in the z -direction of the OCT volume and these shadows can be

used to detect the location of the vessels. The segmentation of vessels in SD-OCT could lead to a more objective diagnosis of diseases, be used for OCT-to-OCT or OCT-to-fundus registration, help remove the influence of vessels (when desirable), and aid in the detection of other retinal anatomic structures.

However, vessel segmentation in SD-OCT volumes of the ONH is not a trivial problem for a number of reasons. 1) The vessel shadows have a weak visibility in OCT due to the 3-D structure of the ONH. 2) Many vessels can overlap in regions, especially the region inside the neural canal opening (NCO), where the individual vessels cannot be discerned. 3) The presence of the NCO boundary, which crosses with the vessels, causes false positives for the vessel segmentation. 4) The vessel shadows have a decreased contrast with the background within the NCO, due to more variable tissue properties in the ONH.

Niemeijer *et al.* [2] previously presented an automated vessel segmentation algorithm for SD-OCT volumes of the ONH. In their algorithm, they first created a 2-D projection image from the 3-D volume and then utilized a pixel classification approach to segment the vessels in the projection image. Features were computed using Gaussian filter banks. However, their approach mainly segmented the vessels in the region outside the NCO and the vessels inside the NCO were excluded from the analysis. In addition, the algorithm exhibited a large numbers of false positives near the NCO. For example, the top row of Fig. 1 shows the NCO boundary (blue arrows) and typical false positives (red arrow) near the NCO from Niemeijer’s vessel segmentation approach.

In this work, we present a 2-D pixel classification algorithm to segment the blood vessels in SD-OCT volumes centered at the ONH, with a special focus on better identifying vessels near the NCO (Section 2.2). We approach this by incorporating pre-segmented NCO location (Section 2.1) information into the classification process. We then utilize these 2-D vessel profiles to segment the 3-D vessels in the SD-OCT volumes (Section 2.2).

2 Methods

2.1 Neural Canal Opening Pre-segmentation

In order to incorporate the NCO information into the 2-D vessel classification process, a graph-theoretic approach is first applied to identify the location of the NCO [3]. More specifically, the ONH-centered raw OCT volume is first flattened by identifying four intraretinal surfaces using a graph-search-based multilayer segmentation algorithm in 3-D, fitting a thin-plate spline to a segmented surface by ignoring the central ONH region, and then translating the columns so that this surface becomes a flat plane [4]. For surface 2, 3, and 4, the average positions outside an estimated NCO region are radially interpolated into the inside region. A 2-D projection image is obtained by computing the mean intensity values from a small number of slices between the interpolated surface 2 (orange) and 4 (yellow) (Fig. 1e). Thus the 3-D NCO segmentation problem is converted to

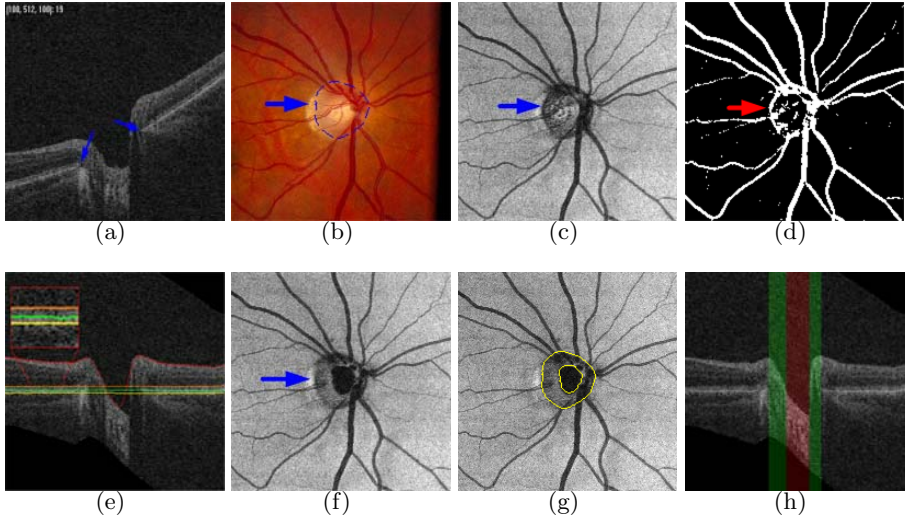


Fig. 1. Niemeijer’s vessel segmentation approach and our neural canal opening segmentation. (Top row) An example illustration of the NCO boundary (indicated with blue arrow) and false positives of NCO (indicated with red arrow) by Niemeijer’s vessel segmentation approach [2]. (a) Central slice of an OCT volume. (b) Corresponding fundus image. Note that the blue-arrow indicated contour is the NCO boundary. (c) OCT projection image. (d) Vessel segmentation result showing the typical false positives near the NCO. (Bottom row) An example illustration of our NCO segmentation algorithm [3]. (e) 3-D four surface segmentation with the interpolation of the radial average positions outside the estimated NCO to inside the NCO for surface 2 (orange), 3 (green), and 4 (yellow). (f) NCO-aimed projection image from the layer between the interpolated surface 2 and 4. Segmented NCO and cup overlapping with (g) the projection image and (h) a cross-sectional slice of the OCT volume.

a 2-D problem. The projection image is referred as the “NCO-aimed projection image” (Fig. 1(f)). The outer boundary in the projection image corresponds to the NCO and the inner boundary corresponds to the cup at the level of the NCO reference plane.

In order to perform the graph search, the NCO-aimed projection image is transformed to polar coordinates by unwrapping from the geometric center of the image. A signed edge-based term favoring a dark-to-bright transition from the transformed projection image is used as the cost function. In our case, the unwrapped cost image can be modeled as a weighted, directed graph similar to the one described by Li *et al.* [5] but reduced to a two-dimensional problem. We use the graph search to simultaneously segment the optimal NCO and cup boundaries. The NCO and cup boundaries are finally smoothed using a B-spline. The bottom row of Fig. 1 illustrates the NCO-aimed projection image creation and NCO pre-segmentation.

2.2 Blood Vessel Segmentation

For the 2-D vessel segmentation, oriented Gabor wavelets around the center of the NCO are utilized to extract features in a 2-D vessel-aimed projection image. The corresponding oriented NCO-based templates are utilized along with the Gabor wavelets to suppress the false positive tendency near the NCO boundary. A supervised pixel classification algorithm is applied to automatically segment the blood vessels in the vessel-aimed projection image. Based on the 2-D vessel location information, the 3-D vessels are detected by applying a triangular-mesh-based graph search to the isotropic SD-OCT volume [6]. These steps are described in more detail below.

Vessel-aimed projection image creation. The main difficulty of vessel segmentation within SD-OCT volumes is the weak visibility of the vessel pattern. In the previous approach [2], they proposed to use a 2-D projection of the vessel pattern from the 3-D volume to segment vessels. They compared two different projection images: the “naive” projection image computed by averaging the whole OCT volume which decreased the contrast between the vessels and background and the “smart” projection image computed by averaging the layer between surface 2 and 4 which provided a good contrast. However, the “smart” projection image also had some disadvantages. For example, near and inside the NCO, the layer between surface 2 and 4 would frequently become very thin (and be subject to layer segmentation errors) and thus the projection image would not necessarily demonstrate an optimal contrast between vessels and background in this region.

In this work, we create a new type of projection image by averaging the layer between surface 2 and an under planar surface defined by the deepest position of the top surface. In columns for which the deepest position is above the interpolated surface 4, the projection image is created as that used for the “smart” projection image. The created new projection is referred as the “vessel-aimed projection image”. Compared with the “naive” and “smart” projection images [2], the vessel-aimed projection image relies less on the surface segmentation and can take advantage of the vessel information inside the NCO. An illustration of the bounding surfaces for the three different projection images is shown in Fig. 2a.

2-D vessel segmentation. In the vessel-aimed projection image, the blood vessels generally radially distribute around the NCO center. Gabor wavelets demonstrate some desirable characteristics: spatial frequency, spatial locality, and orientation selectivity. The nature of the Gabor wavelets makes them well suitable for the feature generation of the blood vessel detection. Recall that a Gabor wavelet $\psi_{\mu,\nu}(z)$ [7] can be defined as:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}} \right], \quad (1)$$

where $z = (x, y)$, $\|\cdot\|$ is the norm operator, μ and ν define the orientation and spatial frequency scale of the Gabor kernel, and σ is related to the standard

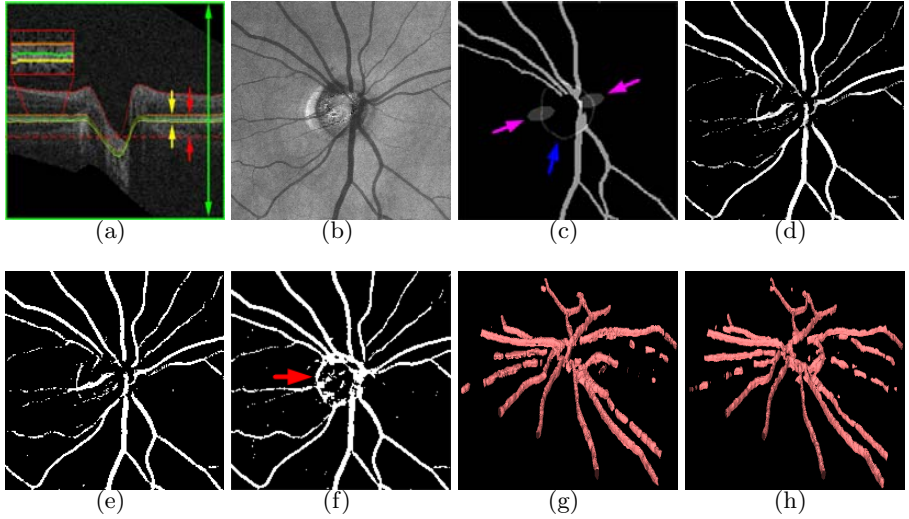


Fig. 2. Illustration of the 2-D vessel segmentation using our approach and Niemeijer’s approach. (a) The bounding surfaces used for the creation of three potential projection images: “naive” (bounding surfaces of whole volume as indicated with green arrows), “smart” (orange surface 2 and yellow surface 4 as indicated by yellow arrows), and “vessel-aimed projection image” (orange surface 2 and dashed red surface where possible as indicated with red arrows; same as “smart” where the red dashed surface is above the interpolated surface 4). (b) Vessel-aimed projection image as used in this work. (c) A schematic illustration of the Gabor wavelet responses and the NCO-based templates oriented at 165 degrees. Blue arrow = NCO contour. Purple arrows = template pair centered on the NCO boundary. (d-e) 2-D vessel segmentation posterior probability and thresholded binary images of our approach. (f) 2-D vessel segmentation of Niemeijer’s approach. (g-h) 3-D vessel segmentation of our approach and Niemeijer’s approach.

derivation of the Gaussian window in the kernel and determines the ratio of the Gaussian window width to the wavelength. The wave vector $k_{\mu,\nu}$ is defined as

$$k_{\mu,\nu} = k_{\nu} e^{i\phi_{\mu}}, \quad (2)$$

where $k_{\nu} = \frac{k_{max}}{f^{\nu}}$ in which k_{max} is the maximum frequency and f^{ν} is the spatial frequency between kernels in the frequency domain. $\phi_{\mu} = \frac{\pi\mu}{n}$ in which n is the total number of orientations. Based on the vessel profiles, in our application, we choose $k_{max} = \frac{\pi}{2}$, $\sigma \in \{1, 2, 3\}$, $f = \sqrt{2}$, $\nu \in \{1, 2, 3\}$, $n=18$, and $\mu \in \{0, \dots, 17\}$.

Additionally, to increase the signal and decrease the noise, three spatial frequency scale additions are applied between $\nu=2$ and $\nu=3$ in $\sigma \in \{1, 2, 3\}$. Together a Gabor wavelet family with 3 Gaussian scales, 3 spatial frequency scales, 3 spatial frequency scale additions, and 18 orientations is generated.

In order to suppress the false positive tendency near the NCO, the oriented templates are utilized along with the corresponding Gabor wavelets in the feature space (Fig. 2c). Specifically a pair of pre-defined templates is first created based

on the previously segmented NCO location information and so-called NCO-based templates. The center of the template pair is that of NCO and the center of each of them lies on the NCO boundary. The shapes of the templates are defined as:

$$\frac{\{x_1 - x_c - r_1\}^2}{W^2} + \frac{\{y_1 - y_c\}^2}{H^2} = 1, \quad (3)$$

and

$$\frac{\{x_2 - x_c + r_2\}^2}{W^2} + \frac{\{y_2 - y_c\}^2}{H^2} = 1, \quad (4)$$

where (x_c, y_c) is the NCO center, r_1 and r_2 are the distances of the center of each template to the center of NCO, and W and H are the maximum width and height of the templates which are defined based on prior knowledge of NCO profiles.

The NCO-based templates rotate in the same orientations with the Gabor wavelets. Wherever the templates rotate, the average background pixel value is assigned to those regions. The main advantage of the NCO-based templates is that it could suppress the false positive tendency from the NCO and at the same time does not affect the true positive of vessels, based on the assumption that the vessels are not parallel with the NCO.

After the feature extraction, each feature is normalized to zero mean and unit variance. A k -NN classifier with $k = 31$ is applied for the pixel classification. Each pixel in the vessel-aimed projection test image is assigned a soft label. Together they form a posterior map. Finally a threshold filter is utilized on the posterior probability map to obtain a binary segmentation image and a voting filter is applied to remove noise and small vessels. Fig. 2d and 2e show an example result of the 2-D vessel segmentation.

3-D vessel segmentation. In order to perform the 3-D vessel segmentation, the flattened SD-OCT is first transformed to an isotropic volume. Surface 1 and 2 are also correspondingly transformed. As mentioned above, the blood vessels themselves in the SD-OCT volume of the ONH are not visible. What we measure in the 2-D vessel-aimed projection image are the vessel shadows. The “true” vessels are located right above the vessel shadows [1]. We assume the vessels are approximately the three middle voxels between surface 1 and 2 in the z -axis of the SD-OCT volume. An initial binary 3-D vessel model is created by projecting the segmented 2-D vessel locations to the layer of the three middle voxels. A marching cube algorithm is applied to the initial model to construct a triangular mesh. The magnitude of the Gaussian derivative of the volumetric SD-OCT is combined with that of the vessel-aimed projection image to create the cost image. The globally optimal surfaces of the 3-D vessels are achieved by solving a maximum flow problem on the constructed triangular-mesh-based graph from the combined cost image. An example of the 3-D vessel segmentation result is shown in Fig. 2g.

3 Experimental Methods and Results

15 ONH-centered SD-OCT volumes from 15 subjects with glaucoma are acquired using a CirrusTM HD-OCT (Carl Zeiss Meditec) device. Each volume has $200 \times 200 \times 1024$ voxels corresponding to $6 \times 6 \times 2$ mm³. Of the 15 volumes, 5 of them are randomly selected as the training set and 10 of them the test set. Each pixel in the vessel-aimed projection image is manually labeled as “vessel” or “non-vessel” with the help of experts. The small vessels by observation are excluded from the “vessel” category and labeled as “non-vessel”.

The performance of the 2-D vessel segmentation is evaluated using Receiver Operating Characteristic (ROC) curves. Our present approach of the 2-D vessel segmentation is compared with Niemeijer’s approach [2] in terms of the areas under the curves (AUC) for the regions around the NCO (± 15 pixels from the NCO boundary) and outside the NCO boundary. An example visual comparison of the 2-D and 3-D vessel segmentation between Niemeijer’s and our approach is illustrated in the bottom row of Fig. 2. Fig. 2e and 2g show the results of the 2-D and 3-D segmentation of our approach and Fig. 2f and 2h of Niemeijer’s approach, respectively. As can be seen, the clear false positive near the NCO exists in the previous approach. However, in the present approach, it is greatly suppressed and the vessels around the NCO are detected. Table 1 demonstrates the 2-D quantitative segmentation results by comparing the AUCs of the region outside the NCO and the region around the NCO (± 15 pixels from the NCO boundary) of the two algorithms. Our present algorithm gives a greater AUC in both regions, especially in the region around the NCO.

Table 1. AUC comparison of 2-D vessel segmentation of our and Niemeijer’s algor.

AUC of the region around NCO		AUC of the region outside NCO	
Our algor.	Niemeijer’s algor.	Our algor.	Niemeijer’s algor.
0.81	0.72	0.84	0.81

4 Discussion and Conclusion

We developed an approach for automatically segmenting the retinal blood vessels by focusing on the region around the NCO in ONH-centered SD-OCT volumes. To our knowledge, this is the first such algorithm to segment the blood vessels by utilizing *a-priori* NCO segmentation information, along with rotated Gabor wavelets and corresponding rotated NCO-based templates to suppress the false positive tendency near the NCO. The result is promising considering the difficulty of the vessel segmentation in ONH-centered SD-OCT volumes. Compared with the closest previous work, the 2-D vessel segmentation results are greatly improved both visually and quantitatively, especially for the region around the NCO. However, the accuracy of the 2-D vessel segmentation is still not perfect.

Some of the potential error sources include the following. 1) The OCT data were from patients with glaucoma and the image quality (noisiness and contrast) was not as good as that from normal scans, thereby causing the broken appearance of some of the smaller vessels. 2) The expert may not have always traced very small vessels they deemed “incomplete,” whereas the algorithm would have found portions of these vessels, thus (incorrectly) causing the identification of these regions as false positives. In the future, for the 2-D vessel pixel classification, the texture information around the NCO might be helpful for more accurately identifying the NCO and blood vessels. For the graph construction, using an arc-weighted graph [8] and considering the graph column crossing problem of vessel bifurcations [9] would also help improve the algorithm performance.

References

1. Wehbe, H., Ruggeri, M., Jiao, S., Gregori, G., Puliafito, C.A., Zhao, W.: Automatic retinal blood flow calculation using spectral domain optical coherence tomography. *Optics Express* 15(23), 15193–15206 (2007)
2. Niemeijer, M., Garvin, M.K., van Ginneken, B., Sonka, M., Abramoff, M.D.: Vessel segmentation in 3D spectral OCT scans of the retina. In: *Proc. of SPIE Medical Imaging: Image Processing*, vol. 6914, p. 69141R (2008)
3. Hu, Z., Abramoff, M.D., Kwon, Y.H., Lee, K., Garvin, M.K.: Automated segmentation of neural canal opening and optic cup in 3-D spectral optical coherence tomography volumes of the optic nerve head. *Invest. Ophthalmol. Vis. Sci.* (2010) (in press)
4. Garvin, M.K., Abramoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans. Med. Imag.* 28(9), 1436–1447 (2009)
5. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images – a graph-theoretic approach. *IEEE Trans. Pattern Anal. Machine Intell.* 28(1), 119–134 (2006)
6. Lee, K., Abramoff, M.D., Niemeijer, M., Garvin, M.K., Sonka, M.: 3-D segmentation of retinal blood vessels in spectral-domain OCT volumes of the optic nerve head. In: *Proc. of SPIE Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 7626, p. 76260V (2010)
7. Liu, C., Wechsler, H.: Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.* 11(4), 467–476 (2002)
8. Song, Q., Wu, X., Liu, Y., Smith, M., Buatti, J., Sonka, M.: Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 827–835. Springer, Heidelberg (2009)
9. Yin, Y., Sonka, M.: Electric field theory based approach to search-direction line definition in image segmentation: Application to optimal femur-tibia cartilage segmentation in knee-joint 3-D MR. In: *Proc. of SPIE Medical Imaging: Image Processing*, vol. 7623, p. 76230W (2010)

Detection of Gad-Enhancing Lesions in Multiple Sclerosis Using Conditional Random Fields

Zahra Karimaghloo¹, Mohak Shah^{1,3}, Simon J. Francis², Douglas L. Arnold³,
D. Louis Collins², and Tal Arbel¹

¹ Centre for Intelligent Machines, McGill University, Canada

² Montreal Neurological Institute, McGill University, Canada

³ NeuroRx Research, Montreal, Canada *

Abstract. Identification of Gad-enhancing lesions is of great interest in Multiple Sclerosis (MS) disease since they are associated with disease activity. Current techniques for detecting Gad-enhancing lesions use a contrast agent (Gadolinium) which is administered intravenously to highlight Gad-enhancing lesions. However, the contrast agent not only highlights these lesions, but also causes other tissues (e.g. blood vessels) or noise in the Magnetic Resonance Image (MRI) to appear hyperintense. Discrimination of enhanced lesions from other enhanced structures is particularly challenging as these lesions are typically small and can be found in close proximity to vessels. We present a new approach to address the segmentation of Gad-enhancing MS lesions using Conditional Random Fields (CRF). CRF performs the classification by simultaneously incorporating the spatial dependencies of data and labels. The performance of the CRF classifier on 20 clinical data sets shows promising results in successfully capturing all Gad-enhancing lesions. Furthermore, the quantitative results of the CRF classifier indicate a reduction in the False Positive (FP) rate by an average factor of 5.8 when comparing to Linear Discriminant Analysis (LDA) and 1.6 comparing to a Markov Random Field (MRF) classifier.

1 Introduction

Multiple Sclerosis (MS) is a disorder of the central nervous system and is a common disease among young adults. Magnetic Resonance Imaging (MRI) is widely used to study this disease and assess its temporal progress. In particular, MRI is well established as the optimal imaging technique for detecting MS lesions which are believed to be highly correlated with the presence of disease activity in the absence of clinical changes. Moreover, using a contrast agent (e.g. Gadolinium) with T1-weighted (T1w) imaging can help to identify areas which represent the initial stage of lesion development. It is generally believed that the number and volume of enhancing lesions are important indicators of disease activity in MS. These indicators have generally been determined by using operator guided segmentation methods. However, in addition to being laborious and time

* This work was supported by NSERC Strategic Grant (350547-07).

consuming, manual detection is prone to intra- and inter-expert variability making the analysis of the results very complicated. Therefore, it is desirable to have a fully automatic segmentation scheme for detection of enhanced lesions in studies with large ensemble of patient data. Unfortunately, automatic identification of enhanced lesions is a challenging task due to the presence of many non-lesion enhancements, which appear as small regions and are very similar to Gad-enhancing lesions (unlike T2-weighted lesions, Gad-enhancing lesions are usually very small).

Even though there have been many studies for MS lesion detection on T2-weighted (T2w) MR images, only a few have investigated assessment of enhanced lesion detection. Bedell *et al.* [1] suggested using a special pulse sequence for T1w with contrast (T1c) imaging which increases the contrast between blood signal enhancement and lesion enhancement. In addition to the need for a special pulse sequence, this algorithm also requires user inputs for initial seed placement for detection of cerebro-spinal fluid to eliminate enhanced areas caused by circumventricular organs particularly in the region of choroid plexus. Miki *et al.* [2] proposed using a fuzzy connectivity to delineate enhanced lesions. Their approach is not fully automatic, as it requires human confirmation after each enhanced area is found by the algorithm. The study by Narayana *et al.* [3] also uses special pulse sequence as in [1]. Although it is a fully automatic approach, it requires prior segmentation of T2w lesions. In this paper, we propose a novel fully automatic probabilistic technique which is one step towards a clinical tool for detection of enhanced lesions. Our framework does not need human intervention and also uses only the commonly acquired MRI sequences (i.e. T1w before and after contrast, T2w and FLAIR). Furthermore, our approach does not rely on pre-segmentation of T2w lesions in other sequences. We model the enhanced lesion identification problem as a probabilistic classification task in which for each instance (i.e. pixel in the MRI) we seek to find the best labeling. This is achieved by computing the posterior distribution over the joint distribution of labels given the observations. A deterministic decision can be sought afterwards in order to have a binary classification result.

Our framework uses a discriminative model based on Conditional Random Fields (CRFs) [4,5] to model the posterior distribution of labels. There are two advantages to CRF-based models. First, they directly model the posterior distribution, and hence avoid modeling the joint distribution of observations and labels. Modeling the joint distribution is a complicated and expensive task and often leads to simplifying assumptions that may not generally be valid (especially in the field of medical imaging). The second benefit of a CRF-based model is that the spatial dependencies among instances can be modeled in a principled manner [5]. In order to model these dependencies, both labels of neighbouring instances, as well as their observations are taken into account. This is in contrast to traditional approaches for modelling data dependencies (e.g. MRF), where typically only the neighbouring labels are considered. Although consideration of label dependencies may be sufficient in some applications, incorporating observation dependencies provides a more powerful source of information that can be

exploited to achieve better classification results. Particularly, in the case of enhanced lesion identification, many False Positives (FPs) which occur due to the enhancement of vascular structures can be avoided by incorporating the neighbouring observations. The experimental results of applying the CRF classifier on a multi-centre clinical data sets from patients with various stages of relapsing remitting MS show significant improvements over traditional classification approaches.

The remains of this paper is organized as follows: Section 2 reviews the background of the CRF and elaborates on our CRF-based classifier. We then present the classification results of Gad-enhanced lesions in Section 3. Section 4 summarizes the paper and discusses the possibilities for further developments.

2 Method

Traditional classification techniques such as Decision Trees, Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) assume data instances are independent and identically distributed. While this assumption is valid for many data mining problems, in the context of image region labeling, it is important to incorporate contextual information in the form of spatial dependencies between pixels in the image. Since labels of neighbouring pixels are typically correlated, integration of spatial information generally yields smoother and more reliable results.

Markov Random Fields (MRF) have been extensively used to capture contextual constraints by allowing the classification of one instance to depend on the labels of neighbouring instances. In fact, MRF is a generative approach which models the joint probability distribution of the image observations and their corresponding labels. Letting $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{1, 0\}$ show the observation vector and the label of the i^{th} instance respectively, MRF models the posterior distribution over all labels $\mathbf{Y} = \{y_1, \dots, y_n\}$ given all observations $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ as:

$$p(\mathbf{Y}|\mathbf{X}) \propto p(\mathbf{X}, \mathbf{Y}) = p(\mathbf{X}|\mathbf{Y})p(\mathbf{Y}) \quad (1)$$

The priori distribution, $p(\mathbf{Y})$, is commonly modeled as a Markov field based on the neighbouring labels. Since this prior does not depend on the observations, the label interactions are modeled neglecting the observed data. This may lead to over smoothing of regions. Furthermore, for tractability purposes, the observations are assumed to be independent given the labels. Thus the likelihood can be expressed with a simple factorized form of $p(\mathbf{X}|\mathbf{Y}) = \prod_i p(\mathbf{x}_i|y_i)$. As a result of this restrictive independence assumption, the interactions among observations are not generally modeled in MRF.

Many discriminative approaches such as CRF [4], and its derivatives (Discriminant Random Field (DRF) [5] and Support Vector Random Field (SVRF) [6]) have been proposed to overcome shortcomings of MRF by integrating dependencies among observations into the model. In a discriminative framework, we directly model the posterior distribution $p(\mathbf{Y}|\mathbf{X})$ without building the joint distribution. Particularly, discriminative approaches relax the conditional independence

assumption of the observations and allow modeling of observation dependent interactions. Incorporating the neighbouring observations should help achieve better classification results. Specifically, in the context of enhanced lesion identification, many FPs can be excluded, due to the fact that many vascular structures (especially those within CSF) have different neighbourhood observations than enhanced-lesions.

As graphical models are widely used to represent probabilistic relationships between instances, we continue using graphical notations to explain the formulation of discriminative methods. In a graphical model $G(V;E)$, instances are shown with a set of nodes (V) and their conditional (in)dependencies are represented by (missing)edges (E). Inspired by the model suggested by Lafferty *et al.* [4], CRF-based models formulate the posterior probability of labels given the observations as:

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z} \prod_{i \in V} \phi_i(y_i|\mathbf{X}) \prod_{i,j \in E} \psi_{i,j}(y_i, y_j|\mathbf{X}) \quad (2)$$

where Z is the normalization term and ϕ_i is the ‘‘association’’ potential which models dependencies between the label of the i^{th} node, y_i , and the set of all (or a part of) observations \mathbf{X} . $\psi_{i,j}$, is the ‘‘interaction’’ potential which represents dependencies between the labels of neighbouring nodes given the observations \mathbf{X} . As it can be observed from the above expression, an important difference between the prior distribution in MRF and the ‘‘interaction’’ potential in CRF is the incorporation of observations for modeling the neighbourhood interactions of labels. This implies that CRF can optimally learn the interaction of adjacent labels based on their observations. One immediate advantage of this is to modify the smoothness among labels by taking into account the observed data and therefore prevent over smoothing [5]. Moreover, we notice that observations are not considered independent in the ‘‘association’’ term, in contrast to the conditional independent assumption for observations that is usually made in MRF models. In the following section we elaborate on how we model ϕ_i and $\psi_{i,j}$ for the problem of enhanced lesion detection.

2.1 Association Potential

Similar to previous work [5,7] we use LDA to capture the interactions of the label of each instance and the observations. Even though ϕ_i can be a function of all the observations in principle, to keep the model simple, we assume that only the observations at each instance are included in the association potential [7,8]. Therefore, the association potential can be written as:

$$\phi_i(y_i = 1|\mathbf{x}_i) = \sigma(\mathbf{w}^T \mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x}_i)} \quad (3)$$

where \mathbf{w} is the ‘‘association’’ parameter vector which is obtained in the training phase (Section 2.3).

2.2 Interaction Potential

The “interaction” potential in CRF allows for learning the relationships between the neighbouring labels given the observed data. Accordingly, we define observation dependent terms for adjacent instances and use a set of parameters to modulate the coupling between the labels of adjacent instances as well as their observed data. In particular, we use the following expression for the interaction potential:

$$\psi_{i,j}(y_i, y_j | \mathbf{x}) = \exp(y_i, y_j \theta^T \mathbf{F}_{i,j}) \quad (4)$$

where θ is the “interaction” parameter vector obtained in the training phase (Section 2.3). $\mathbf{F}_{i,j} = [f(\mathbf{x}_i, \mathbf{x}_j) \ 1]$ consists of a data dependent term, $f(\mathbf{x}_i, \mathbf{x}_j)$, as well as a constant. The data-dependent term computes the similarity of adjacent observations to modulate the smoothness over neighbouring instances accordingly. Moreover, the constant term takes into account the interaction of adjacent labels.

2.3 Parameter Learning and Inference

Simultaneously determining the optimal parameters of the “association” potential and the “interaction” potential can be performed numerically as a convex optimization problem. The parameters of the model $\Theta = \{\mathbf{w}, \theta\}$ are learned in the training phase within a maximum likelihood framework as follows:

$$L(\Theta) = \log p(\mathbf{Y} | \mathbf{X}, \Theta) \quad (5)$$

$$= \sum_{m=1}^M [(\sum_{i \in V} \phi_i(y_i | \mathbf{X}^m) + \sum_{i,j \in E} \psi_{i,j}(y_i, y_j | \mathbf{X}^m)) - \log Z_{\mathbf{x}^m}] \quad (6)$$

$$\Theta^* = \arg \max_{\Theta} L(\Theta) \quad (7)$$

with M being the total number of training sets. Note that since the observation is not modeled, the normalization term is a function of observations X^m for each training set m . Once the parameters are learned during training, the problem of classification is to find the most probable set of labels \mathbf{Y}^* .

$$\mathbf{Y}^* = \arg \max_{\mathbf{Y}} p(\mathbf{Y} | \mathbf{X}, \Theta) \quad (8)$$

The inference problem can be solved with different approaches (e.g. iterated conditional modes, graph cuts, belief propagation technique). We adapted a loopy belief propagation algorithm [9] to solve this problem.

3 Experiments and Results

3.1 Data Pre-processing

We evaluate the performance of the CRF classifier on a multi-centre clinical data set consisting of 25 multidimensional MRI volumes. This data set is acquired from patients with relapsing remitting (RR) MS, all having a heavy load of Gad-enhancing lesions (9 on average). Each acquisition is composed of three

sequences: T1w (before and after contrast) T2w and FLAIR. All volumes are 3mm thick axial slices with $1\text{mm} \times 1\text{mm}$ intra-plane resolution. For each patient, the Gad-enhancing lesions were manually labeled by 3 experts resulting in a silver standard where the “ground truth” lesion pixels result from a consensus among the experts. Prior to classification, MRI data needs to be preprocessed in order to correct for imaging artifacts. These pre-processing steps include bias-field inhomogeneity correction using N3 [10] as well as removal of non-brain regions from the MRI [11]. Furthermore, the intra-subject sequences are registered to their T1-w space [12] and the intensity range of all sequences is normalized [13].

The CRF classifier is trained on 5 MRI volumes and tested on the remaining 20 and its performance is compared against LDA and MRF classifiers. The LDA classifier is defined as in Eq 3 which only includes the “association” potential (i.e. no spatial dependency is captured). Our MRF classifier is formulated similar to CRF (Eq 2) where the only difference is that the “interaction” potential (Eq 4) is modified to be only dependent on the labels (i.e., $f(\mathbf{x}_i, \mathbf{x}_j)$ is removed). With this definition, we highlight the improvements of CRF which are gained due to incorporation of the data dependent term.

3.2 Qualitative Results

Figure 1 illustrates two qualitative examples of the classifier results. In each row we show: (a) a T1c slice along with the classification results of the three classifiers: (b) LDA, (c) MRF and (d) CRF. In each case, green and red colors represent True Positives (TP) and FPs, respectively. The selected slice in the first row has enhanced structures that are both lesions and non-lesions while the example in the second row only has a heavy load of non-lesion enhanced structures. As it can be seen, all three classifiers were able to capture enhanced lesions (TP). However, CRF outperforms LDA and MRF in successfully deleting all the FPs by using the neighbourhood intensity information (in addition to the label information considered by MRF).

3.3 Quantitative Results

The performance of CRF is quantitatively evaluated against LDA and MRF. To that end, the TPs and FPs are defined as follows: if more than 50% of a lesion is captured, the region is counted as a TP, and any falsely detected region with area greater than two pixels is considered a FP. The quantitative results showed that all 141 Gad-enhancing lesions in the cohort of 20 test volumes were detected by all three classification methods. The absolute number of FPs per each patient is depicted in Figure 2(a). Comparing the results, we note that CRF outputs significantly less FPs. In fact, the FP rate in the CRF results is reduced by an average factor of 5.8 and 1.6 upon comparing to LDA and MRF, respectively. It should be noted that many of the FPs, especially in the CRF and MRF results are regions with only two pixels. Due to the presence of Gad-enhancing lesions with an area of two pixels, no further post-processing was applied to the classification results to eliminate FPs with small areas. Moreover, comparison of the dice coefficient (Figure 2(b)) for the three classifiers suggests

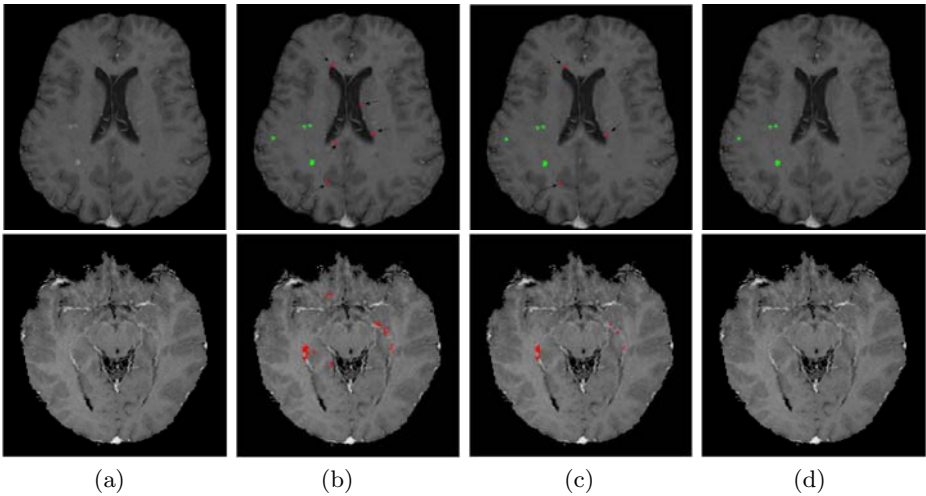


Fig. 1. Two examples of the classification results. (a) T1w image, (b), (c) and (d) classification results for LDA, MRF and CRF respectively. TPs and FPs are shown in green and red, respectively.

an improvement by a factor of 2.1 and 1.3 over LDA and MRF, respectively. However, it should be noted that since Gad-enhancing lesions are very sparse and small, dice coefficient is not suited as a measure of choice for evaluation. This is observed in cases 3, 4, 6, 9 and 17 where low values are observed even though the algorithm performs acceptably in terms of TP and FP. Therefore, TP and FP statistics present a more reasonable measure on classifier performance.

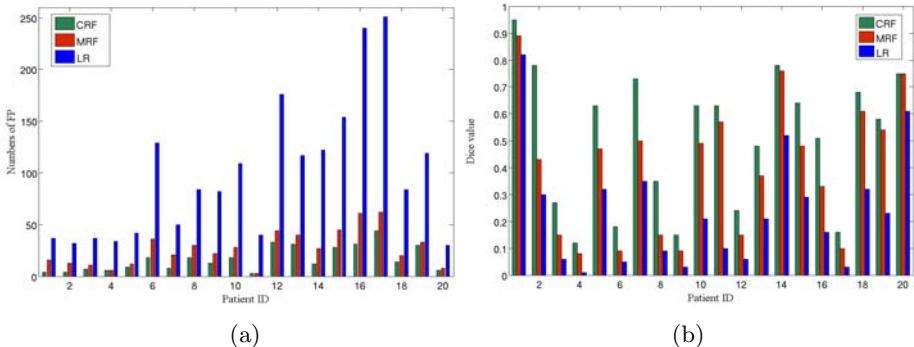


Fig. 2. (a) Numbers of FPs and (b) Dice value for the 20 patients in all three methods

4 Discussion

In this paper we propose a new approach to address the problem of enhanced lesion segmentation in MRI. Our framework uses a CRF-based classification technique to model the spatial relations of the data in a principled manner. Unlike commonly used MRF approaches, where only spatial dependencies of adjacent labels are modeled, the CRF classifier learns the interactions among neighbouring

instances by simultaneously considering their labels and observations. Our classifier is validated on 20 multi-centre real clinical data set from RR MS patients with varying loads of Gad-enhancing lesions. Since there are many non-lesion enhanced structures in the T1c image, excluding FPs is of critical importance in this problem. The experimental results show the advantage of incorporating data spatial dependencies particularly to reduce the number of FPs by an average factor of 1.6 compared to MRF. However, in the current model, both “interaction” and “association” potentials are assumed to have simplified linear models. In the future, we will exploit more sophisticated models to increase the discrimination power of the classifier, by considering shape-based neighbourhood information (inspired by the fact that most vascular structures have line-like shapes). Furthermore, we would also like to investigate the performance of the CRF classifier for the more general problem of lesion segmentation in T2w MRI.

References

1. Bedell, B., Narayana, P.: Automatic segmentation of Gadolinium-enhanced multiple sclerosis lesions. *Magn. Reson. Med* 39, 935–940 (1998)
2. Miki, Y., Grossman, R., Udupa, J., Samarasekera, S., van Buchem, M., Cooney, B., Pollack, S., Kolson, D., Constantinescu, C., Polansky, M., Manon, L.J.: Computer-Assisted Quantitation of Enhancing Lesions in Multiple Sclerosis: Correlation with Clinical Classification. *American Journal of Neuroradiology* 18, 705–710 (1997)
3. He, R., Narayana, P.: Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. *Med. Phys.* 29, 1536–1546 (2002)
4. Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: 18th Int. Conf. on Machine Learning, pp. 282–289 (2001)
5. Kumar, S., Hebert, M.: Discriminative Random Fields. *International Journal of Computer Vision*, 179–201 (2006)
6. Lee, C., Greiner, R., Schmidt, M.: Support vector random fields for spatial classification. In: Jorge, A.M., Torgo, L., Brazdil, P.B., Camacho, R., Gama, J. (eds.) PKDD 2005. LNCS (LNAI), vol. 3721, pp. 121–132. Springer, Heidelberg (2005)
7. Lee, C., Wang, S., Murtha, A., Brown, M., Greiner, R.: Segmenting Brain Tumors Using Pseudo-Conditional Random Fields. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 359–366. Springer, Heidelberg (2008)
8. Plath, N., Toussaint, M., Nakajima, S.: Multi-class image segmentation using Conditional Random Fields and Global Classification. In: *Int. Conf. on Machine Learning* (2009)
9. Murphy, K., Weiss, Y., Jordan, M.: Loopy Belief Propagation for Approximate Inference: An Empirical Study. In: Foo, N.Y. (ed.) *AI 1999*. LNCS, vol. 1747, pp. 467–475. Springer, Heidelberg (1999)
10. Sled, J., Pike, B.: Correction for b(1) and b(0) variations in quantitative t(2) measurements using MRI. *Mag. Reson. Med.* 43, 589–593 (2000)
11. Smity, S.: Fast Robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155 (2002)
12. Collins, D., Neelin, P., Peters, T., Evans, A.: Automatic 3d intersubject registration of MR volumetric data in standardised Talairach space. *Comp. Assisted Tomography* 18, 192–205 (1994)
13. Nyul, L., Udupa, J.: On standardizing the MR Image Intensity Scale. *Comp. Assisted Tomography* 42, 1072–1081 (1999)

Automated Sulci Identification via Intrinsic Modeling of Cortical Anatomy

Yonggang Shi¹, Bo Sun², Rongjie Lai³, Ivo Dinov¹, and Arthur W. Toga^{1,*}

¹ Lab of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA, USA

² Shandong University School of Medicine, Jinan, Shandong, China

³ Department of Mathematics, UCLA, Los Angeles, CA, USA

yshi@loni.ucla.edu

Abstract. In this paper we propose a novel and robust system for the automated identification of major sulci on cortical surfaces. Using multiscale representation and intrinsic surface mapping, our system encodes anatomical priors in manually traced sulcal lines with an intrinsic atlas of major sulci. This allows the computation of both individual and joint likelihood of sulcal lines for their automatic identification on cortical surfaces. By modeling sulcal anatomy with intrinsic geometry, our system is invariant to pose differences and robust across populations and surface extraction methods. In our experiments, we present quantitative validations on twelve major sulci to show the excellent agreement of our results with manually traced curves. We also demonstrate the robustness of our system by successfully applying an atlas of Chinese population to identify sulci on Caucasian brains of different age groups, and surfaces extracted by three popular software tools.

1 Introduction

The automated identification of major sulci on the human cortex is a challenging problem with important applications in brain mapping [1]. While their form and location can vary quite significantly, there is no difficulty for an anatomist to observe the regularity of major sulci based simply on the *geometry* of cortical surfaces regardless of their size, orientation, and the software used to extract them from MRI images. From an engineering point of view, this simplicity is critical for a computational system to achieve the same level of robustness, which essentially lies in its ability of modeling sulcal anatomy with the geometry of cortical surfaces. To this end, we propose in this work a novel system for automated sulci identification by integrating local geometry, i.e., curvature, with global descriptors derived from the Laplace-Beltrami eigen-system [2,3,4] and intrinsic surface mapping [5].

* This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813 entitled Center for Computational Biology (CCB). Information on the National Centers for Biomedical Computing can be obtained from <http://nihroadmap.nih.gov/bioinformatics>.

At the core of our system is an intrinsic atlas of major sulci that represents prior knowledge in training data and enables the integration of curvature information into major sulci. This atlas-based approach is most related to learning-based methods for sulci identification in previous work [6,7,8,9,10]. Principal component analysis were used to model sulcal basins [6] and sulcal lines on the sphere [7]. Boosting methods were used in [9,10]. The Markovian relation of multiple sulci was incorporated with graphical models [8,10]. The main novelty in our system is that the modeling of cortical anatomy is based entirely on intrinsic geometry, which eliminates the use of coordinates in conventional MRI atlases such as the Talairach atlas. This makes our system robust to pose differences and variations across populations. We demonstrate this robustness by using an atlas constructed from a Chinese population to detect twelve major sulci on Caucasian brains of different age groups and surfaces extracted by different software tools.

The rest of the paper is organized as follows. In section 2, we describe the construction of the intrinsic atlas for the modeling of sulcal anatomy. The sulci identification algorithm based on this atlas is developed in section 3. Experimental results are presented in section 4. Finally conclusions are made in section 5.

2 Atlas Construction

As illustrated in Fig. 1, there are two main steps in the construction of the atlas. In the first step, a multiscale representation of the cortical surface is constructed. In the second step, surface maps at a selected scale are computed to bring manually traced sulcal lines to an atlas surface. We describe the two steps in detail next.

2.1 Multiscale Surface Representation

Let $\mathcal{M} = (\mathcal{V}, \mathcal{T})$ denote the triangular mesh representation of a cortical surface, where \mathcal{V} and \mathcal{T} are the set of vertices and triangles. We construct the multiscale representation of \mathcal{M} by using the eigen-system of its Laplace-Beltrami operator $\Delta_{\mathcal{M}}$, which is defined as

$$\Delta_{\mathcal{M}} f = -\lambda f. \quad (1)$$

The spectra of $\Delta_{\mathcal{M}}$ is discrete and we denote the eigenvalues as $\lambda_0 \leq \lambda_1 \leq \dots$ and the corresponding eigenfunctions as f_0, f_1, \dots . To numerically compute the eigenfunctions, we use the finite element method and solve a generalized matrix eigenvalue problem [2,3,4].

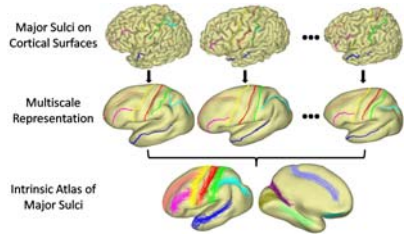


Fig. 1. Atlas construction

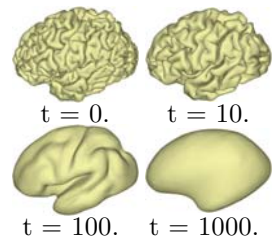


Fig. 2. Multiscale representation of a cortical surface

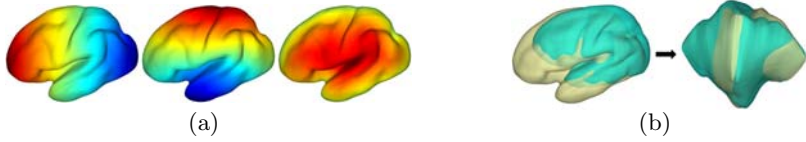


Fig. 3. (a) Intrinsic feature functions. (b) Embedding in the feature space.

Let $X(\cdot, 0) : \mathcal{M} \rightarrow \mathbb{R}^3$ denote the coordinate function on \mathcal{M} , i.e., $X(\mathbf{p}, 0) = \mathbf{p}$ for $\mathbf{p} \in \mathcal{M}$. Using the eigen-system, we can express the heat diffusion of the coordinate function as

$$X(\mathbf{p}, t) = \sum_{n=0}^{\infty} e^{-\lambda_n t} f_n(\mathbf{p}) \int_{\mathcal{M}} f_n(\mathbf{q}) X(\mathbf{q}, 0) d\mathcal{M} \quad (2)$$

By replacing the coordinates of the vertices on \mathcal{M} with $X(\cdot, t)$, we have a multi-scale representation of \mathcal{M} . For numerical implementation, we approximate the diffusion with the first 300 eigenfunctions that are computed efficiently with the spectrum shift technique [11]. As an illustration, we show in Fig. 2 the multi-scale representation of a cortical surface at the scale $t = 0, 10, 100, 1000$. With the increase of the scale, we can see the surface exhibits more regularity that is common across population. However, sulcal landmarks on the original surface might be overly distorted in terms of length and angle if t is too large. In our work, we usually choose $t = 100$ as a tradeoff between surface regularity and landmark distortion.

2.2 Atlas Construction via Intrinsic Surface Mapping

To construct the atlas at a selected scale, we extend the intrinsic surface mapping technique developed for sub-cortical surfaces in [5] to cortical surfaces. Given a pair of surfaces \mathcal{M}_1^t and \mathcal{M}_2^t at a scale t , we compute two maps $u_1 : \mathcal{M}_1^t \rightarrow \mathcal{M}_2^t$ and $u_2 : \mathcal{M}_2^t \rightarrow \mathcal{M}_1^t$ by minimizing the following energy:

$$E = \int_{\mathcal{M}_1^t} \left[\sum_{j=1}^3 \alpha_D^j (\xi_1^j - \xi_2^j \circ u_1)^2 + \alpha_{IC} (I - u_2 \circ u_1)^2 + \alpha_H \|J_{u_1}\|^2 \right] d\mathcal{M}_1^t \\ + \int_{\mathcal{M}_2^t} \left[\sum_{j=1}^3 \alpha_D^j (\xi_2^j - \xi_1^j \circ u_2)^2 + \alpha_{IC} (I - u_1 \circ u_2)^2 + \alpha_H \|J_{u_2}\|^2 \right] d\mathcal{M}_2^t \quad (3)$$

where ξ_i^j ($i = 1, 2; j = 1, 2, 3$) are the feature functions defined on the surfaces that characterize their intrinsic geometry. In each integral, the first term penalizes the difference of feature functions, the second term encourages inverse consistency, and the third term uses the harmonic energy [12] for smoothness regularization, where J_{u_1} and J_{u_2} are the Jacobian of the maps. The regularization parameters $\alpha_D, \alpha_{IC}, \alpha_H$ control the weight of different terms.

To model the sulcal anatomy intrinsically, we define three feature functions on cortical surfaces as shown in Fig. 3(a) by using the Reeb graph of the

first and second eigen-function of its Laplace-Beltrami operator [5]. We can see these functions characterize the frontal-posterior, superior-inferior, and medial-lateral profile of the surface intrinsically. By projecting each point $\mathbf{p} \in \mathcal{M}_i^t$ to $(\xi_i^1(\mathbf{p}), \xi_i^2(\mathbf{p}), \xi_i^3(\mathbf{p}))$, we construct an embedding of each surface in the feature space. As shown in Fig. 3(b), the two cortical surfaces align much better in the embedding space, which enables us to find a good initial map between them by simply looking for the nearest point on the other surface. Starting from the initial maps, we iteratively evolve them by solving a pair of PDEs on the surfaces in the gradient descent directions [5].

By choosing one of the surface $\overline{\mathcal{M}}$ in the training set as the atlas surface, we compute the maps from all other surfaces in the training data to this atlas surface $\overline{\mathcal{M}}$. All the manually traced sulcal lines can then be projected onto the atlas surface as shown in the third row of Fig. 1. Since both the multiscale representation and surface mapping are established via intrinsic geometry, we denote the collection of sulcal lines projected onto the atlas surface $\overline{\mathcal{M}}$ as the intrinsic atlas of major sulci.

3 Automated Identification of Major Sulci

Using the atlas of major sulci, we develop in this section an automated system for their identification on cortical surfaces. Given a new surface \mathcal{M} , the skeletal representation of its sulcal region is first computed based on the mean curvature of \mathcal{M} [13] as shown in Fig 4(a). The skeleton is decomposed into a set of branches $B = \{B_1, B_2, \dots\}$, where each branch is a polyline on \mathcal{M} . The multiscale representation \mathcal{M}_t of \mathcal{M} and the map $u : \mathcal{M}_t \rightarrow \overline{\mathcal{M}}$ is then computed. With the map u , the branches are projected onto the atlas surface and denoted as $\hat{B} = \{\hat{B}_1, \hat{B}_2, \dots\}$. Our goal is to extract anatomically consistent sulcal lines from these skeletal branches.

Let $S_i = \{S_i^1, S_i^2, \dots, S_i^J\}$ denote the set of training curves for the i -th sulcus on the atlas surface. Given the prior model S_i , the challenge is how to model the likelihood of skeletal branches in \hat{B} . The difficulty arises from the fact that B contains only partial observations of major sulci as they frequently cross gyral regions. To bridge the gap between the prior model of complete sulci and the partial observation in skeletal branches, we propose below a projection operator to model the likelihood of a branch on a major sulcus. Given a curve segment C on $\overline{\mathcal{M}}$, its projection onto the training data S_i is defined as:

$$P_{S_i}(C) = \{y(\mathbf{x}) \in S_i^j, x \in C \mid \|x - y\| = \min_{z \in S_i^j} \|x - z\|\} \quad (4)$$

where $S_i^j = \arg \min_j d_H(C, S_i^j)$ and d_H is the Hausdorff distance. Given a branch \hat{B}_k and its projection $P_{S_i}(\hat{B}_k)$, we calculate the Hausdorff distance $D_k = d_H(\hat{B}_k, P_{S_i}(\hat{B}_k))$, and the projection ratio $R_k = \|P_{S_i}(\hat{B}_k)\| / \|\hat{B}_k\|$, where $\|\cdot\|$ denotes the length of a curve. Using the distance D_k and projection ratio R_k , we can pick a set of candidate branches as $\hat{N} = \{\hat{B}_k \in \hat{B} \mid D_k \leq THD_1, R_k \geq THD_2\}$ where THD_1 and THD_2 are thresholds which we choose as $15mm$ and 0.5 in

our practice. The candidate branches are plotted together with the training set over the atlas surface in Fig. 4(b).

Using the candidate branches in \hat{N} , we construct a directed graph to generate a set of sample paths on $\overline{\mathcal{M}}$ as candidate curves for the sulcus. Given two nodes \hat{B}_p and \hat{B}_q in \hat{N} , whose points are ordered according to the indices of their projection on S_i , we form a new curve $C_{p,q} = (\hat{B}_p, \hat{B}_q)$ by connecting the end point \hat{B}_p^e of \hat{B}_p with the start point \hat{B}_q^s of \hat{B}_q . Once again we compute the projection of $C_{p,q}$ onto S_i to evaluate the likelihood of B_p and B_q belonging to the same sulcus. Let

$D_{p,q}$ and $R_{p,q}$ denote the distance and projection ratio of $C_{p,q}$. If $R_{p,q} \geq THD_2$, we add an edge from \hat{B}_p to \hat{B}_q and define the weight as $\frac{1}{D_{p,q} \|\hat{B}_p^e - \hat{B}_q^s\|}$, where the distance $\|\hat{B}_p^e - \hat{B}_q^s\|$ is included to encourage the connection of close branches. Starting from any branch without parents, we perform random walks on the graph to generate sample paths on the atlas surface. The probability of taking an edge during any walk is in proportion to its weight.

Let $C_i = \{C_i^1, C_i^2, \dots\}$ be the set of sample paths for the i -th sulcus. For a candidate curve, its distance to the training data is defined as:

$$d(C_i^k; S_i) = \min_j (\bar{d}(C_i^k, S_i^j) + \bar{d}(S_i^j, C_i^k)) \quad (5)$$

where $\bar{d}(\cdot, \cdot)$ is the average distance from points on a curve to the other curve. We also define the ‘‘sulcality’’ of each curve as:

$$Sulcality(C_i^k) = \sum_{\hat{B}_p \subset C_i^k} \|\hat{B}_p\| / \|C_i^k\| \quad (6)$$

which measures how good the path follows the sulcal regions. We then define the likelihood of each curve as

$$\mathcal{L}(C_i^k | S_i) = e^{-d(C_i^k; S_i) / Sulcality(C_i^k)} \quad (7)$$

and choose the detection result as the sample curve in C_i with the maximum likelihood. As an illustration, we show on $\overline{\mathcal{M}}$ the set of sample paths and the path with the maximum likelihood for the superior frontal sulcus in Fig. 4(c). The branches in this path are then connected via a curvature-weighted geodesic on the original surface to obtain the detected sulcus as plotted in Fig. 4(d).

Let C_{i1} and C_{i2} denote the candidate curves of the $i1$ -th and $i2$ -th sulcus, we define the joint distance from a pair of curves $C_{i1}^p \in C_{i1}$ and $C_{i2}^q \in C_{i2}$ to the training curves S_{i1} and S_{i2} as:

$$d(C_{i1}^p, C_{i2}^q; S_{i1}, S_{i2}) = \min_j (\bar{d}(C_{i1}^p, S_{i1}^j) + \bar{d}(S_{i1}^j, C_{i1}^p) + \bar{d}(C_{i2}^q, S_{i2}^j) + \bar{d}(S_{i2}^j, C_{i2}^q)) \quad (8)$$

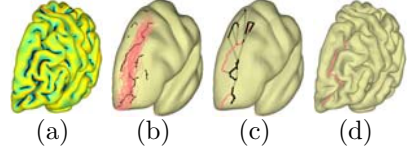


Fig. 4. (a) Skeletal branches of the sulcal region. (b) Candidate branches (black) together with the training curves of the superior frontal sulcus. (c) Sample paths (black) and the most likely path (red). (d) The detected curve on the original surface.

The joint likelihood of the two curves is then defined as:

$$\mathcal{L}(C_{i1}^p, C_{i2}^q | S_{i1}, S_{i2}) = e^{-d(C_{i1}^p, C_{i2}^q; S_{i1}, S_{i2}) / (\text{Sulcality}(C_{i1}^p) * \text{Sulcality}(C_{i2}^q))} \quad (9)$$

The joint detection results are the pair of curves in the sample space $C_{i1} \times C_{i2}$ achieving the maximum likelihood.

4 Experimental Results

In this section, we present experimental results on four different datasets, including surfaces extracted from 3 software tools, to demonstrate our sulci detection system.

4.1 Atlas Construction and Quantitative Validation

In the first experiment, we applied our method to pial surfaces extracted from the 3T MRI images of 65 Chinese young subjects (18 ~ 27 years) by Freesurfer [14]. The left hemispheres were used in this work. Twelve major sulci as listed in Table 1 were manually traced on each surface. We used 40 surfaces as training data to construct the intrinsic atlas of major sulci, which is shown on the third row of Fig. 1, and the algorithm developed in section 3 was used to identify the twelve major sulci on the other 25 surfaces for testing and quantitative validation. For robustness, the central and post-central sulcus were detected jointly by maximizing the joint likelihood in (9). The other 10 sulci were detected separately by maximizing the likelihood in (7).

As an illustration, the results from 4 subjects in the testing data are plotted in Fig. 5. For better visualization, we plotted the detected curves on the surfaces at the scale $t = 10$ in Fig. 5(a) and (b). As shown in Fig. 5(c), the automatically detected sulcal lines align very well with manually traced sulci plotted in black. Quantile statistics of two distances d_{am} and d_{ma} of each sulcus were listed in Table 1, where d_{am} is the distance from points on a detected curve to the corresponding manually traced curve, and d_{ma} denotes the distance from points on each manual curve to the detected sulcus. For example, the first number in the column of

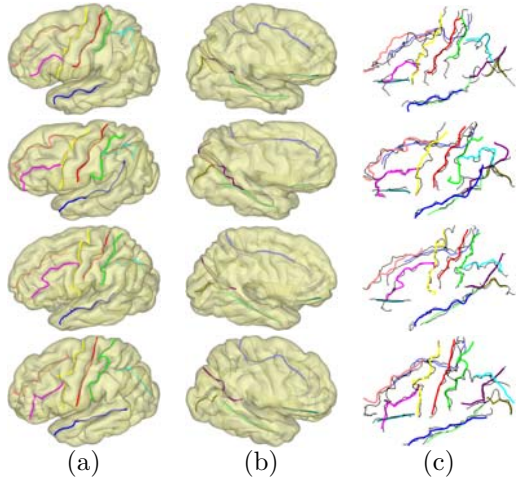


Fig. 5. Detection results on testing data. (a) Lateral view. (b) Medial view. (c) Overlay with manually traced curves (black).

Table 1. Quantile statistics in testing data. (S_1 :central; S_2 :post-central; S_3 :pre-central; S_4 :superior-temporal; S_5 :intraparietal; S_6 :inferior-frontal; S_7 :superior-frontal; S_8 :olfactory; S_9 :collateral; S_{10} :parietal-occipital; S_{11} :cingulate; S_{12} :calcarine.)

		S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
d_{am} (mm)	70%	3.2	5.1	3.6	3.6	3.5	5.2	3.6	3.0	4.7	1.7	3.7	1.7
	80%	3.6	8.2	4.8	4.8	4.6	8.2	5.6	3.3	6.6	2.3	5.1	2.1
	90%	5.2	11.1	7.7	8.4	8.0	10.2	8.7	4.2	8.9	4.4	7.7	3.8
d_{ma} (mm)	70%	3.6	6.6	5.5	4.2	4.4	8.0	4.7	3.3	5.7	2.2	4.6	3.9
	80%	4.6	9.3	7.8	6.4	6.5	9.8	6.9	3.7	7.7	3.4	6.9	7.3
	90%	7.1	12.3	11.2	10.6	10.1	11.7	10.2	5.1	11.2	6.0	9.2	11.5

S_1 means that 70% of the points on the automatically identified central sulcus are within a distance of 3.2mm to the manually traced curve. While there is variability across different sulci, we can see the quantile statistics show the automated results accurately capture the main body of the major sulci.

4.2 Robustness

In the second experiment, we applied the atlas built from the Chinese population in the first experiment to detect sulci on three different datasets of Caucasian brains. The first dataset consists of the left pial surfaces of eight elderly subjects (63 ~ 85 years) extracted by Freesurfer. The second dataset is composed of the left pial surfaces of two young adults extracted by BrainSuite [15]. The third dataset consists of white matter surfaces of two young adults extracted by BrainVisa [16]. For

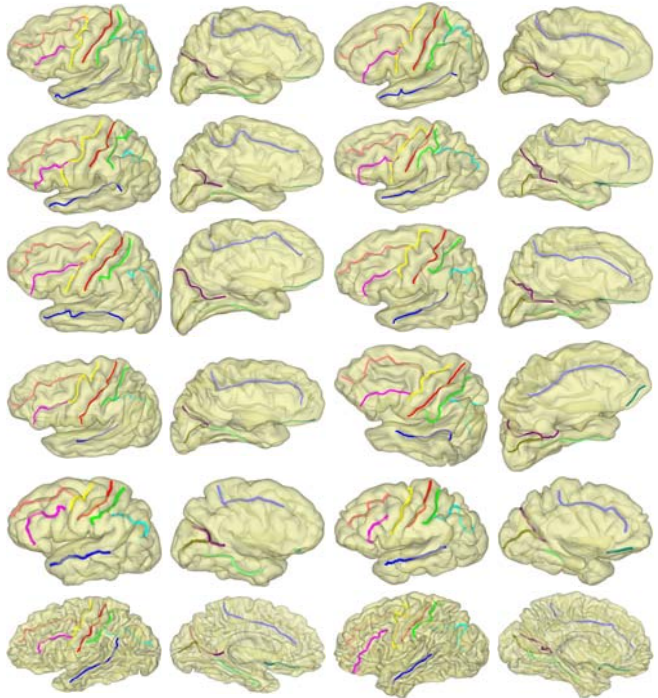


Fig. 6. Row 1-4: results on elderly subjects. Row 5: results on BrainSuite surfaces. Row 6: results on BrainVisa surfaces.

better visualization, we also plot the detected curves on all surfaces at the scale $t = 10$ in Fig. 6. These results demonstrate the robustness of our method across ethnic and age groups, and different software tools for surface extraction.

5 Conclusion

In summary, we have developed a novel system for the automated detection of major sulci on cortical surfaces. Quantitative validations on twelve major sulci showed excellent agreement between our system and manual tracing. We also demonstrated the robustness of our system across different populations and surface extraction tools. For future work, we will augment our system with the intrinsic modeling of gyral landmarks to further improve its performance.

References

1. Thompson, P.M., et al.: Mapping cortical change in alzheimers disease, brain development, and schizophrenia. *NeuroImage* 23, S2–S18 (2004)
2. Qiu, A., Bitouk, D., Miller, M.I.: Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace-Beltrami operator. *IEEE Trans. Med. Imag.* 25(10), 1296–1306 (2006)
3. Reuter, M., Wolter, F., Peinecke, N.: Laplace-Beltrami spectra as Shape-DNA of surfaces and solids. *Computer-Aided Design* 38, 342–366 (2006)
4. Shi, Y., et al.: Anisotropic Laplace-Beltrami eigenmaps: Bridging Reeb graphs and skeletons. In: *Proc. MMBIA*, pp. 1–7 (2008)
5. Shi, Y., et al.: Inverse-consistent surface mapping with Laplace-Beltrami eigenfeatures. In: *Proc. IPMI*, pp. 467–478 (2009)
6. Lohmann, G., Cramon, D.: Automatic labelling of the human cortical surface using sulcal basins. *Med. Image. Anal.* 4, 179–188 (2000)
7. Tao, X., Prince, J., Davatzikos, C.: Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain. *IEEE Trans. Med. Imag.* 21(5), 513–524 (2002)
8. Rivière, D., et al.: Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Med. Image. Anal.* 6, 77–92 (2002)
9. Tu, Z., et al.: Automated extraction of the cortical sulci based on a supervised learning approach. *IEEE Trans. Med. Imag.* 26, 541–552 (2007)
10. Shi, Y., et al.: Joint sulcal detection on cortical surfaces with graphical models and boosted priors. *IEEE Trans. Med. Imag.* 28(3), 361–373 (2009)
11. Vallet, B., Lévy, B.: Spectral geometry processing with manifold harmonics. *Computer Graphics Forum.* 27(2), 251–260 (2008)
12. Mémoli, F., Sapiro, G., Osher, S.: Solving variational problems and partial differential equations mapping into general target manifolds. *Journal of Computational Physics* 195(1), 263–292 (2004)
13. Shi, Y., et al.: Hamilton-Jacobi skeleton on cortical surfaces. *IEEE Trans. Med. Imag.* 27(5), 664–673 (2008)
14. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis i: segmentation and surface reconstruction. *NeuroImage* 9, 179–194 (1999)
15. Shattuck, D., Leahy, R.: BrainSuite: An automated cortical surface identification tool. *Med. Image. Anal.* 8(2), 129–142 (2002)
16. Mangin, J.F., et al.: From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision* 5(4), 297–318 (1995)

In Vivo MRI Assessment of Knee Cartilage in the Medial Meniscal Tear Model of Osteoarthritis in Rats

Zhiyong Xie¹, Serguei Liachenko^{2,*}, Ping-Chun Chiao¹, Santos Carvajal-Gonzalez¹, Susan Bove¹, and Thomas Bocan¹

¹ Pfizer Inc, Groton, CT, United States
{Zhiyong.Xie, Ping-Chun.Chiao}@pfizer.com,
{Santos.Carvajal-Gonzalez, Susan.Bove, Thomas.Bocan}@pfizer.com
² NCTR/FDA, Jefferson, AR, United States
serguei.liachenko@fda.hhs.gov

Abstract. We present a new approach for quantifying the degradation of knee cartilage in the medial meniscal tear (MMT) model of osteoarthritis in the rat. A statistical strategy was used to guide the selection of a region of interest (ROI) from the images obtained from a pilot study. We hypothesize that this strategy can be used to localize a region of cartilage most vulnerable to MMT-induced damage. In order to test this hypothesis, a longitudinal study was conducted in which knee cartilage thickness in a pre-selected ROI was monitored for three weeks and comparisons were made between MMT and control rats. We observed a significant decrease in cartilage thickness in MMT rats and a significant increase in cartilage thickness in sham-operated rats as early as one week post surgery when compared to pre-surgery measurements.

Keywords: Osteoarthritis, Preclinical MRI, Cartilage segmentation, Surface registration, Thickness statistics.

1 Introduction

One of the major hallmarks of osteoarthritis (OA) is degradation of the articular cartilage in weight bearing joints [1]. The morphology of the articular cartilage is considered to be the primary marker of OA progression. Magnetic resonance imaging (MRI) has been used to assess the anatomical/functional integrity of the knee structure and has shown great potentials in clinical studies [2][3][4][5][6]. However, in preclinical research where small animals are widely used to study the pathogenesis of OA and evaluate efficacy of drugs [7], the use of MRI has been limited due to the challenges of small animal in-vivo imaging and image analysis. Because of mediocre image contrast and thin cartilage structure relative to the image resolution, the full extent of the cartilage surface and its spatial morphological variation cannot be accurately determined in MR images of knee joints in small animal preclinical imaging without advanced image manipulation.

This paper presents a MR imaging and data analysis method to assess knee cartilage in the rat medial meniscal tear (MMT) model. The MMT model is a surgically-induced

* All of the work described in this paper was performed while Serguei Liachenko was employed by Pfizer Inc. Nothing should be interpreted as an opinion of NCTR/FDA.

model of OA that results in rapid degradation of cartilage [7]. Results from a previous in-house MRI study indicated that the MMT model induced cartilage degeneration in a similar region on the medial tibia for all rats. However, the degeneration was not noticeable if we use the average thickness of whole cartilage as a metric. To maximize the sensitivity in detecting and quantifying MMT-induced morphological changes in articular cartilage, we propose a statistical strategy to guide the selection of a region of interest (ROI) on the medial tibia and hypothesize that this strategy can localize an area of cartilage most vulnerable to MMT-induced damage. To test this hypothesis, we performed a pilot study to identify the ROI based on the MR images of five MMT and six sham-operated rats. The same ROI was then applied to the images collected in a longitudinal study in which MR images of two groups of rats (MMT vs. sham) were scanned before surgery, and again one and three weeks after surgery. In order to determine repeatability, every rat was re-scanned four times in one day. Average thickness of the cartilage inside the predefined ROI was measured blindly for each image. Results show that the proposed method can detect MMT-induced cartilage degeneration as early as one week post surgery.

2 Materials and Methods

2.1 Animals and Experimental Design

All animal handling procedures were carried out in compliance with the NIH Guide for the Care and Use of Laboratory Animals under a protocol approved by the Pfizer Global Research and Development Animal Care and Use Committee.

The first part of this experiment was a pilot study which was used to identify the regions of the cartilage that were most vulnerable to MMT-induced damage. Eleven male Sprague-Dawley rats (415 ± 22 g) were used in the pilot study. Five rats were randomly selected to receive MMT surgery and the other six rats served as a control (sham-operated) group. OA was surgically induced by transection of the medial collateral ligament and medial meniscus of the femoro-tibial joint as described previously [8]. To reduce the bias caused by the pain of the surgery, rats in control group received sham preparation in which the medial collateral ligament was exposed, but not transected. Knee joint MR images were acquired three weeks after the surgery. A ROI was determined based on the thickness statistics of the medial tibial cartilage in these images.

In the longitudinal scan-rescan study, eight male Sprague Dawley rats (MMT $N = 4$, and sham $N = 4$) underwent MMT or sham surgery. MRI was performed prior to surgery and then again one and three weeks post surgery. In order to determine repeatability of the image acquisition, each animal was re-scanned four times during a single imaging session (one day). Rats were taken out of the magnet and removed from the cradle between the scans, their legs were massaged for 2 minutes and they were repositioned back into the cradle and MRI scanner. The average thickness of the cartilage inside the predefined ROI was computed from the acquired images.

2.2 Image Acquisition

Knee joint MRI was performed on a 7T Bruker Biospec scanner equipped with 12 cm ID gradient insert (up to 20 G/cm). Animals were anesthetized using isoflurane

(3% induction with 1.2-1.5% maintenance) in oxygen. Rats were placed in a supine position on heated cradle, their right knee was flexed to an angle of 105° and secured to the animal holder to prevent motion. An actively decoupled, curved quadrature receive-only surface coil (mouse brain coil, Bruker BioSpin) was placed on the knee and the cradle was then placed inside the magnet. RF excitation was delivered using a 72mm ID birdcage volume resonator. An i.v. bolus of Magnevist® (0.4ml/kg), followed by a constant infusion (0.44 ml/kg/hr) was delivered to improve synovial fluid-cartilage delineation. Optimal slice planning was performed using information obtained from quick orthogonal images in the coronal, axial and sagittal planes. High resolution 3D SPGR anatomical images were acquired in the sagittal plane with the following acquisition parameters: TE = 3.6ms, TR = 25 ms, FA = 30° , NA = 6, MTX = $512 \times 170 \times 64$, resolution of $29 \times 116 \times 231 \mu\text{m}$. The left pane of figure 1 shows the representative MR image of the knee joint of a rat.

2.3 Image Analyses

The tibia and tibial cartilage in collected images were segmented automatically using internally developed tools (see below). Segmentation results were reviewed and errors were manually corrected by a trained expert who was blinded to the treatment group. After the segmentation, a cartilage thickness map on the tibial surface from each image was created and mapped to a template tibial surface. A statistical analysis was performed on aligned cartilage thickness maps to detect the regions where the cartilage thickness was significantly different between the MMT and control groups.

The tibia was segmented for two purposes: first, it provided a spatial reference for the location of the cartilage which helped the segmentation of tibial cartilage; second, it was used to align cartilages from different images to identify the regions which were most vulnerable to MMT-induced damage.

Tibia Segmentation. A surface-to-image registration method was used to segment the tibia on the MR image [9]. Briefly, we selected one MR knee joint image as a template and delineated the tibia in the image using a semi-automatic tool [10]. The segmentation was converted to a surface representation of the tibia. The template tibia surface was fitted to other MR images with weighted external force constraints in which weights and forces were determined by gradients and local intensity profiles obtained from images. To accomplish this fitting, edge points obtained from each MR image set were searched along normal directions of the template tibia surface and the local intensity profiles at these points were extracted and compared to those obtained from the template image. Correlation between intensity profiles was used to select the best-matched edge point. Transformations were then computed to minimize the sum of weighted square distance between transformed atlas surface points and their corresponding best-matched edge points; the weights were again determined by the level of correlation between intensity profiles. To stabilize this fitting process, a sequence of transformations consisting of translation, rigid, affine, and multi-level B-splines, was employed to iteratively morph the bone surface of atlas into the individual tibial structure in the MR image. An example of the segmentation result is presented in the right pane of figure 1.

Cartilage segmentation. Various methods have been proposed for cartilage segmentation [11][12]. Our method segments medial tibial cartilage in three sequential steps: cartilage edge identification, tibia evolution, and cartilage labeling. The first step identifies edge pixels of the cartilage from all edge pixels detected by the Canny-Edge algorithm in the whole image. The edge identification is guided by the prior information about the intensity pattern of the cartilage edge and its location relative to the tibia surface. The model is similar to the one described by Kapur et al. [13]. In the second step, the identified edge pixels attract a curve to the cartilage boundary based on a deformable model. Since the tibial cartilage is a thin tissue connected to the tibia, we expand the tibia boundary to the identified cartilage edge pixels rather than evolving a seed inside the cartilage. The final step labels the new included region of the tibia as tibial cartilage. An example of the segmentation result is presented in the right pane of figure 1.

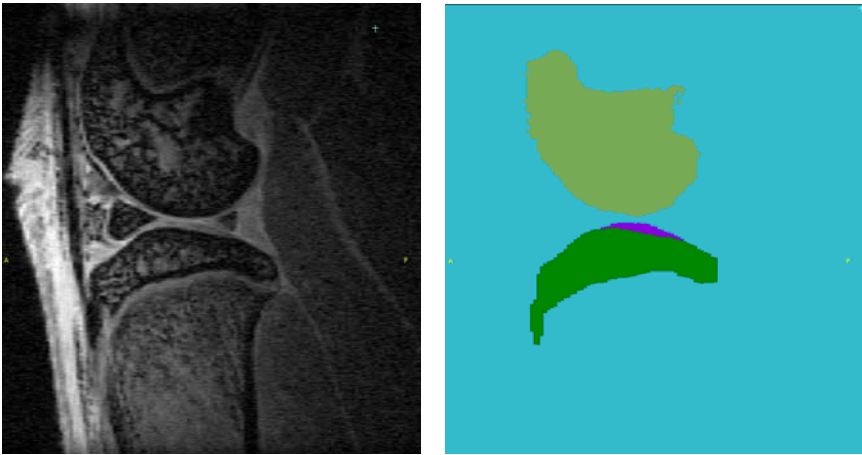


Fig. 1. MRI image of the knee joint (*left*) of a rat and its segmentation of tibia (green), femur (yellow), and tibial cartilage (magenta, *right*)

ROI determination. The region of cartilage most vulnerable to MMT-induced damage was identified using statistical parametric mapping based on the images from 11 rats (MMT $N = 5$, and sham $N = 6$) in the pilot study. After cartilage segmentation, a cartilage thickness map was computed by sampling cartilage thickness estimates on each vertex of the tibial surface. One example of a cartilage thickness map is shown in the left pane of figure 2. All cartilage thickness maps from 11 subjects can be aligned to a template by registering each individual tibial surface to the template tibial surface using the method described by Xie and Farin [14]. Once the thickness map is created and mapped to the template tibial surface, statistical analysis can be performed at each vertex of the tibial surface across the subjects in the same way as proposed by Lerch and Evans or by Worsley et al [15, 16]. The right pane of figure 2 illustrates the procedure of this thickness statistical analysis.

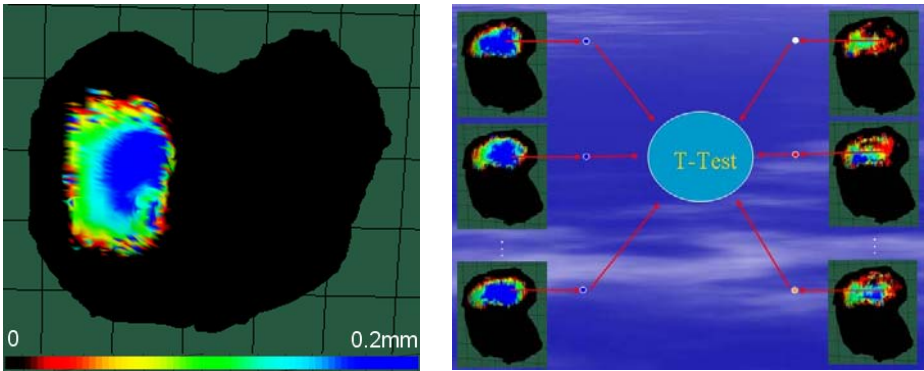


Fig. 2. Cartilage thickness map on the tibial surface (*left*) and thickness statistical analysis (*right*). After aligning all cartilage thickness maps to the same template, cartilage thickness at the same vertex can be compared across the subjects to determine group differences.

Using the thickness statistics, we created a p-value map which shows the significance of the thickness difference between two groups at each vertex. The left picture of figure 3 shows the p-value map on the template tibial surface. The dark region has a p-value less than 0.001, which indicates that the cartilage thickness in these regions is significantly different between the two groups. The dark region in the middle of the cartilage was selected and the ROIs were programmatically delineated for the subsequent analysis of the regional cartilage thickness. The right pane of figure 3 shows the ROI defined on the tibial surface.

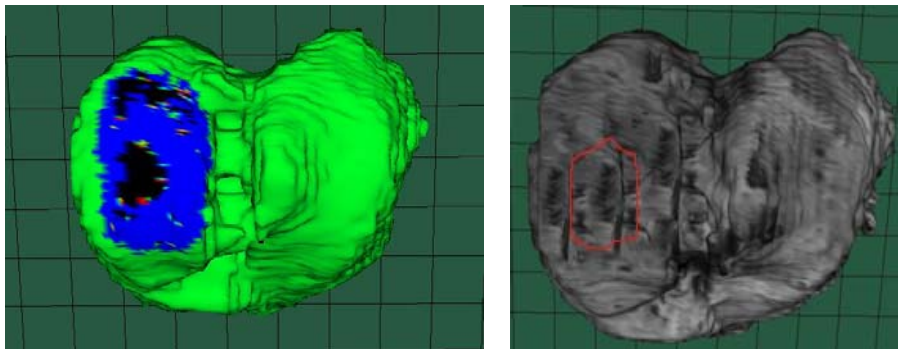


Fig. 3. *Left*, p-value map after thickness statistical analysis between MMT and sham groups. *Right*, defined ROI (region enclosed by the red line) based on thickness statistical analysis.

3 Results

After images in the longitudinal study were collected and segmented, a thickness map was created for each image. The ROI defined from the pilot study was transformed to the thickness map by registering the template tibial surface to the tibial surface of the

subject. The average thickness of the cartilage inside the ROI was computed for every image.

To compare each group, the ANOVA model included terms for Scan, Weeks, Group as fixed effects, and Scan×Weeks×Group as an interaction term. In order to consider the repeated measure, an animal within the scan was introduced as a random effect. Figure 4 shows the least squares mean of cartilage thickness and standard error for each group at each nominal time point (i.e. Weeks 0, 1 and 3).

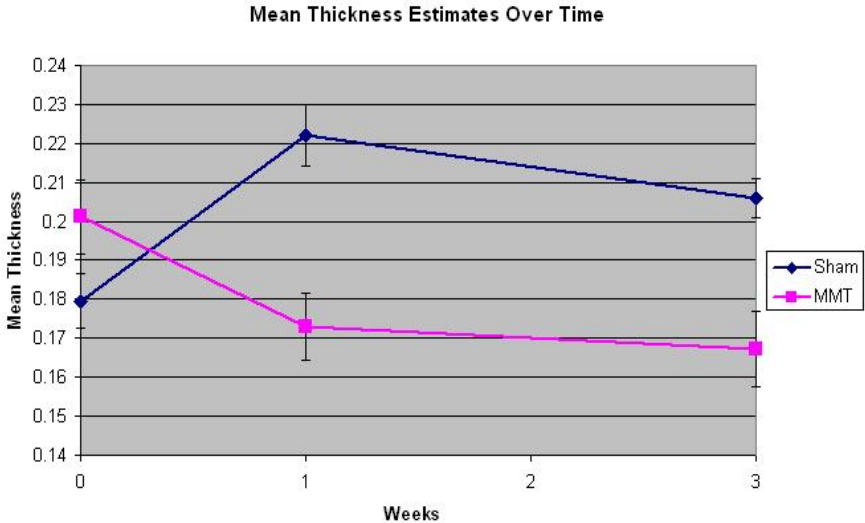


Fig. 4. Least squares means estimates for mean thickness over time and for each group of rats

Based on the statistical analysis, the MMT group had significantly thinner cartilage compared to the sham group at weeks 1 and 3 ($p = 0.0030$ and $p = 0.0181$, respectively). Within the MMT group, the cartilage thickness was significantly decreased at weeks 1 and 3 when compared to the pre-surgery scan at week 0 ($p = 0.0238$ and $p = 0.0073$, respectively). For the sham group, the cartilage thickness was significantly increased at weeks 1 and 3 compared to week 0 ($p = 0.0238$ and $p = 0.0073$, respectively).

The coefficient of variation (CV) of the mean thickness was computed for each rat. For the MMT group the CV varied between 4.87% and 27.53%. For the sham group, the CV was in the range between 4.45% to 20.62%.

To verify that MMT induced damage was located on a similar region as we detected in the pilot study, we performed thickness statistics for the data collected at weeks 1 and 3. Figure 5 shows the p-value map between the MMT and sham groups at week 1 (left) and week 3 (right). We can see the dark region (with p-value less than 0.001) located in a similar region to the one from the pilot study (left pane of the figure 3).

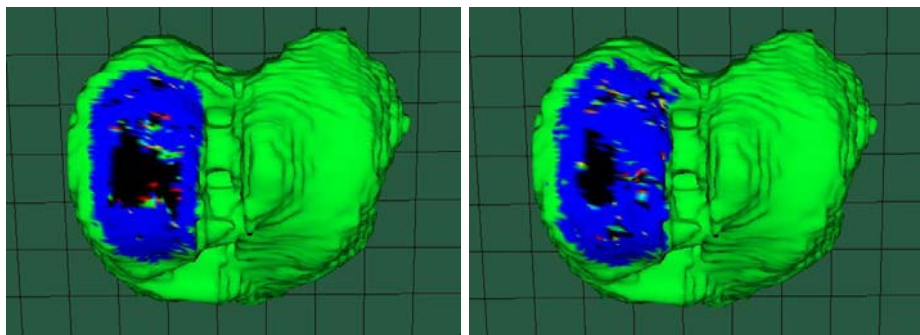


Fig. 5. P-value map generated by statistical comparison between MMT and Sham controls 1 week (*left*) and three weeks (*right*) post surgery

4 Discussion

In this study we demonstrated that MMT surgery induces a consistently located cartilage lesion on the medial tibial plateau of the rat knee joint. We have proposed a method to identify this region based on MR images from a pilot study. The identified region was then used as the region of interest to monitor MMT-induced cartilage degradation in a longitudinal study. Results of the longitudinal study showed for the first time that the proposed method could detect a significant decrease in cartilage thickness as early as one week post surgery, which indicates the increased sensitivity of the proposed approach. Surprisingly, the sham-operated group showed the increase in the cartilage thickness, which may be attributed to several factors. First, the animals used for this study were still growing, and sham data may reflect just this. Second, the surgery by itself may serve as a stimulator of cartilage growth, which means that similar studies should always be performed with sham operated animals as a control group. In our study we showed that the use of sham operation increased the power of statistical analysis as seen in figure 4.

The biggest concern of this study is the repeatability. In the worst case, the CV of the mean thickness of 4 repeat scans was 27.53%. We believe this is due to limited image resolution and contrast to noise ratio. Our in-plane resolution is $29 \times 116 \mu\text{m}$. Considering the average cartilage thickness is $200 \mu\text{m}$, one pixel in the image will account for at least 15% of the cartilage thickness. The situation is even worse if we take into account the big slice thickness which is $231 \mu\text{m}$). A slight difference in imaging position and orientation may cause a significant variation in the final measurement. Higher image resolution is highly desired to improve the repeatability and sensitivity for evaluation of potential protective effects of disease modifying compounds for osteoarthritis in the MMT model.

Theoretically, this approach could be translated to human studies, which will also benefit from the higher resolution (due to larger objects and wide availability of fast parallel imaging techniques) and better contrast in clinical MRI.

References

1. Poole, A.R.: An introduction to the pathophysiology of osteoarthritis. *Frontiers in bioscience* 15(4), 662–670 (1999)
2. Eckstein, F., Reiser, M., Englmeier, K., Putz, R.: In vivo morphometry and functional analysis of human articular cartilage with quantitative magnetic resonance imaging - from image to data, from data to theory. *Anatomy and Embryology* 203(3), 147–173 (2001)
3. Gold, G.E., Beaulieu, C.F.: Future of MR imaging of articular cartilage. *Semin Musculoskelet Radiol.* 5(4), 313–327 (2001)
4. Recht, M.P., Goodwin, D.W., Winalski, C.S., White, L.M.: MRI of Articular Cartilage: Revisiting Current Status and Future Directions. *Am. J. Roentgenol.* 185(4), 899–914 (2005)
5. Lang, P., Noorbakhsh, F., Yoshioka, H.: MR Imaging of Articular Cartilage: Current State and Recent Developments. *Radiol. Clin. N Am.* 43, 629–639 (2005)
6. Peterfy, C., Gold, G., Eckstein, F., Cicuttini, F., Dardzinski, B., Stevens, R.: MRI protocols for whole-organ assessment of the knee in osteoarthritis. *Osteoarthritis and Cartilage* 14, 95–111 (2006)
7. Bendele, A.M.: Animal models of osteoarthritis. *J. Musculoskelet Neuronal Interact.* 1(4), 363–376 (2001)
8. Bove, S., Laemont, K., Brooker, R., Osborn, M., Sanchez, B., Guzman, R., Hook, K., Juneau, P., Connor, J., Kilgore, K.: Surgically induced osteoarthritis in the rat results in the development of both osteoarthritis-like joint pain and secondary hyperalgesia. *Osteoarthritis and Cartilage* 14(10), 1041–1048 (2006)
9. Xie, Z., Tamez-Pena, J., Liachenko, S., Dhamija, S., Gieseg, M., Chiao, P.: Segmentation by surface-to-image registration. In: *Proceedings of SPIE Medical Imaging 2006* (2006)
10. Tamez-Pena, J., Parker, K.J., Totterman, S.: Unsupervised statistical segmentation of multispectral volumetric MR images. In: *Proceedings of SPIE Medical Imaging 1999* (1999)
11. Solloway, S., Taylor, C.J., Hutchinson, C.E., Waterton, J.C.: Quantification of Articular Cartilage from MR Images Using Active Shape Models. In: Buxton, B.F., Cipolla, R. (eds.) *ECCV 1996, Part II. LNCS*, vol. 1064, pp. 400–411. Springer, Heidelberg (1996)
12. Fripp, J., Crozier, S., Warfield, S.K., Ourselin, S.: Automatic Segmentation of Articular Cartilage in Magnetic Resonance Images of the Knee. In: *10th International Conference on Medical Image Computing and Computer Assisted Intervention*, Brisbane, Australia, pp. 186–194 (2007)
13. Kapur, T., Beardsley, P., Gibson, S., Grimson, W., Wells, W.: Model-based segmentation of clinical knee MRI. In: *Proc. IEEE Int'l. Workshop on Model-Based 3D Image Analysis*, pp. 97–106 (1998)
14. Xie, Z., Farin, G.E.: Image Registration Using Hierarchical B-Splines. *IEEE Transactions on Visualization and Computer Graphics* 10(1), 85–94 (2004)
15. Lerch, J.P., Evans, A.C.: Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage* 24(1), 163–173 (2005)
16. Worsley, K.J., Taylor, J.E., Carbonell, F., Chung, M.K., Duerden, E., Bernhardt, B., Lyttelton, O., Boucher, M., Evans, A.C.: SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory. *NeuroImage*, OHBM poster (2009) (accepted)

Construction of Neuroanatomical Shape Complex Atlas from 3D Brain MRI

Ting Chen¹, Anand Rangarajan¹,
Stephan J. Eisenschenk², and Baba C. Vemuri^{1,*}

¹ Department of CISE, University of Florida, Gainesville, FL 32611, USA

² Department of Neurology, University of Florida, Gainesville, FL 32611, USA
{tichen,anand,vemuri}@cise.ufl.edu, stephan.eisenschenk@neurology.ufl.edu

Abstract. This paper proposes a novel technique for constructing a neuroanatomical *shape complex* atlas using an information geometry framework. A shape complex is a collection of shapes in a local neighborhood. We represent the boundary of the entire shape complex using the zero level set of a distance function $S(\mathbf{x})$. The spatial relations between the different anatomical structures constituting the shape complex are captured via the distance transform. We then leverage the well known relationship between the stationary state wave function $\psi(\mathbf{x})$ of the Schrödinger equation $-\hbar^2 \nabla^2 \psi + \psi = 0$ and the eikonal equation $\|\nabla S\| = 1$ satisfied by any distance function $S(\mathbf{x})$. This leads to a one-to-one map between $\psi(\mathbf{x})$ and $S(\mathbf{x})$ and allows for recovery of $S(\mathbf{x})$ from $\psi(\mathbf{x})$ through an explicit mathematical relationship. Since the wave function can be regarded as a square-root density function, we are able to exploit this connection and convert shape complex distance transforms into probability density functions. Furthermore, square-root density functions can be seen as points on a unit hypersphere whose Riemannian structure is fully known. A shape complex atlas is constructed by first computing the Karcher mean $\bar{\psi}(\mathbf{x})$ of the wave functions, followed by an inverse mapping of the estimated mean back to the space of distance transforms in order to realize the atlas. We demonstrate the shape complex atlas computation via a set of experiments on a population of brain MRI scans. We also present modes of variation from the computed atlas for the control population to demonstrate the shape complex variability.

1 Introduction

In the past two decades, human brain MRI analysis has attracted immense attention for the purposes of diagnosis and treatment of neurological diseases. In this context, the construction of neuroanatomical shape atlases of the human brain has been of particular interest and its importance has been emphasized in a number of recent studies [1]. In brief, an atlas provides a reference shape or image for a population of shapes/images which can be useful in numerous applications including but not limited to, statistical analysis of the populations,

* This research is in part supported by the NIH grant RO1 NS046812 to BCV & AR and the NSF grant RI-IIS 0954032 to BCV .

the segmentation of the structures of interest and the detection of the disease regions based on the shape variations between the atlas and the subject etc. Most existing atlases are based on isolated, single anatomical shapes [11,10,12] which do not contain any inter-structural information for example, the spatial relationships among different neighboring structures and the effect of volume shrinkage or expansion of structures in its neighborhood. However, many neurological disorders are diagnosed by the structural abnormalities (e.g. volume change) ascribed to several brain structures rather than a single structure. Mania, which is most often associated with bipolar disorder serves as an example, as in [2] all the brain structures associated with the neural pathways were examined and the authors claimed that the patients with mania have a significant *overall* volume difference in the regions including the thalamus, hippocampi and the amygdala. In [3], Seidman *et al.* concluded that one remarkable vulnerability of schizophrenia, is the structural abnormalities in the thalamus and the amygdala-hippocampus region. Therefore, a neuroanatomical shape complex atlas which captures the structural relationships will be of primary clinical importance.

2 Previous Work

In the context of atlas construction for multiple brain structures, most of the efforts in the past were focussed on building full brain *image* atlases. For instance, in [5,4], several image atlas construction methods for the entire brain were proposed based on 3-D brain MRI. However, it is a nontrivial task to extend these techniques to shape atlas construction. Therefore, we will not be discussing these image based methods any further in this paper and only focus on shape atlas construction instead. One of the most common shape representations in the literature is to represent shapes using feature point sets or landmarks. As in [6], several research articles on point set atlas construction for anatomical structures (e.g. hippocampi) have been published where they model the shapes using mixture of densities and estimate the unbiased shape atlas via information theoretic methods. In [8], Cootes *et al.* developed a diffeomorphic statistical shape model which analyzes the parameters of the deformation field instead of the traditional landmark positions. Other methods that represent shapes in 2D using parametric curves or in 3D using parametric surfaces have also received considerable attention in the literature [7]. Since statistical shape analysis in curve/surface space is very difficult, methods using this representation have usually resorted to computing means etc. of spline parameters. In [9], a characteristic 3D shape model dubbed the m-rep, is proposed and based on this representation, an atlas is constructed via computation of the Karcher mean of the population [10]. Recent work in [11] describes an interesting model for continuous spherical shapes used to analyze the anatomical shape differences in the hippocampus of control group and blind subjects.

To summarize, in all the techniques discussed thus far, the shape atlas is developed only for a single structure. *In this paper, we propose a novel technique for constructing an atlas of a neuroanatomical shape complex of multiple structures.* The novelty lies in the relationship we exploit between the stationary state

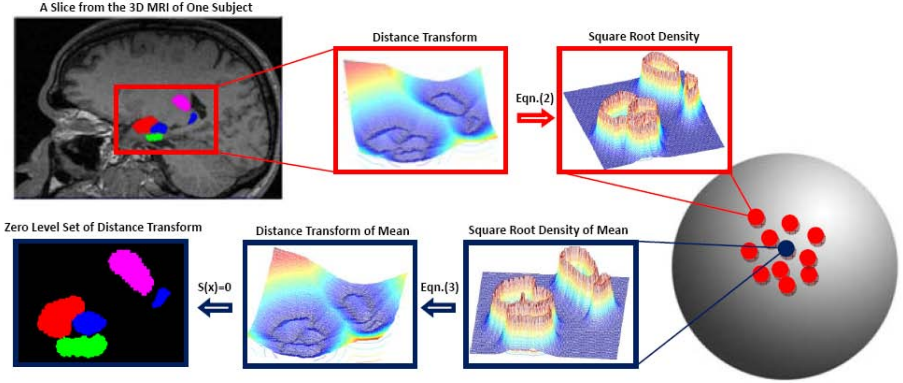


Fig. 1. Illustration of our framework. We visualize the distance transform and square-root density in the 2-D case. Each sample data turns out to be a red point on the high dimensional sphere and the blue point is the Karcher mean.

wave function $\psi(\mathbf{x})$ of the Schrödinger equation $-\hbar^2 \nabla^2 \psi + \psi = 0$ and the eikonal equation $\|\nabla S\| = 1$ for the Euclidean distance transform problem which serves as a “bridge” that connects the distance transform representation of the shape to its square-root-density. The choice of a square-root density representation is motivated by the fact that the manifold of square root densities is a unit Hilbertian sphere and its geometry is well understood. This allows us to use the intrinsic geometry of the sphere to compare shapes represented by square-root densities. Additionally, the inter-structural relationship is well captured in our distance transform representation of the shape complex. In section 4, we demonstrate our technique by presenting examples of atlas construction for the shape complex of 8 structures from a population of 15 3-D brain MRI with all the structures labeled by an expert neurologist.

3 Shape Complex Atlas

3.1 From Distance Transforms to Square-Root Density Functions

In our model, each shape complex data sample is represented by the distance transform function, the zero level set of which gives the individual boundaries of the various shapes constituting the shape complex. At least two decades of effort have gone into level set and distance function representations of shapes [13] - the principal advantage being the ability to combine different shapes into a single scalar field representation. However, since variational and partial differential equation methods are at the foundation of level sets, it is a non-trivial task to employ statistical methods on scalar field distance function representations. Alternatively, there exists a class of methods that perform shape analysis by representing single shapes using probability density functions [6], and obtaining very good results. For instance, despite sacrificing the ability to represent a set of shapes or a shape complex, in this framework, the mean, variance and

principal modes of the shape population are all easily computed. One of the main contributions of this paper is to successfully *bridge* the two disparate domains - variational and level set methods on the one hand and probabilistic methods on the other - and directly obtain the density function of a shape complex from a distance function representation.

In [15], Gurumoorthy and Rangarajan apply the Schrödinger equation to the Euclidean distance transform problem. They solve the Schrödinger wave equation instead of the corresponding static Hamilton-Jacobi equation for the distance transform. While they emphasize the main advantage of their approach to be the linearity of the Schrödinger equation (as opposed to the non-linearity of the Hamilton-Jacobi equation), we wish to draw upon the obvious, historical precedent in quantum mechanics of motivating the Schrödinger wave function as a square-root density [16]. Inspired by this voluminous previous work, we adopt the interpretation of the stationary state Schrödinger wave function for the Euclidean distance function as a square-root density.

Let $\psi(\mathbf{x})$ be the stationary state wave function and let \hbar - Planck's constant - be a free parameter in this model. The static wave equation for the Euclidean distance function problem is¹

$$\hbar^2 \nabla^2 \psi(\mathbf{x}) = \psi(\mathbf{x}). \tag{1}$$

Claim: When $\psi(\mathbf{x}) = \alpha \exp(\frac{-S(\mathbf{x})}{\hbar})$ and satisfies Eqn. (1), $S(\mathbf{x})$ asymptotically satisfies the eikonal equation $\|\nabla S\| = 1$ as $\hbar \rightarrow 0$. Here α is a normalization constant such that $\psi(\mathbf{x})$ is a square-root density, i.e. $\int \psi^2(s) ds = 1$.

Proof: From the definition of a square-root density, $\alpha^2 = \frac{1}{\int \exp(\frac{-2S(\mathbf{x})}{\hbar}) d\mathbf{x}}$, which is a constant for each $S(\mathbf{x})$. Taking the 2-D case as an example, when $\psi(x_1, x_2) = \alpha \exp(\frac{-S(x_1, x_2)}{\hbar})$, we have the second partials for the Laplacian as

$$\begin{aligned} \frac{\partial^2 \psi}{\partial x_1^2} &= \frac{\alpha}{\hbar^2} \exp(\frac{-S}{\hbar}) (\frac{\partial S}{\partial x_1})^2 - \frac{\alpha}{\hbar} \exp(\frac{-S}{\hbar}) \frac{\partial^2 S}{\partial x_1^2} \\ \frac{\partial^2 \psi}{\partial x_2^2} &= \frac{\alpha}{\hbar^2} \exp(\frac{-S}{\hbar}) (\frac{\partial S}{\partial x_2})^2 - \frac{\alpha}{\hbar} \exp(\frac{-S}{\hbar}) \frac{\partial^2 S}{\partial x_2^2}. \end{aligned}$$

From Eqn. (1), we have $(\frac{\partial S}{\partial x_1})^2 + (\frac{\partial S}{\partial x_2})^2 - \hbar(\frac{\partial^2 S}{\partial x_1^2} + \frac{\partial^2 S}{\partial x_2^2}) = 1$ which implies

$$\|\nabla S\|^2 - \hbar \nabla^2 S = 1.$$

Since $\nabla^2 S$ is bounded, we obtain $\|\nabla S\| = 1$ as \hbar goes to 0. □

The derivation above allows us to recover the distance transform function from the square-root density representation by computing the inverse map of

$$\psi(\mathbf{x}) = \alpha \exp(\frac{-S(\mathbf{x})}{\hbar}), \tag{2}$$

that is

$$S(\mathbf{x}) = \hbar \log(\alpha) - \hbar \log(\psi(\mathbf{x})). \tag{3}$$

¹ Please see [15] for a more detailed derivation.

This important relationship builds a direct connection between the two realms, i.e. the level set framework and probability density functions. Hence, a shape complex of complicated topology can be represented using a single distance transform function and further statistical analysis of the shape population can be accomplished in the space of the unit hypersphere as a result of a transition from the distance function to the square-root density representation.

3.2 Space of Square-Root Densities

Note that Eqn. (1) does not entail that the solution is a square-root density. Rather, it merely builds a relationship between exponentiated distance functions and the Schrödinger equation. We further restrict the solution to be in the square-root density space via Eqn. (2) and Eqn. (3). The principal reasons for focusing on the square-root density space rather than the exponentiated function space is as follows.

- Probability density functions are very useful shape representations as shown by several researchers in the literature. For instance, one can compute moments of the density and get global/local shape descriptors [18] while this cannot be done with an un-normalized exponentiated distance function. One can also match either the densities or their moments for the purpose of registration.
- Probability density functions allow us to relate our unknown parameter (Planck’s constant) \hbar to uncertainty.
- Furthermore, the space of exponentiated functions is positive semidefinite while the square-root density space is the hypersphere which leads to a closed form metric (and geodesic) that is efficient to compute.

Due to the fact that the manifold for square-root density functions is a unit sphere in Hilbert space, a variety of Riemannian operations, such as geodesic distance, exponential map and log map [17] are in closed form. Equipped with this basic infrastructure, we are now able to construct an atlas for the shape complex by computing the Karcher mean of the given shape complex population in the space of unit hypersphere. We illustrate the idea of our framework on a simple example in Fig.1. Note that here the notion of atlas corresponds to the mean computed from the the L^2 norm. However, any norm is applicable in our framework, for example, estimating the median of the population via the L^1 norm. As a matter of fact, with the square-root density representation, we are capable of performing many different kinds of statistical analysis.

4 Experimental Results

In this section, we demonstrate the strength of our technique via a set of experiments on a population of real shape complex data. The images are affine registered using an ITK-based mutual information registration algorithm [14]. Before going into the details of the experimental results, we first clarify two empirical issues related to the experiments.

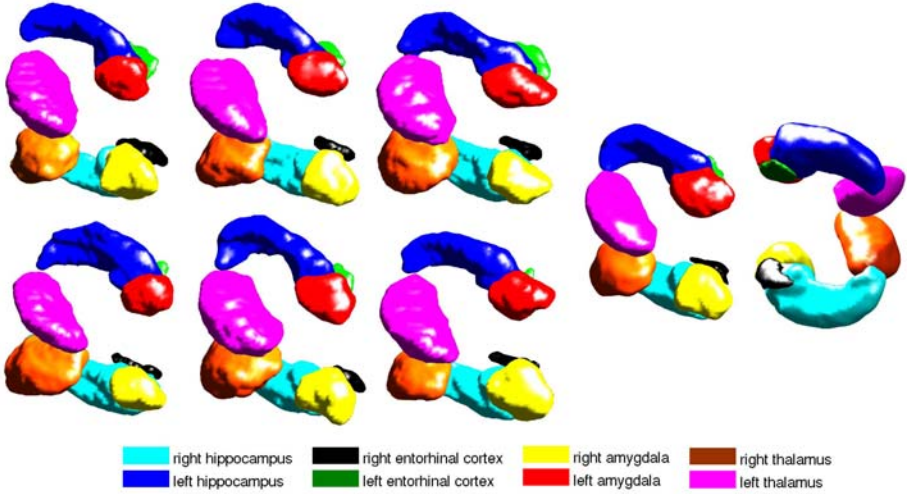


Fig. 2. 6 samples from the group of 15 subjects and the two different views of the atlas with $\bar{h} = 0.4$

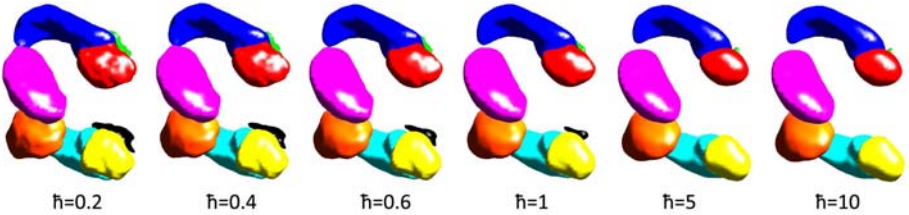


Fig. 3. Atlas corresponding to different \bar{h} values. As \bar{h} increases, the atlas becomes more smooth.

De-normalization of $\bar{\psi}$: Since $\bar{\psi}(\mathbf{x})$ is the geodesic mean on the sphere of the sample square root densities, it is valid to assume that $\psi(\mathbf{x})$ has the same formulation as each shape complex data sample (represented by $\psi_i(\mathbf{x}) = \alpha_i \exp(\frac{-S_i(\mathbf{x})}{\bar{h}}), i = 1, \dots, n$), that is, $\bar{\psi}(\mathbf{x}) = \bar{\alpha} \exp(\frac{-\bar{S}(\mathbf{x})}{\bar{h}})$. Therefore, $\bar{S}(\mathbf{x}) = -\bar{h} \log(\bar{\psi}(\mathbf{x})) + \bar{h} \log(\bar{\alpha})$.

To recover $\bar{S}(\mathbf{x})$, we first need to estimate $\bar{\alpha}$. One approach is to heuristically approximate $\log(\bar{\alpha})$ using the average of $\log(\alpha_1), \dots, \log(\alpha_n)$. Here we describe a more principled solution. Assume $\phi(\mathbf{x}) = \exp(\frac{-S(\mathbf{x})}{\bar{h}})$, is the un-normalized version of ψ (the exponentiated function). Note that we lose one degree of freedom by normalization, hence an extra constraint needs to be imposed. We approximate $\bar{\psi}(\mathbf{x})$ by the linear combination of the un-normalized density ϕ due to the linearity imposed by Eqn. (III). The problem is finally reduced to solve $b^* = \arg \min_{\mathbf{b}} \|\sum_{i=1}^n b_i \phi_i(\mathbf{x}) - \bar{\psi}(\mathbf{x})\|^2$. This assumption is in accordance with the observation that the density has peaks on the locations corresponding to the zero level set of the distance function. The normalization parameter $\bar{\alpha}$ is approximated using $\sum_i b_i^*$ and it scales the un-normalized exponentiated distance

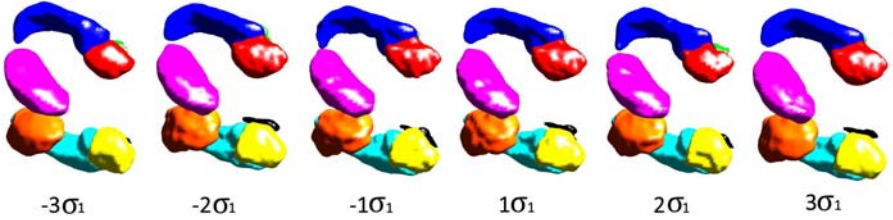


Fig. 4. The shape variation along the first principal direction

function so that we can obtain a zero-level set shape complex atlas. Since this is just one way of obtaining a zero-level set atlas, a numerically more stable flux-based method might be a better future work direction.

Visualization: We transfer the labels of each structure in the shape complex by mapping the label image to our atlas. (This is done only for visualization purposes.) The transformation parameters of the mapping is computed by a non-rigid warping from a binary image of the shape complex template to the binary image estimated from the shape atlas. We leave the automatic labeling of the shape complex atlas for future work.

In Fig. 1, we illustrate the idea of our framework on a simple 2-D example. The 2-D image is taken from one slice of the 3-D MRI of the shape complex. The flowchart shows that we first estimate the distance transform from the shape and then compute the square-root density via Eqn. (2). The shape of the atlas is recovered from the Karcher mean of the densities via Eqn. (3).

We finally apply our proposed framework to the real shape complex data set. The data set contains 15 controls of 3-D brain MRI with the following 8 structures labeled: left/right hippocampus, entorhinal cortex, amygdala and thalamus. We show 6 samples from the group of 15 subjects and the atlas constructed as the mean shape in two different angles of view in Fig. 2. While we have not emphasized this in our presentation, the parameter \hbar acts as a smoothing/regularization term for atlas construction and is expected to act as an uncertainty control - similar to the role played by Planck's constant in physics. We demonstrate the variation of the atlas when different \hbar is used in Fig. 3. As \hbar increases, the atlas becomes more smooth. In this paper, we fix $\hbar = 0.4$ in the experiments. When Principal Geodesic Analysis [10] is applied to our data set, we can recover the modes of deformation and the shape variation along the first principal direction is shown in Fig. 4.

We implemented this method using Matlab® on a 2.33GHZ PC. It takes 4 minutes to construct an atlas from 15 labeled brain MRI with dimension of the ROI being $90 \times 91 \times 87$. This serves to anecdotally illustrate the computational time involved.

5 Conclusions

In this paper, we presented a novel and efficient algorithm that constructs a neuroanatomical atlas for shape complex data with complicated topology. We

derived the relationships between the Euclidean distance transform and the square-root density wave function representation and this successfully builds on a connection between the realms of the level set framework and probability density functions. Our model is not only capable of preserving the spatial relationships among the different structures in the shape complex but also of carrying out a variety of statistical analyses of the shape complex population.

References

1. Yeo, B.T.T., et al.: Effects of Registration Regularization and Atlas Sharpness on Segmentation Accuracy. *Medical Image Analysis* 12(5), 603–615 (2008)
2. Strakowski, S.M., et al.: Brain Magnetic Resonance Imaging of Structural Abnormalities in Bipolar Disorder. *Arch. Gen. Psychiatry* 56(3), 254–260 (1999)
3. Seidman, L.J., et al.: Thalamic and Amygdala Chippocampal Volume Reductions in First-Degree Relatives of Patients with Schizophrenia: An MRI-Based Morphometric Analysis. *Biological Psychiatry* 46(7), 941–954 (1999)
4. Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R.: On the Manifold Structure of the Space of Brain Images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5761, pp. 305–312. Springer, Heidelberg (2009)
5. Sabuncu, M.R., Shenton, M.E., Golland, P.: Joint Registration and Clustering of Images. In: *MICCAI Stat. Reg. Workshop*, pp. 47–54 (2007)
6. Chen, T., et al.: Group-Wise Point-Set Registration Using a Novel CDF-Based Havrda-Charvat Divergence. *IJCV* 86(1), 111–124 (2010)
7. Sebastian, T.B., Crisco, J.J., Klein, P.N., Kimia, B.B.: Constructing 2D Curve Atlases. In: *IEEE Workshop on MMBIA*, pp. 70–77 (2000)
8. Cootes, T.F., Twining, C.J., Babalola, K.O., Taylor, C.J.: Diffeomorphic Statistical Shape Models. *Image and Vision Computing* 26, 326–332 (2008)
9. Styner, M., et al.: Statistical Shape Analysis of Neuroanatomical Structures Based on Medial Models. *Medical Image Analysis* 7, 207–220 (2003)
10. Fletcher, P., Lu, C., Pizer, S., Joshi, S.: Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape. *IEEE Trans. Med. Imaging* 23(8), 995–1005 (2004)
11. Liu, X., et al.: Models of Normal Variation and Local Contrasts in Hippocampal Anatomy. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008*, Part II. LNCS, vol. 5242, pp. 407–415. Springer, Heidelberg (2008)
12. Wang, L., et al.: Abnormalities of Hippocampal Surface Structure on Very Mild Dementia of the Alzheimer Type. *NeuroImage* 30(1), 52–60 (2006)
13. Malladi, R., Sethian, J.A., Vemuri, B.C.: Shape Modeling with Front Propagation: A Level Set Approach. *PAMI* 17(2), 158–175 (1995)
14. Thevenaz, P., Unser, M.: Optimization of Mutual Information for Multiresolution Image Registration. *IEEE Trans. Image Processing* 9, 2083–2099 (2000)
15. Gurumoorthy, K., Rangarajan, A.: A Schrödinger Equation for the Fast Computation of Approximate Euclidean Distance Functions. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) *SSMV 2009*. LNCS, vol. 5567, pp. 100–111. Springer, Heidelberg (2009)
16. Born, M.: Zur Quantenmechanik der Stoßvorgänge. *Zeitschrift für Physik A Hadrons and Nuclei* 37(12), 863–867 (1926)
17. Peter, A., Rangarajan, A., Ho, J.: Shape L'Âne Rouge: Sliding Wavelets for Indexing and Retrieval. In: *CVPR* (2008)
18. Ho, J., Peter, A., Ranganranjan, A., Yang, M.: An Algebraic Approach to Affine Registration of Point Sets. In: *ICCV* (2009)

Non-parametric Iterative Model Constraint Graph min-cut for Automatic Kidney Segmentation

M. Freiman¹, A. Kronman¹, S.J. Esses^{2,3}, L. Joskowicz¹, and J. Sosna³

¹ School of Eng. and Computer Science, The Hebrew Univ. of Jerusalem, Israel

² Mount Sinai School of Medicine, New York, NY

³ Dept. of Radiology, Hadassah Hebrew University Medical Center, Jerusalem, Israel
freiman@cs.huji.ac.il

Abstract. We present a new non-parametric model constraint graph min-cut algorithm for automatic kidney segmentation in CT images. The segmentation is formulated as a maximum a-posteriori estimation of a model-driven Markov random field. A non-parametric hybrid shape and intensity model is treated as a latent variable in the energy functional. The latent model and labeling map that minimize the energy functional are then simultaneously computed with an expectation maximization approach. The main advantages of our method are that it does not assume a fixed parametric prior model, which is subjective to inter-patient variability and registration errors, and that it combines both the model and the image information into a unified graph min-cut based segmentation framework. We evaluated our method on 20 kidneys from 10 CT datasets with and without contrast agent for which ground-truth segmentations were generated by averaging three manual segmentations. Our method yields an average volumetric overlap error of 10.95%, and average symmetric surface distance of 0.79mm. These results indicate that our method is accurate and robust for kidney segmentation.

1 Introduction

Kidney segmentation and volumetric measurement from Computed Tomography (CT) datasets has been proven to be an effective and accurate indicator for renal function in many clinical situations. These include urological treatment decision-making, radiotherapy planning, and estimation of the glomerular filtration rate of living donors [1,2,3]. CT imaging is widely used for kidney analysis and diagnosis since it provides essential anatomical information, including kidney morphology and renal vessel characteristics.

Automatic kidney segmentation is a challenging task. The main factors are unclear borders between the kidney, the liver and the spleen, image acquisition artifacts, image noise, and various pathologies, such as tumors and nephrolithiasis. The interactive graph min-cut based segmentation method provides globally optimal segmentation based on both weighted voxel adjacencies and prior models of the object and the background [4]. However, extensive user interaction

is required to provide estimates of the prior intensity models, and to prevent “segmentation leaks” with predefined spatial constraints. This user interaction limits its routine clinical use for kidney segmentation [5].

A variety of methods have been proposed to incorporate fixed parametric shape information to spatially constrain the graph-min cut optimization. Slabaugh and Unal [6] constrain the optimization to a narrow band of a predefined ellipse. Freedman and Zhang [7] use a level-set representation of the shape prior to compute the probability of each voxel to belong to the object class. Ali et al. [8] use the Poisson distribution and distance maps to compute the shape term of the graph for the segmentation of 2D kidney slices from DCE-MRI. Freiman et al. [9] use a local tubular descriptor to adapt the graph-cut segmentation to vascular structures. All these methods require the explicit formulation of a specific shape model which is inadequate as a general solution for kidney segmentation due to the large inter-patient shape variability. In addition, an interactive initialization is often required to properly position the shape model in the image domain [6,7]. Although Freiman et al. [9] use an automatic initialization, their method is designed specifically for carotid arteries segmentation, and cannot be directly applied to the segmentation of other organs.

Recently, Kumar et al. [10] and Malcolm et al. [11] proposed to use adaptive parametric shape models to constrain the min-cut optimization. In their method, the model parameters are considered as a latent variable, and an iterative approach is used to simultaneously estimate the shape model parameters and to compute the segmentation. The drawback of parametric shape models is that they are less suitable for medical images in which the inter-patient organ shape variability is relatively large, and where various types of pathologies are present [12,13].

In this paper we present a non-parametric global shape constrained graph min-cut approach for the automatic segmentation of kidneys from CT images. Our approach defines a non-parametric hybrid model that couples both shape and patient specific intensity information as a latent variable. An Expectation Maximization (EM) algorithm is then used to simultaneously estimate the latent model and to produce the kidney segmentation. The main advantages of our method are that: 1) it iteratively refines both the shape and the intensity models during the segmentation to increase its accuracy; 2) it uses a non-parametric shape representation that allows weighting of the training instances, and; 3) it uses the graph min-cut framework to perform global optimization on the entire volume in each iteration, as opposed to level-sets [13] or Maximum A-Posteriori (MAP) estimation [12], which perform only local optimizations. Experimental evaluation of our method on 20 kidneys from both contrasted and non-contrasted CT datasets yield an average volumetric overlap error of 10.95% (std: 3.7%).

2 Method

The input of our method is a set of training images $\{I_i\}$ with their corresponding segmentations $\{M_i\}, i \in \{1 \dots N\}$ and the image I to segment. The output is the kidney segmentation M . The EM approach is applied as follows:

1. Initially, each training dataset is registered to the current image.
2. **E-step:** Non-parametric shape and intensity models are computed with adaptive weighting of each training dataset.
3. **M-step:** The kidney segmentation is computed with the graph min-cut technique with edges weights derived from the kidney model.

The kidney model is refined with respect to the current image by iterating over steps 2 and 3 until convergence. Next, we formalize our framework and describe each step of the algorithm in detail.

The graph min-cut approach [4] defines segmentation as a voxel labeling problem. The optimal labeling is defined as the Maximum A-Posteriori (MAP) estimation of a Markov Random Field (MRF) M that minimizes an energy function $E(M)$ incorporating both a prior model and voxel neighborhood information:

$$E(M) = \sum_{\mathbf{x}} \left(\phi(\mathbf{x}|m(\mathbf{x})) + \sum_{\mathbf{y}} \psi(m(\mathbf{x}), m(\mathbf{y})) \right) \quad (1)$$

where \mathbf{x} is the voxel coordinates vector, \mathbf{y} is a neighboring voxel of \mathbf{x} , and $m(\mathbf{x}) \in \{m_{Obj}, m_{Bkg}\}$ are the object and background voxel labels, respectively. The prior model likelihood term $\phi(\mathbf{x}|m(\mathbf{x}))$ is defined as:

$$\phi(\mathbf{x}|m(\mathbf{x})) = \begin{cases} -\log(p(\mathbf{x} \in \text{Obj}|\Theta_{Obj})) & \text{if } m(\mathbf{x}) = m_{Obj}, \\ -\log(p(\mathbf{x} \in \text{Bkg}|\Theta_{Bkg})) & \text{if } m(\mathbf{x}) = m_{Bkg}. \end{cases} \quad (2)$$

where $p(\mathbf{x} \in \text{Obj}|\Theta_{Obj})$ is the likelihood of voxel \mathbf{x} to belong to the object or background classes given a prior model Θ . In the original graph min-cut approach [4], this likelihood term is computed using a fixed prior intensity model, which is usually acquired via user interaction. The voxel neighborhood term $\psi(m(\mathbf{x}), m(\mathbf{y}))$ penalizes neighboring voxels with different labels. It is usually proportional to a contrast-based term that reduces the penalty when neighboring voxels have large intensity differences.

Instead of using a fixed intensity model, we define the prior model Θ as a combination of both shape and intensity information. This prior model is pre-computed from the training datasets. We refine the initial estimate Θ by considering it to be a latent variable in the energy functional, which now takes the form of $E(M, \Theta)$. An EM approach is then used to simultaneously estimate the unknown model Θ and to find the segmentation M [10,11]. We describe the EM steps and the iterative process in detail next.

2.1 E-Step: Non-parametric Hybrid Model Estimation

Given a set of training images $\{I_i\}$ with their corresponding segmentations $\{M_i\}, i \in \{1 \dots N\}$ and a new image to segment I , our goal is to construct a prior model Θ that combines both shape S_{Obj} and intensity I_{Obj} information.

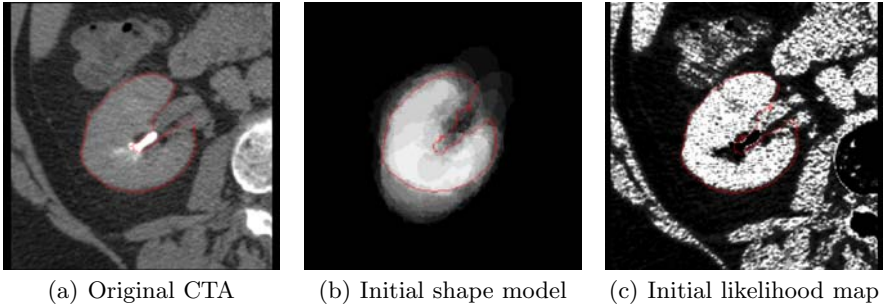


Fig. 1. Model components. The red contour denotes the ground-truth segmentation. (a) Axial slice from the kidney CT dataset to segment, (b) The shape model, and; (c) Intensity based Likelihood image. Bright voxels represent high probability to belong to the object class based on the shape/intensity information. Note that using each component solely, does not provide enough information to obtain accurate segmentation. The presented model components were taken from a first iteration of the algorithm.

The shape information is computed from the object segmentation in the training datasets $\{I_i\}$. The probability of voxel \mathbf{x} to belong to the object class based on the shape model is defined as:

$$p(\mathbf{x}|S_{Obj}) = \frac{1}{N} \sum_{i=1}^N \alpha_i \chi(\mathbf{x}, M_i) \quad (3)$$

where N is the number of training datasets, α_i is a weighting parameter, and the function χ is:

$$\chi(\mathbf{x}, M_i) = \begin{cases} 1 & \text{if } M_i(\Phi_i(\mathbf{x})) \in Obj, \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where Φ_i is a geometrical transformation that maps training dataset I_i to the current image I . The transformations Φ_i are computed using an intensity-based B-Spline registration algorithm [14]. The parameter α_i indicates how close is dataset i to the current patient image I . In the first iteration, α_i is proportional to the Mutual Information (MI) between $M_i(\Phi_i)$ and I . Subsequently, α_i is proportional to the Dice coefficient between current segmentation and the training segmentation M_i .

The intensity model is defined as a patient-specific non-parametric estimation of the Intensity Probability Distribution Function (IPDF), which is estimated using intensity histogram of the voxels in current image I , that belong to the object. In the first iteration, only voxels with high confidence to belong to the object based on the shape model are included in the histogram computation. In subsequent iterations, all the voxels that were labeled as object in the previous iteration are used. The intensity-based background likelihood is set to be its complement, i.e., $1 - p(I(\mathbf{x})|Obj)$. Fig. 1 illustrates the different components of our model.

2.2 M-Step: Graph min-cut MAP-MRF Optimization

Given an estimation of the prior model Θ , the goal now is to compute the MAP-MRF that best model the given image I . The graph min-cut formulation for MAP-MRF estimation is as follows.

Let $G = (V, E)$ be the image graph. Graph nodes $V = \{v_1, \dots, v_n, v_s, v_t\}$ are defined such that node v_x corresponds to voxel \mathbf{x} and terminal nodes v_s and v_t correspond to the object and background classes. Graph edges $E = \{(v_x, v_s), (v_x, v_t), (v_x, v_y)\}$ consist of three groups: 1) edges (v_x, v_s) from voxels to the object terminal node; 2) edges (v_y, v_t) from voxels to the background terminal node, and; 3) edges (v_x, v_y) between adjacent voxels. The cost of a cut that divides the graph into the object class and the background class is defined as the sum of the cut edges' weights. The segmentation is the bipartite graph partition that minimizes the cut cost.

Edge weights are assigned as follows. Edge weights $w(v_x, v_s)$ represent the likelihood of voxel \mathbf{x} to belong to the kidney (object) based on hybrid intensity and geometric model Θ :

$$w(v_x, v_s) = -\log(p(\mathbf{x} \in \text{Obj}|\Theta_{Obj})) = -\log(p(I(\mathbf{x})|I_{Obj}) \cdot p(\mathbf{x}|S_{Obj})) \quad (5)$$

Edge weights $w(v_x, v_t)$ represent the likelihood of each voxel to belong to the background class. We define it as the complement of the object prior model:

$$w(v_x, v_t) = -\log(1 - (p(\mathbf{x} \in \text{Obj}|\Theta_{Obj}))) \quad (6)$$

Edge weights $w(v_x, v_y)$ penalize for nearby voxels \mathbf{x}, \mathbf{y} that have different labels. This edge weight is a combination of the local intensity difference and the spatial location of the edge with respect to the prior shape model S_{Obj} :

$$w(v_x, v_y) = \exp\left(-\frac{(I(\mathbf{x}) - I(\mathbf{y}))^2}{\sigma}\right) \cdot \psi(m(\mathbf{x}), m(\mathbf{y})|S_{Obj}) \quad (7)$$

where σ is a normalization constant that represents the standard deviation of the intensity values inside the object, and $\psi(m(\mathbf{x}), m(\mathbf{y})|S_{Obj})$ is computed using Eq. 3. The minimal cut of the graph is then computed as described in 4.

2.3 The Iterative Solution

The iterative process starts with the computation of the pairwise registration transformations Φ_i that map each training dataset to the current image domain. Since the pairwise registrations are independent, they can be computed in parallel by spanning simultaneous registration processes on computer clusters 15. The method described in 16 can be used as an alternative, to reduce the number of required registrations. Although this initialization step may be computationally intensive, it prevents the accumulation of individual registrations errors that pervade in existing parametric methods 12.

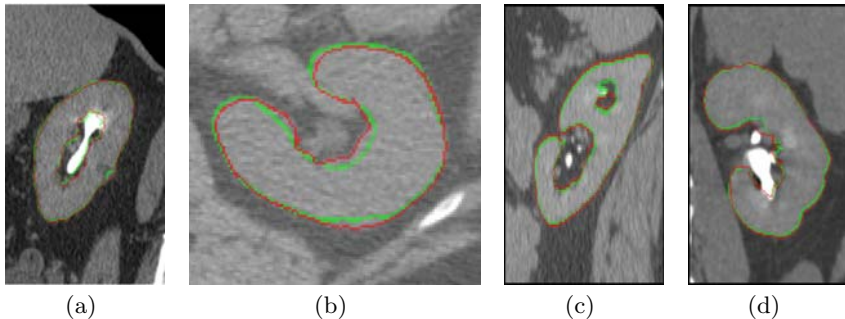


Fig. 2. Representative results from four kidneys with comparison to the ground truth. Our algorithm result (red contour) and the ground truth segmentation (green contour) are overlaid on the CT slices. (a) Coronal view of a left kidney from CTA, (b) Axial view of a right kidney from CT, (c) Sagittal view of a right kidney from CTA, and (d) Coronal view of a right kidney from CTA. Additional images and 3D movies are available on: http://www.cs.huji.ac.il/~freiman/kidney_seg

In the E-step, the shape and intensity model Θ is computed. The shape model is first computed with Eq. 3 and then the intensity model is computed as described in Sec. 2.1. In the M-step, the updated estimation of the segmentation M is computed based on these models. Both steps are iteratively repeated until Θ and M remain stable.

3 Experimental Results

We evaluated our method on 20 kidneys from 10 CT datasets of size $512 \times 512 \times 350 - 500$ voxels, $0.5 - 1.0 \times 0.5 - 1.0 \times 1.0 - 1.5\text{mm}^3$, with and without contrast agent administration. The datasets were acquired on a 64-row CT scanner (Brilliance 64 - Phillips Healthcare, Cleveland, OH) and were chosen randomly from the hospital archive to represent wide variety of patients with different ages and pathologies. Three different observers annotated both the left and right kidney on each dataset. The STAPLE algorithm [17] was used to estimate the ground-truth from the three manual segmentations. In addition we measured the intra-observer performance with respect to the estimated ground-truth.

We evaluated our method using the Leave-One-Out (LOO) approach, where all datasets except the one tested were used for training. Fig. 2 presents representative results of kidney segmentation.

Both volumetric and surface based measures were computed. Table 1 summarizes the average (std) results for the evaluated metrics as compared to the human observers performance. Our method yields an average volumetric overlap error of 10.95% (std: 3.7%) which is better than the previously published error of 17% for semi-automatic kidney segmentation from CT images [2]. The average volumetric overlap error of the observers was 4.67% (std: 0.28%). The average running time using computer cluster to perform the registrations in parallel was 13:54 min (std: 0:27 min). Although our automatic method did not achieve

Table 1. Comparison metrics for kidney segmentation. The first column is the observer number. The second column is the Absolute Volume Difference (AVD) from the ground-truth in %. The third column is the Average Symmetric Surface Distance (ASSD) in mm. The fourth is the Root Mean Square Symmetric Surface Distance (RMS SSD) in mm. The fifth column is the Maximal Symmetric Surface Distance (MSSD) in mm. The sixth column is the Volumetric Overlap Error (VOE) in %. The first row presents the results of our method. The additional three rows present the observers performance compared to the ground-truth.

	AVD (%)		ASSD (mm)		RMS (mm)		MSSD (mm)		VOE (%)	
	mean	std	mean	std	mean	std	mean	std	mean	std
Our	6.2	4.09	0.79	0.34	1.52	0.73	10.46	3.96	10.95	3.7
Observer 1	2.72	2.53	0.32	0.18	0.72	0.28	6.76	2.19	4.88	2.11
Observer 2	2.37	2.13	0.27	0.17	0.66	0.38	5.87	3.31	4.36	2.15
Observer 3	2.45	2.11	0.3	0.18	0.71	0.39	6.29	3.35	4.78	2.39

the human performance, it produces clinically acceptable results for volumetric measurements and surgical planning with no need for user interaction.

4 Conclusion

We have presented a new iterative non-parametric model-based graph min-cut approach for kidney segmentation in CT images. Given a set of training images with their segmentations, our method computes the kidney segmentation based on MAP-MRF estimation of the current image. The algorithm treats the model parameters as a latent variable in a discrete energy functional. An Expectation-Maximization approach is then used to iteratively estimate the model and to obtain the MAP-MRF estimation using the graph min-cut technique. We evaluated the performance of our method on 20 kidneys. Our results show that the proposed method is accurate, robust, easy to use, and provides relevant clinical measurements for many applications. In the future we plan to apply the proposed method to other organ segmentations from various imaging modalities.

The datasets that were used in this work are now publicly available for further evaluation and comparison with other algorithms on: <http://www.cs.huji.ac.il/~caslab/kidneyEval>.

Acknowledgment

This research is supported in part by MAGNETON grant 38652 from the Israeli Ministry of Trade and Industry. M. Freiman is also supported in part by the Hebrew University Hoffman Leadership and Responsibility Fellowship Program.

References

1. Herts, B.R., et al.: Estimating glomerular filtration rate in kidney donors: a model constructed with renal volume measurements from donor CT scans. *Radiology* 252(1), 109–116 (2009)
2. Rao, M., et al.: Comparison of human and automatic segmentations of kidneys from CT images. *Int. J. Rad. Onc. Bio. Phys.* 61(3), 954–960 (2005)
3. Hidas, G., et al.: Estimating relative renal function from relative parenchymal volume—a feasibility study. *J. Endourol.* 22(11), 2527–2530 (2008)
4. Boykov, Y., Funka-Lea, G.: Graph Cuts and Efficient N-D Image Segmentation. *Int. J. of Comp. Vis.* 70(2), 109–131 (2006)
5. Shim, H., et al.: Semiautomated segmentation of kidney from high-resolution multidetector computed tomography images using a graph-cuts technique. *J. Comput. Assist. Tomogr.* 33(6), 893–901 (2009)
6. Slabaugh, G., Unal, G.: Graph cuts segmentation using an elliptical shape prior. In: *IEEE ICIP 2005*, vol. 2, pp. 1222–1225 (2005)
7. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: *IEEE CVPR 2005*, vol. 1, pp. 755–762 (2005)
8. 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)
9. Freiman, M., et al.: Vessels-Cut: a graph based approach to carotid arteries patient-specific modeling. In: Magnenat-Thalmann, N. (ed.) *Modelling the Physiological Human. LNCS*, vol. 5903, pp. 1–12. Springer, Heidelberg (2009)
10. Kumar, M., Torr, P., Zisserman, A.: OBJ CUT. In: *IEEE CVPR 2005*, pp. 18–25 (2005)
11. Malcolm, J., Rathi, Y., Tannenbaum, A.: Graph Cut Segmentation with Nonlinear Shape Priors. In: *IEEE ICIP 2007*, pp. 365–368 (2007)
12. Sabuncu, M.R., et al.: Supervised Nonparametric Image Parcellation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 1075–1083. Springer, Heidelberg (2009)
13. Wimmer, A., Soza, G., Hornegger, J.: A Generic Probabilistic Active Shape Model for Organ Segmentation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 26–33. Springer, Heidelberg (2009)
14. Klein, S., et al.: elastix: a toolbox for intensity based medical image registration. *IEEE Trans. Med. Imaging* 29(1), 196–205 (2010), <http://elastix.isi.uu.nl/>
15. Barak, A., Shiloh, A.: The MOSIX2 Management System for Linux Clusters, Multi-Clusters and Clouds, http://www.mosix.org/pub/MOSIX2_wp.pdf
16. Commowick, O., Malandain, G.: Efficient Selection of the Most Similar Image in a Database for Critical Structures Segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II. LNCS*, vol. 4792, pp. 203–210. Springer, Heidelberg (2007)
17. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23(7), 903–921 (2004)

Synthetic MRI Signal Standardization: Application to Multi-atlas Analysis

Juan Eugenio Iglesias¹, Ivo Dinov², Jaskaran Singh²,
Gregory Tong², and Zhuowen Tu²

¹ Medical Imaging Informatics, University of California, Los Angeles
jeiglesias@ucla.edu

² Laboratory of Neuroimaging, University of California, Los Angeles
kearny89@ucla.edu, gregory.tong@ucla.edu, ivo.dinov@loni.ucla.edu,
zhuowen.tu@loni.ucla.edu

Abstract. From the image analysis perspective, a disadvantage of MRI is the lack of image intensity standardization. Differences in coil sensitivity, pulse sequence and acquisition parameters lead to very different mappings from tissue properties to image intensity levels. This presents challenges for image analysis techniques because the distribution of image intensities for different brain regions can change substantially from scan to scan. Though intensity correction can sometimes alleviate this problem, it fails in more difficult scenarios in which different types of tissue are mapped to similar gray levels in one scan but different intensities in another. Here, we propose using multi-spectral data to create synthetic MRI scans matched to the intensity distribution of a given dataset using a physical model of acquisition. If the multi-spectral data are manually annotated, the labels can be transferred to the synthetic scans to build a dataset-tailored gold standard. The approach was tested on a multi-atlas based hippocampus segmentation framework using a publicly available database, significantly improving the results obtained with other intensity correction methods.

1 Introduction

Magnetic resonance imaging (MRI) is the modality of choice for brain imaging due to its excellent contrast in soft tissue. MRI images are a function of three properties of the tissue: the spin-lattice relaxation time T_1 , the spin-spin relaxation time T_2 , and the proton density ρ [1]. The way in which these physical properties are mapped to image intensities depends heavily on the imaging pulse sequence. The acquisition is typically designed to enhance the contribution of one of the properties and minimize the impact of the other two, leading to the well-known T_1 -weighted (T_1 -w), T_2 -w and ρ -w imaging. Acquisition of real T_1 and T_2 maps [2] is possible but not very extended yet. Automated analysis of T_1 -w, T_2 -w and ρ -w images is difficult because the statistical distribution of voxel intensities can change significantly from scan to scan. Intensity standardization methods can alleviate this problem by trying to match the histogram of one scan

to that of a target volume. However, they cannot completely solve the problem because the image intensity of a scan is a function of three physical properties of the tissue, and the relationship between gray levels in two volumes is in general not one-to-one. A mapping can be found if three or more channels are available for both the source and target images[4], but this is seldom the case.

Meanwhile, multi-atlas segmentation is becoming increasingly popular in brain image analysis[5]. The main idea is to register a number of manually-labeled volumes to a target scan, propagate the annotations with the resulting transforms, and fuse the propagated labels to generate a probabilistic map that can be thresholded at a certain level (typically 0.5) to yield the final segmentation. If the atlases and the target scan have been acquired using different protocols, it is possible to use mutual information (MI) as the metric for the registration and then use a simple fusion scheme in which the mode of the labels at each location is taken[5]. If the image intensities are matched, the segmentation results can be improved by estimating the local success of the registration as the gray level difference between each registered atlas and the target volume at each point, and then using these differences to give a higher weight to the more accurately registered atlases in the label fusion scheme.

In this study, a MRI standardization method based on synthetic MRI is presented. A number of scans for which three channels (T_1 -w, T_2 -w and ρ -w) are available are used to synthesize T_1 -w volumes matched to a specific target dataset. Manual annotations of the hippocampus made on the original T_1 -w scans can be made, propagated to the synthetic volumes and finally used to segment the hippocampus in the target dataset in a multi-atlas setup. The hippocampus was selected as the target of the segmentation because its morphometry is relevant in many diseases (Alzheimers, Parkinsons, schizophrenia...) and, as a consequence, there are publicly available datasets with manual annotations that can be used to validate the technique. The performance of the proposed method in the multi-atlas framework is compared with the cases in which other intensity standardization algorithms are used.

2 Materials and Methods

2.1 Data

Two datasets are used in this study; one for training and one for evaluation. The training dataset consists of 10 volumes acquired with a Siemens 1.5T scanner using two different pulse sequences: 3D MP-RAGE for the T_1 -w scans ($1 \times 1 \times 1 \text{ mm}^3$ resolution, TR/TE/TI = 1,900/4.38/1,100ms., flip angle $\alpha = 15^\circ$) and double-echo spin echo for the T_2 -w and ρ -w (TR/TE1/TE2 = 3,300/17/100ms., 1.4 mm slices with $1 \times 1 \text{ mm}^3$ resolution). The left and right hippocampus were manually delineated in the T_1 -w scans by two trained neuroscientists. These scans play the role of atlases in this study.

Fifteen scans from the publicly available Hippocampus Segmentation Database (HSD, www.radiologyresearch.org/HippocampusSegmentationDatabase) were used with testing purposes. Ten of them belong to patients with temporal

lobe epilepsy and may have atrophic hippocampi, which makes the segmentation harder. These 15 scans were acquired with a 1.5T GE scanner using an inversion recovery spoiled gradient echo sequence (IR-SPGR) with $TR/TE/TI = 7.6/1.7/500$ ms, flip angle $\alpha = 20^\circ$. Two mm coronal slices were acquired with a $0.78 \times 0.78 \text{ mm}^2$ in-plane resolution. The same two neuroscientists who annotated the three-channel dataset also delineated the hippocampus in a subset of coronal slices of 10 scans with inter-reader variability evaluation purposes.

2.2 Synthetic MRI Generation

Estimation of true T_1 , T_2 and ρ : The first step in the generation of synthetic MRI images is to calculate the true physical properties of the tissue i.e. the true T_1 , T_2 and ρ maps of the training dataset. In an ideal situation, a large number of volumes would be acquired with different parameters (typically TE/TR) and then fed to a least-squares or expectation maximization algorithm to compute a robust estimate of the physical values [6]. However, in this study we assume the typical clinical scenario in which the available volumes are T_2 -w and ρ -w scans acquired with a double-echo sequence and a T_1 -w scan acquired with a fast T_1 sequence such as SPGR or MP-RAGE. To minimize the impact of noise on our estimates, the volumes were first denoised with an implementation of the Perona-Malik filter [7]. Then, the scans were skull-stripped with the BET algorithm [8] and subsequently bias field corrected with the N3 method [9]. Finally, the T_1 -w scans were registered to the T_2 and ρ -w scans. The software package ITK (www.itk.org) was used to maximize a MI metric over a rigid transform (i.e. rotation and translation), which suffices to align the scans because they are images from the same patient acquired a few minutes apart.

Now, assuming the pulse sequences for the training dataset (i.e. double-echo spin echo, MP-RAGE), the signal models for the T_2 -w, ρ -w and T_1 -w scans are [1]:

$$S_2(\mathbf{r}) = k_2 \cdot \rho(\mathbf{r}) \cdot (1 - e^{-\frac{TR_2}{T_1(\mathbf{r})}}) e^{-\frac{TE_2}{T_2(\mathbf{r})}} \quad (1)$$

$$S_\rho(\mathbf{r}) = k_\rho \cdot \rho(\mathbf{r}) \cdot (1 - e^{-\frac{TR_\rho}{T_1(\mathbf{r})}}) e^{-\frac{TE_\rho}{T_2(\mathbf{r})}} \quad (2)$$

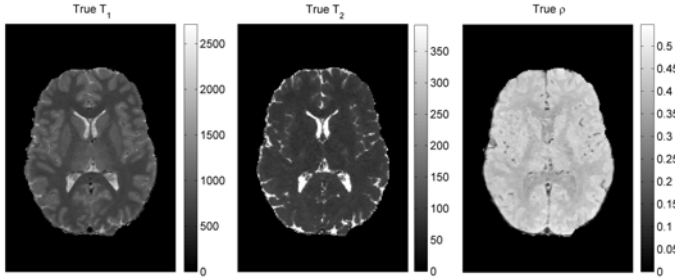
$$S_1(\mathbf{r}) = k_1 \cdot \rho(\mathbf{r}) \cdot f(T_1(\mathbf{r})) \quad (3)$$

where \mathbf{r} is the position vector, TR_1 and $TR_2 = TR_\rho$ are the repetition times and TE_1 , TE_2 and TE_ρ are the echo times. The constants k_1 , k_2 and k_ρ account for global gains due to coil sensitivity, digital signal stretching and other constant multiplicative factors. The function $\rho(\mathbf{r})$ is proportional to the magnetization, which is in turn proportional to the proton density, but also affected by other factors such as flow attenuation. The function $\rho(\mathbf{r})$ is arbitrarily rescaled so that its maximum is one. The function $f(T_1)$, which neglects T_2^* effects, has a complex expression which is not reproduced here; the reader is referred to [10] for the details. Nevertheless, $f(T_1)$ is strictly increasing and hence invertible, which will be useful in the proposed algorithm.

The system of equations can be solved if prior knowledge on the characteristic T_1 and T_2 of some type of tissue is used; otherwise k_1 , k_2 and k_ρ cannot be

Table 1. Gauss-Siedel algorithm to estimate the true T_1 , T_2 and ρ maps

<p>1. Estimate $T_2(\mathbf{r})$ from equations 1 and 2 as $T_2(\mathbf{r}) = (TE_2 - TE_\rho) / \log \frac{k_2 \bar{S}_\rho(\mathbf{r})}{k_\rho S_2(\mathbf{r})}$ where the ratio $\frac{k_2}{k_\rho}$ can be obtained as $\frac{k_2}{k_\rho} = \frac{\bar{S}_{2,WM}}{S_{\rho,WM}} \exp\left(\frac{TE_2 - TE_\rho}{T_{2,WM}}\right)$</p> <p>2. Assuming $T_1(\mathbf{r}) = 0$, use equation 1 to initialize $\rho'(\mathbf{r}) \leftarrow S_2(\mathbf{r}) \exp\left(\frac{TE_2}{T_2(\mathbf{r})}\right)$ $k_2 \leftarrow \max_{\mathbf{r}} \rho'(\mathbf{r})$ and $\rho(\mathbf{r}) \leftarrow \rho'(\mathbf{r}) / k_2$</p> <p>3. Estimate k_1 using the current values of $\rho(\mathbf{r})$: $k_1 \leftarrow \bar{S}_{1,WM} / [f(T_{1,WM}) \cdot \bar{\rho}_{WM}]$ where $\bar{\rho}_{WM}$ is the median value of $\rho(\mathbf{r})$ in the white matter mask.</p> <p>4. Update the real T_1 volume using equation 3: $T_1(\mathbf{r}) \leftarrow f^{-1}(S_1(\mathbf{r}) / k_1 \rho(\mathbf{r}))$</p> <p>5. Update $\rho'(\mathbf{r}) \leftarrow S_2(\mathbf{r}) \exp[TE_2 / T_2(\mathbf{r})] / [1 - \exp(-TR_2 / T_1(\mathbf{r}))]$, as well as $k_2 \leftarrow \max_{\mathbf{r}} \rho'(\mathbf{r})$ and $\rho(\mathbf{r}) \leftarrow \rho'(\mathbf{r}) / k_2$</p> <p>6. Go to step 3 until convergence; seven or eight iterations often suffice. Figure 1 shows the T_1, T_2 and ρ maps for an axial slice of one of the training cases.</p>
--

**Fig. 1.** T_1 , T_2 and ρ maps for an axial slice of one of the cases in the training dataset

estimated. We propose using the white matter, which can be segmented in the T_1 -w scan with the method from [11](#). Even if the segmentation is not perfect, we can use the median value (a robust estimate) of the images in the mask $\bar{S}_{1,WM}$, $\bar{S}_{2,WM}$ and $\bar{S}_{\rho,WM}$ to match the predicted and theoretical values of T_1 and T_2 in the white matter: $T_{1,WM}=785$ ms (at 1.5T), $T_{2,WM}=92$ ms [12](#). The physical parameters are then estimated with the Gauss-Siedel method (Table [1](#)).

Synthesizing dataset-matched volumes: Once the real T_1 , T_2 and ρ maps have been estimated, it is straightforward to compute how the volumes would look like if they had been acquired with the T_1 -w IR-SPGR pulse sequence of the HSD dataset. The approximate signal equation for IR-SPGR is [11](#):

$$S_{HSD} \propto \rho(\mathbf{r}) [1 - 2e^{-\frac{TI}{T_1(\mathbf{r})}} + e^{-\frac{TR}{T_1(\mathbf{r})}}]$$

The intensities can be computed up to a multiplicative constant. The constant is estimated by matching the median intensity of the brain (a very consistent intensity landmark) in the synthesized and test volumes.

2.3 Multi-atlas Segmentation

The multi-atlas segmentation framework from [13] was used in this study. The atlases are registered to the (skull-stripped, bias field corrected) scan and the absolute difference image is calculated for each registered atlas. These difference images are blurred and inverted (previous addition of a small constant ϵ to prevent division by zero) to yield the weight images $\lambda_i(\mathbf{r})$, $i \in \{1, \dots, N_{atlas}\}$, where \mathbf{r} is again the position vector. Then, for each structure s to segment (i.e. the left and right hippocampus), the probability volume is given by:

$$P_s(\mathbf{r}) = \frac{\sum_{i=1}^{N_{atlas}} \lambda_i(\mathbf{r}) L_i(\mathbf{r})}{\sum_{j=1}^{N_{atlas}} \lambda_j(\mathbf{r})} \quad (4)$$

where $L_i(\mathbf{r})$ represents the propagated labels from atlas i (one if inside, zero if outside). The probability volume is blurred and thresholded at 0.5 to generate a binary map, out of which the largest connected component is extracted as the final segmentation. The performance depends on the quality of the weight images, which in turn depends on the image standardization.

3 Experiments and Results

3.1 Experiment Setup

Two experiments were designed in this study. In both of them, the first step was to register and match the intensities of the atlases to the test images. ITK was used to optimize a B-spline nonrigid transform using a MI metric between the T_1 -w volumes. Then, the image intensity of the atlases was matched to that of the test image using our approach (which uses three channels) as well as three other methods (based solely on the T_1 -w volume): 1) histogram stretching, disregarding the top and bottom 1% of the histogram; 2) histogram equalization; and 3) landmark-based matching [3] using eight landmarks (including the median white and gray matter intensities given by the method from [11]).

The first experiment quantizes the influence of the method on the mean joint histogram of the samples and the registered, intensity-corrected atlases. In the second experiment, the impact of the presented approach on a multi-atlas method to segment the hippocampus is assessed. The dependence of the performance on the intensity-matching strategy is analyzed (the registration was the same for all the methods) and compared to the inter-user variability in the manual annotations, which defines the boundary for the performance of the system. In both experiments, the width of the Gaussian kernels to smooth the difference image and the likelihood map was set to $\sigma=1$ mm, whereas the regularization parameter to invert the difference image was set to $\epsilon = 10^{-3}$.

3.2 Results

Joint histogram: The success of an intensity-matching strategy can be evaluated by building the mean joint histogram of the test images and registered,

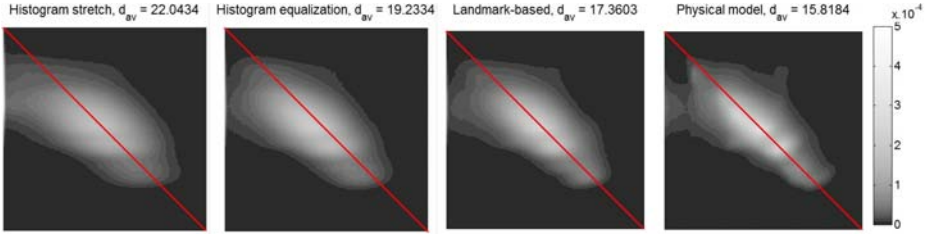


Fig. 2. Mean joint histogram of the images of the HSD dataset (rows) and the corresponding atlases (columns) after registration and intensity matching using different strategies. The color map follows a logarithmic scale. The weighted average distance from each bin to the diagonal is displayed above each joint histogram.

intensity corrected atlases. If the registration and matching were perfect, the histogram would be a diagonal matrix, but in practice the result is a cloud around the diagonal. The weighted average distance from each bin to the diagonal can thus be used as a performance metric. This measure is displayed, along with the histograms, in Figure 2. Our physical model outperforms the landmark-based approach, which in turn outperforms histogram equalization and stretching.

Multi-atlas segmentation: In order to evaluate the performance of the segmentation with an overlap measure (the Dice coefficient in this study), it is important to analyze the impact of the inter-reader variability on the measure first. For that purpose, the neuroscientists who annotated the scans in this study also delineated the hippocampus in a subset of slices of the test dataset (see section 2.1). The neuroscientists were not shown the ground truth annotations from the test dataset in order to prevent bias in their delineations. The average Dice coefficient was 0.783 ± 0.086 for the left hippocampus and 0.825 ± 0.033 for the right; these values are not significantly different according to a proportion equality test ($p \approx 0.45$). These values for the overlap measure are consistent with other values reported in the literature [14], and represent a boundary for the performance our algorithm can achieve.

Table 2 displays the Dice coefficients achieved by the multi-atlas scheme for the original fusion method from [5] (i.e. taking the mode of the labels) and for the intensity-based fusion from equation 4 when the different intensity matching methods are used. The algorithms outperform one another in the same order as in the previous experiment: our method provides again the highest performance. A paired t-test shows that our method is better than the landmark-based approach (which provides the second largest overlap) at a level of significance $\alpha = 0.05$. In absolute terms, the mean Dice coefficient provided by the proposed method (0.755) is in the range of results of other recent methods in the literature in which the training and test datasets have been acquired with different scanners and/or pulse sequences (for example [15]). The segmentations for the first two volumes of the HSD dataset are shown in figure 3.

Table 2. Dice coefficients for the multi-atlas segmentation of the left and right hippocampus. The last column is the p-value of a paired t-test comparing our method with the landmark-based approach.

Method	Mode 5	Hist. stret.	Hist. equal.	Landmark-based	This study	p
Left	0.678±0.226	0.702±0.117	0.713±0.142	0.718±0.133	0.748±0.126	0.0306
Right	0.695±0.212	0.711±0.089	0.722±0.125	0.733±0.106	0.761±0.096	0.0562
Both	0.686±0.215	0.706±0.102	0.718±0.132	0.725±0.119	0.755±0.110	0.0403

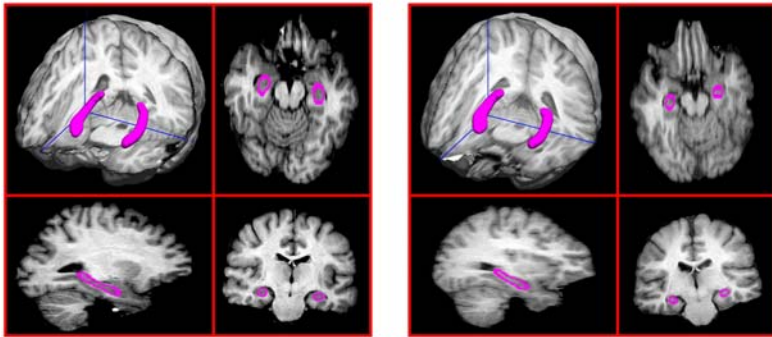


Fig. 3. Segmentation of the left and right hippocampus in the first two volumes of the test dataset: three orthogonal planes and 3D rendering

4 Discussion

A MRI image intensity standardization method based on synthetic MRI and its application to multi-atlas segmentation have been presented in this paper. Despite approximations such as neglecting T_2^* effects and despite the fact that only three channels are used to estimate the T_1 , T_2 and ρ maps (unlike other synthetic MRI methods that use many more), the method provides a fairly accurate intensity mapping between different acquisition protocols. Though a multi-atlas framework was used to test the method, any other image analysis application could in principle benefit from the algorithm. The results for the hippocampus segmentation are reasonable given that only ten atlases were utilized, and they should improve if more atlases were available and a proper atlas selection process could be carried out. The performance is also limited by the fact that the annotations of our dataset and HSD are not very consistent; this inter-user variability gives an idea of how difficult automated hippocampus segmentation is.

A potential disadvantage of the proposed algorithm is that, whereas T_2 is relatively independent of the magnetic field strength of the scanner, T_1 presents a dependence on this parameter. This study is based on images acquired at 1.5T. As of today, most scans are acquired at 1.5T or 3T, which are the standard clinical field strengths. It would therefore be possible to annotate a set of atlases acquired at 3T to cover most of the clinical scans currently acquired in the World. Another possibility would be to modify the T_1 map of the atlases to

account for the field strength difference using a model of the relationship between them. Addressing this problem, testing the method on more scans from different datasets and utilizing more atlases in the segmentation remain as future work.

Acknowledgements

This work is funded by NSF grants IIS-0844566 and N000140910099, and NIH grant U54RR021813. The authors would like to thank Dr. D. Ennis and Dr. Y. Natsuaki from UCLA Biomedical Physics for their help with MRI signal models. The first author would also like to thank the Fulbright program for the funding.

References

1. Bernstein, M., King, K., Zhou, X.: Handbook of MRI Pulse Sequences. Elsevier, Academic Press (2004)
2. Deoni, S., Rutt, B., Peters, T.: Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magnetic Resonance in Medicine* 49(3), 515–526 (2003)
3. Nyul, L., Udupa, J., Zhang, X.: New variants of a method of MRI scale standardization. *IEEE T. Med. Imaging* 19(2), 143–150 (2000)
4. Jager, F., Hornegger, J.: Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. *IEEE T. Med. Imaging* 28(1), 137–150 (2009)
5. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33(1), 115–126 (2006)
6. Maitra, R., Riddles, J.J.: Synthetic magnetic resonance imaging revisited. *IEEE T. Med. Imaging* 29(3), 895–902 (2010)
7. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE T. Pattern Anal.* 12(7), 629–639 (1990)
8. Smith, S.: Fast robust automated brain extraction. *Hum. Brain Mapp.* 17(3), 143–155 (2002)
9. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in mri. *IEEE T. Med. Imaging* 17, 87 (1998)
10. Deichmann, R., Good, C.D., Josephs, O., Ashburner, J., Turner, R.: Optimization of 3-D MP-RAGE for structural brain imaging. *Neuroimage* 12(1), 112–127 (2000)
11. Shattuck, D., Leahy, R.: Brainsuite: An automated cortical surface identification tool. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 50–61. Springer, Heidelberg (2000)
12. Nishimura, D.: Principles of Magnetic Resonance Imaging. Stanford U., Stanford (1996)
13. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion - application to cardiac and aortic segmentation in ct scans. *IEEE T. Med. Imaging* 28, 1000–1010 (2009)
14. Jeukens, C., Vlooswijk, M., Marian Majoie, H., de Krom, M., Aldenkamp, A., Hofman, P., Jansen, J., Backes, W.: Hippocampal MRI Volumetry at 3 Tesla: Reliability and Practical Guidance. *Invest. Radiol.* 44(9), 509 (2009)
15. Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE T. Med. Imaging* 27, 495–508 (2008)

Multi-organ Segmentation from Multi-phase Abdominal CT via 4D Graphs Using Enhancement, Shape and Location Optimization

Marius George Linguraru, John A. Pura, Ananda S. Chowdhury,
and Ronald M. Summers

Imaging Biomarkers and Computer-Aided Diagnosis Laboratory,
Radiology and Imaging Sciences, Clinical Center,
National Institutes of Health, Bethesda, MD
lingurarum@mail.nih.gov

Abstract. The interpretation of medical images benefits from anatomical and physiological priors to optimize computer-aided diagnosis (CAD) applications. Diagnosis also relies on the comprehensive analysis of multiple organs and quantitative measures of soft tissue. An automated method optimized for medical image data is presented for the simultaneous segmentation of four abdominal organs from 4D CT data using graph cuts. Contrast-enhanced CT scans were obtained at two phases: non-contrast and portal venous. Intra-patient data were spatially normalized by non-linear registration. Then 4D erosion using population historic information of contrast-enhanced liver, spleen, and kidneys was applied to multi-phase data to initialize the 4D graph and adapt to patient specific data. CT enhancement information and constraints on shape, from Parzen windows, and location, from a probabilistic atlas, were input into a new formulation of a 4D graph. Comparative results demonstrate the effects of appearance and enhancement, and shape and location on organ segmentation.

Keywords: multi-phase CT, segmentation, 4D graph, shape, enhancement.

1 Introduction

In current CT-based clinical abdominal diagnosis, radiologists rely on analyzing multi-phase CT data, as soft tissue enhancement can be an indicator of abnormality. This makes multi-phase data (with/without contrast) readily available. Diagnosis also relies on the comprehensive analysis of groups of organs and quantitative measures of soft tissue, as the volumes and shapes of organs can be indicators of disorders.

Computer-aided diagnosis (CAD) and medical image analysis traditionally focus on organ- or disease-based applications. However there is a strong incentive to migrate toward the automated simultaneous segmentation and analysis of multiple organs for comprehensive diagnosis or pre-operative planning and guidance. Additionally, the interpretation of medical images should benefit from anatomical and physiological priors, such as shape and appearance; synergistic combinations of priors were seldom incorporated in the optimization of CAD.

The segmentation of abdominal organs was initialized from probabilistic atlases in [10] using relationships between organs and manual landmarks. Alternatively, multi-dimensional contrast-enhanced CT data were employed in [5,7,13]. In [5,13] the segmentation used independent component analysis in a Bayesian framework. A 4D convolution was proposed in [7] constrained by a historic model of abdominal soft tissue enhancement. These intensity-based methods are hampered by the high variability of abdominal intensity and texture. More recently, a hierarchical multi-organ statistical atlas was developed [9]; the analysis was restricted to the liver area due to large variations to be statistically modeled for inter-organ relationships.

On a different note, graph cuts [2] have become popular for image segmentation, due to their ability to handle highly textured data via a numerically robust global optimization. A major drawback remains the manual initialization of such applications [4,8,16]. In [1,6] model-based information was included for the heart and kidney; however the models were aligned using markers. Compact shape priors were used in [4], but medical data often involves complex shapes. A shape model was also integrated in [15] as a density estimation for shape priors, initially proposed for level sets in [3], but a symmetric shape distance can be biased if shape initialization is poor.

We propose a new formulation of a 4D directional graph to automatically segment abdominal organs, at this stage the liver, spleen, and left and right kidneys using graph cuts. The approach is optimized to medical images through the use of location probabilistic priors that are intrinsic to medical data, an enhancement constraint characteristic to the clinical protocols using abdominal CT, and an asymmetric shape distance that avoids shape bias to build Parzen windows. The method is optimized globally and starts with historic (entire patient population) 4D intensity data to automatically initialize the graph, then migrating to patient specific information for better specificity. Comparative results at different stages of the algorithm show the effects of appearance, shape and location on the accuracy of organ segmentation.

2 Methods and Materials

2.1 Data, Preprocessing and Model Initialization

Eight random abdominal CT studies (normal and abnormal) were obtained with two temporal acquisitions. The first image was obtained at non-contrast phase (NCP) and a second at portal venous phase (PVP) using fixed delays. The CT data were collected on LightSpeed Ultra and QX/I scanners [GE Healthcare] at multiple time points. Image resolution ranged from 0.62 to 0.82 mm in the axial view with a slice thickness of 5 mm. The algorithm was trained and tested with a leave-one-out strategy.

The liver, spleen, and left and right kidneys were manually segmented (by two research fellows supervised by a board-certified radiologist) in the 8 CT cases using the PVP CT volumes to provide a gold standard for testing the method. Histograms of the segmented organs (objects) and background in NCP and PVP were computed and modeled as sums of Gaussians, as in Figure 1. While there are partial overlaps between the object and background distributions (especially at NCP), the combination of multi-phase data ensures a better separation.

Although images were acquired during the same session and intra-patient, there was small, but noticeable abdominal inter-phase motion, especially associated with

breathing. The preprocessing follows the work in [7]. Data were smoothed using anisotropic diffusion [12]. NCP data were registered to the PVP images. The demons non-linear registration algorithm was employed [14], as the limited range of motion ensures partial overlaps between organs over multiple phases. The deformation field F of image I to match image J is governed by the optical flow equation

$$F = \frac{(I - J)\nabla J}{\|\nabla J\|^2 + (I - J)^2}; \quad (1)$$

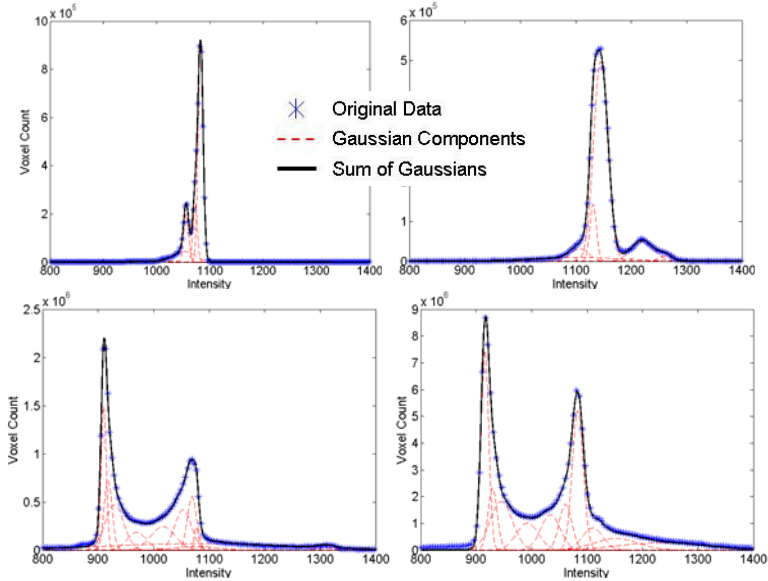


Fig. 1. Fitted sums of Gaussians to historic data of organs/objects (top row) and background (bottom row). NCP data is shown on the left column and PVP data on the right. Historic data we refer to the training cases in the leave-one-out strategy.

A probabilistic atlas (PA) was constructed from a different set of 10 non-contrast CT from healthy cases, independent from the data above, with manually segmented liver, spleen and kidneys. Organ locations were normalized to an anatomical landmark (xiphoid) to preserve spatial relationships and model organs in the anatomical space. The tip of the xiphoid (an ossified cartilaginous extension below the sternal notch) was extracted manually in the data used in the location model. A random image set was used as reference and the remaining images registered to it. Structural variability including the size of organs was conserved by a size-preserving affine registration. The location bias was minimized by the normalization by the xiphoid. The 10 unprocessed CT data were further used to build shape constraints via a Parzen window distribution, as explained in the construction of the 4D graph.

4D Convolution

From smoothed historic data of contrast-enhanced CT, the min and max intensities for the organs were estimated: $min_{i,t} = \mu_{i,t} - 3\sigma_{i,t}$ and $max_{i,t} = \mu_{i,t} + 3\sigma_{i,t}$, where $i=1..3$ for

liver, spleen and kidneys, $\mu_{p,t}$ and $\sigma_{p,t}$ represent the mean and standard deviation, and $t=1,2$ for NCP and PVP. As in [7], a 4D array $K(x,y,z,t)=I_t(x,y,z)$ was created from multi-phase data. A convolution with a 4D filter f labeled only regions for which all voxels in the convolution kernel satisfied the intensity constraints. L represents the labeled image and l_j the labels ($j=1..4$ for liver, spleen, left kidney and right kidney).

$$L(x, y, z) = (K \circ f)(x, y, z, t) = \begin{cases} l_j, & \text{if } \bigcap_t (\min_{j_t} \leq K(x, y, z, t) \leq \max_{j_t}); \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The labeled organs in L appear eroded as a result of the 4D convolution. In our method, L provided seeds for objects (I_o) in the 4D graph and was used to estimate the patient-specific histograms. The eroded inverted L provided the background (I_b) seeds and the related histograms. To report the segmentation results by 4D convolution (see Results), L was dilated to compensate for the undersegmentation of organs.

4D Graph

Graph cuts (GC) were chosen for the inherent capability to provide a globally optimal solution [2]. The input to our problem is two sets of registered abdominal CT scans per patient: the NCP and PVP sequences. Hence every voxel p in the graph has two intensity values I_{ncp}^p and I_{pvp}^p . Let $A = (A_1, A_2, \dots, A_p, \dots, A_p)$ be a binary vector with components A_p that can be either objects of interest (i.e. liver, spleen and kidneys) denoted by O or background B , where $B \cap O = \emptyset$. Typical graphs perform data labeling (t-links), via log-likelihoods based solely on 2D or 3D interactive histogram fitting, and penalize neighborhood changes (n-links) through likelihoods from the image contrast [2]. We first extend the formulation to analyze 4D data, and then incorporate penalties from the contrast enhancement of CT soft tissue, Parzen shape windows, and location from a priori probabilities. While location knowledge was incorporated in the labeling of objects, shape information penalized boundaries not resembling the references. The cost function E to minimize becomes

$$E(A) = E_{data}(A) + E_{enhance}(A) + E_{location}(A) + \sum_{i=1}^4 (E_{boundary}(A) + E_{shape}(A)); \quad (3)$$

The first three terms define the objects (t-links) and the last two energies find the cuts (n-links) with $i=1..4$ for liver, spleen, left kidney and right kidney. In this application, E_{data} is a regional term that computes penalties based on 4D histograms of O and B ; the probabilities P of a voxel to belong to O or B are computed from patient specific histograms of NCP and PVP data.

$$E_{data}(A) = \lambda \sum_{p \in O} R_p(O) + (1 - \lambda) \sum_{p \in B} R_p(B); \quad (4)$$

$$R_p(O) = -\ln \left(\frac{\sqrt{P_{ncp}(I_{ncp}^p | O) P_{pvp}(I_{pvp}^p | O)}}{\sqrt{P_{ncp}(I_{ncp}^p | O) P_{pvp}(I_{pvp}^p | O)} + \sqrt{P_{ncp}(I_{ncp}^p | B) P_{pvp}(I_{pvp}^p | B)}} \right);$$

$E_{enhance}$ penalizes regions that do not enhance rapidly during the acquisition of NCP-PVP CT data (i.e. muscles, ligaments and marrow). σ_{ncp} and σ_{pvp} are the standard deviations of noise associated with NCP and PVP.

$$E_{enhance}(A) = \sum_{p \in P} 1 / (1 + E_p^2), \quad \text{with } E_p = \frac{(I_{pvp}^p - I_{ncp}^p)^2}{2\sigma_{ncp}\sigma_{pvp}}. \quad (5)$$

Similarly, location constraints from a normalized probabilistic atlas (PA) were implemented in $E_{location}(A) = -\sum_{p \in P} \ln(S_p(p|O))$, where S_p represents the probability of p to belong to O . S_p was obtained by registering PA to the test images by a sequence of coarse-to-fine affine registrations.

$E_{boundary}$ assigns penalties for 4D heterogeneity between two voxels p and q , with $q \in N_p$ a small neighborhood of p . λ , μ and δ are constants and weigh the contribution from object/background, and the directionality of the graph at boundaries/shape, respectively (all set to value 0.5 for equal contributions). $dist(p, q)$ is the Euclidean distance between p and q .

$$E_{boundary}(A) = \mu \sum_{\{p,q\} \in N_p} w_{\{p \rightarrow q\}} + (1 - \mu) \sum_{\{p,q\} \in N_p} w_{\{q \rightarrow p\}}; \quad (6)$$

$$Initialize \quad w_{\{p \rightarrow q\}} = w_{\{q \rightarrow p\}} = \begin{cases} 0 & , if A_p = A_q \\ \exp\left(-\frac{|I_{ncp}^p - I_{ncp}^q| \cdot |I_{pvp}^p - I_{pvp}^q|}{2\sigma_{ncp}\sigma_{pvp}}\right) \frac{1}{dist(p,q)} & , otherwise \end{cases};$$

$$IF((I_{pvp}^p - I_{pvp}^q) > \sigma_{pvp} \text{ OR } (I_{ncp}^p - I_{ncp}^q) > \sigma_{ncp}) \text{ THEN } w_{\{q \rightarrow p\}} = 1; \text{ ELSE } w_{\{p \rightarrow q\}} = 1.$$

The last condition in (6) penalizes transitions from dark (less enhanced) to brighter (more enhanced) regions considering image noise, to correct the edges of O . This is an intrinsic attribute of medical data (e.g. the abdominal muscles are darker than O). Additional penalties were implemented from the seeds for O and B from I_o and I_b .

Shape constraints were introduced using Parzen windows [11] estimated from the reference liver shapes from the 10 non-contrast CT data. First, the result of the 4D convolution (L) was used to align the shape references using scaling, rotation and the location of the centroids. An asymmetric normalized dissimilarity measure D was used to avoid the bias introduced by L ; H is the Heaviside step function

$$D(s_1, s_2) = \int (H(s_1) - H(s_2))^2 H(s_1) dx / \int H(s_1) dx.$$

The Parzen shape probability PS of s given n shape references was calculated [3] to encourage cuts that minimize the shape dissimilarity

$$PS(s) = \sum_{i=1}^n \exp(-D(s, s_i) / 2\sigma^2) / n; \quad \sigma^2 = \sum_{i=1}^n \min_{j \neq i} D(s_j, s_i) / n;$$

$$\text{then } E_{shape}(A) = \delta \sum_{\{p,q\} \in N_p} v_{\{p \rightarrow q\}} + (1 - \delta) \sum_{\{p,q\} \in N_p} v_{\{q \rightarrow p\}}; \quad \text{with} \quad (7)$$

$$v_{\{p \rightarrow q\}} = v_{\{q \rightarrow p\}} = \begin{cases} 0 & , if A_p = A_q \text{ or } PS(s)^p = PS(s)^q \\ \max(PS(s)^p, PS(s)^q) / dist(p, q) & , otherwise \end{cases};$$

$$IF(PS(s)^p > PS(s)^q) \text{ THEN } v_{\{q \rightarrow p\}} = 1; \text{ ELSE } v_{\{p \rightarrow q\}} = 1.$$

We compared results obtained after the 4D convolution to those achieved using intensity-based 4D GC, and after including shape and location correction. The influence of patient specific versus population (historic) statistics on the accuracy of organ segmentation was also analyzed. We computed the Dice coefficient, volume error, root mean square error, and average surface distance. Non-parametric statistical tests (Wilcoxon paired test) were performed to assess the significance of segmentation improvement at different steps of the algorithm at 95% confidence interval.

3 Results

Quantitative results from applying our method to the segmentation of liver, spleen and kidneys are shown in Table 1 at different stages of the algorithm. Figure 2 presents a typical example of liver, spleen and kidneys segmentation. Another example is shown in 3D in Figure 3 along with the errors between manual and automated segmentations.

The use of 4D graph-cuts (GC) improved the results significantly over those of the 4D convolution for all organs, as seen in Table 1. Employing shape and location information brought a further significant improvement for the segmentation of the spleen and liver. Significantly better segmentations by using patient specific data over historic data were noted for both kidneys (not shown in Table 1).

Table 1. Statistics (mean \pm std) for the liver, spleen, left kidney and right kidney segmentation results from data of low resolution (5mm slice thickness). Columns present the Dice coefficient (DC), volume estimation error (VER), root mean square (RMSE) error and average surface distance (ASD). 4D C represents the convolution, GCI is GC based solely on image intensity (including 4D appearance and enhancement) and 4D GCSL includes shape and location constraints. Highlighted cells mark the organs where a significant improvement was obtained relative to the previous step of the segmentation algorithm ($p < 0.05$).

	ORGAN	DC (%)	VER (%)	RMS (mm)	ASD (mm)
4D C (Historic Data)	LKidney	88.7 \pm 3.7	10.9 \pm 8.9	2.3 \pm 0.4	1.1 \pm 0.3
	RKidney	89.6 \pm 3.4	13.6 \pm 6.8	2.1 \pm 0.5	1.1 \pm 0.3
	Spleen	79.9 \pm 10.1	14.9 \pm 16.9	4.5 \pm 1.9	2.7 \pm 1.7
	Liver	89.1 \pm 3.7	7.3 \pm 4.6	6.7 \pm 1.5	3.4 \pm 1.0
4D GCI (Patient Data)	LKidney	92.6 \pm 2.4	5.4 \pm 6.9	1.8 \pm 1.2	0.8 \pm 0.6
	RKidney	92.8 \pm 1.9	5.6 \pm 5.8	1.8 \pm 0.8	0.8 \pm 0.4
	Spleen	89.6 \pm 2.7	11.4 \pm 6.9	3.0 \pm 1.4	1.5 \pm 0.9
	Liver	94.0 \pm 1.2	6.2 \pm 2.8	4.4 \pm 2.0	1.8 \pm 0.7
4D GCSL (Patient Data)	LKidney	91.9 \pm 2.4	4.5 \pm 4.6	1.7 \pm 0.5	0.8 \pm 0.3
	RKidney	92.6 \pm 1.3	5.2 \pm 2.9	1.7 \pm 0.4	0.8 \pm 0.3
	Spleen	90.7 \pm 1.4	8.2 \pm 5.0	2.2 \pm 1.0	1.2 \pm 0.5
	Liver	94.9 \pm 0.8	4.1 \pm 2.0	3.4 \pm 1.8	1.4 \pm 0.4

4 Discussion

We proposed a new formulation for a 4D graph-based method to segment abdominal organs from multi-phase CT data. The method extends basic graph cuts by using: 1) temporal acquisitions at two phases and enhancement modeling; 2) shape priors from Parzen windows; and 3) location constraints from a probabilistic atlas. Enhancement information allowed improving regional bias within tissues, thereby better modeling the biological properties. Location probabilistic priors, intrinsic to medical data, and shape information from the asymmetric computation of Parzen shape windows (to avoid shape bias) supplied additional constraints for the global optimization of the graph. A Parzen distribution was preferred as a non-parametric probability model that converges to the true density with increasing number of samples.

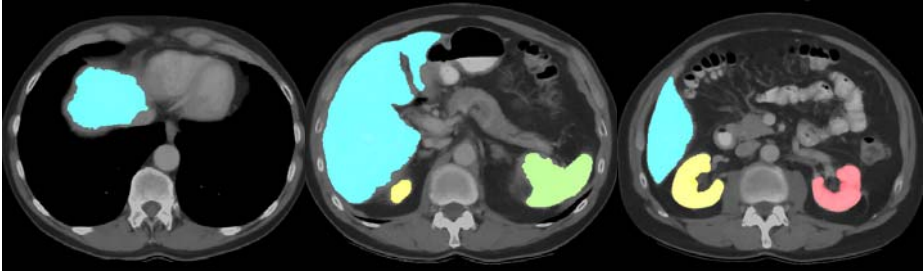


Fig. 2. A typical example of liver (blue), spleen (green), right kidney (yellow) and left kidney (red) automated segmentation on 2D axial views of the CT data

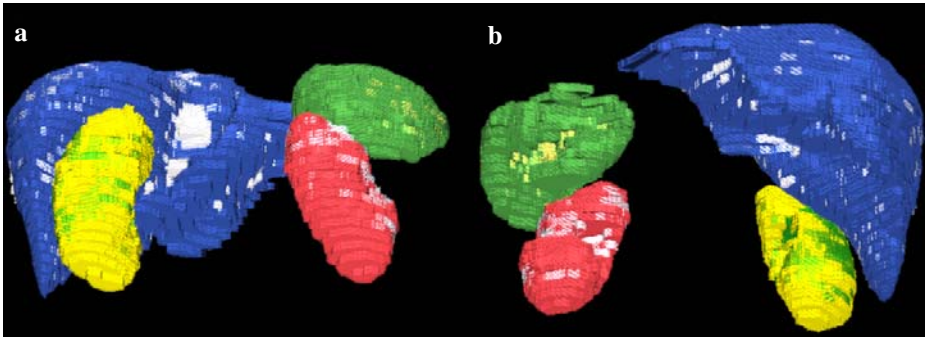


Fig. 3. 3D images of the automatically segmented abdominal organs; (a) is a posterior view and (b) an anterior view. The liver ground truth is blue with segmentation errors in white; spleen is green with errors in yellow; right kidney is yellow with errors in green; left kidney is red with errors in white. The pixilation is due to image low resolution (5mm slice thickness).

Livers, spleens and kidneys were segmented from multi-phase clinical data following the typical acquisition protocol of abdominal CT images. An automated initialization of the graph was employed. Historic data from a patient population were used to initialize the graph based on an adaptive 4D convolution. Then patient specific image characteristics were estimated for improved specificity and input into the directional graph. Results from image data with low spatial resolution showed overlaps over 90% and average surface distances less than 1.5mm for all organs.

The method avoided the inclusion of heart segments in the segmentation of liver, but had the tendency to underestimate organ volumes, in particular that of the spleen. Parts of the inferior vena cava may be erroneously segmented in the mid-cephalocaudal liver region, especially when contrast enhancement is low, and represent one of the sources of error in the liver segmentation (Figure 3). Partial volume effects (low image resolution), small registration errors, and the estimation of object and background distributions may also contribute to undersegmentation. Results are expected to be superior on data with high spatial resolution.

As expected, using graph cuts based only on intensity significantly improved the segmentation of all four abdominal organs over the 4D convolution. However, moving from historic to patients specific statistics only improved the segmentation of kidneys, probably due to the prevalence of liver and spleen statistics in the object (O) histogram. Optimizing the graph with shape and location constraints brought a

significant improvement only in the segmentation of spleen and liver, as kidneys, already well segmented at the previous step of the algorithm due to strong image contrast at edges from fast enhancement, vary less in shape. In the future we will include more shape/location references and variation to improve the segmentation.

Acknowledgements

This work was supported by the Intramural Research Program of the National Institutes of Health, Clinical Center. The authors thank Jesse K. Sandberg and Javed Aman for helping with the data analysis.

References

- [1] Ali, A.M., Farag, A.A., El-Baz, A.S.: Graph cuts 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] Boykov, Y., Jolly, M.-P.: Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: *Int. Conf. Comp. Vis.*, pp. 105–112 (2001)
- [3] Cremers, D., Osher, S.J., Soatto, S.: Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. Comp. Vis.* 69(3), 335–351 (2006)
- [4] Das, P., et al.: Semiautomatic segmentation with compact shape priors. *Image and Vision Computing* 27(1-2), 206–219 (2008)
- [5] Hu, X., Shimizu, A., Kobatake, H., Nawano, S.: Independent analysis of four-phase abdominal CT images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 916–924. Springer, Heidelberg (2004)
- [6] Lin, X., Cowan, B., Young, A.: Model-based graph cut method for segmentation of the left ventricle. *Proc. IEEE Eng. Med. Biol. Soc.* 3, 3059–3062 (2005)
- [7] Linguraru, M.G., Summers, R.M.: Multi-organ segmentation in 4D contrast-enhanced abdominal CT. In: *IEEE Symposium on Biomedical Imaging* 2008, pp. 45–48 (2008)
- [8] Liu, L., Raber, D., et al.: Interactive separation of segmented bones in CT volumes using graph cut. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 296–304. Springer, Heidelberg (2008)
- [9] Okada, T., et al.: Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 502–509. Springer, Heidelberg (2008)
- [10] Park, H., Bland, P.H., Meyer, C.R.: Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Trans. Med. Imaging* 22(4), 483–492 (2003)
- [11] Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Stat.* 33, 1065–1076 (1962)
- [12] Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence* 12, 629–639 (1990)
- [13] Sakashita, M., et al.: A method for extracting multi-organ from four-phase contrasted CT images based on CT value distribution estimation using EM-algorithm. In: *SPIE*, vol. 6509, pp. 1C1–12 (2007)
- [14] Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis* 2(3), 243–260 (1998)
- [15] Vu, N., Manjunath, B.S.: Shape prior segmentation of multiple objects with graph cuts. In: *Proceedings CVPR* 2008 (2008)
- [16] Zheng, Y., et al.: Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 393–401. Springer, Heidelberg (2007)

A Semi-automatic Method for Segmentation of the Carotid Bifurcation and Bifurcation Angle Quantification on Black Blood MRA

Hui Tang^{1,3}, Robbert S. van Onkelen^{1,2}, Theo van Walsum¹, Reinhard Hameeteman¹,
Michiel Schaap¹, Fufa. L. Tori¹, Quirijn J.A. van den Bouwhuijsen^{1,2},
Jacqueline C.M. Witteman², Aad van der Lugt¹, Lucas J. van Vliet³,
and Wiro J. Niessen^{1,3}

¹ Departments of Radiology and Medical Informatics

² Department of Epidemiology

Erasmus MC- University Medical Center Rotterdam, The Netherlands

³ Department of Image Science and Technology

Faculty of Applied Science, Delft University of Technology, The Netherlands

`h.tang@tudelft.nl`

Abstract. Quantitative information about the geometry of the carotid artery bifurcation may help in predicting the development of atherosclerosis. A geodesic active contours based segmentation method combining both gradient and intensity information was developed for semi-automatic, accurate and robust quantification of the carotid bifurcation angle in Black Blood MRA data. The segmentation method was evaluated by comparing its accuracy to inter and intra observer variability on a large dataset that has been acquired as part of a longitudinal population study which investigates the natural progression of carotid atherosclerosis. Furthermore, the method is shown to be robust to initialization differences. The bifurcation angle obtained from the segmented lumen corresponds well with the angle derived from the manual lumen segmentation, which demonstrates that the method has large potential to replace manual segmentations for extracting the carotid bifurcation angle from Black Blood MRA data.

1 Introduction

Carotid atherosclerosis, i.e. plaque build-up in the arterial wall, is one of the major causes of stroke. To resolve the mechanisms behind plaque formation, many research institutes are investigating plaque formation, plaque growth and the factors affecting plaque formation and growth [1]. Thomas et al. showed that the inter subject variation in geometry of the carotid bifurcation significantly increase with age and early atherosclerotic disease progression [1]. Whether an individual's geometry will predict the development and progression of atherosclerosis is still unclear and needs to be investigated. An automated method for accurate and robust quantification of carotid geometry is valuable, especially when studying the relation between carotid artery geometry and disease progression in large clinical trials or population studies.

MRI is a promising modality for imaging diseased arteries and the arterial wall due to its capability to visualize both the arterial lumen as well as the plaque composition [2]. The current work is carried out in the context of a population study[3], where Black Blood MRA (BBMRA) is used for lumen visualization as it allows contrast agent free, non-invasive imaging.

Several authors investigated the segmentation of carotid artery bifurcation in BBMRA images, using explicit contour or surface deformation schemes [4][5][6][7]. However, these papers lack quantitative evaluation or the evaluation is limited to a few patients. Yuan et al. employed a closed contour snake and a weighted distance transform to detect the edge of both the lumen and the outer wall in BBMRA, but quantitatively evaluated their method on only five subjects [4]. Ladak et al. [5] applied a semi-automatic technique based on discrete dynamic contours to outline the inner and outer boundaries of vessels and evaluated the accuracy on the measurement of vessel wall area for 12 subjects. More recently, Jin and Ladak [6] extended their work with an explicit 3D deformable model to segment the lumen of the carotids using a manually drawn centerline as initialization. The method was quantitatively evaluated on five datasets. Adame et al [7] segmented the carotid lumen in BBMRA using fuzzy c-means clustering and a minimum cost path as a preprocessing step towards characterizing plaque composition. However, the accuracy and reproducibility of this lumen segmentation method was not evaluated quantitatively.

We present a method that uses a 3D geodesic active contour approach to segment the lumen from BBMRA, and show how the resulting surface can serve as basis to subsequently extract and quantify the bifurcation angle describing the carotid artery geometry. In contrast to most of the previously presented approaches, we utilize an implicit surface representation. Additionally, we combine both image gradient and image intensity terms in the energy function that steers the level set segmentation. The image intensity term is based on an estimate of the local fore- and background intensities, using k-means clustering. Moreover, we perform an extensive quantitative evaluation of both the segmentation accuracy, reproducibility on a large set of subjects. Finally, we quantify the bifurcation angle for both the semi-automatically segmented and the manually annotated lumen to study the accuracy of quantification.

2 Segmentation and Quantification Method

The segmentation and quantification method consists of three steps. First, we preprocess the BBMRA to correct for intensity inhomogeneity and to reduce noise. Second, we apply a geodesic active contour approach to segment the lumen, and finally we extract the carotid bifurcation angle from the resulting segmentation. The parameters of the preprocessing steps are listed in the next section.

2.1 Preprocessing: Bias Correction, Noise Suppression and Normalization

Preprocessing is applied to alleviate segmentation errors that are caused by severe intensity inhomogeneity or the low SNR in MR sequences. We applied a bias field correction approach for BBMRA as proposed by Sled et al. [8] to correct for intensity inhomogeneity. After bias correction, an anisotropic diffusion filter [9] is employed to

reduce image noise while preserving edges and hence the geometry. Finally, we apply intensity normalization to the images such that the mean intensity is zero and the standard deviation is one. In this way the intensity variation between different scans can be ignored. Fig.1(a-c) shows the original image and images after bias correction and noise reduction.

2.2 Segmentation

After preprocessing we apply a geodesic active contour (GAC) [10] to segment the carotid lumen from the BBMRA. The GAC evolution is initialized with a centerline with an initial surface that is defined by a tube with a radius of 1 mm that is generated along the lumen centerline. The centerlines of both the internal and external carotid arteries were annotated by an experienced observer, using an in-house developed annotation tool[11].

As the low image resolution in MRI, image gradient is not sufficient for defining the lumen boundary, thus the energy function optimized with the GAC approach incorporates both intensity and gradient information and is defined as follows:

$$E = \oint [g(|\nabla I(C(p))|) + \alpha s(I(C(p)))] |C'(p)| dp \quad (1)$$

where $C(p)$ is a contour parameterized by p , α a weighting factor that balances the influence of both terms. All parameters including weighing factors are optimized in an exhaustive parameter optimization (see section 3.2).

The first term of the stopping function in Eq.(1) is a gradient-based term which is commonly used as energy function for geodesic active contours. The function $g(x)=1/(1+cx)$ is monotonically decreasing. The second term of Eq.(1) incorporates the intensity information through an intensity potential term $s(i)$, which is defined as a monotonically increasing function of the intensity dissimilarity to the intensity of the estimated lumen boundary I_b , and goes to zero if the dissimilarity goes to zero:

$$s(i) = |i - I_b| \quad (2)$$

with i an intensity value. Shown in Fig1.(d), the boundary intensity I_b is estimated with the following linear function of the lumen fore- and background intensity:

$$I_b = I_{center} + \beta (I_{vicinity} - I_{center}) \quad (0 < \beta < 1) \quad (3)$$

where I_{center} is the estimated mean intensity within a circle with radius of 1mm centered around centerline points in every slice and $I_{vicinity}$ is estimated via k-means clustering using two clusters and is selected to be the cluster center with the highest value. This clustering is performed slice-based, to account for slight inter slice intensity variation that remain after bias-correction. This equation is inspired by the work of Hoogeveen et al [12], where it is shown that the lumen boundary in time-of-flight MR lies at $\beta = 0.5$ and in phase-contrast MR at $\beta = 0.1$. We will optimize β in the training session for BBMRA.

The level-set equation describing the evolving surface u that corresponds to minimizing the energy function of evolution of Eq.(1) is Eq.(4) where κ_{min} is the local minimum curvature commonly used for segmenting tubular structures[11].

$$u_t = (g + \alpha s)\kappa_{min}|\nabla u| + \nabla(g + \alpha s) \cdot \nabla u \quad (4)$$

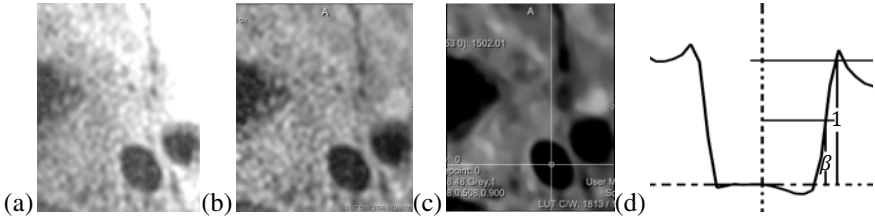


Fig. 1. (a) An original BBMRI, (b) unbiased BBMRI, (c) denoised BBMRI, (d) the corresponding cross-sectional intensity profile of denoised BBMRI

2.3 Quantification

We quantified the carotid bifurcation angle according to the definition of Thomas [1] and Antiga [14] using the VMTK (Vascular Modeling Tool Kit) implementation [15].

3 Experiments and Results

3.1 Data Acquisition and Reference Standard

Proton Density Weighted (PDW) BBMRI images were obtained from 49 subjects that were randomly chosen from participants in a population study [3]. All the subjects had one healthy and one diseased carotid artery (as assessed with ultrasound). The MRI data was acquired with an in-plane resolution of $1.10 \times 0.81 \text{ mm}^2$, a slice thickness of 0.90 mm , a field of view of $13 \times 13 \text{ cm}^2$ and a matrix size of 160×128 . The images were interpolated on the scanner to a pixel size of $0.51 \times 0.51 \text{ mm}^2$ before processing. The image quality was visually assessed by a trained observer and the datasets were grouped into four classes according to their image quality, i.e. bad (1), normal (17), good (29), and excellent (2). The single image with bad image quality was excluded from the study, as the image quality did not permit quantification of the bifurcation angle. All other scans from 48 subjects were included in the experiments.

We randomly selected around 20% of these images (4 normal and 6 good) for training. We also randomly selected another 30% (5 normal, 9 good and 1 excellent) to obtain the inter and intra observer variability as well as the robustness to initialization using different centerlines. The intra and inter annotation was done three months later than the first annotation.

3.2 Parameters Selection and Optimization

The parameters of the bias field correction method were set as proposed in [16] (i.e. shrinking factor 2 and fitting levels 4), except for the number of iterations, which was set to 600 after visual assessment. The parameters for the edge enhancing diffusion were optimized by visual inspection and were chosen as follows: Gaussian gradient scale 1 mm , and 10 iterations with a time step of 0.15.

The remaining parameters of the method, α , β and c were optimized in three steps. We used the Dice Similarity Coefficient (DSC) [17] between the reference

standard and our segmentation as optimization criterion. First we optimized β , by performing a geodesic active contour evolution on the intensity information alone:

$$E = \oint |I(C(p)) - I_b| |C'(p)| dp \tag{5}$$

The DSC was maximized at values for β around 0.55, corresponding more or less to the criterion proposed by Hooegeveen et al. for time-of-flight MR images [12]. The second step was to optimize c according to:

$$E = \oint \frac{1}{1+c|I'(C(p))|} |C'(p)| dp \tag{6}$$

The DSC was maximized when c equals 10. Third, we optimized α according to Eq.(1), where we fixed β to 0.55 and c to 10. The DSC was maximal at $\alpha=0.12$.

To confirm the selection of these three parameters, we fixed α to 0.12 and β to 0.55 and varied parameter c and we fixed parameters α and c at their optimum and varied β . These two optimizations resulted in exactly the same optimum, confirming the validity of the chosen parameters. Fig.2 shows the BBMRA after preprocessing, the gradient edge potential, the intensity dissimilarity as well as the combined stopping function for optimized α, β as well as c .

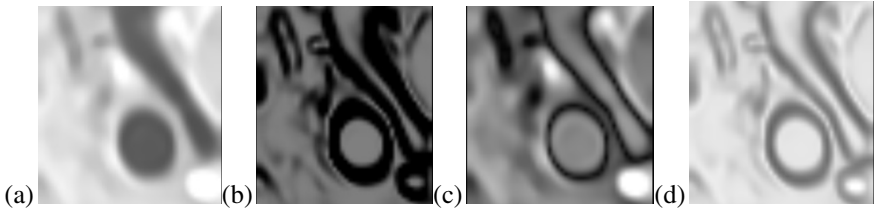


Fig. 2. (a) A preprocessed BBMRA and various stopping functions based on: (b) gradient only, (c) intensity only, and (d) a combination of intensity and gradient using the optimized parameters.

3.3 Segmentation Performance

The accuracy and reproducibility of the segmentation approach was quantitatively evaluated by comparing the segmentation results with annotated surfaces by one observer using DSC (0.87), average absolute surface distance (0.38mm) and maximum absolute surface distance (1.41mm) for 38 subjects (76 carotid arteries).

Furthermore, for the randomly selected 15 subjects (30 carotid arteries) we compared the manual annotations of the first observer to manual annotations of a second observer to quantify the inter-observer variance and we compared the manual annotations of the first observer to a second annotation by the first observer in order to quantify the intra-observer variability. Moreover, we quantified the reproducibility of the method by initializing the method with two different centerlines. The results are summarized in Tab.1(a-c). An example segmentation is shown in Fig. 3.

Tab. 1(a-c) shows that the DSC and average mean absolute surface distance as well as the maximum absolute surface distance between the semi automatic segmentation (A1) and the three manual ones (O11, O12 and O2) are close to the inter observer

variability (O11 and O12 vs O2). The results also show that, as expected, the intra observer variability is marginally better than the inter observer variability.

The segmentation reproducibility, which was evaluated by running the segmentation using a second centerline initialization (manually annotated by the first observer), was high, with a DSC of 0.98 (A1 vs A2 in Tab.1). This demonstrates that the method is robust to initialization differences.

Table 1. Pair-wise comparisons (a) DSC and Pearson coefficient between bifurcation angles(rAngle). (b)Average mean (meanAD) and max abs distance(maxAD).

rAngle \ DSC	A1	A2	O11	O12	O2	meanAD \ maxAD	A1	A2	O11	O12	O2
	A1	1	0.98	0.88	0.88		0.89	A1	0.00	0.06	0.36
A2	0.95	1	0.88	0.89	0.89	A2	0.84	0.00	0.36	0.35	0.30
O11	0.85	0.81	1	0.94	0.92	O11	1.47	1.46	0.00	0.13	0.18
O12	0.83	0.81	0.67	1	0.93	O12	1.58	1.41	0.87	0.00	0.16
O2	0.89	0.89	0.82	0.73	1	O2	1.16	1.34	0.89	0.88	0.00

A paired t-test was used to determine the significance of DSC increase for the combined energy function compared to a gradient only and intensity only function. In both cases, the segmentation was significantly improved, from respectively 0.86 ($p=0.002$) and 0.85 ($p=10^{-5}$) to 0.87 when evaluated on all the 76 testing datasets.

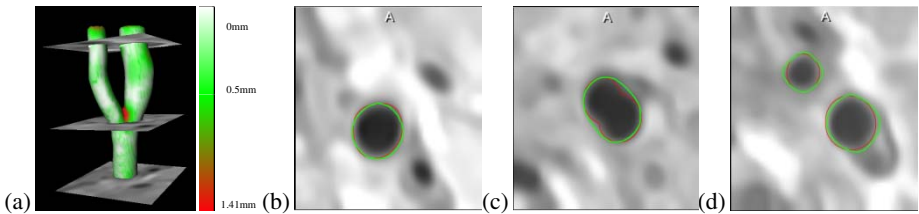


Fig. 3. (a) A 3D segmentation result (DSC: 0.92) (white: perfectly segmented; green: abs surface distance>0.2mm; red: abs surface distance>0.5mm) (b-d) Cross sectional view of segmentation in three slices (red: manual segmentation; green: our method)

3.4 Quantification of Bifurcation Angle

Carotid bifurcation angles are quantified using both the manual segmentation and the automatic segmentation method. Fig.4(a) shows the scatter plot of the angles obtained from segmentations of our method and the first observer for 76 carotid arteries. The Pearson coefficient between the angles by our method and the first observer for 76 carotid arteries is 0.896. Fig.4(b) shows that the bifurcation angles derived from our method is slightly under-quantified but 72 out of 76 cases are within $\text{mean} \pm 1.96\text{std}$.

For 30 datasets we measured the inter and intra observer variability. Tab.1(d) indicates that the correlation between the bifurcation angle obtained with the presented method, and the bifurcation angle extracted from both manual segmentations, is higher than the correlation in bifurcation angles between the two annotations of the observer

due to the large variability between the two annotations for the same observer in a few carotid arteries. The result suggests that the automated method can replace manual segmentation for extracting the bifurcation angle.

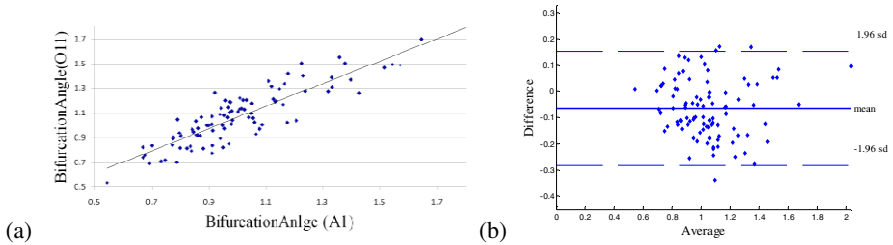


Fig. 4. (a) Correlation (b) Bland Altman plot between the bifurcation angle obtained from segmentations by our method and the first manual annotation of the first observer

4 Discussion and Conclusion

We presented a semi-automatic level-set based segmentation method for extracting the carotid artery bifurcation from BBMRA data. Both gradient and intensity information have been used to steer the level set. Combining both gradient and intensity information is shown to improve the segmentation accuracy compared to using only gradient or intensity information by paired t-test.

The threshold parameter for the intensity term has been optimized in a training stage. The resulting optimal value indicates that the FWHM criterion, which was previously found to be optimal in time-of-flight MRA, also holds for BBMRA.

Manually annotated centerlines were used for initialization, but several methods for automated centerline extraction exist [18]. Extending the current method with automatically extracted centerlines is part of future work.

An extensive quantitative evaluation on 76 carotid arteries showed that the accuracy of the segmentation is close to the manual segmentations. The accuracy of the presented method approaches the inter observer variability (evaluated on 30 carotid arteries). Furthermore, the carotid bifurcation angle can be assessed as accurately from the automated segmentation method as from the manually obtained segmentation. Finally, the reproducibility of the segmentation and quantification of the bifurcation angle for different initial centerlines is high. The present paper demonstrates that automated carotid artery segmentation in BBMRA has a high potential to replace manual segmentation for quantifying the carotid bifurcation angle.

References

1. Thomas, J.B., Antiga, L., Che, S.L., et al.: Variation in the Carotid Bifurcation Geometry of Young Versus Older Adults: Implications for Geometric Risk of Atherosclerosis. *Stroke* 36, 2450–2656 (2005)

2. Skinner, M.P., Yuan, C., Mitsumori, L., et al.: Serial Magnetic Resonance Imaging of Experimental Atherosclerosis Detects Lesion Fine Structure, Progression and Complications in vivo. *Nat. Med.* 1, 69–73 (1995)
3. Hofman, A., Breteler, M.M.B., et al.: The Rotterdam Study: 2010 objectives and design update. *Eur. J. Epidemiol.* 24, 553–572 (2009)
4. Yuan, C., Lin, E., Millard, J., Hwang, J.N.: Closed Contour Edge Detection of Blood Vessel Lumen and Outer Wall Boundaries in Black-Blood MR Images. *Magn. Reson. Imaging* 17, 257–266 (1999)
5. Ladak, H.M., Thomas, J.B., Mitchell, J.R., et al.: A Semiautomatic Technique for Measurement of Arterial Wall from BBMRI. *Med. Phys.* 28(6), 1098–1107 (2001)
6. Jin, Y., Ladak, H.M.: Software for Interactive Segmentation of the Carotid Artery From 3D Black blood Magnetic Resonance Images. *Comput. Meth. Programs. Biomed.* 75(9), 31–43 (2004)
7. Adame, I.M., van der Geest, R.J., Wasserman, B.A., et al.: Automatic Segmentation and Plaque Characterization in Atherosclerotic Carotid Artery MR Images. *Magma. Magn. Reson. Mater. Phys. Biol. Med.* 16(5), 227–234 (2004)
8. 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)
9. Perona, P., Malik, J.: Scale Space and Edge Detection Using Anisotropic Diffusion. *IEEE Trans. Pattern. Anal. Mach. Intell.* 12, 629–639 (1990)
10. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. *Int. J. Comput. Vis.* 22(1), 61–79 (2004)
11. Hameeteman, K., Freiman, M., et al.: Carotid Lumen Segmentation and Stenosis Grading Challenge (2009)
12. Hoogeveen, R.M., Bakker, C.I.G., Viergever, M.A.: Limits to the Accuracy of Vessel Diameter Measurement in MR Angiography. *J. Magn. Reson. Imag.* 8(6), 1228–1235 (1998)
13. Lorigo, L.M., Faugeras, O.D., Grimson, W.E.L., et al.: CURVES: Curve evolution for vessel segmentation. *Med. Image. Anal.* 5(3), 195–206 (2001)
14. Antiga, L., Steinman, D.A.: Robust and objective decomposition and mapping of bifurcating vessels. *IEEE Trans. Med. Imag.* 23, 704–713 (2005)
15. The Vascular Modeling Tool Kit, <http://www.vmtk.org>
16. Tustison, N.J., Gee, J.C.: N4ITK: Nick's N3 ITK Implementation For MRI Bias Field Correction. *Insight Journal* (2009)
17. Dice, L.R.: Measures of the Amount of Ecologic Association Between Species. *Ecology* 26(3), 297–302 (1945)
18. Schaap, M., Metz, C.T., van Walsum, T., et al.: Standardized Evaluation Methodology and Reference Database for Evaluating Coronary Artery Centerline Extraction Algorithms. *Med. Image. Anal.* 13(5), 701–714 (2009)

Standing on the Shoulders of Giants: Improving Medical Image Segmentation via Bias Correction

Hongzhi Wang¹, Sandhitsu Das¹, John Pluta¹, Caryne Craige¹,
Murat Altinay¹, Brian Avants¹, Michael Weiner²,
Susanne Mueller², and Paul Yushkevich^{1,*}

¹ Departments of Radiology, University of Pennsylvania

² Department of Veterans Affairs Medical Center, San Francisco, CA

Abstract. We propose a simple strategy to improve automatic medical image segmentation. The key idea is that without deep understanding of a segmentation method, we can still improve its performance by directly calibrating its results with respect to manual segmentation. We formulate the calibration process as a bias correction problem, which is addressed by machine learning using training data. We apply this methodology on three segmentation problems/methods and show significant improvements for all of them.

1 Introduction

Automatic image segmentation plays an important role in medical applications. Due to the limitations of the imaging process and the difficulty of transferring manual segmentation protocols into algorithms, automatic segmentation is challenging. We show that without deeply understanding the limitations of an existing segmentation method, one easy/straightforward way to make improvements is through a calibration process to directly transfer its results closer to manual segmentations. To this end, we propose to use machine learning techniques to correct segmentation errors.

From a theoretical perspective, the segmentation errors produced by a segmentation algorithm can be categorized into two classes: 1) random errors and 2) consistent bias. The random errors are caused by random effects, e.g. imaging noises or random anatomical variations. They can be reduced by averaging techniques such as multi-atlas based segmentation. In this paper, we focus on addressing the other type of errors, consistent bias¹. Bias are systematic errors mostly caused by mistranslating manual segmentation protocols into the criteria followed by the automatic segmentation method. By definition, bias occurs consistently across different segmentation trials when certain conditions are met.

* This work was supported by the Penn-Pfizer Alliance grant 10295 (PY) and NIH awards K25 AG027785, R21 NS061111, R01 AG010897, and P01 AG12435.

¹ The meaning of *bias* in this paper is different from its common use to describe MRI field inhomogeneity. By bias, we mean those errors in the initial segmentation that are systematic, i.e., follow a pattern from training subject to training subject.

For example, a manual segmentation protocol may assign a specific label to a voxel if and only if a certain criterion, e.g. the voxels next to it all have low intensities, is met. However, because of the translation error an automatic method may follow a slightly different criterion, e.g. the average intensity of its neighbors is low. In this example, the automatic segmentation method makes errors whenever a voxel’s neighbors have a low average intensity but have at least one bright voxel.

Since bias occurs consistently, it is feasible to detect and correct them. Although it may be difficult to figure out the exact cause behind each bias, it is relatively easy to capture the patterns that are strongly correlated to the bias. Hence, one can detect bias via capturing the correlated patterns. For example, the example above demonstrates a simple bias whose correlated appearance pattern, i.e. a voxel’s neighbors have a low average intensity but have at least one bright voxel, can be learned using training images. In reality, the bias may appear in more complex and less intuitive patterns. Although it may be difficult for the human to identify such bias, most machine learning techniques are capable of providing satisfactory solutions.

In related work, Morra et al [3] use machine learning to directly learn how to perform segmentation. During training, they use intermediate classification results to improve the classifier’s performance. The main difference with our method is that they do not use any other segmentation methods and train the classifier from scratch. By contrast, our contribution lies in proposing the idea of improving the performance of existing segmentation algorithms relative to a specific manual segmentation protocol via learning-based bias correction. Our approach takes full advantage of other segmentation algorithms to simplify learning. To validate our method, we apply it to three segmentation problems/methods and show significant improvements for all of them.

2 Learning-Based Explicit Bias Correction (EBC)

To improve segmentation results produced by a segmentation method, we propose a two-step procedure for bias correction (see Fig. 1): 1) bias detection that finds the mislabeled voxels produced by the host segmentation method and 2) bias correction that corrects the mislabeled voxels found by bias detection.

2.1 Bias Detection as a Binary Classification Problem

Given a segmentation produced by a host segmentation method, our goal is to identify mislabeled voxels. With manual segmentations, it is straightforward to formulate the bias detection problem as a binary classification problem. For each label we train one classifier using all voxels assigned to this label to separate correctly labeled voxels from mislabeled.

To train classifiers, we use AdaBoost [2]. For effective learning, abundant informative features are crucial. The simplest feature is the raw image appearance, i.e. pixel-wise intensities. For more discriminative representations, textures are

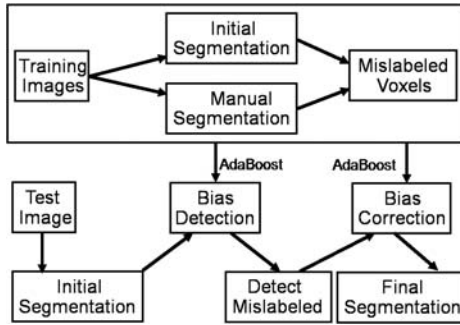


Fig. 1. Flow chart of our explicit bias correction approach

often used as well. One common approach to construct texture features is to use an over-complete description for each voxel and its neighborhood by convolving the image with a filter bank. In our experiment, for more efficiency we use the following features. We denote $A^{\Delta X}(i) = I(X_i + \Delta X) - \bar{I}$ to be the appearance feature for voxel i with coordinate X_i at the relative location ΔX . I is intensity. To compensate for different intensity ranges, we normalize the intensities by the average intensity of the region of interest (ROI), \bar{I} . To train a bias detection classifier for a label, the ROI contains all voxels assigned to the label by the host method (see experiments). More robust features with scale and rotation invariance can be used as well. Since the brain image data used in our experiments all have similar scales and orientations, we use these simple features.

Low level texture features can capture image related bias, e.g. the host segmentation method always makes errors when a certain appearance pattern occurs. To capture non-image related bias, e.g. the host method always shifts the segmentation a few voxels, we include the segmentation produced by the host segmentation method for learning. We denote these features by $L^{\Delta X}(i) = s(X_i + \Delta X)$, where s is the segmentation produced by the host method.

To include spatial information, we use the coordinate feature $S_X(i) = X_i - \bar{X}$, where \bar{X} are the average coordinates of the ROI. To enhance the spatial correlation, we include the joint feature obtained by multiplying spatial features with appearance and label features, i.e. $A^{\Delta X}(i)S_X(i)$ and $L^{\Delta X}(i)S_X(i)$. Overall, we use ~ 1000 features in all experiments.

For each feature, a weak classifier is constructed using a simple threshold. AdaBoost combines these weak classifiers into a single strong classifier.

2.2 Learning-Based Bias Correction

Bias detection outputs candidate mislabeled voxels. We need to reassign labels to them. Again, we use a learning-based method. Given mislabeled voxels in the initial segmentations, applying the same learning algorithm with the same features used for bias detection we train a binary classifier for each label to separate it from others. To assign a new label to a candidate mislabeled voxel detected

by bias detection, we re-evaluate the voxel by each bias correction classifier and assign the label whose corresponding classifier gives the strongest response to the voxel. Since bias correction is only for detected candidate mislabeled voxels, the computational cost is much lower than re-evaluating the whole image when the host segmentation method can produce accurate results.

2.3 Variants of the Learning Algorithm

In our method, we explicitly perform bias detection and bias correction. This strategy is efficient because for bias correction only the potentially mislabeled voxels need to be relabeled. One variant of our bias correction method is that we skip the bias detection step and directly perform bias correction on the initial segmentation. Instead of only using mislabeled voxels, we use all voxels in ROI for training. We call this method implicit bias correction (IBC). Note that IBC has higher computational complexity for both training and testing. IBC is closely related to [3], where instead of segmentation results produced by other segmentation methods the segmentation labels produced by the learning algorithm itself are included in the learning process.

One way to view the segmentation feature produced by other host segmentation methods is that like the low level texture filters any host segmentation method can be considered as a high level, task specific filter. If the host segmentation method works reasonably well, i.e. better than random guesses, the produced segmentation provides useful information for the segmentation task. To demonstrate the usefulness of host segmentation methods for learning, we compare with a variant of IBC that each classifier is learned without using segmentation results produced by any other segmentation methods. We call this variant the direct learning (DL) approach. So given training images and their manual segmentations, we train one classifier for each label to separate voxels belonging to this label from other voxels. The features used for DL, is only image and spatial features. For IBC and DL, the ROI is the whole segmentation produced by the host method plus some dilation. Dilation is necessary only when the background label needs to be corrected (see experiments for examples).

3 Experiments

We apply our methods to three segmentation problems. The problems are image registration based hippocampal segmentation, whole brain extraction using BET [7] and brain tissue segmentation using FAST [8].

3.1 Hippocampal Segmentation

The hippocampus plays an important role in memory function [6]. Macroscopic changes in brain anatomy, detected and quantified by magnetic resonance imaging (MRI), consistently have been shown to be highly predictive of AD pathology and highly sensitive to AD progression [5]. Compared to clinical measures and

neuropsychological testing, MRI-derived biomarkers require an order of magnitude smaller cohort size to detect disease-related changes over time. Accordingly, automatic hippocampus segmentation from MR images has been widely studied e.g. [13,4]. In this section, we test our methods with one semi-automatic hippocampal segmentation method [4].

We use the data in the Alzheimer’s Disease Neuroimaging Initiative (ADNI, www.loni.ucla.edu/ADNI). Our study is conducted using only 3 T MRI and only includes data from mild cognitive impairment (MCI) patients and controls. Overall, the data set contains 139 images. The image has 1.00×1.00 mm in plane resolution and 1.2 mm slice thickness. For cross validation evaluation, 70 subjects are randomly selected for training, and the remaining 69 for testing. The reported results are the average of 10 cross-validation experiments.

A landmark-guided atlas-based segmentation method [4] is applied to segment the hippocampi for each image. This method is designed to minimize user efforts while maximizing the benefit of human input to the algorithm. It requires a user to approximately label six key landmarks of the hippocampus through a user-interface. The partial labeling is combined with image similarity terms to guide volumetric diffeomorphic normalization between an individual brain and an unbiased template, with fully labeled hippocampi. It is shown that such simple human interactions help increase minimum performance levels relative to fully-automatic segmentation algorithms and provides high inter-rater reliability.

Whole hippocampal segmentation is a binary segmentation problem (we do experiments on left side and right side separately). Once the mislabeled voxels are identified we can fix them by simply switching their labels and vice versa. Hence, for the binary segmentation problem EBC is equivalent to IBC.

Since the results produced by [4] are accurate, we define the ROI for bias correction to be the initially segmented hippocampi plus one voxel dilation. On average, this ROI includes 99.5% hippocampal voxels. By contrast, the ROI obtained from the initial segmentation plus two voxel dilation covers 99.9% hippocampal voxels but also includes significantly more irrelevant voxels, which increases the chances for our bias correction to make mistakes. Since DL does not use the results produced by [4], DL should take the whole image as ROI. However, for direct comparison with our methods, we apply the same ROI for DL. Since the ROI excludes significant non-hippocampus distracters, using ROI simplifies the learning problem. Hence, in this experiment DL partially benefits from the results produced by [4].

On average, each hippocampus contains 1603 voxels. [4] produces 465 mislabeled voxels. Note that the errors include hippocampal voxels mislabeled as background and background voxels mislabeled as hippocampi. Our bias correction method achieved 35.7% fewer errors (299 mislabeled voxels). Using the larger ROI, i.e. initial segmented hippocampi plus two voxel dilation, results in slightly worse results of 305 mislabeled voxels. By contrast, DL produces worse segmentations with 523 mislabeled voxels. Fig. 2 shows example segmentation results. In terms of average Dice overlaps, [4], DL and IBC/EBC resulted in 0.862, 0.832 and 0.903 respectively.

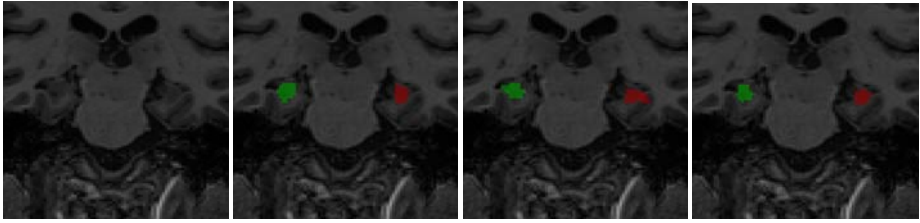


Fig. 2. Hippocampal segmentation. Left to right: original image, manual segmentation, segmentation produced by [4], after bias correction.

3.2 Brain Extraction/Segmentation in MR Images

In this section, we test brain segmentation. The data set contains 18 T1-weighted MR brain images and their manual segmentations, which are available at the Internet Brain Segmentation Repository (IBSR). The manual segmentation contains labels for gray and white matter and ventricles. These images have been positionally normalized into the Talairach space (rotation only) and have been preprocessed by intensity inhomogeneity correction routines. These images have the same slice thickness of 1.5 mm with three in plane resolutions: eight have 0.94×0.94 mm; six have 1.0×1.0 mm; four have 0.84×0.84 mm.

Using this data, we test two methods: the Brain Extraction Tool (BET) [7], and the FMRIB’s Automated Segmentation Tool (FAST) [8]. For cross validation evaluation, 9 subjects are randomly selected for training, and the remaining 9 for testing. The results are the average of 10 cross-validation experiments.

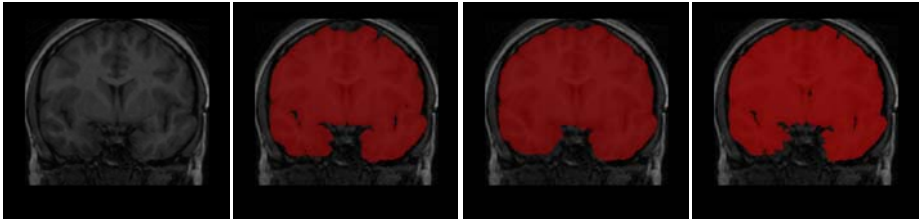


Fig. 3. Brain extraction. Left to right: original image, manual brain extraction, initial brain extraction by BET, after bias correction.

Brain extraction. The BET algorithm is applied with the default parameter setting to segment the images into brain and non-brain regions. Again, for this binary segmentation problem EBC is equivalent to IBC. Since the BET algorithm is relatively accurate and most segmentation errors are mislabeling background voxels as brain tissues, we define a ROI for bias correction by performing a one-voxel dilation of the BET result, similar to how the ROI was defined in the binary hippocampus segmentation experiment. On average, this ROI covers 99.3% manually labeled brain. DL still partially benefits from BET’s results by using the same ROI.

On average, each brain contains 9.7×10^5 voxels. BET produces 1.1×10^5 mislabeled voxels. Our bias correction method achieved 29% fewer errors (8.0×10^4 mislabeled voxels). By contrast, DL produces worse segmentations with 9.1×10^4 mislabeled voxels. In terms of average Dice overlaps, BET, DL and IBC/EBC resulted in 0.948, 0.956 and 0.961 respectively.

Brain tissue segmentation. In this experiment, the FAST [8] algorithm is applied for segmenting gray matter, white matter and cerebrospinal fluid (CSF) for all 18 subjects used in the previous experiment. To apply FAST, the binary brain segmentation is assumed to be provided.

Since the manual segmentation in IBSR merges CSF outside ventricles into gray matter (see Fig. 4), the CSF produced by FAST that overlaps gray matter in manual segmentation is also considered correct. For quantitative evaluations, we merge the CSF into gray matter for both manual and automatic segmentation and compare the consistency of white matter and merged gray matter. See Fig. 4 for segmentation examples.

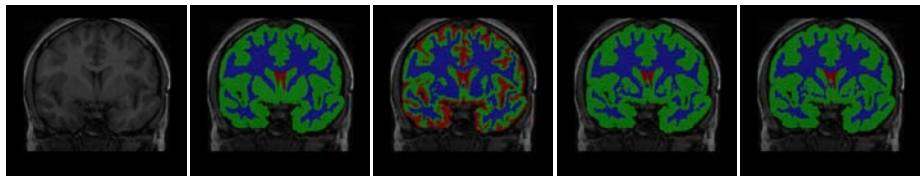


Fig. 4. Brain tissue segmentation. left to right: original image, manual, initial segmentation produced by FAST, after bias correction by IBC, and EBC.

Out of the average brain volume, 9.7×10^5 voxels, the FAST algorithm produces 8.9×10^4 mislabeled voxels. For EBC, the bias detection step achieved the precision($\frac{\# \text{ of correct detection}}{\# \text{ of detection}}$) of 92% with the recall($\frac{\# \text{ of correct detection}}{\# \text{ of true bias}}$) of 84%. The bias correction step correctly classified 91% of the detected mislabeled voxels. Overall, EBC achieved 21% fewer errors (7.0×10^4 mislabeled voxels). IBC achieved 17% fewer errors (7.4×10^4 mislabeled voxels). Note that EBC outperforms IBC with even fewer computational costs. By contrast, DL produces worse segmentations with 8.1×10^4 mislabeled voxels. Table 1 reports the average Dice overlaps. Like in the previous experiments, our bias correction methods outperformed DL and the host segmentation method.

Table 1. Brain tissue segmentation results in Dice overlap

method	FAST(Dice)	DL(Dice)	IBC(Dice)	EBC(Dice)
gray	0.936	0.944	0.948	0.951
white	0.862	0.891	0.899	0.905

4 Discussion

Combining segmentations/methods has been proven to be a good strategy to improve performance. One can view our bias correction as a combining method that integrates a pure machine learning based segmentation method with the host segmentation method. One main difference from previous combining methods is that the machine learning method can automatically adapt itself through training to optimally combine with the host segmentation method. As demonstrated in our experiments, as long as the machine learning algorithm uses complementary information to the host segmentation methods the combined results consistently outperform the host segmentation methods and the machine learning method when applied separately. The information integration interpretation also suggests that using the same machine learning algorithm used in bias correction to improve the results produced by our bias correction may not give as much improvement because of the significant information overlap. However, a learning method using different features or learning models may still help.

References

1. Carmichael, O.T., Aizenstein, H.A., Davis, S.W., Becker, J.T., Thompson, P.M., Meltzer, C.C., Liu, Y.: Atlas-Based Hippocampus Segmentation In Alzheimers Disease and Mild Cognitive Impairment. *NeuroImage* 27(4), 979–990 (2005)
2. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi, P.M.B. (ed.) *EuroCOLT 1995*. LNCS, vol. 904, pp. 23–27. Springer, Heidelberg (1995)
3. Morra, J., Tu, Z., Apostolova, L., Green, A., Toga, A., Thompson, P.: Automatic subcortical segmentation using a contextual model. In: *Proceedings of the 11th international Conf. on Medical Image Computing and Computer-Aided Intervention*, pp. 194–201 (2008)
4. Pluta, J., Avants, B., Glynn, S., Awate, S., Gee, J., Detre, J.: Appearance and Incomplete Label Matching for Diffeomorphic Template Based Hippocampus Segmentation. *Hippocampus* 19(6), 565–571 (2009)
5. Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C.: Mapping the evolution of regional atrophy in Alzheimer’s disease: unbiased analysis of fluidregistered serial MRI. *Proc. Natl. Acad. Sci. U.S.A* 99(7), 4703–4707 (2002)
6. Squire, L.R.: Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99, 195–231 (1992)
7. Smith, S.: Fast robust automated brain extraction. *Human Brain Mapping* 17(3), 143–155 (2002)
8. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. on Medical Imaging* 20(1), 45–57 (2001)

Layout Consistent Segmentation of 3-D Meshes via Conditional Random Fields and Spatial Ordering Constraints

Alexander Zouhar^{1,2}, Sajjad Baloch¹, Yanghai Tsin¹,
Tong Fang¹, and Siegfried Fuchs²

¹ Siemens Corporate Research, Inc., Princeton, USA

² Dresden University of Technology, Dresden, Germany

Abstract. We address the problem of 3-D Mesh segmentation for categories of objects with known part structure. Part labels are derived from a semantic interpretation of non-overlapping subsurfaces. Our approach models the label distribution using a Conditional Random Field (CRF) that imposes constraints on the relative spatial arrangement of neighboring labels, thereby ensuring semantic consistency. To this end, each label variable is associated with a rich shape descriptor that is intrinsic to the surface. Randomized decision trees and cross validation are employed for learning the model, which is eventually applied using graph cuts. The method is flexible enough for segmenting even geometrically less structured regions and is robust to local and global shape variations.

1 Introduction

Surface segmentation involves the partitioning of a surface mesh into non overlapping sub-meshes, each representing a part of the underlying 3-D object. Such decomposition into parts has traditionally been addressed through metrics derived from geometric properties of a surface, such as curvature, dihedral angles, and protrusion [1,2,3]. Such raw geometric properties are typically local or regional in nature, and completely ignore the compositional arrangement of constituent parts. Consequently, the accuracy of these methods gets challenged in two major ways: (1) previously unseen shape variations may, in general, result in erroneous labeling if the descriptor is not strong enough, (2) neighboring surface regions with similar geometric properties could actually belong to different parts, i.e., they are interpreted differently. These issues may be resolved through explicit constraints on the relative spatial arrangement of neighboring labels.

In this paper, we present an object model-based approach for the segmentation of categories of 3-D objects with known compositional structure. Our method employs a labeling of various parts, which densely covers a 3-D surface mesh. The parts are derived from a semantic interpretation of compact non-overlapping regions on a surface. For instance, typical parts of a car include hood, trunk, roof, and so on.

The label distribution is modeled using a CRF whose undirected graph is isomorphic to the mesh topology. Each label variable is associated with a rich local shape descriptor that is intrinsic to the surface. Prior knowledge about the object structure is incorporated via constraints on the spatial layout of neighboring labels in a mesh graph. For instance, the windshield and the roof are neighbors, but the roof must be detected “above” the windshield, not vice-versa. In short, the proposed layout constraints directly address the limitations of the aforementioned methods by penalizing the assignment of ambiguous and/or spatially incorrect labels.

The capability of the model is demonstrated on the example of human outer ear 3-D surface meshes. Mesh segmentation is highly significant in digital hearing aid design [4]. It serves as a pre-requisite for automatic surface manipulation in order to reduce the amount of human intervention in the design process [5].

1.1 Related Work and Our Contribution

An extensive review of algorithms for 3-D mesh segmentation can be found in [6]. Most of these methods achieve segmentation exclusively via raw geometric information derived from a surface. Instead, our method utilizes a rich descriptor coupled with spatial ordering constraints. There has been little work that exploits explicit constraints on the spatial arrangement of parts. Coupled with our descriptor, they help in resolving over- and under-segmentation even if the parts have boundaries in geometrically less structured regions. [7] used a CRF model for partially occluded objects in 2-D images, which may extend directly to 3-D volumetric data. However, its application to 3-D meshes is more complicated due to the absence of natural (grid) directions on a curved 2-manifold embedded in 3-D space.

2 Layout Consistent Segmentation of Surfaces

Given a surface representation of a 3-D object in the form of a polygonal mesh $X := (V, E)$, where $V = \{v_i\}$ is the set of vertices, and $E = \{e_k\}$ is the set of edges. Our goal is to infer a labeling $h : V \rightarrow L, v_i \mapsto h_i := h(v_i)$, that assigns a label h_i to each vertex $v_i \in V$.

Inference is usually based on some characteristic observations on a mesh X , in the form of local shape descriptors $\{x_i\}$. In this paper, we develop a model that structures the labels $H := \{(v_i, h_i) | v_i \in V\}$ in the form of an undirected graph G . We assume that G is isomorphic to X .

The relationship between neighboring labels is directly associated with the relationship between the corresponding local mesh characteristics. This allows the compositional constraints to be imposed through the joint probability distribution of (X, h) , which in turn may be modeled as a *Markov Random Field* (MRF) [8]. However, this requires that the local mesh characteristics are statistically independent from those of their immediate neighbors. In this paper, we allow a richer shape representation involving a complex neighborhood. Consequently, optimizing the MRF model becomes difficult, if not intractable.

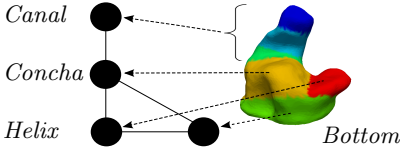


Fig. 1. Part adjacency graph and corresponding mesh representation of a human outer ear. Parts (circles) are linked (solid) according to adjacency of anatomical regions, i.e., *Canal*, *Concha*, *Helix*, and *Bottom*. The colors indicate the anatomical interpretation (dashed) of the parts. The *Canal* is composed of 3 subparts.

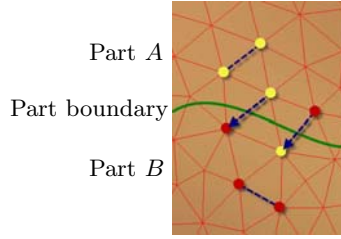


Fig. 2. Neighboring labels on G . Two nodes are connected via a link (dashed). Part membership of the nodes is indicated by the color, i.e., yellow means part *A* and red means part *B*. The labeling along the directed link from red to yellow is incorrect.

The problem can be resolved by the conditional distribution model of CRFs [9]. The joint distribution over elementary events (X, h) may be written as

$$P(X, h; \theta) = P(X; \theta)P(h|X; \theta). \quad (1)$$

The conditional distribution $P(h|X; \theta)$ is sufficient for label estimation, so that the marginal $P(X; \theta)$ can be dropped. Because the model is conditional, complex dependencies among input variables do not need to be explicitly represented which allows the use of rich local shape descriptors. Assuming pairwise statistical dependencies between neighboring labels a model for the conditional distribution $P(h|X; \theta)$ is given by

$$P(h|X; \theta) = \frac{1}{Z(\theta, X)} \prod_{(i,j)} \xi_{ij}(h_i, h_j, X; \theta) \prod_i \nu_i(h_i, X; \theta), \quad (2)$$

where ν_i is a unary potential function that models the posterior distribution of the labels at vertex v_i in a mesh X . ξ_{ij} is the pair-wise potential that constrains the spatial arrangement over the neighboring labels. Note that the partition function Z can be dropped, since we are not interested in actual probabilities. Consequently, Eqn. (1) reduces to an energy functional whose maximization leads to the optimal labeling. A “price” to pay here is that the potential functions tend to be more complex than those in a pair-wise MRF. On the other hand, the dependency on X allows us to exploit the mesh geometry for explicitly enforcing constraints on the spatial layout of neighboring labels. In the following discussion, we develop these layout constraints through a part adjacency graph.

Definition 1 Part Adjacency Graph (PAG). A PAG for a class of objects is an undirected graph with nodes representing constituent object parts. Undirected links between nodes indicate which parts are adjacent to which other parts. The underlying binary relation is symmetric and reflexive.

Fig. 1 depicts a simplified PAG of the human outer ear together with the corresponding anatomical interpretation of the parts. A natural constraint for neighboring labels in a mesh graph G can be derived from its PAG.

Definition 2 Soft Layout Consistency (SLC). Let G be a graph representing the structure between label variables, and let h_i and h_j , ($i \neq j$) be the labels of two neighboring vertices. Let G_{adj} be the PAG for the class of objects under consideration. The labels h_i and h_j are soft layout consistent if $(h_i, h_j) \in G_{adj}$.

Definition 2 explicitly ensures local spatial coherence, and implicitly encourages regional compatibility of the labels. However, the constraint is “soft” in the sense that it does not capture the relative spatial arrangement of neighboring parts (e.g., above, below, left, right, back, front). For instance, an inferred labeling on a human body surface may put the neck on top of the head, since this arrangement will still be consistent according to Definition 2. Figure 2 illustrates the main idea on the example of two parts. Note that while moving along the directed link, a change from label $h_i = A$ to a label $h_j = B$ is correct. However, a change from label $h_i = B$ to $h_j = A$ is incorrect.

For two neighboring labels h_i and h_j , $i \neq j$ there are three possible types of transitions:

Part interior: $h_i = h_j$,

Adjacent parts: $(h_i, h_j) \in G_{adj}$, $h_i \neq h_j$. Note the symmetry here.

Inconsistent: $(h_i, h_j) \notin G_{adj}$, $h_i \neq h_j$.

Erroneous label assignments as shown in Fig. 2 should be penalized, but this requires an understanding of the spatial order or arrangement of the neighboring labels. Similar to [7] the value of the pairwise potential varies according to transition type as follows:

$$-\log \xi_{ij}(h_i, h_j, X; \theta) = \begin{cases} 0 & \text{Part interior,} \\ \gamma_1 & \text{Adjacent parts,} \\ \gamma_2 & \text{Inconsistent.} \end{cases} \quad (3)$$

Let $g(h_i, h_j) \in \{0, 1\}$ be denoted as *directional consistency function* that characterizes the consistency of soft-layout consistent labels h_i and h_j in a mesh graph G (assuming $h_i \neq h_j$). We define the behavior of $g(\cdot)$ as follows. With reference to the spatial order defined by the arrow in Fig. 2, if $h_i = A$, $h_j = B$ then $g(A, B) = 1$, and if $h_i = B$, $h_j = A$ then $g(B, A) = 0$. We discuss the choice of the directional consistency function in Section 3.3.

Definition 3 Strict Layout Consistency (StLC). Let h_i and h_j , ($i \neq j$) be two neighboring labels in a mesh graph G . Furthermore, let G_{adj} be the PAG for the class of objects under consideration. The labels h_i and h_j are strict layout consistent if one of the following conditions holds, (1) $h_i = h_j$, (2) $(h_i, h_j) \in G_{adj} \wedge g(h_i, h_j) = 1$.

For two neighboring labels h_i and h_j , $i \neq j$ we define transitions to be one of the following types:

Part interior: $h_i = h_j$,

Adjacent parts: $(h_i, h_j) \in G_{adj} \wedge g(h_i, h_j) = 1$,

Inconsistent: $(h_i, h_j) \notin G_{adj} \vee g(h_i, h_j) = 0$.

The consistency function $g(\cdot)$ needs to know the spatial arrangement of the labels h_i and h_j on a mesh X . Hence, for StLC, the pairwise potentials in Eqn. (3) make use of the mesh geometry X (see Section 3.3).

3 Learning and Inference

In this section, we learn the potentials in Eqn. (1) using a supervised algorithm. This requires that all training surfaces are pre-labeled. Once the potentials are learned, the negative logarithm of Eqn. (1) may be efficiently minimized using the α -expansion algorithm [10].

3.1 Local Shape Descriptors

The unary potentials use “local” information of a mesh. As mentioned earlier a mesh X is represented by a set of local shape descriptors $\{x_i\}$. Each local shape descriptor x_i is associated with a vertex v_i . The choice of descriptor is driven by invariance requirements to geometric transformations of a surface and the need for robustness to non-ideal conditions, such as noisy 3-D scans.

We propose the *Geodesic Shape Context* (GSC) which is invariant to rigid transformations and scale. A GSC is obtained by binning the geodesic distances measured from a vertex v_i to all other vertices on a mesh [11].

We now define two types of features at each vertex: (1) the cumulative mean curvature of vertices inside each bin. This yields a function of cumulative mean curvature versus the bin index, (2) a difference between the cumulative curvatures of the neighboring bins. Both features are normalized between 0 and 1. Each vertex is hence represented by two feature vectors, which we collectively refer to as the GSC. Note that the GSC at a vertex v_i has a global support region, i.e., it covers the entire mesh relative to v_i [1]. The GSC leads to a very rich representation of the underlying geometry compared with [12] capturing variations in curvature such as the amount of bending.

3.2 Unary Potentials

Local shape descriptors, including GSCs, normally reside in a high dimensional space which makes learning of a model for the unary potentials challenging due to the amount of required training data. We use randomized binary decision trees for the unary potentials. The procedure for learning a binary decision tree and inference of $\nu_i(h_i, X; \theta)$ are similar to [7]. A set of trained decision trees (a forest) returns a distribution over the labels at a vertex $v_i \in V$.

¹ We call it local shape descriptor only to reflect that it is computed at a location v_i and is not a global representation of the entire shape.

3.3 Pairwise Potentials

The parameters of the pairwise potentials, γ_1, γ_2 are learned using cross validation by a search over a reasonable range of positive values of the parameters.

Spatial ordering of neighboring parts. We characterize the spatial order of neighboring labels h_i, h_j , ($h_i \neq h_j$) in terms of the relative spatial arrangement of neighboring parts on a mesh. First, we define the *geodesic center* of a part on a mesh as the average geodesic distance of all vertices in this part from a pole. To this end, we assume that a consistent pole can be detected on each mesh. The average of this measure across all training surfaces is denoted as *expected geodesic center* of the part.

The notion of ordering is established by noting that a configuration of neighboring labels is more likely when each label is assigned closer to its expected geodesic center. Formally, let q_i, q_j denote the geodesic distances of v_i and v_j from the pole, and let q_A and q_B denote expected geodesic centers of parts A and B . q_A and q_B are pre-computed across training surfaces as mentioned in the previous paragraph. We define the directional consistency function $g(h_i, h_j)$ for neighboring vertices v_i , and v_j as:

$$g(h_i, h_j) = \begin{cases} 1 & |q_i - q_A| + |q_j - q_B| < |q_i - q_B| + |q_j - q_A|, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Note, that the mesh X is required as input of the pairwise potentials in order to evaluate the directional consistency function among neighboring labels.

3.4 Inference

The α -expansion algorithm was used for approximate MAP inference of the labels due to its convergence properties (see [10,13,14] for details).

4 Evaluation

We validated the proposed method on a data set of 216 outer ear impressions, which in turn were laser scanned to reconstruct 3-D triangular mesh representations. The resulting meshes had open boundaries and were composed of roughly 5000 vertices. An expert was asked to manually label the meshes along anatomical lines using a CAD software system. In this way 6 compact regions are obtained as illustrated in Fig. 1. Such anatomical regions play a significant role in the design of personalized hearing aid devices [4]. Note that various boundaries divide geometrically less structured regions into parts, which makes segmentation challenging.

The data set was randomly divided into a training set of 180 meshes and a test set of 36 meshes. First, the unary potentials were learned on the training set. A resolution of 20 bins turned out to be reasonable for the GSCs. Next, the pairwise potential parameters were found via cross-validation against the

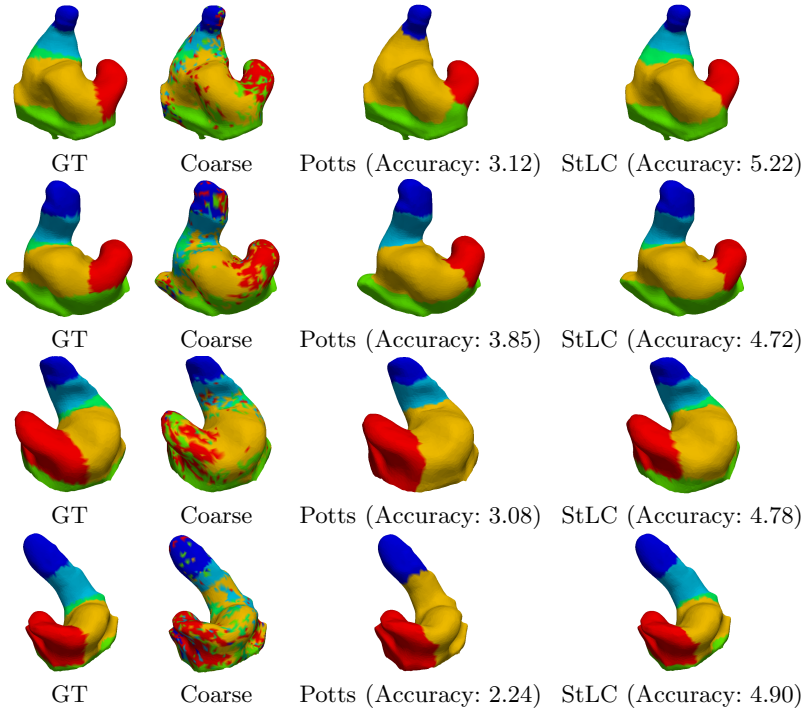


Fig. 3. Example segmentations. “Coarse” means: $\gamma_1 = \gamma_2 = 0$. For comparison the segmentation accuracy is depicted for the Potts model and the StLC model.

training set. The final parameters for StLC were: $\gamma_1 = 1$, $\gamma_2 = 4$. We provide comparison with the Potts model [15] with the cost of dissimilar neighboring labels equal to one and zero otherwise.

Some examples are depicted in Fig. 3. The first column shows the ground-truth (GT). The second column presents the coarse segmentation result, achieved as the MAP estimate from the decision trees. This corresponds to $\gamma_1 = \gamma_2 = 0$, which amounts to no smoothness or layout constraints. The third column represents the Potts model with $\lambda = 1$. Notice how various regions have become smoother. The fourth column represents the StLC model. The results indicate that StLC yields the best agreement with the GT.

Quantitative comparison was carried out as follows. For a label, first the intersection and the union of the estimated region and the corresponding GT is computed. The ratio of area of the former to that of the latter yields a measure of segmentation accuracy per part. The *segmentation accuracy* of a test candidate is defined as the sum of the part scores. The average of this measure across all test individuals using the coarse model was 2.31, for the Potts model 2.97, and for the StLC model 4.52 with 6 being the best score.

5 Conclusions

We have presented an object model-based 3-D mesh segmentation algorithm for 3-D objects with known part structure, and evaluated its performance on a data set of outer ear 3-D meshes with promising results. The method is particularly attractive for the segmentation of organs where the layout of different anatomical regions is already known, such as functional segmentation of the brain.

References

1. Golovinskiy, A., Funkhouser, T.: Randomized cuts for 3D mesh analysis. *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia)* 27 (2008)
2. Shapira, L., et al.: Consistent mesh partitioning and skeletonisation using the shape diameter function. *Visual Computing* 24 (2008)
3. Liu, R., et al.: A part-aware surface metric for shape analysis. *Eurographics* 28 (2009)
4. Slabaugh, G., et al.: 3D shape modeling for hearing aid design. *IEEE Signal Processing Magazine* (2008)
5. Zouhar, A., et al.: Freeform shape clustering for customized design automation. In: *ICCV Workshop on 3-D Digital Imaging and Modeling* (2009)
6. Agathos, A., et al.: 3D mesh segmentation methodologies for CAD applications. *Computer-Aided Design and Applications* 4 (2007)
7. Winn, J., Shotton, J.: The layout consistent random field for recognizing and segmenting partially occluded objects. In: *CVPR* (2006)
8. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI* 6 (1984)
9. Lafferty, J., et al.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *ICML* (2001)
10. Boykov, Y., et al.: Efficient approximate energy minimization via graph cuts. *PAMI* (2001)
11. Surazhsky, V., et al.: Fast exact and approximate geodesics on meshes. *SIGGRAPH* (2005)
12. Shi, Y., et al.: Direct mapping of hippocampal surfaces with intrinsic shape context. *Neuroimage* (2007)
13. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *PAMI* (2002)
14. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI* (2004)
15. Potts, R.: Some generalized order-disorder transformation. In: *Proceedings of the Cambridge Philosophical Society* (1952)

Cross-Visit Tumor Sub-segmentation and Registration with Outlier Rejection for Dynamic Contrast-Enhanced MRI Time Series Data

G.A. Buonaccorsi^{1,3}, C.J. Rose^{1,3}, J.P.B. O'Connor^{1,3}, C. Roberts^{1,3}, Y. Watson^{1,3},
A. Jackson^{1,3}, G.C. Jayson^{2,3}, and G.J.M. Parker^{1,3}

¹ Imaging Science and Biomedical Engineering, School of Cancer and Imaging Sciences, University of Manchester, Manchester, United Kingdom

² CRUK Dept of Medical Oncology, Christie Hospital, Manchester, United Kingdom

³ The University of Manchester Biomedical Imaging Institute, University of Manchester, Manchester, United Kingdom

Abstract. Clinical trials of anti-angiogenic and vascular-disrupting agents often use biomarkers derived from DCE-MRI, typically reporting whole-tumor summary statistics and so overlooking spatial parameter variations caused by tissue heterogeneity. We present a data-driven segmentation method comprising tracer-kinetic model-driven registration for motion correction, conversion from MR signal intensity to contrast agent concentration for cross-visit normalization, iterative principal components analysis for imputation of missing data and dimensionality reduction, and statistical outlier detection using the minimum covariance determinant to obtain a robust Mahalanobis distance. After applying these techniques we cluster in the principal components space using k -means. We present results from a clinical trial of a VEGF inhibitor, using time-series data selected because of problems due to motion and outlier time series. We obtained spatially-contiguous clusters that map to regions with distinct microvascular characteristics. This methodology has the potential to uncover localized effects in trials using DCE-MRI-based biomarkers.

Keywords: DCE-MRI, PCA, k -means, tracer kinetic modeling, image registration, imputation, outlier detection, minimum covariance determinant.

1 Introduction

Biomarkers derived from quantitative Dynamic Contrast-Enhanced MRI (DCE-MRI) data are used in clinical trials to support early decisions on the viability of emerging anti-angiogenic and vascular-disrupting agents [1]. A common approach is to acquire a time series of 3-D images at regular time intervals bracketing the injection of a contrast agent and then apply voxel-by-voxel analyses to generate 3-D parameter maps, e.g. using tracer-kinetic modeling [2]. Standard practice [1] is to report statistics for a volume of interest (VOI) that identifies the target tissues, e.g. a whole tumor. This practice discards the spatially-heterogeneous information in the parametric maps. Several recent studies have presented quantitative analyses that aim to describe this heterogeneous structure [3-5].

Our proposal is to partition tumors into 3-D sub-regions using a novel segmentation based on the DCE time series data. To obtain a consistent partitioning for multiple scan visits, we pool data from all visits for a given patient. Thus we may apply statistical analyses to tumor sub-regions to reveal localized variations in treatment response that would be masked in a whole-tumor VOI analysis. Our method requires no prior knowledge or assumptions about the form of the time series data.

A preliminary study [6] found that segmentation using dimensionality reduction by standard principal components analysis (PCA) and clustering by standard k -means was not feasible with motion-corrupted or missing data and was highly susceptible to outliers. In this study we present a procedure incorporating registration, missing value imputation and outlier detection and we demonstrate its robustness.

2 Methods

2.1 Data Acquisition and Data Set Selection

In a clinical trial of an angiogenesis inhibitor [7], patients had 6 MRI visits: 2 pre-treatment scans, within 7 days prior to treatment; and 4 post-treatment scans (4 hours, 2, 8 and 12 days after treatment). All lesions were liver metastases arising from primary colorectal tumors. At each visit we acquired 3-D spoiled gradient echo (SPGR) images on a Philips 1.5 T Intera scanner for baseline T_1 estimation (3 images with flip angles of 2° , 10° and 30°) and for the dynamic time series (75 images: flip angle 20° , temporal resolution 4.97 s, voxel matrix $128 \times 128 \times 25$). Omniscan (Amersham Health, Amersham) was injected as a single bolus (dose 0.1 mmol/kg) after the 5th dynamic image, at a rate of 3 ml/s using a power injector (Spectris MR). To compensate for visit-to-visit variations in bolus arrival time we removed all time series images prior to the first appearance of contrast agent in the aorta in a central slice, giving a variable number (T) of time points (range 65-68). A research radiographer manually defined 3-D tumor VOIs on co-localized T_1 - and T_2 -weighted images.

To test the segmentation, we selected 3 tumors in 3 patients with specific data characteristics: Patient 1 had significant motion corruption but few time series that were sufficiently different from the norm to be classed as outliers; Patient 2 had low motion but several outliers and Patient 3 had both motion and outliers.

2.2 Motion Correction

DCE-MRI presents difficulties for established registration methods, which readily deal with contrast variations but may fail when new image features arise due to contrast enhancement. We used tracer-kinetic model-driven registration (TKMDR) for the time-series images—for details of the registration procedure see [8]. Standard TKMDR does not correct motion of the pre-contrast variable flip-angle images relative to the time series, leaving the possibility of misregistration of $T_1(t)$ maps and leading to corruption of the cross-visit normalization (Section 2.3). We therefore additionally used a linear registration [9] to align each variable flip-angle image to the pre-contrast TKMDR synthetic target images.

2.3 Cross-Visit Normalization

Raw DCE-MRI data from different patient visits are subject to unpredictable variations in scanner gain and potentially in native tumor relaxation time characteristics, so before pooling we normalized the time-series data to ensure that the scaling differences did not cause voxels from different visits to map to distinct clusters. Tracer kinetic models are typically expressed in terms of the tracer concentration $[CA](t)$ [2], which we used to normalize the data based on the physics of the data acquisition.

The conversion required the equation:

$$[CA](t) = \frac{1}{r_1} \left(\frac{1}{T_1(t)} - \frac{1}{T_1(0)} \right), \quad (1)$$

where r_1 is the spin-lattice relaxivity constant, and t is time. We used the standard SPGR equation [10] to estimate the pre-contrast longitudinal relaxation time, $T_1(0)$, from the variable-flip angle images and to derive $T_1(t)$ from the dynamic images.

2.4 Dimensionality Reduction with Missing Value Imputation

Each single-voxel $[CA](t)$ series is a vector in a T -dimensional data space ($T \geq 65$), and it was desirable to reduce dimensionality prior to clustering. In addition, particularly in the presence of noise, low T_1 values may give rise to physiologically unfeasible $[CA](t)$ values from equation (1)—filtering to remove the most extreme values resulted in missing data for the corresponding vector elements. We employed a modified PCA to address both issues.

The $N \times T$ data matrix for PCA pooled the time series data for each voxel of the tumor VOI for each visit for the given patient (N is the number of voxels in the pooled VOIs). Missing-value imputation was based on approximate reconstruction of the data via PCA, using the equation:

$$\bar{y} = \sum_{t=1}^r w_t \bar{p}_t + \sum_{t=r+1}^T w_t \bar{p}_t = \hat{y} + \bar{\epsilon} \quad (2)$$

where \bar{y} is a T -dimensional time series data vector, the subscript t is the time point index, T is the number of time points, \bar{p}_t is a T -dimensional principal component (PC) i.e. an eigenvector (obtained via singular value decomposition) of the sample covariance matrix for the data, w_t is a scalar weighting factor (derived from the PCA), \hat{y} is the estimate of \bar{y} obtained using r PCs (see below) and $\bar{\epsilon}$ is a T -dimensional residual vector. For each missing element of each \bar{y} , imputation required an initial estimate that was iteratively refined by replacement with \hat{y} until the sum of the elements of $\bar{\epsilon}$ for all missing values became arbitrarily small [11]. Standard practice initializes using the data matrix row or column means [12], but as time-series data are correlated we initialized using the mean of the adjacent non-missing time points. To obtain a close approximation, we selected r such that the PCA explained 97.5% of the variance. We then back-substituted the imputed values into the original time series in place of the missing values. We used the resulting PCs for dimensionality reduction,

retaining the subset of d principal components indicated by examination of log-eigenvalue plots ($d \neq r$: d ranged from 3 to 7 PCs and r was typically around 40 PCs).

2.5 Outlier Detection

The filtering of erroneous $[CA]$ values (Section 2.4) was extremely conservative, to avoid unintended distortion of the data. The remaining artefactually high $[CA]$ values resulted in time series that deviated sufficiently from the norm to be considered outliers. These outliers may be detected in the multivariate space defined by the retained PCs by thresholding the Mahalanobis distance (M) at a critical value (M_C) obtained from the χ^2 distribution with degrees of freedom (d) equal to the number of PCs retained for dimensionality reduction at a chosen percentile α (typically 97.5%) [13]:

$$M = (p - \bar{p})' \Sigma^{-1} (p - \bar{p}); \quad M_C = \text{inv} \left(\sqrt{\chi_d^2(\alpha)} \right) \quad (3)$$

where Σ is the sample covariance matrix and the prime indicates matrix transposition. The vectors p and \bar{p} are the truncated d -dimensional PCs from Section 2.4.

If \bar{p} and Σ are estimated as the classical mean and sample covariance matrix then each will be strongly affected by the outliers we wish to detect (the masking problem [13]), so we employed a robust Mahalanobis distance using the minimum covariance determinant (MCD) estimators of \bar{p} and Σ [14], as obtained with the FAST-MCD algorithm from the Matlab Library for Robust Analysis [15] (we increased α to 99.99% to reduce the potential misclassification of valid data as outliers). For the purposes of this study, any voxel time series with $M_{\text{voxel}} > M_C$ was considered to be an outlier, and the entire $[CA](t)$ series was removed from further analysis.

2.6 Clustering

We performed 10 repetitions of multi-start k -means clustering [16] in the data space of the retained principal components. Initial cluster centers were chosen randomly—different randomizations could converge to different solutions, so we retained the solution with the minimum sum of the squared Euclidean distance from each data point to its cluster mean vector. Clustering was done separately for each patient, but the data from all visits for the same patient were pooled. We set $k = 7$ to reflect typical tumor structure in a DCE-MRI data set: voxels could correspond to tumor enhancing rim, tumor non-enhancing core or surrounding liver and to partial voluming among these tissue types. Alternative arrangements are clearly possible. Only the retained principal components of the $[CA](t)$ series data were used in the clustering procedure—no spatial regularization or other spatial (e.g. neighborhood or connectivity) information was used, and clusters were therefore not constrained to be spatially-contiguous. For visualization purposes, we gave each cluster an integer label.

3 Results

Figure 1 shows sample imputation results for two voxel time series, each of which had 3 missing values. The imputed values generally lie within the range of the time series data, and do not significantly alter its observed pattern.

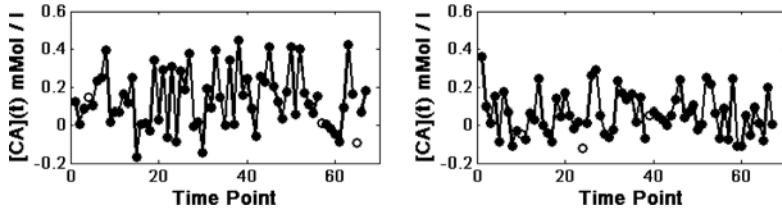


Fig. 1. Contrast agent concentration time series before (line with solid circles) and after (open circles) imputation of missing data for two voxel time series (Patient 3, post-registration)

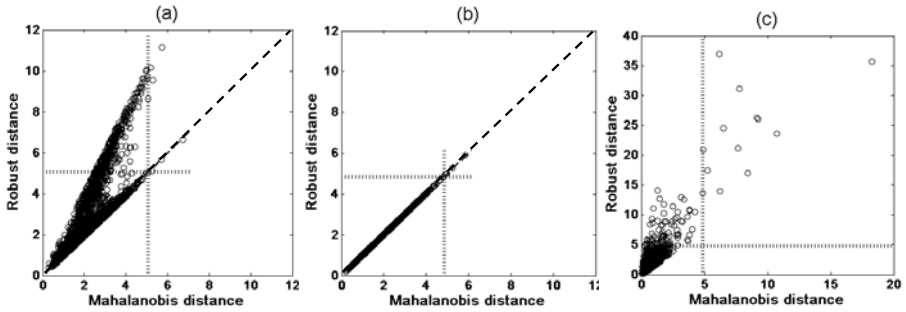


Fig. 2. Distance-Distance scatterplots of robust versus classical Mahalanobis distance: (a) pre- and (b) post-registration for Patient 1; (c) post-registration for Patient 3. Hatched lines show M_C values calculated using Equation (3); dashed lines in (a) and (b) show the line of equality. The data have been pooled for all visits.

The pre-registration Distance-Distance plot for Patient 1 (Figure 2a) shows two data sub-populations. The post-registration plot (Figure 2b) indicates that these were due to motion, which has been adequately corrected. The close correspondence of robust and classical Mahalanobis distance also indicates that the Patient 1 data had very few outliers. Figure 2c illustrates that outliers persist after registration for Patient 3. The Distance-Distance plot for Patient 2 was similar to Figure 2c and is not shown.

Figure 3 illustrates the effect of TKMDR on the segmentation results. While the differences are relatively subtle, the post-registration cluster images are visibly less fragmented in nature (though the effect is subtle). After treatment, the typical enhancing-rim pattern of tumors is evident in that the blue and blue-green clusters correspond to low contrast agent uptake (clusters 1 and 7 in Figure 4d) while the red and orange clusters correspond to higher uptake (clusters 3 and 5 in Figure 4d). This is consistent with a post-treatment reduction in angiogenesis and/or microvessel density.

Table 1 provides further evidence of the presence of outliers in the very small clusters among the “Raw” data for Patient 2 ($\leq 0.15\%$ of voxels) and Patient 3 ($\leq 0.35\%$ of voxels). The very small clusters for Patient 2 had physiologically unfeasible mean $[CA](t)$ series (cf Figure 4a and 4b). Outlier removal eliminated the very small clusters (Table 1 “Patient 2: MCD”) and returned all cluster mean $[CA](t)$ series to fall within the physiological range (Figure 4c). For Patient 2, the clusters after registration and

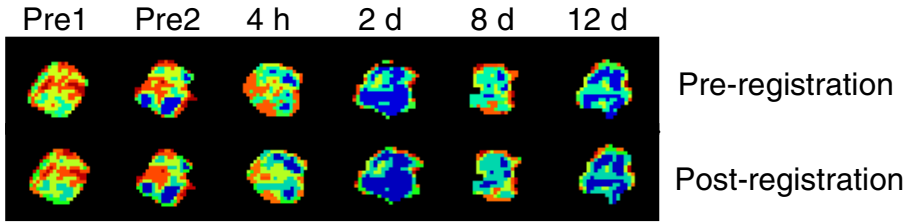


Fig. 3. Segmentation maps for clustering after outlier removal for Patient 1. Clusters are individually colour-coded using integer labels.

Table 1. Number of voxels per cluster for each patient before and after outlier removal using the robust MCD-based Mahalanobis distance. The data have been pooled for all visits.

Cluster	Patient 1		Patient 2		Patient 3	
	Raw	MCD	Raw	MCD	Raw	MCD
1	661	664	12	6383	764	462
2	654	757	7	6327	819	291
3	603	649	12603	3533	346	123
4	1073	471	23283	1251	8	260
5	758	619	1	6582	2	85
6	494	1015	56	4426	1	509
7	1054	1116	1	5300	344	452

outlier removal were stratified in mean $[CA](t)$ (Figure 4c), but for Patient 1 the clustering resulted in different mean $[CA](t)$ curve shapes (compare solid lines with dashed lines in Figure 4d), reflecting more complex enhancement characteristics.

4 Discussion

The methodology of this paper used TKMDR for motion correction, conversion of MR signal to $[CA](t)$ for cross-visit normalization, iterative PCA for missing data imputation and dimensionality reduction, and a robust Mahalanobis distance via MCD for outlier detection. After applying these techniques to DCE-MRI $[CA](t)$ series data, tumor sub-segmentation could be performed using standard k -means clustering.

The iterative PCA imputation provided reasonable $[CA](t)$ estimates (Figure 1), although one imputed value in each example fell lower than its neighbors—this may reflect the true data structure but could indicate a need to fine-tune the procedure by retaining more PCs, or increasing the number of iterations.

As the clustering procedure did not incorporate spatial regularization or any other spatial information, it is encouraging that the segmentation results showed a high level of spatial contiguity, in that the voxels of any given cluster were generally found grouped in the same spatial location (Figure 3) rather than being loosely scattered through the image volumes.

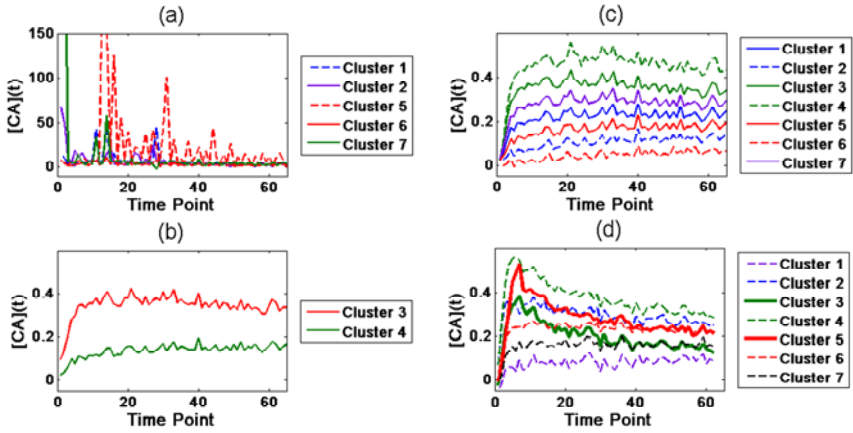


Fig. 4. Cluster mean $[CA](t)$ series for Patient 2 before (a, b) and after (c) outlier removal, and for Patient 1 after outlier removal (d). Note the scale change in (a), for which all clusters have $[CA]$ values that exceed the maximum of c.20 mmol/l observed in the aorta. The data have been pooled for all visits.

The clustering results can be skewed by even a small number of outliers if they lie far enough from the bulk of the data, as illustrated by the unphysiological mean $[CA](t)$ series for the very small clusters for Patient 2 (Table 1 and Figure 4a). Identification and removal of the outliers restored the mean $[CA](t)$ series to the physiological range (Figures 4c and 4d).

In the absence of outliers, Distance-Distance plots take the form of Figure 2b because the MCD and classical estimates for the data centroid and covariance matrix match, as do the corresponding Mahalanobis distances. Outliers pull the classical estimates towards themselves [13], so their robust Mahalanobis distances are greater than the classical ones, as in Figure 2a and 2c. Registration could reduce the number of outlier time series (Figure 2a and 2b) but not in all cases (Figure 2c) and its effect on cluster structure was subtle (Figure 3). Note that our additional registration of the variable flip-angle images to the pre-contrast TKMDR synthetic target images (Section 2.2) ensured that $T_i(0)$ and $T_i(t)$ in Equation (1) arose from the same voxels when applying cross-visit normalization.

Outliers have been defined as data that are sufficiently different from the norm to suggest a different generative mechanism [12]. For this study we simply removed outliers to robustify the clustering, but it would be informative to gather outliers into additional clusters, which could be examined to understand their origin. While some outliers will arise from artifactually low $T_i(t)$ values (Section 2.4), some may also provide insight into treatment effects. Finally, as ground truth is impossible to establish with patient data, simulation studies would provide further validation.

5 Conclusions

We have described a method for unsupervised, data-driven segmentation for DCE-MRI time series data, applicable to multi-visit clinical trials. The method uses registration for motion-correction, imputation of missing data values and statistical outlier

identification and rejection. The resulting clusters are in general spatially contiguous and have physiologically reasonable mean $[CA](t)$ series. Our procedure will allow statistical analyses to be applied to sub-regions of tumors to reveal localized variations in treatment response that would be masked in a whole-tumor VOI analysis. Such results could, for example, influence decisions on whether or not to proceed with the development of a particular drug.

Acknowledgments. This project was partially-funded by CRUK grants C237/A6295 and C19221/A6086.

References

1. O'Connor, J.P.B., Jackson, A., et al.: DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. *Br. J. Cancer* 96, 189–195 (2007)
2. Jackson, A., Buckley, D.L., Parker, G.J.M. (eds.): *Dynamic contrast-enhanced magnetic resonance imaging in oncology*. Springer, Berlin (2005)
3. Rose, C.J., Mills, S., et al.: Quantifying spatial heterogeneity in Dynamic Contrast-Enhanced MRI Parameter Maps. *Magn. Reson. Med.* 62, 488–499 (2009)
4. Berry, L.R., Barck, K.H., et al.: Quantification of viable tumor microvascular characteristics by multispectral analysis. *Magn. Reson. Med.* 60, 64–72 (2008)
5. Carano, R.A.D., Ross, A.L., et al.: Quantification of tumor tissue populations by multispectral analysis. *Magn. Reson. Med.* 51, 542–551 (2004)
6. Buonaccorsi, G.A., Roberts, C., et al.: Cross-visit tumor sub-segmentation reveal localised response to anti-angiogenic treatment in DCE-MRI data. *Proc. Int. Soc. Magn. Reson. Med.* 18, 4813 (2010)
7. O'Connor, J.P.B., Carano, R.A.D., et al.: Quantifying antivascular effects of monoclonal antibodies to vascular endothelial growth factor: insights from imaging. *Clin. Cancer Res.* 15, 6674–6682 (2009)
8. Buonaccorsi, G.A., O'Connor, J.P.B., et al.: Tracer kinetic model-driven registration for dynamic contrast-enhanced MRI time-series data. *Magn. Reson. Med.* 58, 1010–1019 (2007)
9. Jenkinson, M., Bannister, P., et al.: Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841 (2002)
10. Haase, A.: Snapshot FLASH MRI: application to T_1 , T_2 and chemical shift imaging. *Magn. Reson. Med.* 13, 77–89 (1990)
11. Stanimirova, I., Daszykowski, M., et al.: Dealing with missing values and outliers in principal component analysis. *Talanta* 72, 172–178 (2007)
12. Walczak, B., Massart, D.L.: Dealing with missing data: Part I. *Chemometr. Intell. Lab. Syst.* 58, 15–27 (2001)
13. Hawkins, D.: *Identification of Outliers*. Chapman & Hall, London (1980)
14. Rousseeuw, P.J., van Zomeren, B.C.: Unmasking multivariate outliers and leverage points. *J. Am. Stat. Assoc.* 85, 633–659 (1990)
15. Verboven, S., Hubert, M.: LIBRA: a Matlab library for robust analysis. *Chemometr. Intell. Lab. Syst.* 75, 127–136 (2005)
16. Everitt, B.S.: *Cluster Analysis*. Edward Arnold, London (1993)

Nonlocal Patch-Based Label Fusion for Hippocampus Segmentation

Pierrick Coupé¹, José V. Manjón², Vladimir Fonov¹, Jens Pruessner¹,
Montserrat Robles², and D. Louis Collins¹

¹ McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University,
Montreal, Canada. University, 3801 University Street, Montreal, Canada H3A 2B4

² Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones
Avanzadas (ITACA), Universidad Politécnica de Valencia,
Camino de Vera s/n, 46022 Valencia, Spain

Abstract. Quantitative magnetic resonance analysis often requires accurate, robust and reliable automatic extraction of anatomical structures. Recently, template-warping methods incorporating a label fusion strategy have demonstrated high accuracy in segmenting cerebral structures. In this study, we propose a novel patch-based method using expert segmentation priors to achieve this task. Inspired by recent work in image denoising, the proposed nonlocal patch-based label fusion produces accurate and robust segmentation. During our experiments, the hippocampi of 80 healthy subjects were segmented. The influence on segmentation accuracy of different parameters such as patch size or number of training subjects was also studied. Moreover, a comparison with an appearance-based method and a template-based method was carried out. The highest median kappa value obtained with the proposed method was 0.884, which is competitive compared with recently published methods.

Keywords: hippocampus segmentation, nonlocal means estimator.

1 Introduction

The crucial role of magnetic resonance (MR) imaging has been demonstrated in the detection of pathology, the study of brain organization and in clinical research. However, the vast amount of data produced everyday in clinical settings prevents the use of manual approaches to data analysis. The development of accurate, robust and reliable segmentation techniques for the automatic extraction of anatomical structures is becoming an important challenge in quantitative MR analysis. To achieve this task, several automatic methods have been proposed, such as deformable models or region growing [1-2], appearance-based models [3-4] and atlas/template-warping techniques [5-9]. Recently, template-warping techniques that use a library of templates (i.e., MR images with manual expert-based segmentation) have been investigated intensively because of their high accuracy to segment anatomical structures. Barnes et al. [5] proposed to register the most similar template from a library of pre-labeled subjects to segment the hippocampus (HC). However, the use of only one template may result in a biased segmentation. To avoid this problem, it is possible to use several similar

templates [6-7, 9-11]. This requires a label fusion strategy [6, 8-9, 11] to efficiently merge the information derived from the selected templates. In such approaches, two main assumptions are made. First, constraints on structure shape are used implicitly due to the one-to-one correspondence between the voxels of the image to be segmented and those of the warped templates. This presents the advantage of forcing the resulting segmentation to have a similar shape to those of expert-labeled structures in the template library. However, according to the regularization used during registration, some details can be lost and local high variability cannot be captured. Second, label fusion techniques usually assign the same weight to all samples during a vote procedure and consider only the absolute number as a criterion. This approach is sensitive to registration error, since it does not take into account the anatomical relevance of each sample [10]. Therefore, we propose a patch-based scheme with a weighted label fusion, where the weight of each sample is only driven by the similarity of intensity between patches.

In this study, we describe a fully automated patch-based method and the different steps required for its utilization, such as the library construction. Our method is applied to the HC segmentation of healthy subjects. During experiments the influences of different parameters were studied, and a comparison with two other methods was performed. Finally, we discuss further improvements and questions revealed by this new approach.

2 Materials and Methods

2.1 Overview

First, the patch library is constructed by removing image acquisition variability and linearly aligning the subjects of the library in a common space. For each voxel of the image to be segmented, the surrounding patch is then compared to the patches contained in the library. Finally, the expert-based information is merged using a nonlocal means approach [12] to produce the final segmentation.

2.2 Dataset

The HC dataset consists of 80 subjects selected from 152 T1-weighted (T1w) MR images of young, healthy individuals acquired in the context of the International Consortium for Brain Mapping (ICBM) project. The volunteers were scanned with a 1.5T Philips GyroScan imaging system (1 mm thick slices, TR = 17 ms, TE = 10 ms, flip angle = 30 °, 256 mm field of view). The local ethics committee approved the study and informed consent was obtained from all participants. The 80 selected subjects were manually segmented by an expert. This data subset comprises 39 males and 41 females of similar ages (mean age: 25.09 ± 4.9 years). The resulting segmentations obtained an intra-class reliability coefficient (ICC) of 0.900 for inter- (4 raters) and 0.925 for intra-rater (5 repeats) reliability.

2.3 Patch Library Construction

Denoising: All images in the database were first denoised with the three-dimensional (3D) block-wise nonlocal means filter proposed for MR images in [13].

Inhomogeneity correction: To ensure that each tissue type has the same intensity within a single image, the well-known N3 intensity non-uniformity correction of Sled et al. [14] was used.

Linear registration to stereotaxic space: Each subject was linearly registered to the MNI 152 template into the stereotaxic space using ANIMAL [15].

Intensity normalization: Finally, the intensities of the images were set in [0-100] and were normalized together by using the method proposed in [16]. Since our method involves the matching of a sub-region of anatomical structures based on intensity, the contrast and the luminance information are preserved by performing the global normalization of the entire 3D image.

2.4 Search Strategy within the Library

Initialization mask: Instead of performing the segmentation of the entire image under study, we define an initialization mask around the structure of interest. Many different strategies can be used to propose an accurate initialization, such as the matching of the best subject [5] followed by a morphological dilation of the mask. In this study, we chose a very fast and simple approach that consists in using the union of all the expert segmentations in the training database as the initial mask. In this way, we ensure that the structure is completely included in the mask and demonstrate the robustness of our method to coarse initialization (median Dice kappa of initial mask was around 0.4).

Subject selection: A selection is also performed at the subject level. This strategy is similar to the selection of best subjects in label fusion methods [7]. In our method, we use the sum of the squared difference across the initialization mask instead of using normalized mutual information over the entire image, as suggested in [7]. This strategy was chosen because our patch comparison is based on the L2-norm. Thus, we want to prioritize subjects with similar anatomy as well as similar luminance and contrast. The N closest subjects are finally retained for use during the patch comparison.

Patch pre-selection: As proposed for denoising purposes [13], we pre-select the patches to be compared. In fact, the main part of computational time is dedicated to computing the intensity-based distance between patches. By using simple statistics such as mean or variance, it is possible to discard a priori the most dissimilar patches [13]. In the proposed approach, we use luminance and contrast criteria to achieve this pre-selection. Based on the first and second terms of the well-known structural similarity measure (SSIM) [17], the pre-selection procedure can be written as follows:

$$ss = \frac{2 \cdot \mu_i \cdot \mu_{s,j}}{\mu_i^2 + \mu_{s,j}^2} \times \frac{2 \cdot \sigma_i \cdot \sigma_{s,j}}{\sigma_i^2 + \sigma_{s,j}^2}, \quad (1)$$

where μ represents the means, and σ , the standard deviations of the patches centered on voxel x_i (voxel under consideration) and voxel $x_{s,j}$ at location j in the subject s . If the value ss is superior to a given threshold th (0.95 for all experiments), the intensity distance between patches i and j is computed. The patch mean and variance are pre-computed as maps of local means and local variances, thus avoiding multiple computations.

Search area definition: Initially, the nonlocal means denoising filter was proposed as a weighted average of all the pixels in the image, with patch-based similarity used to assign the weights [12]. For computational reasons, the entire image cannot be

used and the number of pixels involved has to be reduced. As for denoising [12-13], we use a limited search area V_i , defined as a cube centered on the voxel x_i under study. Thus, within each subject, we search similar patches in a cubic region around the location under study. This search area can be viewed as the inter-subject variability of the structure of interest in stereotaxic space. This variability can increase for a subject with pathology or according to the structure under consideration.

2.5 Nonlocal Patch-Based Label Fusion

Nonlocal means estimator: For all voxels x_i of the image to be segmented (included in the initialization mask), the estimation of the final label is based on a weighted label fusion $v(x_i)$ of all labeled samples in the selected library (i.e., inside the search area V_i for the N considered subjects):

$$v(x_i) = \frac{\sum_{s=1}^N \sum_{j \in V_i} w(x_i, x_{s,j}) \cdot y_{s,j}}{\sum_{s=1}^N \sum_{j \in V_i} w(x_i, x_{s,j})}, \quad (2)$$

where $y_{s,j}$ is the label given by the expert to voxel $x_{s,j}$ at location j in subject s . The weight $w(x_i, x_{s,j})$ is computed as:

$$w(x_i, x_{s,j}) = \begin{cases} \exp^{-\frac{\|P(x_i) - P(x_{s,j})\|_2^2}{h}} & \text{if } ss > th \\ 0 & \text{else} \end{cases}, \quad (3)$$

where $\|\cdot\|_2$ is the L2-norm, normalized by the number of elements, computed between each intensity of the elements of the patches $P(x_i)$ and $P(x_{s,j})$. If the structure similarity ss between patches is less than th , the weight is not computed and is set directly to zero. Finally, by considering the labels y defined as $\{0,1\}$, the final label $L(x_i)$ is computed as:

$$L(x_i) = \begin{cases} 1 & v(x_i) > 0.5 \\ 0 & v(x_i) < 0.5 \end{cases}. \quad (4)$$

In the event that all patches in the library have $ss < th$, -1 is returned to indicate that the selected library does not allow a decision to be made.

Local adaptation of h : As usual in estimation problems using a robust function, the tuning of the decay parameter h plays a crucial role. When h is very low, only a few samples are taken into account. When h is very high, all samples tend to have the same weight and the estimation is similar to a classical average. The value of h should depend on the distance between the patch under consideration and the library content. To automatically achieve this local adaptation of h , we propose an estimation of $h(x_i)$, based on the minimal distance between the patch under consideration and the selected subpart of the library:

$$h(x_i) = \arg \min_{x_{s,j}} \|P(x_i) - P(x_{s,j})\|_2 + \varepsilon, \quad (5)$$

where ε is a small constant to ensure numerical stability in case the patch under consideration is contained in the library.

2.6 Validation Framework

A leave-one-out procedure was performed for the 80 subjects. Dice's kappa was then computed by comparing the expert-based segmentations with the segmentations obtained with our method. The impact of the patch size, search area size and number of training subjects was studied. Moreover, the proposed patch-based method was compared with an appearance-based approach using level-set shape constraints [3], and a template-based technique inspired by the work of Barnes et al. [5] that uses ANIMAL [15] for the nonlinear registration of the best subject. For the appearance-based method, only one modality was used during the process. We used the 79 remaining subjects to construct the training dataset involved in PCA computation. For the template-based method inspired by [5], the best subject was selected using the normalized mutual information, as suggested in [7], and then nonlinearly warped to the subject under study with ANIMAL within a multi-resolution framework until a resolution of 2 mm.

3 Results

The Dice kappa values obtained with the initial standard mask was 0.44, which corresponds to coarse initialization.

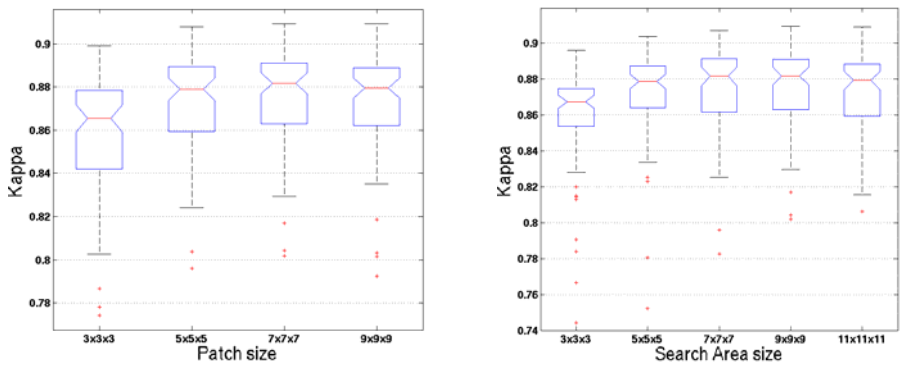


Fig. 1. Dice Kappa values according to the patch size (left) and the search area size (right). The results are obtained with 20 training subjects.

Impact of the patch size: First, we studied the impact of patch size on segmentation accuracy. The kappa results are presented at left in Fig. 1. The best median Dice kappa was obtained with a patch size of $7 \times 7 \times 7$ voxels for the HC dataset ($K = 0.882$).

Impact of the search area size: We also studied the impact of the search area size on segmentation accuracy. The kappa results are presented at right in Fig. 1. The best median kappa was obtained with a search area of $9 \times 9 \times 9$ voxels ($K = 0.882$).

Impact of the number of training subjects: The last important parameter of the proposed method is the number of selected training subjects. During this experiment, segmentation accuracy was studied for $N=2$ to $N=30$ selected training subjects. As described previously, the N best training subjects are selected from the 79 remaining ones. The results are presented in Fig. 2. The median kappa value was 0.848 for 2 subjects and 0.884 for 30 subjects. As expected, increasing the number of selected training subjects increased the accuracy of the segmentation.

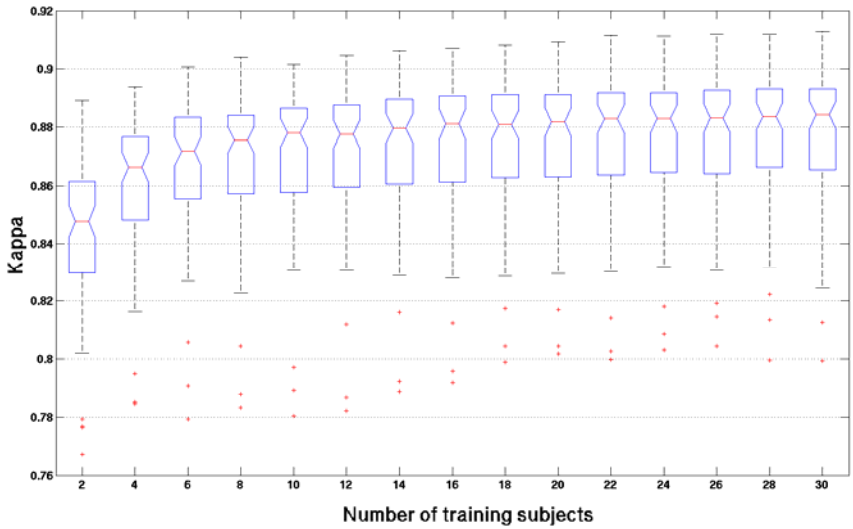


Fig. 2. Dice Kappa values according to the number of training subjects obtained with a patch size of $7 \times 7 \times 7$ voxels and search area size of $9 \times 9 \times 9$ voxels

Comparison with appearance-based and template-based methods: Finally, the proposed patch-based method was compared with two other methods. Figure 3 presents the kappa values obtained for each method. The results presented for our method were obtained with $N = 20$. The appearance-based method obtained a median kappa value of 0.800; the best template approach obtained 0.837, whereas the proposed method obtained 0.882. One can note that by using only 2 training subjects ($K=0.848$) our method already outperforms the two other methods. Figure 4 shows HC segmentations obtained by the methods compared.

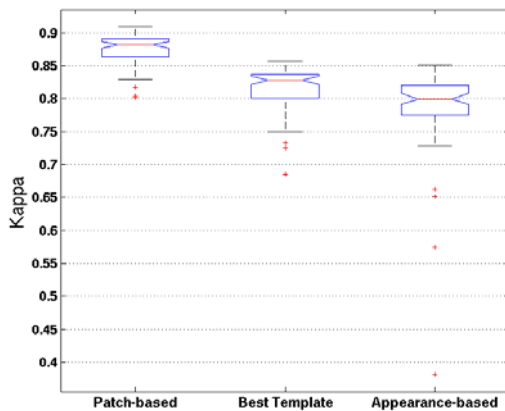


Fig. 3. Kappa values obtained by the three methods. The patch-based approach obtained significantly better results compared to the two others methods with a p -value $\ll 0.001$ in both cases using Kruskal-Wallis tests.

Computational time: The computational time was proportional to the number of subjects; for each subject, around 40 seconds were required. Compared with other approaches, the appearance-based method [3] took around 1 minute to provide the segmentation of the HC. The best template-based approach inspired by [5] required around 6 minutes to achieve the nonlinear registration of the cropped images already linearly registered into stereotaxic space. However, the comparison of computational time is difficult since our method was coded in MATLAB[®] and not in C like the other two methods.

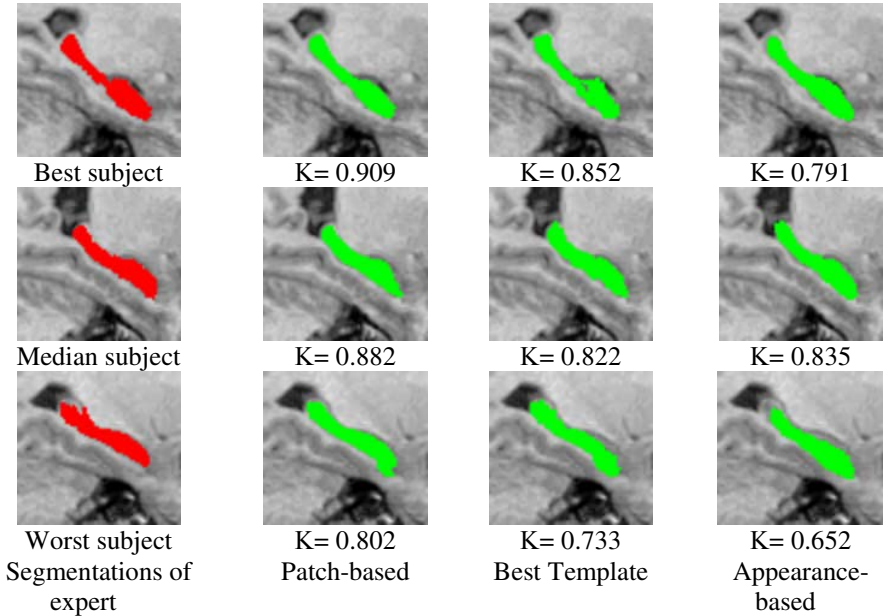


Fig. 4. Segmentation results with the three methods compared, for the best (top), a median (middle) and the worst (bottom) subjects obtained with our method

4 Conclusion

We proposed a novel patch-based approach to automatically segment anatomical structures using expert segmentation priors. Despite its simplicity, the accuracy of the proposed method has been demonstrated within our validation framework. The highest Dice kappa values obtained during experiments were 0.884 for $N = 30$ training subjects. Moreover, comparison with an appearance-based [3] and a template-based method [5] highlighted the competitive results obtained by the proposed nonlocal patch-based approach. Comparing published methods is always difficult due to differences within the databases used for validation, the studied populations, the quality of expert segmentations and the reported quality metrics. However, tendencies in method evolution and their respective performances can be obtained by studying published results. Recently published results [1, 5, 7] indicated kappa values lower than 0.88. To the best of our knowledge, only the methods based on nonlinear warping of the best templates and involving a label fusion step [9-11] obtained a kappa value equal to or greater than 0.88. Gousias et al. [11] reported a mean kappa of 0.88 with the use of a b-spline-based nonlinear registration on a 2-year-old brain. Lotjonen

et al. [10] proposed two intensity-based models to improve label fusion. With the graph-cut-based method, they obtained a kappa of 0.88, and with the EM-based algorithm, a kappa of 0.885. Collins and Pruessner [9] obtained a median kappa of 0.886 by using nonlinear registration of 11 best templates and a classical voting scheme for label fusion. By comparison, our proposed method offers the main advantages of its simplicity for similar segmentation accuracy ($K = 0.884$). As a result of the proposed automatic adaptation of the robust function parameter, our approach can be used easily and implemented in a fully automatic manner.

References

1. Chupin, M., et al.: Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease. *NeuroImage* 34(3), 996–1019 (2007)
2. Shen, D., et al.: Measuring size and shape of the hippocampus in MR images using a deformable shape model. *NeuroImage* 15(2), 422–434 (2002)
3. Hu, S., Collins, D.L.: Joint level-set shape modeling and appearance modeling for brain structure segmentation. *NeuroImage* 36(3), 672–683 (2007)
4. Duchesne, S., Pruessner, J., Collins, D.L.: Appearance-based segmentation of medial temporal lobe structures. *NeuroImage* 17(2), 515–531 (2002)
5. Barnes, J., et al.: A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage* 40(4), 1655–16571 (2008)
6. Heckemann, R.A., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33(1), 115–126 (2006)
7. Aljabar, P., et al.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738 (2009)
8. Hammers, A., et al.: Automatic detection and quantification of hippocampal atrophy on MRI in temporal lobe epilepsy: a proof-of-principle study. *NeuroImage* 36(1), 38–47 (2007)
9. Collins, D., Pruessner, J.: Towards Accurate, Automatic Segmentation of the Hippocampus and Amygdala from MRI. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 592–600. Springer, Heidelberg (2009)
10. Lotjonen, J.M., et al.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage* 49(3), 2352–2365 (2010)
11. Gousias, I.S., et al.: Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *NeuroImage* 40(2), 672–684 (2008)
12. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 60–65 (2005)
13. Coupé, P., et al.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Trans. Med. Imaging* 27(4), 425–441 (2008)
14. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging* 17(1), 87–97 (1998)
15. Collins, D.L., et al.: Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping* 3(3), 190–208 (1995)
16. Nyul, L.G., Udupa, J.K.: Standardizing the MR image intensity scales: making MR intensities have tissue specific meaning. *Medical Imaging 2000: Image Display and Visualization* 1(21), 496–504 (2000)
17. Wang, Z., et al.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612 (2004)

Cellular Automata Segmentation of Brain Tumors on Post Contrast MR Images

Andac Hamamci¹, Gozde Unal¹, Nadir Kucuk², and Kayihan Engin²

¹ Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey
gozdeunal@sabanciuniv.edu

² Department of Radiation Oncology, Anadolu Medical Center, Kocaeli, Turkey

Abstract. In this paper, we re-examine the cellular automata(CA) algorithm to show that the result of its state evolution converges to that of the shortest path algorithm. We proposed a complete tumor segmentation method on post contrast T1 MR images, which standardizes the VOI and seed selection, uses CA transition rules adapted to the problem and evolves a level set surface on CA states to impose spatial smoothness. Validation studies on 13 clinical and 5 synthetic brain tumors demonstrated the proposed algorithm outperforms graph cut and grow cut algorithms in all cases with a lower sensitivity to initialization and tumor type.

1 Introduction

Segmentation of tumors on medical images is not only of high interest in serial treatment monitoring of "disease burden" in oncologic imaging, but also gaining popularity with the advance of image guided surgical approaches [1]. Outlining the tumor contour is a major step in planning spatially localized radiotherapy (e.g. Cyberknife, iMRT) which is done manually on post contrast T1 MRI in current clinical practice. On T1 images acquired after administration of a contrast agent (gadolinium), blood vessels and the parts of the tumor, where the contrast can pass the blood-brain barrier are observed as hyper intense areas.

Region-based active contour models are widely used in image segmentation [2]. In general, these region-based models have several advantages over gradient-based techniques for segmentation, including greater robustness to noise. However, classical snakes had the problem of being "only as good as their initialization", even when using level-set snakes in 3D. Because the tumor class does not have a strong spatial prior, many small structures, mainly blood vessels, are classified as tumor as they also enhance with contrast. Ho et.al. used fuzzy classification of pre and post contrast T1 images to obtain a tumor probability map to evolve a level-set snake [3]. Liu et.al. have adapted the fuzzy connectedness framework for tumor segmentation by constructing a rectangular volume of interest selected through identifying the first and last slice of the tumor and specifying a set of voxels in the tumor region [4].

Interactive algorithms have become popular for image segmentation problem in recent years. Graph based seeded segmentation framework has been generalized such that graph cuts (GC) [5], random walker (RW) [6], shortest paths,

and power watersheds [7] have been interpreted as special cases of a general seeded segmentation algorithm, which solves a minimization problem involving a graph's edge weights constrained by adjacent vertex variables or probabilities. In [8], the connection between GC, RW, and shortest paths was shown to depend on different norms: L_1 (GC); L_2 (RW); L_∞ (shortest paths), in the energy that is optimized. Although it was reported that the shortest paths and RW produce relatively more seed-dependent results, it can be argued that the global minimum of an image segmentation energy is worth as good as the ability of its energy to capture underlying statistics of images [9], and a local minimum may produce a solution closer to the ground truth than that of a global minimum. Hence, with good prior information provided as in the case of a seeded image segmentation problem, efficiently finding a good local minima becomes meaningful and worthwhile.

On the other hand, cellular automata (CA) algorithm motivated biologically from bacteria growth and competition, is based on a discrete dynamic system defined on a lattice, and iteratively propagates the system states via local transition rules. It was first used by Vezhnevets et.al. [10] (grow-cuts) for image segmentation, which showed the potential of the CA algorithm on generic medical image problems.

In this paper, we re-examine the CA algorithm to establish the connection of the CA-based segmentation to the graph-theoretic methods to show that the iterative CA framework converges to the shortest path algorithm, for the first time, to our knowledge. Next, as our application is in the clinical radiotherapy planning, where manual segmentation of tumors are carried out on CT fused post contrast T1-MR images by a radio-oncology expert, we modify the CA segmentation towards the nature of the tumor properties undergoing radiation therapy by adapting relevant transition rules. Finally, a smoothness constraint using level set active surfaces is imposed over the resulting CA states. We present our framework for brain tumor segmentation in Section 2, and demonstrate its performance via validation studies on both synthetic, and radiation therapy planning expert-segmented data sets in Section 3, followed by discussions and conclusions.

2 Method

2.1 Cellular Automata: Its Connection to Graph Theoretic Methods

A graph consists of a pair $G = (V, E)$ with vertices (nodes) $v \in V$ and edges $e \in E \subseteq V \times V$. The weight of an edge, e_{ij} , is denoted by w_{ij} and is assumed here to be nonnegative and undirected (i.e., $w_{ij} = w_{ji}$). We will use closed neighborhood $N_G[v]$ where $v_i \in N_G(v_i)$. The edge weights are similarity measures calculated using measured data (e.g. voxel intensity) for vertices: $w_{ij} = f(I_i, I_j) \in (0, 1]$ and self-similarity $w_{ii} = 1$. State of a vertex $s(v_i) = s_i$ is specified with a real value $x(v_i) = x_i \in [0, 1]$ and a label $l_i \in \{BG, FG, \dots\}$ pair. Starting with

initial states of vertices, in each iteration, vertices of graph G is updated by the following rule:

$$l_i^{t+1} = l_{i^*}^t \text{ and } x_i^{t+1} = w_{i^*i} x_{i^*}^t \text{ where } i^* = \arg \max_{j \in N_G[v_i]} w_{ji} x_j \quad (1)$$

Note that since the vertex itself is also included in its neighborhood, Eq. (1) also covers the static case:

$$s_i^{t+1} = s_i^t \text{ if } x_i \geq w_{ji} x_j \text{ for } \forall v_j \in N_G[v_i] \setminus v_i \quad (2)$$

Vertex states are initialized by user supplied seeds $p_i \in P$ such as:

$$s^0(v_i) = (1, l(p_i)) \text{ for } v_i \in P \text{ and } s^0(v_i) = (0, \emptyset) \text{ for } v_i \notin P \quad (3)$$

This map converges since $\sum_i x_i$ is upper-bounded and monotonically increasing:

$$\lim_{t \rightarrow \infty} s_i^{t+1} = s_i^t \text{ for } \forall v_i \in V \quad (4)$$

Now, let us derive some properties on the final map. Consider any vertex v_i of a graph G , and assume that a latest update occurred on this vertex at time t_i . The vertex which updates v_i is v_{i^*} . Final state for v_i is:

$$s_i^{t \geq t_i} = (w_{i^*i} x_{i^*}^{t_i}, l_{i^*}^{t_i}) \quad (5)$$

If any update occurs on v_{i^*} at time $t_{i^*} \geq t_i$ by $v_{i^{**}}$, this should satisfy the condition:

$$x_{i^*}^{t_{i^*}} = w_{i^{**}i^*} x_{i^{**}}^{t_{i^*}} > x_{i^*}^{t < t_{i^*}} \text{ that gives } w_{i^*i} x_{i^*}^{t_{i^*}} > w_{i^*i} x_{i^*}^{t < t_{i^*}} \quad (6)$$

However, this will also cause an update on v_i at $t > t_{i^*} > t_i$, which violates the condition in (5). Then, at the converged map, there exists a neighbor v_{i^*} for each vertex v_i such that:

$$s_i = (w_{i^*i} x_{i^*}, l_{i^*}) \quad (7)$$

If we go one step further:

$$s_{i^*} = (w_{i^{**}i^*} x_{i^{**}}, l_{i^{**}}) \text{ and } s_i = (w_{i^*i} w_{i^{**}i^*} x_{i^{**}}, l_{i^{**}}) \quad (8)$$

We can follow this path for any vertex until we reach a seed which is never updated:

$$s(v_i) = \left(\prod_{\Omega(p_i \rightarrow v_i)} w_{jk}, l(p_i) \right) \quad (9)$$

Therefore, this algorithm cuts the graph G to independent subgraphs for each seed, consisting of spanning trees with seeds at root nodes.

If we set edge weights depending on similarity of image ($I : R^3 \rightarrow R$) neighborhoods as:

$$w_{jk} = e^{-B \|\nabla_{jk} I\|} \quad (10)$$

where $\|\nabla_{jk}I\|$ denotes a Euclidean norm on the difference between intensities of two adjacent vertices v_j and v_k . Maximization the product of w_{jk} 's along the path Ω becomes equivalent to minimization of the summation of $\|\nabla_{jk}I\|$'s along the same path. $\sum_{\Omega(p_i \rightarrow v_i)} \|\nabla_{jk}I\|$ is a discrete approximation to a geodesic or shortest path between the seed p_i to a voxel v_i . Each voxel is then assigned to the foreground label if there is a shorter path from that voxel to a foreground seed than to any background seed, where paths are weighted by image content. With this interpretation, cellular automata algorithm solves the shortest paths energy form formulated in [8].

The equivalence, which we showed, between CA updates by Eq. (II) and shortest path algorithm is illustrated in Fig. II.

The main advantage of using CA algorithm is its ability to obtain a multilabel solution in a simultaneous iteration. Another advantage is that the local transition rules are simple to interpret, and it is possible to impose prior knowledge, specific to the problem, into the segmentation algorithm.

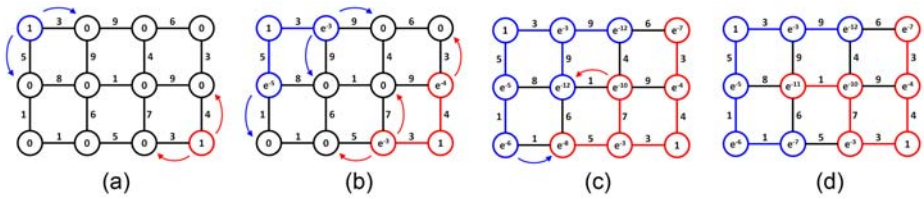


Fig. 1. (a) The graph is initialized with similarities as edge weights and vertex values 1 for seeds, 0 elsewhere; (b-c) intermediate propagation steps for CA; (d) shows the final vertex values obtained from CA which can also be obtained as the shortest path from each vertex to a seed

2.2 Seed Selection Based on Tumor Response Measurement Criteria

As each path, defining the labeling of a vertex ends at a seed, the efficiency of the algorithm can be increased by choosing the background seeds on a closed surface around the volume of interest (VOI) because the result of labeling inside the VOI is equivalent to using the whole data set.

Robustness to seed selection is an important property of a segmentation algorithm, as it is natural to expect similar results for the same tumor while allowing the user to guide the segmentation process interactively by imposing constraints in different way. In RECIST tumor response criteria [11], a general procedure to follow-up tumor progress is to measure the maximum observable tumor diameter. Our seed selection algorithm employs the same idea to follow the familiar clinical routine to which the clinicians are used to. Focusing on tumor segmentation problem, we utilize the following seed selection procedure (see Fig. 2a, 2b):

- Ask user to draw a line along the maximum visible diameter of the tumor.
- Crop the line 15% from each end and thicken to 3 pixels wide to obtain fg seeds.

- Choose bounding box of the sphere having 20% longer of the line as VOI.
- Use the 1 voxel wide border of this VOI as background seeds.

One obvious drawback is that the input seed information is obtained from only a single slice of the tumor volume, hence it is not guaranteed that the depth of the tumor will also coincide with the VOI. However, our experimental studies revealed that spherical assumption for the tumor is mostly valid.

2.3 Adapting Transition Rule to Tumor Characteristics

In the seeded tumor segmentation application for heterogeneous tumors, which mostly consist of a ring enhancing region around a dark necrotic core (and also irregular borders), most of the foreground seeds fall in the necrotic region. This causes the segmentation algorithm to get stuck at necrotic to active transition borders. To overcome such problems, a prior knowledge is added to the edge weight function as follows:

$$w_{jk} = e^{-\beta_{\text{tumor}}^{l, \text{sgn}(I_j - I_k)} \|I_j - I_k\|} \text{ where sgn denotes sign function.} \quad (11)$$

Enhancing tumor cells are brighter than the normal tissue, and more centrally located necrotic core is darker, hence by adjusting β parameter, the weight reduction (strength loss) of a tumor state while passing through a ramp up gradient is adjusted to be lower than other cases:

$$\beta_{\text{tumor}}^{l, \text{sgn}(I_j - I_k)} = \begin{cases} 0.7 & \text{if } l_k \text{ is foreground and } \text{sgn}(I_j - I_k) = +1 \\ 1.0 & \text{otherwise} \end{cases} \quad (12)$$

Although, some of the properties we derived for this algorithm is no more valid, and due to asymmetric edge values, we can no more interpret the algorithm in the undirected graph framework, our experimental results revealed that the new tumor CA (tCA) algorithm significantly improved the results obtained, especially on glioblastomas.

2.4 Using Level Set on Strength Maps

Smoothing is an important prior in segmentation of brain tumors from post contrast T1 images, because of three main reasons: First, an area surrounded by tumor tissue is considered as a tumor region even the intensity characteristics likely to be healthy. Secondly, it is possible to include misclassified necrotic regions to tumor region, which are usually surrounded by enhanced tissue. Finally, it is possible to exclude nearby structures such as arteries that are enhanced by administration of the contrast agent.

As described in Section 2.1, cellular automata algorithm assigns a label l , and a likelihood value x_i in the interval $(0,1]$ to each voxel v_i . The latter indicates how much it is likely to assign one of the labels to the voxel. Remapping values of the final map $X = \{x_i\}_{i \in V}$ to the interval $(-1,1)$ for all voxels in V , we obtain a new map M :

$$M_i = \begin{cases} \frac{x_i - \min(X)}{\max(X) - \min(X)} & \text{if } l_i \text{ is foreground} \\ -\frac{x_i - \min(X)}{\max(X) - \min(X)} & \text{if } l_i \text{ is background} \end{cases} \quad (13)$$

with values M_i at a voxel i . Finally, a level set snake is evolved on map M with a piecewise constant region assumption of [2], however by using a local Gaussian kernel to define inner and outer regions around the propagating surface, to obtain the final tumor segmentation map.

Steps of the proposed cellular automata based tumor segmentation algorithm is shown in Fig. 2. First, the user draws a line over the largest visible diameter of the tumor (a); using this line, a VOI is selected with foreground-background seeds (b); tCA algorithm is run on the VOI to obtain a label map and strengths at each voxel (c); label maps and strengths are combined to obtain the signed strength values, i.e. map M , such that contours have value of zero (d). The map M is used to evolve a level-set snake. In (d), initial level set contour is depicted in white, and final evolved contour is shown in black. Comparison to expert segmentation (blue) is visualized in (e), overlaid with tCA result (red), and tCA-Level set result (yellow).

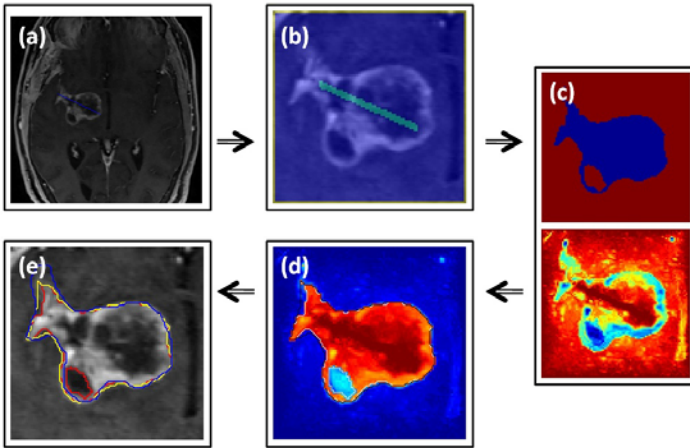


Fig. 2. Steps of the proposed tumor segmentation method: see text for explanations

3 Results and Discussion

An expert-segmentation during a radiation therapy planning session is compared against results of Graph Cut (GC), Random Walker (RW), Grow-cut, and the tCA over a slice (see Fig 3). It can be observed that highly varying necrotic and enhancing tumor characteristics present challenges to all computer algorithms, which fail to capture the expert segmented boundaries. Cellular-automata based algorithms grow-cut and tCA could propagate further towards the enhanced tumor margins.

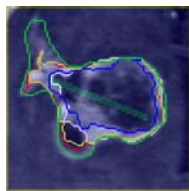


Fig. 3. Comparison of graph based algorithms: Expert in Green; RW in Blue (Dice: %70), GC in White (Dice: %80); Grow cut in Yellow (Dice: %87), tCA in Red (Dice: %89)

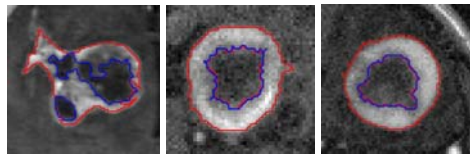


Fig. 4. Segmentation of enhancing and necrotic regions of the tumor using multi-label cellular automata

3.1 Enhancing/Necrotic Core Segmentation Qualitative Results

Cellular automata segmentation algorithm is applied on heterogeneous tumors with enhancing and necrotic regions, whose delineation is important especially in assessment of radiotherapy response. In the first step, tCA-LS method is applied to obtain a total tumor mask. Tumor and necrotic seeds are chosen by applying a threshold to intensity histogram of the segmented region. For background seeds, a one voxel boundary around the VOI is used. tCA algorithm is initialized with these 3 label seeds on a single slice of three tumors and the results are given in Fig 4.

3.2 Validations on Synthetic Data

Dice similarity measure, $Dice(A, B) = 2 \times s(A \cap B) / (s(A) + s(B))$, is used to quantify the overlap between obtained segmentation maps and expert manual segmentations extracted from radiotherapy planning sessions for each tumor. To measure the robustness of the method, for each tumor case, overlap for 5 different initialization lines are calculated and mean and standard deviation of the overlap are given, and the performance is compared between GC, Grow-cut, tCA, tCA-LevelSet(LS).

Five synthetic brain tumor datasets, available online from University of Utah² are used for validation and the dice measures are reported in Table 1. Synthetic Tumor 5, which is not enhanced with contrast agent and out of scope of the proposed algorithm, is included for the completeness of the Utah dataset (see Fig 5g).

3.3 Validations on Tumors That Undergo Radiation Therapy Planning

Validations on clinical data set were carried out over high resolution ($\approx 0.5 \times 0.5 \times 1.0$ mm) post Gd T1 weighted 3D FLASH MRI scans of 13 tumors of

¹ Due to unavailability of RW method in 3D, it was not included in the validation tests.

² <http://www.ucnia.org/softwaredata/5-tumordata/10-simtumordb.html>

Table 1. Dice overlap \pm std deviations over 5 different initial seed lines for each tumor for synthetic tumor data set from [12]

	Graph cut	Grow cut	tCA	tCA-Level Set
Synthetic Tumor 1	6.6 ± 2.5	83.8 ± 1.1	87.4 ± 0.9	90.4 ± 0.7
Synthetic Tumor 2	58.0 ± 32.3	77.8 ± 3.2	81.4 ± 3.6	84.6 ± 4.4
Synthetic Tumor 3	96.5 ± 0.0	96.2 ± 0.3	96.3 ± 0.3	97.6 ± 0.2
Synthetic Tumor 4	91.1 ± 0.7	89.3 ± 1.0	91.9 ± 0.8	93.0 ± 0.9
Synthetic Tumor 5	11.6 ± 7.4	73.6 ± 2.7	73.1 ± 3.3	69.1 ± 5.7
Average Overlap	52.8 ± 42.5	84.1 ± 9.0	86.0 ± 9.1	86.9 ± 11.0

Table 2. Dice overlap \pm std deviations over 5 different initial seed lines for each tumor demonstrate improved overlap with the proposed method

		Graph cut	Grow cut	tCA	tCA-Level Set
Tumor 1	Metastasis	76.8 ± 0.0	79.5 ± 2.0	80.2 ± 1.6	83.5 ± 0.3
Tumor 2	Gliosarcoma; Grade IV	15.0 ± 5.5	53.5 ± 7.4	57.6 ± 6.0	69.8 ± 5.5
Tumor 3	Grade II Astrocytoma	34.5 ± 16.0	76.9 ± 3.1	83.2 ± 1.0	89.1 ± 1.2
Tumor 4	Metastasis	17.0 ± 37.1	72.6 ± 5.8	74.6 ± 4.0	79.5 ± 3.2
Tumor 5	Metastasis	39.0 ± 6.5	44.4 ± 5.1	46.5 ± 3.0	51.5 ± 2.6
Tumor 6	Metastasis	5.1 ± 8.6	51.7 ± 5.3	54.6 ± 4.9	60.5 ± 3.7
Tumor 7	Metastasis	76.6 ± 2.5	73.8 ± 1.9	74.8 ± 1.5	81.3 ± 1.8
Tumor 8	Metastasis	69.3 ± 0.3	76.6 ± 0.9	76.9 ± 1.0	81.6 ± 0.9
Tumor 9	Metastasis	55.3 ± 1.9	63.3 ± 5.1	65.2 ± 4.2	68.4 ± 4.0
Tumor 10	Meningioma	71.6 ± 10.0	61.1 ± 6.8	65.5 ± 6.3	76.9 ± 3.9
Tumor 11	Meningioma	83.0 ± 0.1	69.8 ± 3.4	73.1 ± 2.7	83.5 ± 1.3
Tumor 12	Meningioma	44.9 ± 24.5	49.0 ± 8.8	52.7 ± 7.5	64.1 ± 5.9
Tumor 13	Meningioma	68.6 ± 1.7	67.7 ± 1.6	68.0 ± 1.8	71.7 ± 1.8
Average Overlap		50.5 ± 26.5	64.6 ± 11.7	67.1 ± 11.4	74.0 ± 10.8

7 patients obtained from Anadolu Medical Center. As the ground truth for segmentation, we used the tumor contours outlined manually by a radio-oncologist for radiotherapy planning. The clinical classification of tumors along with the segmentation performances are tabulated in Table 2.

The results we observed with the GC approach exhibit similar problems reported before in [7] such as shrinking bias due to minimum cut optimization. The shortest path algorithms, equivalently CA, showed lack of the shrinking bias problem. The proposed tCA-LS algorithm exhibit a lower coefficient of variation (std/mean) on the average compared to the other methods used in validation.

3.4 Qualitative Results

We present qualitative results of both synthetic and real tumors using the proposed tCA-LS algorithm in Figure 5. The result on a synthetic tumor with a non-enhanced region having no boundary to healthy tissue is given in Fig 5(e). The metastasis (Tumor 6) in Fig 5(f) is a small tumor (1.4cc) with weak boundaries, which produces a low overlap score even for small errors on the boundaries.

The synthetic tumor in Fig 5(g) is not enhanced by the contrast agent, and the result obtained leaks outside due to the lower intensities than surrounding tissue and weak boundaries. For the metastasis in Fig 5(h), surrounding bright tissue is misclassified as tumor, even after smoothing with a level set.

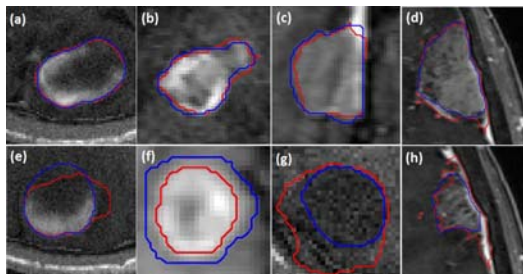


Fig. 5. Examples of typical (top row) and challenging cases (bottom row) obtained by tCA-LS: Expert segmentation in Blue, tCA-LS in Red

4 Conclusion

The proposed segmentation algorithm for the problem of tumor delineation, has only two main parameters: $\beta_{tumor}^{l,+}$, $l \in \{fg, bg\}$ and mean curvature term weight in the level set evolution. One future work includes optimizing both curvature term and the tumor sensitivity parameter β_{tumor}^l over a larger tumor database, although the results over 18 tumors of varying degrees showed that the algorithm performs with high overlap ratios even with the fixed heuristic values. Another item is to investigate the issues related to the VOI and seed selection procedure. Our current work includes assessment of the tumor response to therapy, which is built on the given segmentation framework.

Acknowledgement. This work was partially supported by TUBITAK Grant No:108E126, and EU FP7 Grant No: PIRG03-GA-2008-231052.

References

1. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M.C., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index. *Academic Radiology* 11(2), 178–189 (2004)
2. Chan, T.F., Vese, L.: Active contours without edges. *IEEE Transactions on Image Processing* 10(2), 266–277 (2001)
3. Ho, S., Bullitt, E., Gerig, G.: Level-set evolution with region competition: Automatic 3-d segmentation of brain tumors. In: *ICPR*, vol. 1, p. 10532 (2002)
4. Liu, J., Udupa, J.K., Odhner, D., Hackney, D., Moonis, G.: A system for brain tumor volume estimation via mr imaging and fuzzy connectedness. *Computerized Medical Imaging and Graphics* 29, 21–34 (2005)

5. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In: ICCV, pp. 105–112 (2001)
6. Grady, L.: Random walks for image segmentation. PAMI 28(11), 1768–1783 (2006)
7. Couprie, C., Grady, L., Najman, L., Talbot, H.: Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest. In: ICCV (2009)
8. Sinop, A., Grady, L.: A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In: ICCV (2007)
9. Szeliski, R., et al.: A comparative study of energy minimization methods for markov random fields with smoothness-based priors. PAMI 30(6) (2008)
10. Vezhnevets, V., Konouchine, V.: Growcut - interactive multi-label n-d image segmentation by cellular automata. In: Graphicon, Novosibirsk Akademgorodok, Russia (2005)
11. Therasse, P.: Evaluation of response: new and standard criteria. *Annals of Oncology* 13, 127–129 (2002)
12. Prastawa, M., Bullitt, E., Gerig, G.: Synthetic ground truth for validation of brain tumor mri segmentation. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 26–33. Springer, Heidelberg (2005)

Agreement-Based Semi-supervised Learning for Skull Stripping

Juan Eugenio Iglesias¹, Cheng-Yi Liu², Paul Thompson², and Zhuowen Tu²

¹ Medical Imaging Informatics, University of California, Los Angeles
jeiglesias@ucla.edu

² Laboratory of Neuroimaging, University of California, Los Angeles
chengyiliu@ucla.edu, thompson@loni.ucla.edu, zhuowen.tu@loni.ucla.edu

Abstract. Learning-based approaches have become increasingly practical in medical imaging. For a supervised learning strategy, the quality of the trained algorithm (usually a classifier) is heavily dependent on the amount, as well as quality, of the available training data. It is often very time-consuming to obtain the ground truth manual delineations. In this paper, we propose a semi-supervised learning algorithm and show its application to skull stripping in brain MRI. The resulting method takes advantage of existing state-of-the-art systems, such as BET and FreeSurfer, to sample unlabeled data in an agreement-based framework. Using just two labeled and a set of unlabeled MRI scans, a voxel-based random forest classifier is trained to perform the skull stripping. Our system is practical, and it displays significant improvement over supervised approaches, BET and FreeSurfer in two datasets (60 test images).

1 Introduction

Supervised learning approaches have become increasingly popular and practical in brain MRI segmentation [1,2,3]. These algorithms produce classifiers that utilize a large number of features by applying modern learning algorithms. However, supervised learning often demands large amounts of training data with consistent manual labeling, which are difficult to obtain. Recent semi-supervised learning approaches [4,5,6,7] have provided new mechanisms to take advantage of the information in unlabeled data to train a better system.

In this paper, we propose a semi-supervised approach to skull stripping, which is the first element in most neuro image pipelines, and therefore critical for their overall performance. The goal of skull stripping is to segment the brain from non-brain matter in MRI in a robust manner. Skull stripping is expected to follow the major folds on the surface; if the deeper sulci are to be extracted for surface analysis, subsequent post-processing techniques can be used. Automated skull stripping is challenging due to the large variations in image intensity and shape in MRI scans. Expert systems exist in this domain (e.g. BET [8], FreeSurfer [9]), but none of them offer a fully satisfactory performance.

We propose taking advantage of these expert systems and unlabeled data to train a voxel classifier to segment the brain by: 1) training on the labeled data;

and 2) iteratively re-training the classifier including samples from the unlabeled data for which the expert systems agree but the classifier is not confident yet. This approach is related to the tri-training algorithm [10] from the co-training family [7]. Co-training requires having two or more conditionally independent views (sets of features) for the data, which is often difficult [11], whereas tri-training works on single-view data by simultaneously training three classifiers. The system described in this study can be seen as a special case of tri-training in which two well-developed skull stripping algorithms (BET and FreeSurfer) play the role of two of the classifiers in the framework. The output for a test image is based solely on the trained classifier, so running BET and FreeSurfer on the image to be analyzed is not necessary.

2 A Semi-supervised Skull Stripping Algorithm

2.1 Proposed Method

Training: The training process (see top half of flowchart in Figure 1) is divided in four stages: registration, preprocessing of unlabeled data, feature extraction and learning. The first step is to coarsely align all the volumes, labeled and unlabeled, to a template brain scan. The first volume in the dataset was arbitrarily chosen to be the template. This alignment makes it possible to use position features in the posterior classification. ITK (www.itk.org) was used to optimize an affine transform using a mutual information metric and a multi-resolution scheme. Using a nonlinear registration method could make the classifier rely too much on the registration through location features, making the method less robust.

The next step is to preprocess the unlabeled volumes. The brain is first segmented using BET and FreeSurfer. The binary outputs of the two methods are then “softened” using a signed distance transform (positive if inside the brain, negative if outside). The distance map is mapped to the template space using the transforms from the registration step. The warped maps are used to calculate preliminary brain masks in the unlabeled scans by averaging the two maps for each volume and thresholding the result at zero, and they will also be used in the posterior semi-supervised learning step.

The third step in the training stage is feature extraction. A pool of 58 image features is used in this study: (x, y, z) position, Gaussian derivatives of order up to two at five different scales ($\sigma = \{1.0, 2.0, 4.0, 8.0, 16.0\}$, in mm), and gradient magnitudes at the same scales. A subset of voxels from the training volumes is randomly selected for training purposes under the constraints that: 1) all scans contribute the same number of voxels; 2) 50% of the voxels have to be positives according to the annotated boundary (for the labeled scans) or the preliminary mask from the previous step (for the unlabeled); and 3) 50% of the voxels have to lie within 5mm of the boundary and 75% within 25mm. The features are normalized to zero mean and unit variance.

Finally, a classifier can be trained using the labeled and unlabeled data. Breiman’s random forests [12] were used as the base classifier because they

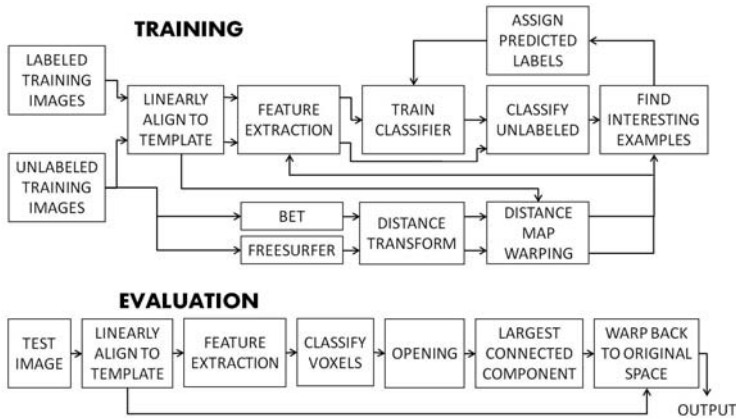


Fig. 1. Flowchart for the training and test stages of the proposed algorithm

compare favorably to other state-of-the-art algorithms [13]. Feature selection is performed through training a preliminary classifier with all the features and 20% of the available data and then retraining with the features that provide the highest mean decrease in accuracy in the out-of-bag data. For the semi-supervised learning, the random forest is first trained solely on the labeled scans, and then updated in an iterative fashion using “interesting” voxels from the unlabeled volumes i.e. those for which both distance maps (BET and FreeSurfer) are greater than a given positive threshold (5 mm in all the experiments in this study). Among those voxels, the ones for which the classifier predicts a negative label with highest probability (i.e. those with fewest trees voting for positive) are shifted from the unlabeled set to the labeled data with positive labels. Then the procedure is repeated with negative voxels using the opposite of the previous threshold. The iteration concludes with retraining the random forest.

Testing: The testing pipeline (see bottom half of flowchart in Figure 1) is similar to the training process, with the important difference that BET and FreeSurfer do not need to be run on the data in this stage. When a new volume is presented to the system, the first step is to register it to the template. The optimized transform is stored to warp the final mask back to the original space later on. From the aligned scan, features are extracted from every voxel and fed to the random forest. The output is a volume with the number of trees that have voted positive at each location. Upon division by the total number of trees, this volume can be interpreted as a probability map for the voxels being part of the brain. The map can then be smoothed and thresholded at 0.5 to binarize the data and obtain the preliminary mask. After applying a morphological opening operator to smooth this mask, the largest connected component is extracted, holes in it are filled, and then it is warped back to the original space using the inverse of the affine transform to obtain the final output.

Theoretical justification: Even though a formal theoretical analysis of the proposed method is out of the scope of this paper, in this section we try to provide a brief justification of why it works. Theoretically, the analysis is very similar to that of tri-training [10], which is in turn very related to [14] and mostly follows the PAC (probably approximately correct) learning theory. Let H^* and H be the ground-truth and our classifier, respectively. It can be shown that, for $Pr[d(H, H^*) \geq \epsilon] \leq \delta$, where $d(\cdot)$ is the difference between H and H^* , one needs to have a sequence of m samples where $m \geq \frac{2}{\epsilon^2(1-2\eta)^2} \ln(\frac{2N}{\delta})$. $\eta < 0.5$ is an upper bound on the error rate by the expert systems, N is the number of possible hypotheses and δ is the confidence in PAC learning.

The error bounds on the unlabeled data for BET and FreeSurfer can be directly estimated using labeled data. Our task is then to design a new rule by combining the experts to make a joint decision that achieves a small error rate η . This can be translated into a function that weights each unlabeled data point with a positive/negative label as:

$$w_i^\pm(X|H_i, F_1, F_2) \propto A^\pm[F_1(X), F_2(X)]$$

where i denotes the i -th iteration, H_i is the trained classifier at iteration i , $F_1(X)$ and $F_2(X)$ give a measure of how likely X is to be a positive according to the two expert systems, and A^\pm measures the level of agreement and confidence of F_1 and F_2 for positive/negative samples. In the proposed method, $F_1(X)$ and $F_2(X)$ are the distance maps provided by BET and FreeSurfer. For the agreement function, success in pilot experiments led to the use of ($H[\cdot]$ is the Heaviside function): $A^\pm = H(\pm F_1(X) - 5mm) \cdot H(\pm F_2(X) - 5mm)$.

In this setting, one iteration would ideally suffice because two of the classifiers (BET and FreeSurfer) in the tri-training framework are fixed. The iterative scheme we propose to train the classifier can be seen as a bootstrapping process [15]: m is large but only a limited number of training samples are used. We define an empirical criterion to select samples which will potentially improve the classifier: agreement between BET and FreeSurfer but not with the classifier.

3 Experiments and Results

3.1 Data

Two different datasets are used in this study. The first one, henceforth dataset A, consists of 10 T1-weighted volumes from the LPBA40 dataset [16]. The brain surface was annotated by an expert radiologist in all the scans. The second dataset, henceforth dataset B, consists of 152 T1-weighted scans from healthy subjects acquired with an inversion recovery rapid gradient echo sequence on a Bruker 4T system. Manual delineations of the brain by an expert physiologist are available for 52 of the scans.

3.2 Setup of Experiments

Impact of the number of annotated scans (dataset A): In this experiment, the 10 volumes from dataset A are segmented using a cross-validation

scheme. For scan $i \in \{1, \dots, 10\}$, and assuming N_{lab} annotated examples: 1) 10 random subsets of N_{lab} elements are extracted from the pool $P = \{1, \dots, 10\} \setminus \{i\}$; 2) for each subset, a classifier is trained using the N_{lab} extracted elements as labeled data and the remaining $N_{unlab} = 9 - N_{lab}$ elements in P as unlabeled; 3) volume i is processed with the 10 resulting classifiers; and 4) the performance metric for scan i is the average of the performances of the 10 outputs for that scan. Four different metrics are used in this study: classification error rate, Jaccard index of the segmentation (related to the Dice coefficient as $D^{-1} = (1 + J^{-1})/2$), mean surface-to-surface distance, and the 95% quantile of this distance, which is a measure of robustness. The error rate and the Jaccard index are computed using only the voxels that are within 12.5 mm of the annotated boundary for easier interpretation of the results. The settings of the parameters were the following: 100,000 training voxels for the supervised classifier (independently of the number of images); 10 loops of the semi-supervised update with 2,500 voxels each; 500 trees for the random forest; $\sigma = 1$ mm to smooth the likelihood map; and a 2 mm radius spherical element for the opening.

Evaluation on a larger dataset and impact of the number of unlabeled scans (dataset B): In this experiment, two randomly selected scans of the 52 labeled volumes from dataset B are used to train the initial classifier, and the 100 scans without annotations play the role of unlabeled data. The remaining 50 scans are used for evaluation. Then, the experiment was repeated by randomly removing elements from the pool of unlabeled scans in order to assess the impact of the amount of available unlabeled instances N_{unlab} . The parameters were all set to the same values as in the previous experiment. No cross validation was performed, which is reasonable given the size of the dataset.

3.3 Results

The results from the first experiment (dataset A) are shown in Figure 2, which compares the semi-supervised strategy with a supervised version (i.e. same algorithm with no semi-supervised update), BET and FreeSurfer. Though FreeSurfer generally outperforms BET, the latter works better in this particular dataset. The semi-supervised approach outperforms all the others at every value of N_{lab} , whereas the supervised method requires $N_{lab} = 2 \sim 3$ to improve the results of BET, except for the robustness measure. As expected, the difference between the semi-supervised and supervised methods decreases as N_{lab} increases, since the number of unlabeled volumes $N_{unlab} = 9 - N_{lab}$ also becomes lower.

Table 1 shows the p-values for a paired t-test comparing the results given by the semi-supervised method and BET, and also for a test comparing the semi-supervised version against the supervised. Compared with BET, all the metrics improve significantly for $N_{ann} = 2$ at $\alpha = 0.05$ (also for $N_{ann} > 2$, not displayed due to lack of space). Compared with the supervised approach, the difference is significant at $N_{ann} = 1$ but not at $N_{ann} \geq 2$ except for the 95% quantile, due to the low robustness of the supervised method with a small number of

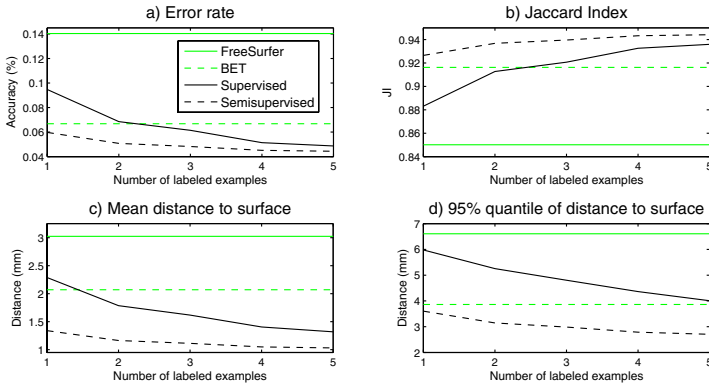


Fig. 2. Different performance metrics for dataset A depending on the number of labeled volumes used in the training N_{lab} , as well as results provided by BET and FreeSurfer

Table 1. Paired t-test comparing the performance metrics provided by the semi-supervised learning and the supervised version / BET

Method	Error rate	Jaccard index	Mean distances	95% quantile
BET ($N_{lab} = 1$)	$2.7 \cdot 10^{-1}$	$2.3 \cdot 10^{-1}$	$1.4 \cdot 10^{-3}$	$2.7 \cdot 10^{-1}$
BET ($N_{lab} = 2$)	$4.0 \cdot 10^{-5}$	$6.2 \cdot 10^{-5}$	$1.3 \cdot 10^{-19}$	$5.0 \cdot 10^{-2}$
Supervised ($N_{lab} = 1$)	$2.9 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$
Supervised ($N_{lab} = 2$)	$1.9 \cdot 10^{-1}$	$1.5 \cdot 10^{-1}$	$8.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$

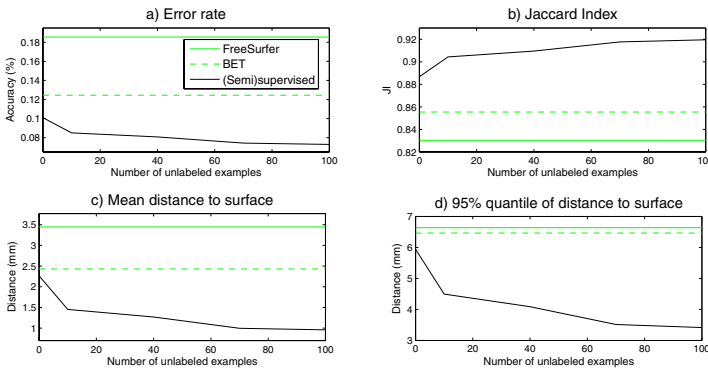


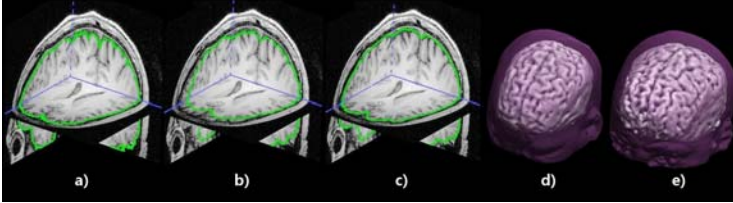
Fig. 3. Different performance metrics for dataset B depending on the number of unlabeled volumes N_{unlab} used in the training, as well as results by BET and FreeSurfer. The values of the metrics at zero give the performance for the supervised method.

training scans. This lack of significance is caused by the low number of available unlabeled volumes, as confirmed by the second setup.

Figure 3 shows the results for the experiments on dataset B, where the number of annotated scans is two. The performance keeps growing until $N_{unlab} = 100$,

Table 2. Paired t-test comparing the performance metrics provided by the semi-supervised learning and the supervised version / BET for $N_{unlab} = 100$

Method	Error rate	Jaccard index	Mean distances	95% quantile
BET	$2.8 \cdot 10^{-6}$	$3.6 \cdot 10^{-6}$	$1.4 \cdot 10^{-15}$	$2.3 \cdot 10^{-5}$
Supervised	$8.3 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$1.2 \cdot 10^{-11}$	$2.4 \cdot 10^{-4}$

**Fig. 4.** Brain surfaces for a sample scan from dataset B: (a) ground truth, (b) BET, (c) semi-supervised. 3-D renderings for two scans: (d) dataset A, two labeled and nine unlabeled; (e) dataset B, two labeled and 50 unlabeled.

which, next to having a larger test data sample, provides a larger statistical significance for the improvement with respect to the supervised version (Table 2). The significance with respect to BET from the first experiment is preserved. Figure 4(a-c) shows three orthogonal slices of a fairly difficult scan from dataset B segmented with BET, the unsupervised algorithm (100 unlabeled examples) and the manual annotations. The output from BET is unacceptable in the frontal lobe, whereas the more robust semi-supervised version is still able to detect the correct boundary. Figure 4(d-e) displays renderings of two segmentations by the proposed semi-supervised algorithm, one from each dataset.

4 Discussion

A semi-supervised method for skull stripping which utilizes a small amount of labeled data has been presented in this paper. We take advantage of two factors to boost the performance of a classifier: the existence of expert systems and the abundance of unlabeled data. This situation is not uncommon in medical imaging, since large amounts of unlabeled scans and well-developed segmentation methods are often accessible. A key parameter of the system is the threshold of the distance transform at which unlabeled data are allowed to be sampled from. This depends on the performance of the expert systems and is justified in section 2.1: the success of a semi-supervised learning approach relies on the classification error by the other classifiers (experts in our case). If all the experts have high agreement but wrong predication, using unlabeled data may even degrade the performance. Future work related to this study includes: 1) testing the supervised algorithm on a large number of scans with more labeled examples; 2) designing more robust features that do not rely as much on the pure voxel intensities and

allow the algorithm to segment the brain precisely in cases where the gray levels are not very consistent with the training data; and 3) assessing the performance on scans from patients with pathology.

Acknowledgements

This work is funded by grants NSF IIS-0844566, N000140910099 and partly NIH U54 RR021813 and China863 2008AA01Z126. The authors would like to thank Cornelius Hojakshtani for the acquisition and annotation of dataset A, G. de Zubicaray, K. McMahon and M. Wright for the acquisition of dataset B, and M. Barysheva for the annotation of the dataset B. The first author would also like to thank the U.S. Department of State's Fulbright program for his funding.

References

1. Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE T. Med. Imaging* 27, 495–508 (2008)
2. Yi, Z., Criminisi, A., Shotton, J., Blake, A.: Discriminative, semantic segmentation of brain tissue in mr images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 558–565. Springer, Heidelberg (2009)
3. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: Supervised non-parametric image parcellation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 1075–1083. Springer, Heidelberg (2009)
4. Zhu, X.: Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, U. of Wisconsin-Madison (2005)
5. Bonwell, C., Eison, J.: Active learning: Creating excitement in the classroom. *AEHE-ERIC Higher Education Report No.1* (1991) ISBN 1–87838–00–87
6. Zhu, X.: Semi-supervised learning with graphs. PhD thesis, CMU (2005)
7. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: *Proc. COLT*, pp. 92–100 (1998)
8. Smith, S.: Fast robust automated brain extraction. *Hum. Brain Mapp.* 17(3), 143–155 (2002)
9. Segonne, F., Dale, A., Busa, E., Glessner, M., Salat, D., Hahn, H., Fischl, B.: A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22(3)
10. Zhou, Z.H., Li, M.: Tri-training: Exploit unlabeled data using three classifiers. *IEEE T. Knowl. Data En.* 17(11), 1529–1541 (2005)
11. Nigam, K., Ghani, R.: Analyzing the effectiveness and applicability of co-training. *Proc. Info. Knowl. Manag.*, pp. 86–93 (2000)
12. Breiman, L.: Random forests. *Mach. Learn.* 45(1), 5–32 (2001)
13. Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: *Proc. 23 Int. Conf. Mach. Learn.*, pp. 161–168. ACM, New York (2006)
14. Angluin, D., Laird, P.: Learning from noisy examples. *Mach. Learn.* 2(4)
15. Efron, B.: Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7(1), 1–26 (1979)
16. Shattuck, D., Prasad, G., Mirza, M., Narr, K., Toga, A.: Online resource for validation of brain segmentation methods. *NeuroImage* 45(2), 431–439 (2009)

Construction of Patient Specific Atlases from Locally Most Similar Anatomical Pieces

Liliane Ramus^{1,2}, Olivier Commowick³, and Grégoire Malandain¹

¹ INRIA Sophia Antipolis - Asclepios Team, France

² DOSIsoft S.A., France

³ INRIA Rennes - Visages Team, France

Abstract. Radiotherapy planning requires accurate delineations of the critical structures. To avoid manual contouring, atlas-based segmentation can be used to get automatic delineations. However, the results strongly depend on the chosen atlas, especially for the head and neck region where the anatomical variability is high. To address this problem, atlases adapted to the patient's anatomy may allow for a better registration, and already showed an improvement in segmentation accuracy. However, building such atlases requires the definition of a criterion to select among a database the images that are the most similar to the patient. Moreover, the inter-expert variability of manual contouring may be high, and therefore bias the segmentation if selecting only one image for each region. To tackle these issues, we present an original method to design a piecewise most similar atlas. Given a query image, we propose an efficient criterion to select for each anatomical region the K most similar images among a database by considering local volume variations possibly induced by the tumor. Then, we present a new approach to combine the K images selected for each region into a piecewise most similar template. Our results obtained with 105 CT images of the head and neck show that our method reduces the over-segmentation seen with an average atlas while being robust to inter-expert manual segmentation variability.

1 Introduction

The purpose of radiotherapy planning is to optimize the dose received by the tumor while controlling the dose on the surrounding Organs At Risk (OARs). This requires the accurate delineation of the Clinical Target Volume (CTV) and the OARs. In clinical routine, this task is often performed manually, which is tedious and prone to inter-expert variability. To ease this task, atlas-based segmentation may be used to get automatic delineations, and showed satisfying results for the brain [1] and promising results for the head and neck region [2].

In the head and neck, the anatomical variability among patients is high, mainly due to corpulence and neck flexion. Previous studies showed that an average atlas has difficulties to cope with this high variability, and may result in over-segmentation for some structures [2]. Utilizing an atlas that is specifically adapted to the anatomy of the patient to delineate may help to improve

the registration quality, and therefore the accuracy of the segmentation. To this end, one solution is to compute population-specific atlases, for example by clustering the database into homogeneous sub-groups [3] and computing an average atlas for each sub-group. To be even more specific to the patient (and not only to a given population), other approaches [4,5] have been developed to consider each manually delineated image of a database as a potential atlas, and to select the most appropriate one for each new query image to segment. By extension, and to enhance robustness, it has been proposed to select several of the most appropriate images, register them independently to the patient and combine the segmentation results [6]. All these approaches bring up two questions: how to select the most appropriate images for a given patient and how to fuse them.

The selection criterion must be able to account for the anatomical variability in the database (various corpulence, neck flexion, various tumor size and grade), and it must be fast enough to be used in clinical routine. Selection criteria based on meta-information (e.g. age [6]) have been used, but they are not suitable when dealing with anatomical variability independent of simple meta-information. Therefore, criteria based on intensities [6,4] have been proposed. However, our database is composed of pathological images, which may corrupt intensity based criteria. Commowick et al. proposed to estimate the amount of deformation needed to warp each image onto the patient image, using the average atlas to reduce computation time [5]. This criterion is computationally interesting but it still requires inverting and composing many deformation fields. Our first contribution is to propose an efficient selection criterion based on the degree of contraction and dilation of the structures. This criterion is well-suited for our case as it may account for the local volume variations caused by the tumor.

Regardless of the nature of the selection criterion, it may be applied globally on the images [6,5], or locally in order to cope with the local changes of each region [7,8,9,10,4]. Because of the high anatomical variability and as our database is composed of pathological images, a local selection seems more appropriate to consider the local impact of the tumor on the surrounding anatomical structures.

Once the most appropriate images have been selected for each region of interest, the fusion step has to be performed. In [9], a framework was proposed to build a piecewise most similar atlas from a set of images selected on predefined regions. This showed an improvement in segmentation accuracy with respect to an average atlas. However, it was restricted to the selection of a single image for each region, which makes it more sensitive to the selection step (e.g. outliers may exist in the selection process). Moreover, it may also be sensitive to the relatively high inter-expert variability in the head and neck region. Our second contribution is then to provide a framework to combine K_l selected images for each region R_l into one template for segmentation, taking into account the relative values of the selection criterion to weight each selected image accordingly.

We illustrate the capacities of our framework with 105 CT images of the head and neck region, showing its ability to reduce the over-segmentation seen with an average atlas while being less sensitive to inter-expert segmentation variability than a piecewise atlas computed using only one image per region.

2 Method

We present a new method to design an atlas locally adapted to the patient P to delineate on predefined regions. We assume that a database of N manually delineated images $\{I_j\}_{j \in [1 \dots N]}$ is available. Moreover, we suppose that an average atlas M has been built from this database. The average atlas construction provides for each image I_j a transformation warping it on M . We denote by $T_{I_j \leftarrow M}$ the non-linear part of the transformation allowing to resample I_j on M , and $J_{I_j \leftarrow M}$ the corresponding image of the Jacobian determinant values.

2.1 Efficient Local Selection of the Most Similar Images through Volume Variation Estimation

We wish to select among the images $\{I_j\}_{j \in [1 \dots N]}$ the ones that are the most similar to the query patient P on predefined regions $\{R_l\}_{l \in [1 \dots L]}$. The regions R_l are defined once and for all on the average atlas M . Typically, one may define them as a dilation of the anatomical structures of interest. For a given region R_l in M , we define our criterion as a comparison of the average degree of contraction/dilation when deforming I_j on M and when deforming P on M . To do this, we first average on R_l the logarithms of the determinants of the Jacobian matrices for each non-linear deformation $T_{I_j \leftarrow M}$, as described below:

$$\bar{J}_{R_l}(I_j \leftarrow M) = \frac{1}{\text{card}(R_l)} \sum_{x \in R_l} \log(J_{I_j \leftarrow M}(x)) \quad (1)$$

In the same way, after registering M and P , we can estimate $\bar{J}_{R_l}(P \leftarrow M)$ from $T_{P \leftarrow M}$. Then, the images $\{I_j\}_{j \in [1 \dots N]}$ can be ranked from the most similar to the least similar to the patient P on R_l according to the distance $d_{R_l}(I_j, P) = \|\bar{J}_{R_l}(P \leftarrow M) - \bar{J}_{R_l}(I_j \leftarrow M)\|$. This criterion is well-suited for the local selection of the most similar images. Our images indeed present tumors of various sizes and grades that can induce local volume variations of the CTV and of the surrounding OARs. Moreover, it is very efficient as the $\bar{J}_{R_l}(I_j \leftarrow M)$ are pre-computed. It only requires performing one non-linear registration between P and M and computing $\bar{J}_{R_l}(P \leftarrow M)$. By comparison, other methods either require multiple registrations [34] or many inversions and compositions of deformation fields [5].

2.2 Construction of a Piecewise Most Similar Atlas Incorporating Selection Weights

For each region R_l , the K_l images of the database having the lowest distances $d_{R_l}(I_j, P)$ are selected to build the piecewise most similar atlas and are denoted $\{\tilde{I}_{l,n}\}_{n \in [1 \dots K_l]}$. Further, we associate each image $\tilde{I}_{l,n}$ with a selection weight $\alpha_{l,n}$, based on $d_{R_l}(\tilde{I}_{l,n}, P)$, that reflects its relative degree of similarity to P on R_l . To compute $\alpha_{l,n}$, we used the Gaussian kernel, i.e. $\alpha_{l,n} = G_{\mu,\sigma}(d_{R_l}(\tilde{I}_{l,n}, P))$, as it allows us to discriminate distances that are very large. The Gaussian can be centered either on zero, or on the minimum distance found for the region

R_l (we chose the second solution). As to the standard deviation σ , it controls the rejection of images with a large distance and was computed from the whole distribution of distances on R_l . The weights are then normalized for each region, so that for each l , $\sum_{n=1}^{K_l} \bar{\alpha}_{l,n} = 1$. In addition, we also consider spatial weights to allow a smooth transition when interpolating between the regions R_l in the construction of the piecewise atlas. The spatial weight of the region R_l at location x is defined as $w_l(x) = 1/(1+\beta \text{dist}(x, R_l))$ where $\text{dist}(x, R_l)$ refers to the minimal distance to R_l at location x . It is then normalized so that $\sum_{l=1}^L \bar{w}_l(x) = 1$.

Construction of the Piecewise Most Similar Image. The construction process may be seen as a classical atlas construction [11] where the images have varying weights depending on the spatial location of each voxel ($\bar{w}_l(x)$) and on the selection distances ($\bar{\alpha}_{l,n}$). We iterate over the following steps ($\tilde{M}_0 = M$):

1. Register the images $\tilde{I}_{l,n}$ on the current reference \tilde{M}_k . This step provides affine transformations $A_{\tilde{I}_{l,n} \leftarrow \tilde{M}_k}$ and non-linear transformations $T_{\tilde{I}_{l,n} \leftarrow \tilde{M}_k}$
2. Compute the new average image M_{k+1} by interpolating the intensities of the warped $\tilde{I}_{l,n}$ using the two sets of weights $\bar{w}_{l,k}(x)$ and $\bar{\alpha}_{l,n}$
3. Compute an average diffeomorphism \bar{T}_k from the $T_{\tilde{I}_{l,n} \leftarrow \tilde{M}_k}$ and the weights
4. Apply \bar{T}_k^{-1} to M_{k+1} to get the new reference $\tilde{M}_{k+1} = M_{k+1} \circ \bar{T}_k^{-1}$
5. Update the regions of interest by applying \bar{T}_k^{-1} to $R_{l,k}$: $R_{l,k+1} = R_{l,k} \circ \bar{T}_k^{-1}$, and update the spatial weights $\bar{w}_{l,k+1}(x)$ accordingly

This process is similar to [9]. However, it is much more general as it allows the combination of several images for each region R_l . This is achieved by the following equations for steps 2 and 3. First, the intensities are interpolated by:

$$M_{k+1}(x) = \sum_{l=1}^L \left[\bar{w}_{l,k}(x) \left(\sum_{n=1}^{K_l} \bar{\alpha}_{l,n} \left(\tilde{I}_{l,n} \circ A_{\tilde{I}_{l,n} \leftarrow \tilde{M}_k} \circ T_{\tilde{I}_{l,n} \leftarrow \tilde{M}_k} \right) (x) \right) \right) \right] \quad (2)$$

The inner term (sum over n) computes a weighted average of the selected images for a region R_l , while the outer term uses the spatial weights to combine the contributions from each region R_l . Similarly, in step 3, we compute an average polydiffeomorphism \bar{T}_k using the Log-Euclidean framework [12]. This framework ensures to remain on the manifold of diffeomorphisms and leads to an autonomous Ordinary Differential Equation that can be easily integrated: $\dot{x} = \sum_{l=1}^L \left[\bar{w}_{l,k}(x) \left(\sum_{n=1}^{K_l} \bar{\alpha}_{l,n} \log \left(T_{\tilde{I}_{l,n} \leftarrow \tilde{M}_k} \right) (x) \right) \right]$.

Construction of the Associated Segmentation. After building the piecewise most similar template, we need to compute its associated segmentation from the delineations of the selected images. The images of our database have been delineated for a clinical purpose, and some contours are missing for some structures. To deal with this difficulty, we chose to define one region R_l for each anatomical structure in the construction of the template image.

¹ The deformations in the head and neck region are close enough to the identity, ensuring that the computed logarithms are correct, as specified by Arsigny et al.

The construction of the associated segmentation is then achieved in two steps. First, we compute a probability map for each structure independently using the selected manual segmentations and the selection weights $\bar{\alpha}_{l,n}$. Then, we assign each voxel of the template image to the structure that has the highest probability.

3 Evaluation

We evaluated the proposed framework with $N = 105$ CT images of the head and neck region. On these images, the CTVs and OARs were manually delineated following the guidelines in [13]. The structures involved are the lymph node levels II, III and IV (CTVs), the parotids, the spinal cord, and the brainstem (OARs). We performed a Leave-One-Out analysis, each patient being successively excluded from the database and delineated with each of the three following atlases built from the $N - 1$ remaining images: (1) AVE: average atlas built as in [2], (2) PW_1: piecewise most similar atlas built with $K_l = 1$ image for each region, and (3) PW_10: piecewise most similar atlas built with $K_l = 10$ images for each region. As registration algorithm, we used the framework described in [2].

3.1 Qualitative Results

Fig. 1 shows the three different atlases (b,c,d) computed for a given patient (a) whose neck flexion is above average. The spinal cord contours show that the average atlas (b) and the piecewise atlas PW_1 (c) both have a relatively low neck flexion, whereas the neck flexion of PW_10 (d) looks much more similar to the patient's one (see arrows). When registering head and neck images, a different neck flexion between the atlas and the patient is a common issue, often leading to registration errors and low segmentation accuracy. Therefore, our method's ability to provide a correct neck flexion may increase segmentation quality.

Fig. 2 illustrates some qualitative segmentation results on the parotids and on the lymph nodes levels III-IV using the three atlases. Compared to the manual contours (a), the automatic contours provided by the average atlas (AVE) (b) are too large, which was already observed in [2]. As mentioned in [9], PW_1 (image (c)) allows to reduce the over-segmentation. However, it was built from only

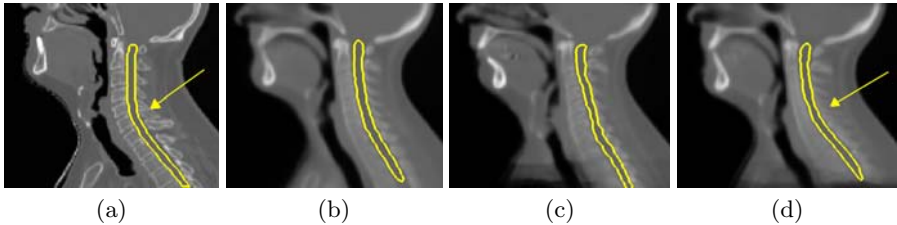


Fig. 1. Illustration of the atlases used for a given patient. For the given patient (a), comparison between the average atlas (AVE) (b), the piecewise most similar atlases PW_1 (c) and PW_10 (d). The atlases shown were affinely registered on the patient.

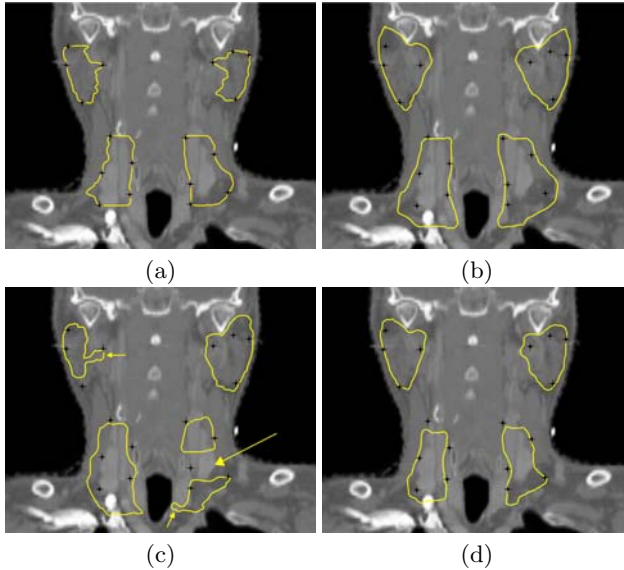


Fig. 2. Qualitative segmentation results of each method. (a) Manual contours. Automatic contours with AVE (b), with PW_1 (c) and with PW_10 (d). Black landmarks attached to the manual contours are also shown to draw the comparison.

one image for each region, and it is therefore likely to be biased by the inter-expert variability of delineation. The two small arrows on image (c) show the influence of local specificities of the selected segmentations on each region. Moreover, by construction, PW_1 segmentations can present some discontinuities. For instance, the large arrow on image (c) shows some non-connected lymph node levels III and IV, which is anatomically inconsistent. The automatic contours obtained with PW_10 (image (d)) are much less dependent on the inter-expert variability as 10 segmentations were fused for each structure. Moreover, the obtained contours are closer to the manual contours than both contours from AVE (b) and PW_1 (c), which results in shorter correction time for the clinician.

3.2 Quantitative Results

We now compare the performance of the three atlases AVE, PW_1 and PW_10 in terms of segmentation accuracy. To this end, sensitivity and specificity were averaged for each structure over all the Leave-One-Out tests. The results are presented in Fig. 3. First, as observed in [9], PW_1 shows an improvement of the specificity with respect to AVE, which is related to the reduction of the over-segmentation. However, this improvement is achieved at the expense of the sensitivity. With PW_10, the specificity is even higher than with PW_1 and the decrease in sensitivity is lower. For all structures, we also performed paired t-tests on the Dice values for each pair of methods. Whereas PW_1 has significantly lower Dice than AVE and PW_10 ($P < 0.05$), the differences between AVE and PW_10 are statistically

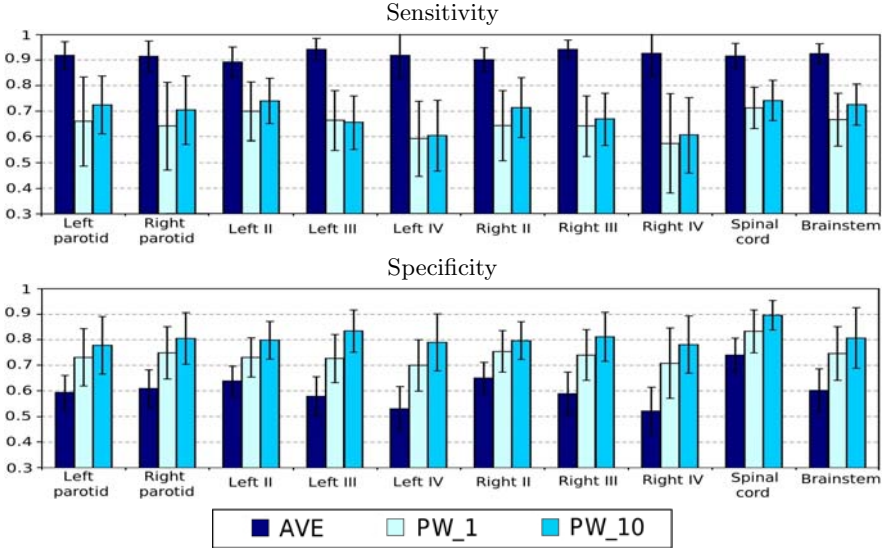


Fig. 3. Quantitative segmentation results. Average sensitivities and specificities for atlas-based segmentation using the atlases AVE, PW_1, and PW_10.

not significant ($P > 0.05$), illustrating that the overall overlap is similar while PW_10 significantly reduces the over-segmentation. Therefore PW_10 combines the advantages of both PW_1 (avoiding over-segmentation) and AVE (avoiding errors due to inter-expert variability).

4 Conclusion

We presented a new approach to build a piecewise most similar atlas to the patient. We first introduced an efficient criterion to select among a database the images that are the most similar to the patient for each region. This criterion is well adapted to model the impact of the tumor on the CTVs and the OARs as it is based on the local degree of contraction/dilation. Then, we presented a novel approach to build from the selected images a piecewise atlas and its associated segmentation. We applied our algorithm with 105 CT images of the head and neck region. The proposed approach was compared to other atlas-based approaches (single average atlas and piecewise most similar atlas built from a single image per region). We showed that our approach combines the advantages of both techniques. It indeed enables reducing the over-segmentation observed with the average atlas, and it is less dependent on the inter-expert segmentation variability than the piecewise atlas built from a single image per region.

The number K_l of images selected for each region plays an important role, as well as the standard deviation σ of the Gaussian in the selection weights. Here we used arbitrarily $K_l = 10$ mainly for computational reasons, but we will study the influence of these two parameters to find out the optimal solution between

the average atlas ($K_l = N$, infinite σ) and the method proposed in [9] ($K_l = 1$). Future work will also include a separate evaluation of the selection criterion and the piecewise atlas construction method. Finally, we will assess our framework on different groups of patients, e.g. on corpulent patients or patients with high neck flexion for which the average atlas provides low segmentation accuracy.

Acknowledgments. This work was undertaken in the framework of the MAE-STRO project (IP CE503564) funded by the European Commission, and was also partially funded by ANRT. The authors gratefully acknowledge Pr. V. Grégoire for the manually delineated database and for his expertise.

References

1. Bondiau, P.Y., Malandain, G., Chanalet, S., et al.: Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. *IJROBP* **61**(1) (2005) 289–98
2. Commowick, O., Grégoire, V., Malandain, G.: Atlas-based delineation of lymph node levels in head and neck computed tomography images. *Rad. Oncol.* **87**(2), 281–289 (2008)
3. Blezek, D.J., Miller, J.V.: Atlas stratification. *MedIA* **11**(5), 443–457 (2007)
4. Wu, M., Rosano, C., Lopez-Garcia, P., et al.: Optimum template selection for atlas-based segmentation. *Neuroimage* **34**(4), 1612–1618 (2007)
5. Commowick, O., Malandain, G.: Efficient selection of the most similar image in a database for critical structures segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II*. LNCS, vol. 4792, pp. 203–210. Springer, Heidelberg (2007)
6. Aljabar, P., Heckemann, R.A., et al.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage* **46**(3), 726–738 (2009)
7. Isgum, I., Staring, M., Rutten, A.: Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. *IEEE TMI* **28**(7), 1000–1010 (2009)
8. van Rikxoort, E.M., Isgum, I., et al.: Adaptive local multi-atlas segmentation: application to heart segmentation in chest CT scans. *MedIA* **14**(1), 39–49 (2010)
9. Commowick, O., Warfield, S.K., Malandain, G.: Using Frankenstein’s creature paradigm to build a patient specific atlas. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II*. LNCS, vol. 5762, pp. 993–1000. Springer, Heidelberg (2009)
10. Shi, F., Yap, P.T., Fan, Y., et al.: Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *Neuroimage* **51**(2), 684–693 (2010)
11. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: A convergence study. *CVIU* **77**(2), 192–210 (2000)
12. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: Larsen, R., Nielsen, M., Sparring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006)
13. Grégoire, V., Levendag, P., Ang, K.K., et al.: CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines. *Rad. Oncol.* **69**(3), 227–236 (2003)

Automatic Lung Lobe Segmentation Using Particles, Thin Plate Splines, and Maximum a Posteriori Estimation

James C. Ross^{1,2,3}, Raúl San José Estépar^{2,3}, Gordon Kindlmann⁴, Alejandro Díaz^{5,6}, Carl-Fredrik Westin^{2,3}, Edwin K. Silverman^{1,6}, and George R. Washko⁶

¹ Channing Laboratory, Brigham and Women's Hospital, Boston, MA

² Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

³ Surgical Planning Lab, Brigham and Women's Hospital, Boston, MA

⁴ Computer Science Department and Computation Institute, University of Chicago, Chicago, IL

⁵ Pontificia Universidad Católica de Chile, Chile

⁶ Pulmonary and Critical Care Division, Brigham and Women's Hospital, Boston, MA*

Abstract. We present a fully automatic lung lobe segmentation algorithm that is effective in high resolution computed tomography (CT) datasets in the presence of confounding factors such as incomplete fissures (anatomical structures indicating lobe boundaries), advanced disease states, high body mass index (BMI), and low-dose scanning protocols. In contrast to other algorithms that leverage segmentations of auxiliary structures (esp. vessels and airways), we rely only upon image features indicating fissure locations. We employ a particle system that samples the image domain and provides a set of candidate fissure locations. We follow this stage with *maximum a posteriori* (MAP) estimation to eliminate poor candidates and then perform a post-processing operation to remove remaining noise particles. We then fit a thin plate spline (TPS) interpolating surface to the fissure particles to form the final lung lobe segmentation. Results indicate that our algorithm performs comparably to pulmonologist-generated lung lobe segmentations on a set of challenging cases.

1 Introduction

Anatomically, the lungs consist of distinct lobes: the left lung is divided into upper and lower lobes, while the right lung is divided into upper, middle, and lower lobes. Each lobe has airway, vascular, and lymphatic supplies that are more or less independent of those supplies to other lobes. Fissures (left oblique, right oblique, and right horizontal) define the boundaries between the lobes and present as 3D surfaces that have greater attenuation (i.e. are brighter) than the surrounding lung parenchyma in CT datasets.

Several pulmonary diseases preferentially affect specific lobes: tuberculosis and silicosis tend to affect the upper lobes while interstitial pulmonary fibrosis tends to occur in

* This work was supported by NIH grants U01-HL089897, U01-HL089856, K23HL089353-01A1, R01MH074794, P41RR013218, Harvard Catalyst (NIH UL1 RR 025758-02), and the Parker B. Francis Foundation.

the lower lobes [1]. In the case of chronic obstructive pulmonary disease (especially emphysema), there are ongoing efforts to produce clinically relevant disease subtypes for better diagnosis and patient management. Performing lobe-based quantitative analysis can assist such efforts, especially in the context of epidemiological studies. Additionally, lobe specific measurements can help determine whether patients are good candidates for procedures such as lung volume reduction surgery [2]. These issues motivate the need for automatic and reliable lobe segmentation algorithms.

Advanced disease states (e.g. emphysema), atelectasis, and certain imaging protocols (low-dose, and expiratory acquisitions) can make it difficult to detect fissures in certain regions, and so-called incomplete fissures are not uncommon [3]. In [4] the authors address the issue of missing fissures. They use contextual information drawn from segmentations of the lung, fissure, and bronchial tree in conjunction with a multi-atlas selection mechanism to segment datasets that exhibit incomplete fissures. In a similar vein [5] does not rely on the presence of fissures but instead rely on the absence of vessels in the vicinity of the fissures (leveraging the dedicated blood supplies to each lobe). [1] uses vascular and airway tree segmentations to provide contextual clues for fissure locations. These approaches, however, rely on reliable segmentations of auxiliary structures (although [4] do not require such structures to be segmented, segmentation accuracy diminishes when they are not used in the presence of missing fissures).

[6] has shown that by selecting a small set of points along each of the three fissures, thin plate spline (TPS) interpolation is sufficient to accurately delineate the major lobes. However, the points in this study were manually selected. We build on this work and automatically identify fissure locations by adopting and extending the particle system for ridge surface extraction previously described by [7]. Other fissure identification and enhancement schemes have been proposed, namely [8]. We choose the particle approach because it is a fast and flexible way in which to sample likely fissure locations from the image data, and it fits seamlessly into the TPS surface fitting stage.

Following particle convergence, *maximum a posteriori* (MAP) estimation is used to both estimate the parameters of a smoothly fit TPS surface and to classify the particles as being members of either fissure or non-fissure classes. Making the assumption that fissures are locally planar, a post-processing stage is used to reject any remaining non-fissure points. We then use the methodology described in [6] to acquire the final lung lobe segmentation.

The paper is outlined as follows. In section 2 we describe the steps in our approach: lung segmentation, particles based sampling, MAP estimation, post-processing, and TPS fitting. In section 3 we describe the data used in our study and report the results of our algorithm on these datasets as compared to pulmonologist-established ground truth. In section 4 we discuss our results and draw conclusions about our methodology.

2 Methods

In this section we describe our approach to lobe segmentation. We begin by performing lung segmentation, which produces a labeling of the left and right lungs. Lung segmentation algorithms have been described elsewhere ([6], [9]), so we do not discuss this step here. Below we briefly review the method described by [6], by which interpolating surfaces are fit through a sparse set of points to define lobe boundaries. Next we discuss

the sampling method used to identify candidate fissure points, which is based on [7] specialized for ridge features. Following this we describe the MAP estimation that enables the elimination of non-fissure points from the sampling stage. Finally, we present a post-processing stage that eliminates remaining noise points. The final set of points is then used to define the lobe boundaries as in [6].

Thin Plate Spline Interpolation. Our goal is to identify a set of points along each fissure through which a TPS surface can be fit that accurately captures the lobe boundaries. [6] suggested that a small set of such points is sufficient to identify these surfaces. The TPS interpolating surface is the minimally curved surface that passes through all the selected points. The equation is given by

$$f(x, y) = a_1 + a_2x + a_3y + \sum_{i=1}^n w_i U(|P_i + (x, y)|) \quad (1)$$

where $U(r) = r^2 \log r$ is the radial basis function. The coefficient vector, $\mathbf{a} = (a_1, a_2, a_3)$, and the weight vector, $\mathbf{w} = (w_1, \dots, w_n)$ are determined from the n identified fissure points, P , such that the height function's bending energy is minimized [10]. In the procedure we will outline below, a TPS surface is fit independently for each of the three fissures of interest.

Particles Sampling. Once the lung field is segmented we rescale the intensity from $[-1024HU, -200HU]$ to $[0, 2^{16}]$ to maximize contrast and then compute a sampling of features that include the fissures. We adopt a particle system for feature extraction described by [7]. As the fissure surface between lung lobes has higher radio-opacity than the lobes themselves, the fissure can be isolated as a ridge surface, defined by [11] as the loci of points where the gradient of the image is orthogonal to the minor eigenvector of the Hessian. For the particle system, a smooth image domain is computed by cubic B-Spline interpolation, and the particles are constrained to lie within ridge surfaces by a Newton optimization that maximizes the image intensity, restricted to motion along the Hessian minor eigenvector \mathbf{e}_3 (with corresponding negative eigenvalue λ_3).

The iterative solution of the particle system involves adding, moving, or removing particles to minimize the collective energy of the system, which is the sum of energy from all pair-wise particle interactions, governed by a single rotationally symmetric potential $\phi(r)$ around each point. As described in [7], having a small negative well in $\phi(r)$ allows the particle system's population control to be formulated in the same energy minimization that moves particles into a uniform close-packing onto ridge surfaces. Particles are also removed when the local image properties suggest that they are no longer within a well-defined ridge surface, for which we adopted two heuristics. First, ridge surface strength $-\lambda_3$ (how concave-down the image is, cutting across the fissure) had to satisfy $-\lambda_3 > 4,000$, an empirically determined threshold that depends linearly on image contrast. Second, the mode of the Hessian eigenvalues γ_1 (the third standardized moment of the three eigenvalues) had to satisfy $\gamma_1 < -0.3$; a perfect ridge surface might have Hessian eigenvalues $\{0, 0, -1\} \Rightarrow \gamma_1 = -1$. The system is initialized by seeding 6,000 particles randomly throughout the lung field, and then run for 100 iterations. Due to the population control mechanism, the final number of particles ranges from 15,000 to 20,000 depending on the case. We select parameters for the particle

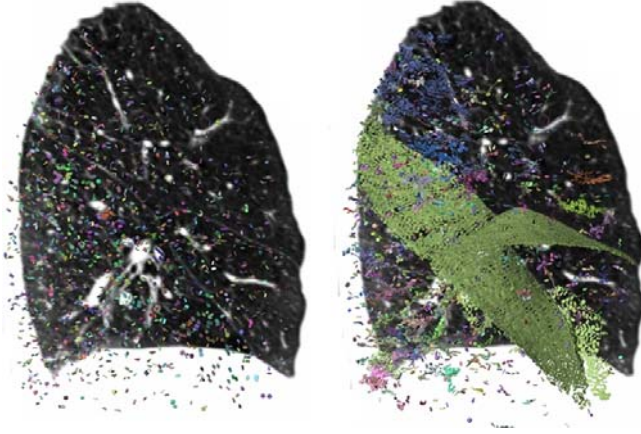


Fig. 1. Sagittal CT slice (right lung) and glyphs illustrating particles-based image sampling. The system is initialized with a random distribution of particles (left image). Upon convergence, the fissures are well sampled (pale green glyphs in right image). Additional, non-fissure locations that locally behave like ridges are also detected.

system that tend to make it very sensitive at the price of specificity. That is, we attempt to find a dense sampling of true fissure locations and permit a great deal of other locations that locally behave like ridge surfaces to be detected. Figure 1 illustrates the sampling procedure.

MAP Estimation. After point selection, we use MAP estimation to separate true fissure particles from noise. We formulate the problem as follows. Let $\mathcal{S} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ be the set of n particle points detected in our image where each particle point, \mathbf{p} , is given by $(\mathbf{x}, \mathbf{e}_3)$, \mathbf{x} being the spatial location, (x, y, z) , of the particle and \mathbf{e}_3 being the Hessian eigenvector in the direction normal to the local planar image feature at \mathbf{x} . We seek the parameters, Θ , of a probability distribution that best explains the subset of particles, $\mathcal{D} \subset \mathcal{S}$, that lie on the true fissure (note that these steps are carried out for each of the three fissures). We assume the elements of \mathcal{D} are independent and identically distributed (*i.i.d.*) random variables. The parameters, $\Theta = (z_1, \dots, z_l)$, that we seek are the heights of l control points (we use $l = 10$) of a “smoothed” TPS surface that is loosely fit to our data, and it is with respect to this surface that the probability that a particle point is a fissure point ($p(\mathbf{p}|\omega)$, where ω is the class of true fissure particles) is defined. We will refer to this surface as STPS and denote it as f_s . It is important to distinguish between the STPS surface and the TPS surface that will be fit through our final set of fissure points. (The construction of the STPS surface will be described below). We represent the class-conditional probability distribution given a specific choice of Θ as

$$p(\mathbf{p}|\omega, \Theta) = N(\mu_h, \sigma_h|\Theta)N(\mu_\theta, \sigma_\theta|\Theta) \quad (2)$$

where N represents the normal distribution. Here h indicates the height difference in the z -direction between the particle and the STPS surface: $z - f_s(x, y)$, and θ is the

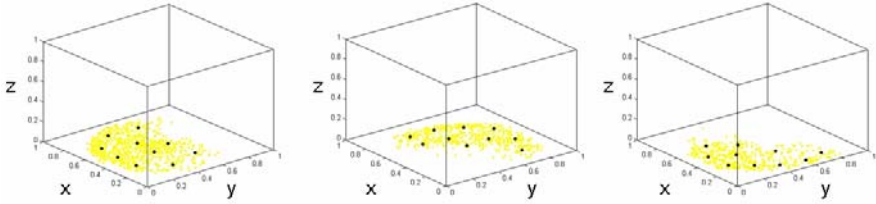


Fig. 2. STPS control points for each of the three fissures: right oblique (left), left oblique (middle), and right horizontal (right). Yellow indicates the pulmonologist-selected points on the training scans, mapped to the unit cube and projected onto the X - Y plane. Black indicates the ten control points used for our MAP estimation.

angle formed between the particle’s eigenvector, e_3 , and the surface’s normal vector at (x, y) . The intuitive description of this distribution is that if a particle is far away from the surface and/or not parallel in orientation with respect to the surface normal, it is unlikely to be a fissure particle. Note that we assume independence between h and θ .

To construct the STPS, we use a parameterization scheme that enables MAP estimation. Each parameter is a control point in our domain (the X - Y plane). By selecting a dispersed set of control points we can fit a surface to the data by controlling their heights (z -values). For each fissure, a different set of control points is needed, and the representation needs to be coordinate frame agnostic. We begin by generating manually segmented CT datasets following [6]: two pulmonologists manually select points on each of the three fissures to segment the lobes. We then map the manually selected points to a common coordinate frame by applying the transform needed to map the corresponding lung’s bounding box to the unit cube (via scaling and translating); this is done independently for the left and right lung. Once all the points are in the unit cube’s coordinate frame, we project their coordinates onto the X - Y plane, which serves as the domain for the STPS function. Note that our approach to placing all points in the same coordinated frame is in lieu of a more elaborate registration scheme. Given that the datasets were acquired with similar patient position (head-first, supine scans) this scheme is sufficient for our purposes. Once all the manually selected points across all the training datasets are projected on the unit cube’s X - Y plane, we have a good idea of where the control points need to be for each fissure STPS. Performing simple k-means clustering then allows us to choose a small set of points that will then be used for subsequent MAP estimation. Figure 2 illustrates the control point selection within the unit cube.

After the control points are chosen, priors for each fissure, $p(\theta)$ are constructed. This is done by determining the parameter vectors for each of the three fissures across all the training datasets. We compute the z -value at each control point using the TPS surface for each fissure in our training set. Thus, for a given fissure the parameter vector can be represented as $\theta = (z_1, \dots, z_{10})$. All of this is done within the unit cube, and it is done independently for each fissure. Once we have our population of parameter vectors, we compute the mean and covariance matrix and represent the priors with a multivariate normal distribution.

With all these terms defined, we can now represent the function we wish to maximize as $\ln p(\mathcal{D}|\Theta) + \ln p(\Theta)$ where

$$p(\mathcal{D}|\Theta) = \prod_{k=1}^n p(\mathbf{p}_k|\omega, \Theta) \tag{3}$$

is the likelihood of Θ given our data.

This is the standard MAP formalism, but we note that this mechanism assumes that all of our data, \mathbf{p} , are drawn from the same distribution. In our case this is not true given that we initially compute the likelihood function across all data points in \mathcal{S} . However, we perform an iterative search for Θ using a gradient descent optimizer (applied to the negative of the log likelihood), and at the end of each convergence we prune away the most unlikely candidate points and then continue the search for Θ . For the pruning step we evaluate the probability of each point, $p(\mathbf{p}|\omega, \Theta)$, and cast away a percentage of the points least likely to be fissure points (we use 30% for our experiments). This “enriches” our dataset over several iterations, increasing the ratio of true fissure particles to non fissure particles until what remains is very close to \mathcal{D} , the set of true fissure points.

Post-Processing. After the MAP estimation stage, we are left with a set of particles along the given fissure of interest. However, some non-fissure particles survive owing to their high probability of being fissure particles as measured by our likelihood function. In order to eliminate these, we perform a final post-processing stage. We make the assumption that fissures are locally planar structures and construct connected components in the following manner. For each remaining particle we compute the vector between its position, \mathbf{p} , and the positions of every other particle. If the magnitude of this vector is above a threshold (chosen to be 5mm in our study), no direct connection is said to exist between them. If the particles are in close proximity to one another, we compute the angle formed between the vector that connects them and each of their eigenvectors, \mathbf{e}_3 . If the angles are sufficiently different than right angles (we use a tolerance of 20°), we assume that they represent different planar structures, and we do not form a connection between them. This is illustrated in the figure below. After all possible connections are made, small connected components are rejected, and we are left with our final set of points through which we compute the TPS that represents the lobe boundaries as in [6].

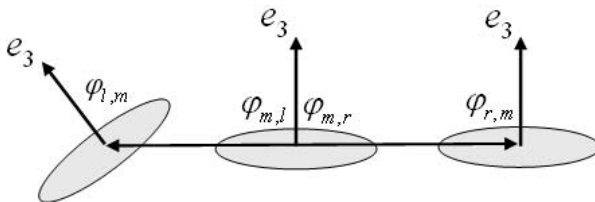


Fig. 3. Connected components analysis. The middle particle, m , is considered connected to the right particle, r , given their proximity and orientation ($\varphi_{m,r} \approx \varphi_{r,m} \approx \pi/2$), but the leftmost particle, l , and the middle particle would not be connected, given their discrepancy in orientation ($\varphi_{l,m} > \pi/2$).

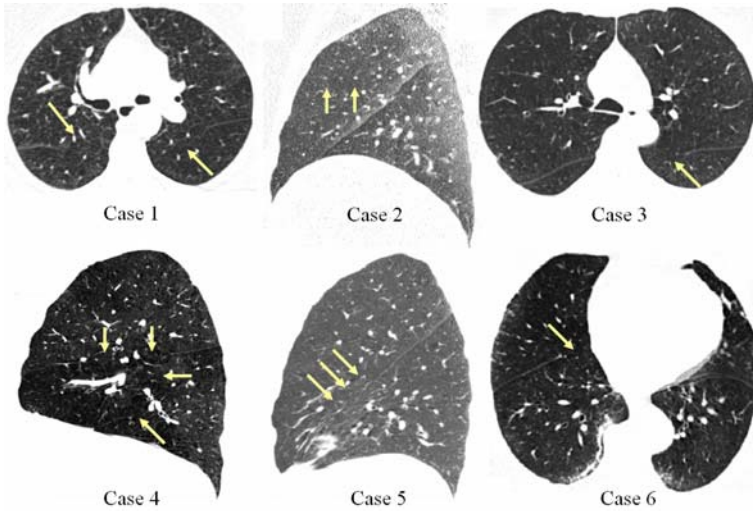


Fig. 4. CT slices of the datasets used in this study. Arrows indicate partial fissures. Case 1: low dose, expiratory scan, severe emphysema; Case 2: low dose, expiratory scan, high BMI, right horizontal fissure nearly absent (feature directly below arrows *appears* as fissure in this slice); Case 3: low dose, expiratory scan, moderate emphysema; Case 4: inspiratory scan, moderate emphysema; Case 5: low dose, expiratory scan, marked fissure aberration; Case 6: inspiratory scan, high BMI, mild diffuse interstitial abnormalities.

3 Results

Here we provide results on a set of challenging cases. For our study we have chosen cases that exhibit incomplete fissures (as determined by pulmonologists and defined as the absence of clearly visible fissure delineating adjacent lobes) and a range of other factors which make the lobe segmentation task difficult (expiratory acquisitions, low dose, high body mass index (BMI), disease presence). The scans have in-plane spacing ranging from 0.52mm to 0.74mm and z-spacing ranging from 0.625mm to 0.75mm. More case details are presented in figure [1](#).

Table [1](#) shows the results of our algorithm on the test cases, and figure [5](#) illustrates an output segmentation mask superimposed on a slice of the CT image from which it was derived. We compared our algorithm's performance to segmentations produced by two pulmonologists (readers 1 and 2), and we additionally compared the readers

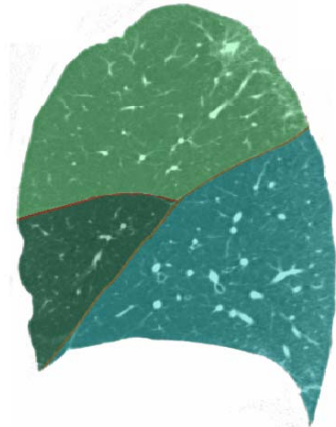


Fig. 5. Sagittal view of right lung (case 3) with algorithm output displayed semi-transparently. Note agreement between segmented boundaries and fissure locations, visible just below the red boundary lines.

Table 1. Algorithm performance comparison to pulmonologist segmentations. Units are in mm. LO = left oblique, RO = right oblique, RH = right horizontal. Note also comparison between readers 1 and 2.

	Reader 1 vs. Alg.			Reader 2 vs. Alg.			Reader 1 vs. Reader 2		
	LO	RO	RH	LO	RO	RH	LO	RO	RH
Case 1	1.63 ± 1.56	0.86 ± 0.96	2.95 ± 2.67	1.82 ± 1.90	2.12 ± 2.01	3.54 ± 5.78	2.75 ± 2.72	2.58 ± 2.36	4.08 ± 3.69
Case 2	3.39 ± 3.19	2.34 ± 3.05	1.21 ± 1.33	1.68 ± 1.69	2.83 ± 3.63	1.02 ± 1.03	2.75 ± 2.59	1.10 ± 1.32	0.80 ± 0.70
Case 3	2.11 ± 2.57	1.81 ± 2.08	1.08 ± 2.96	2.38 ± 2.57	1.29 ± 1.30	1.25 ± 1.43	1.32 ± 1.23	1.29 ± 1.24	0.66 ± 0.81
Case 4	1.82 ± 1.97	4.35 ± 4.65	1.87 ± 2.38	1.84 ± 1.70	6.07 ± 6.99	2.93 ± 5.72	2.83 ± 2.69	2.68 ± 2.72	2.39 ± 2.15
Case 5	0.82 ± 0.92	1.50 ± 1.50	2.92 ± 2.57	1.28 ± 1.71	1.37 ± 1.43	4.68 ± 4.95	1.44 ± 1.54	0.81 ± 0.82	6.44 ± 8.14
Case 6	1.04 ± 1.28	2.52 ± 2.87	1.35 ± 1.27	1.69 ± 1.76	2.29 ± 2.28	4.27 ± 3.99	0.99 ± 1.07	4.01 ± 3.75	4.32 ± 7.73
Average	1.80 ± 1.92	2.23 ± 2.52	1.90 ± 2.20	1.78 ± 1.89	2.66 ± 2.94	2.95 ± 3.82	2.01 ± 1.97	2.08 ± 2.04	3.12 ± 3.87

to each other. To measure segmentation agreement, we computed the average Euclidean distance between the fissures in question. As can be seen, the algorithm performs comparably to the pulmonologists.

4 Discussion and Conclusion

The algorithm execution time was on the order of 30–45 minutes, although we did not attempt to measure this exactly. This was a proof of concept study, and we made no attempt to optimize the algorithm. The most time consuming stage is the MAP estimation, which involves a series of fitting and pruning stages. The fitting stage uses gradient descent to find the optimal heights of our (10) control points. It may be possible to generate equally good results with many fewer control points; we did not investigate this. Reducing the dimensionality of the search space would improve execution time. We also continued the fitting-pruning until we were left with a user-specified fraction of the original number of points. Post-processing was then applied. The fraction that we used was arbitrarily chosen, but terminating the fitting-pruning earlier would reduce execution time and may not adversely affect results.

In summary, we present a fully automatic method for lung lobe segmentation that performs very well on challenging cases. Whereas other methods tend to rely on segmentations of auxiliary structures (vessels, airways) to augment algorithm performance, we only rely on existing fissure image features, even though they may be scarce. We do not deny that other structures can provide contextual clues for fissure localization, but generating segmentations of these structures, especially for patients with advanced disease, presents its own challenges. Furthermore, we believe that obtaining these structures is not necessary and that sufficient fissure information is directly obtainable provided these image features are carefully selected.

References

1. Ukil, S., Reinhardt, J.: Anatomy-guided lung lobe segmentation in x-ray ct images. *IEEE Transactions on Medical Imaging* 28, 202–214 (2009)
2. National Emphysema Treatment Trial Research Group: A randomized trial comparing lung-volume reduction surgery with medical therapy for severe emphysema. *N. Engl. J. Med.* 348, 2059–2073 (2003)

3. Aziz, M., Ashizawa, K., Nagaoki, K., Hayashi, K.: High resolution CT anatomy of the pulmonary fissures. *Journal Thoracic Imaging* 19, 186–191 (2004)
4. van Rikxoort, E., Prokop, M., de Hoop, B., Viergever, M., Pluim, J., van Ginneken, B.: Automatic segmentation of the pulmonary lobes from fissures, airways, and lung borders: Evaluation of robustness against missing data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I. LNCS*, vol. 5761, pp. 263–271. Springer, Heidelberg (2009)
5. Kuhnigk, J., Dicken, V., Zidowitz, S., Bornemann, L., Kuemmerlen, B., Krass, S., Peitgen, H., Yuval, S., Jend, H., Rau, W., Achenbach, T.: New tools for computer assistance in thoracic ct part 1. functional analysis of lungs, lung lobes and bronchopulmonary segments. *Radiographics* 25, 525–536 (2005)
6. Ross, J., San José Estépar, R., Díaz, A., Westin, C., Kikinis, R., Silverman, E., Washko, G.: Lung extraction, lobe segmentation and hierarchical region assessment for quantitative analysis on high resolution computed tomography images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 690–698. Springer, Heidelberg (2009)
7. Kindlmann, G., Estépar, S.J., Smith, R., Westin, S., C.: Sampling and visualizing creases with scale-space particles. *IEEE Transactions on Visualization and Computer Graphics* 15, 1415–1424 (2009)
8. van Rikxoort, E., van Ginneken, B., Klik, M., Prokop, M.: Supervised enhancement filters: application to fissure detection in chest ct scans. *IEEE Transactions on Medical Imaging* 27, 1–10 (2008)
9. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric x-ray CT images. *IEEE Transactions on Medical Imaging* 20, 490–498 (2001)
10. Bookstein, F.: Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1989)
11. Eberly, D.: *Ridges in Image and Data Analysis*. Kluwer Academic Publishers, Dordrecht (1996)

Graph Search with Appearance and Shape Information for 3-D Prostate and Bladder Segmentation^{*}

Qi Song¹, Yinxiao Liu¹, Yunlong Liu¹, Punam K. Saha¹,
Milan Sonka^{1,2,3}, and Xiaodong Wu^{1,2}

¹ Department of Electrical & Computer Engineering,

² Department of Radiation Oncology,

³ Department of Ophthalmology & Visual Sciences, University of Iowa,
Iowa City, IA 52242, USA

{qi-song, xiaodong-wu, milan-sonka}@uiowa.edu

Abstract. The segmentation of soft tissues in medical images is a challenging problem due to the weak boundary, large deformation and serious mutual influence. We present a novel method incorporating both the shape and appearance information in a 3-D graph-theoretic framework to overcome those difficulties for simultaneous segmentation of prostate and bladder. An arc-weighted graph is constructed corresponding to the initial mesh. Both the boundary and region information is incorporated into the graph with learned intensity distribution, which drives the mesh to the best fit of the image. A shape prior penalty is introduced by adding weighted-arcs in the graph, which maintains the original topology of the model and constraints the flexibility of the mesh. The surface-distance constraints are enforced to avoid the leakage between prostate and bladder. The target surfaces are found by solving a maximum flow problem in low-order polynomial time. Both qualitative and quantitative results on prostate and bladder segmentation were promising, proving the power of our algorithm.

1 Introduction

Automated segmentation of medical image has increasingly become a valuable tool for medical diagnosis and treatment planning. Though intensive research has been done, accurate segmentation of 3-D soft tissues is still a challenging problem. The main difficulties lie in the following aspects [12]: First, soft tissues often present a large variation in both shape and size. Second, the target objects often lack strong boundaries and have similar intensity information. Third, many soft tissues have serious mutual influence between each other.

To overcome those difficulties, many methods have been proposed to employ both learned shape and appearance information. One of the widely used methods

^{*} This work was supported in part by NSF grants CCF-0830402 and CCF-0844765, and NIH grants R01-EB004640 and K25-CA123112.

is the AAM model of Cootes *et al.* [3]. The shape and appearance of an object is modeled using principal component analysis (PCA). Freedman *et al.* [1] presented a segmentation algorithm based on matching probability distributions from a learned model of shape and appearance to the image. Rousson *et al.* developed a method to incorporate both shape and appearance model using bayesian formulation [4]. The deformable shape model with shape prior information was employed in [5] and [6] for prostate and bladder segmentation. Recently, graph-based methods with a global optimality guarantee have attracted a lot of attention. Freedman *et al.* [7] proposed an interactive shape prior segmentation based on graph cut algorithms. Malcolm *et al.* [8] employed an iterative graph cut framework with the prior shape information learned from kernel PCA. Besbes *et al.* [9] used an incomplete graph representation for shape modeling and aimed to optimize the connectivity of the graph nodes.

In this paper, we introduce a novel approach incorporating both shape information and object intensity distribution information for prostate and bladder segmentation. Our method is developed based on the framework reported by Kang *et al.* [10,11] and Qi *et al.* [12]. In [12], two coupled geometric graphs are constructed for the prostate and the bladder, respectively. The weights of both graph nodes and arcs are utilized to represent edge-based image energy and a linear soft smoothness penalty. To avoid the possible overlapping of the prostate and the bladder, the hard surface interacting constraints are enforced in the partially interacting regions. The surface segmentation problem is then transformed into that of finding a minimum-cut in a transformed graph.

While the proposed method in [12] was successfully applied for simultaneous segmentation of the prostate and the bladder, it still has several limitations. First, only gradient-based edge weight is employed in the original formulation, which often performs poorly in the presence of weak boundaries, e.g., multiple adjacent objects with similar intensity profiles. Second, only a linear soft smoothness penalty is considered, which limits the flexibility of the shape constraints. In this work, we show how to incorporate the learned intensity distribution information into the graph search framework by using both boundary energy and region energy terms. A convex shape penalty function is also enforced, which allows the incorporation of a wider spectrum of constraints.

2 Optimal Graph Search with Appearance and Shape Information

2.1 Graph Modeling

Our method is mainly based on the graph search framework proposed in [12]. For completeness, we briefly review the framework. As the first step, two initial models are constructed for the prostate and the bladder using the same method as described in [12]. Based on these two initial models, two triangulated meshes $M_1(V_1, E_1)$ and $M_2(V_2, E_2)$ are constructed respectively, where V_i ($i \in 1, 2$) denotes the vertex set of M_i and E_i denotes the edge set of M_i . Fig. 1(a),(b)

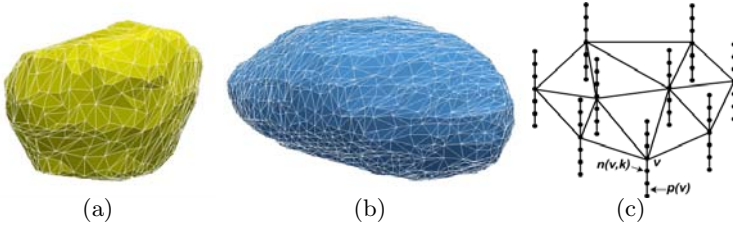


Fig. 1. (a) Triangulated mesh for the prostate. (b) Triangulated mesh for the bladder. (c) Corresponding graph construction. $p(v)$ represents the column with respect to the vertex v . Dots represent nodes $n \in G_i$.

show the constructed meshes for the prostate and the bladder. The weighted graph $G_i(N, A)$ is built from the mesh M_i as follows. For each vertex $v \in V_i$, a column of K nodes $n(v, k)$ is created in G_i , denoted by $p(v)$ (Fig. 1(c)). The positions of nodes reflect the positions of corresponding voxels in the image domain. The length of the column is set according to the required search range. The number of nodes K on each column is determined by the required resolution. The direction of the column is set as the triangle normal. The nodes on the same column are connected by the directed arc from $n(v, k)$ to $n(v, k - 1)$ with infinity weight. Each column also has a set of neighbors, i.e., if $(v_x, v_y) \in E_i$, then $p(v_x)$ and $p(v_y)$ are neighboring columns. The feasible surface S_i in the graph G_i is defined as the surface containing exact one node in each column. To avoid the overlapping of two target surfaces, a “partially interacting area” is defined according to the distance between two meshes, which indicates that the two target surfaces may mutually interact each other at that area. To model the interaction relation, the two graphs G_1 and G_2 “share” some common node columns in that partially interacting area, and the target surfaces S_1 and S_2 both cut those columns, as shown in Fig. 2a. In addition, the distance between the two surfaces at the interacting area is required to be at least δ_l voxels and at most δ_h voxels, which prevents the leakage between prostate and bladder. This surface distance constraint is incorporated by adding inter-surface arcs between corresponding nodes on those columns at the interacting region (Fig. 2b).

The optimal set S of two surfaces corresponding to the prostate and the bladder can then be found by minimizing the following energy:

$$E(S) = \sum_{i=1}^2 E_{boundary}(S_i) + \sum_{i=1}^2 E_{shape}(S_i) \quad (1)$$

The boundary energy term serves as an external force, which drives the mesh towards the best fit to the image data. The shape energy term functions as an internal force, which keeps the shape of the original model and restricts the flexibility of the mesh. In the original framework, a gradient-based edge cost is employed for boundary information, which may fail for surface detection when no strong boundary exists. For shape energy term, a linear soft smoothness

shape compliance is employed. The linear function limited the flexibility of the constraints. In the subsequent sections, we show how to incorporate the learned appearance information using both the boundary energy term and an additional region energy term. Specifically, two surfaces for the prostate and the bladder naturally divide the volume into 3 regions denoted by R_0 , R_1 and R_2 , which corresponds to the region enclosed by the prostate surface S_1 , one between S_1 and the bladder surface S_2 at the partially interacting area, and the region enclosed by S_2 , respectively. Our region energy term $E_{region}(R_i)$ reflects the region property of all voxels inside R_i . In addition, we incorporate a shape prior penalty using a convex function. Thus, the energy function used in this paper is, as follows.

$$E(S) = \sum_{i=1}^2 E_{boundary}(S_i) + \sum_{i=0}^2 E_{region}(R_i) + \sum_{i=1}^2 E_{shape}(S_i) \quad (2)$$

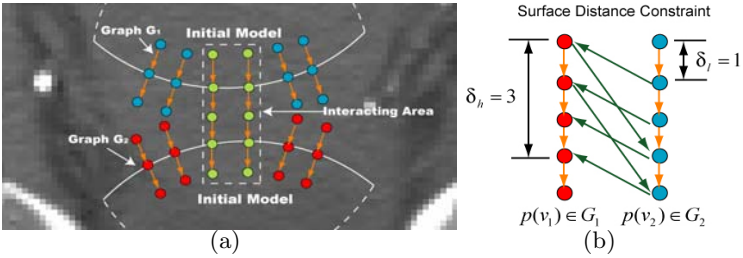


Fig. 2. (a) Graph construction for mutually interacting objects. An example 2-D slice is presented. Note that in the interacting region, for each column with green nodes, there actually exists two columns with the same position, one for Graph G_1 , one for Graph G_2 . (b) Inter-surface arcs are added between the column $p(v_1) \in G_1$ and the column $p(v_2) \in G_2$. $p(v_1)$ and $p(v_2)$ have the same position in the interacting region. The surface distance constraint is set as $1 \leq S_1 - S_2 \leq 3$.

2.2 Boundary Energy Design

For each node $n(v, k) \in G_i$, we have an on-surface cost $c_{S_i}(v, k)$ with respect to surface S_i , which is inversely related to the possibility that the node belongs to the target surface S_i . The boundary energy term $E_{boundary}(S_i)$ is defined as the summation of the on-surface costs associated with all nodes on surface S_i , i.e., $E_{boundary}(S_i) = \sum_{n \in S} c_{S_i}(v, k)$. The on-surface weight of node $n(v, k)$, denoted by $w_{S_i}(v, k)$, is then assigned such that the total weight of a closed set in the graph G_i equals to the boundary energy $E_{boundary}(S_i)$ of the corresponding surface S_i :

$$w_{S_i}(v, k) = \begin{cases} c_{S_i}(v, k) & \text{if } k = 0 \\ c_{S_i}(v, k) - c_{S_i}(v, k - 1) & \text{if } 0 < k \leq K - 1 \end{cases} \quad (3)$$

The on-surface cost function design plays a key role in accurate surface delineation. In soft-tissue segmentation, there is often no clear boundary information

between adjacent tissues. The intensity gradient at the interface of adjacent tissues is also low. To overcome the difficulty, we incorporate the object class-uncertainty information into our cost function. Given *a priori* knowledge of intensity probability distributions of each object, the class-uncertainty can be computed for each intensity value, which yields the uncertainty level of the classification. Suppose the image is divided into m objects, denoted by o_i , $i \in [0, 1, \dots, m - 1]$. For any given node $n(v, k)$ with image intensity $g(v, k)$, the posteriori probability $p(o_i|g)$ is obtained from the training set using Bayes rules. The uncertainty measure for the classification that $n(v, k)$ with an intensity of g falls into any object is the entropy of all posteriori probability values, with the form: $h(g) = \sum_i -p(o_i|g)\log(p(o_i|g))$. As demonstrated in [13][14], intensities with high class uncertainty measures are tend to appear close to object boundaries rather than inside homogeneous regions. Combined with the gradient information, the uncertainty measure is expected to provide an enhanced on-surface cost at the locations absent a clear boundary. Let $\nabla(v, k)$ denote the gradient magnitude at node $n(v, k)$, our boundary cost function has the following form:

$$c_{S_i}(v, k) = -(\nabla(v, k) + \alpha h(g(v, k))) \quad (4)$$

where α is a constant parameter. Here we set $\alpha = 0.5$ according to the experiments on the training set.

2.3 Incorporation of Region Information

The graph search framework allows easy incorporation of region information by assigning proper weights for the graph nodes. As described in Section 2.1, R_0 , R_1 , and R_2 denote the region enclosed by the prostate surface S_1 , one between S_1 and the bladder surface S_2 at the partially interacting area, and the region enclosed by S_2 , respectively. For each node $n(v, k)$, the in-region cost associated with R_i is assigned as $c_{R_i}(v, k)$ ($i = 0, 1, 2$). The region energy term is then defined as the following form: $E_{region}(R_i) = \sum_{n(v, k) \in R_i} c_{R_i}(n(v, k))$. The in-region weight of node $n(v, k)$ in graph G_i , denoted by $w_{R_i}(v, k)$, is assigned such that the total weight of a closed set in graph G equals to the region energy term $\sum_{i=0}^2 E_{region}(R_i)$ (with a constant difference) [15]:

$$w_{R_i}(v, k) = c_{R_{i-1}}(n(v, k)) - c_{R_i}(n(v, k)) \quad (5)$$

For region cost design, the posterior probability learned from the training set is used with the form:

$$c_{R_i}(n(v, k)) = -p(n(v, k) \in R_i | g(v, k)), i \in \{0, 1, 2\} \quad (6)$$

2.4 Shape Constraints

The shape energy keeps the topology of the original mean shape. In our framework, two types of shape constraints are enforced: the hard shape constraint and the shape-prior penalties.

Suppose $p(v_x)$ and $p(v_y)$ are neighboring columns in the graph. If $n(v_x, k_1)$ of $p(v_x)$ and $n(v_y, k_2)$ of $p(v_y)$ are nodes on the surface S_i , we let $S_i(v_x) = k_1$ and $S_i(v_y) = k_2$. Then the hard shape constraint is defined as follows: $\Delta_l \leq S_i(v_x) - S_i(v_y) \leq \Delta_h$. Δ_l and Δ_h are specified shape constraint parameters between $p(v_x)$ and $p(v_y)$. To enforce the hard shape constraint, a directed arc with $+\infty$ weight is put from each node $n(v_x, k)$ to the node $n(v_y, k - \Delta_h)$. On the other hand, we have a directed arc with $+\infty$ from the node $n(v_y, k)$ to $n(v_x, k + \Delta_l)$.

The shape-prior penalties are set as $f(S_i(v_x) - S_i(v_y))$, where $f(\cdot)$ is a convex function penalizing the shape changes of S_i between neighboring column $p(v_x)$ and $p(v_y)$. Based on this definition, our shape energy term has the form:

$$E_{shape}(S_i) = \sum_{(v_x, v_y) \in E_i} f(S_i(v_x) - S_i(v_y)) \tag{7}$$

To incorporate the shape energy, additional inter-column arcs are employed. The main idea is to distribute the convex shape prior penalty $f(S_i(v_x) - S_i(v_y))$ to the corresponding cuts between the columns in G based on the (discrete equivalent of) second derivative of $f(\cdot)$, $f''(h) = [f(h + 1) - f(h)] - [f(h) - f(h - 1)]$, as described in [16,17]. Since $f(h)$ is a convex function, $f''(h) \geq 0$. For each $h = \Delta_l + 1, \Delta_l + 2, \dots, \Delta_h - 1$, if $f'(h) \geq 0$, an arc from $n(v_x, k)$ to $n(v_y, k - h)$ is assigned with an arc-weight of $f''(h)$. If $f'(h) \leq 0$, an arc from $n(v_y, k)$ to $n(v_x, k + h)$ is assigned with the weight of $f''(h)$ (Fig. 3(b)). Using this construction, the total weight of the arcs spanning between two neighboring columns $p(v_x)$ and $p(v_y)$ equals to the shape prior penalty $f(S_i(v_x) - S_i(v_y))$.

For prostate and bladder segmentation, a second order shape prior penalty is employed with the form: $f(h) = \beta \cdot h^2$, where β is a constant parameter learned from the training set. In this project $\beta = 5$.

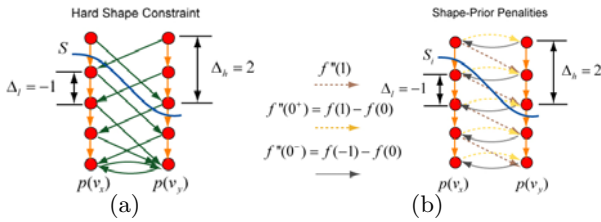


Fig. 3. Arc construction for shape constraints. (a) Hard shape constraints. (b) Shape-prior penalties. Note that here we suppose $f'(0) = 0$. S cuts the arcs with a total weight of $f(2)$.

2.5 Optimization

Once the graph is constructed, an optimal cut $\mathcal{C}^* = (B^*, \bar{B}^*)$ ($B^* \cup \bar{B}^* = N$) can be found in $G(N, A)$, minimizing the total weight of nodes in B^* and the total arc weight of \mathcal{C}^* . The optimal cut actually corresponds to the optimal set S of

two surfaces S_1 and S_2 , which minimizes the total energy $E(S)$. As described in [18], the optimal cut can be found by solving a maximum flow problem in a transformed graph G' in a low-order polynomial time.

3 Experiments and Results

The experiments were conducted for simultaneous segmentation of prostate and bladder. 3-D CT images from different patients with the prostate cancer are used. The image size ranges from $80 \times 120 \times 30$ to $190 \times 180 \times 80$ voxels. The image spacing resolution ranged from $0.98 \times 0.98 \times 3.00 \text{ mm}^3$ to $1.60 \times 1.60 \times 3.00 \text{ mm}^3$. Out of 21 volumes, 8 were randomly selected as the training data and our segmentation was performed on the remaining 13 datasets. Our workflow mainly consisted of three major steps. *Step 1.* Initial model construction. We use the similar method as described in [12]. Since the prostate often shows a good statistical coherency in shape among different datasets [4], a mean shape is obtained from the training set and roughly fitted into the image as the initial model. For bladder, a 3-D geodesic method is employed, the result provides the basic topological structure information. *Step 2.* Iterative single mesh deformation for prostate and bladder, separately. Using iterative graph search, we can find target surfaces which are far away from the initial model. In our experiment, we set iteration times as 3, which is proved enough according to the experimental result. *Step 3.* Simultaneous graph search for bladder and prostate. The possible overlapping was avoided by adding surface distance constraint between the two meshes.

For quantitative validation, the result was compared with the expert-defined manual contours. For volumetric error measurement, the Dice similarity coefficient (DSC) was computed using $D = 2|V_m \cap V_c|/(|V_m| + |V_c|)$, where V_m denotes the manual volumetric result and V_c denotes the computed result. For surface distance error, both mean and the maximum unsigned surface distance error were computed for the bladder and the prostate surfaces between the computed result and the manual delineation. The result is shown in Table 1. Compared with the result reported in [12], the unsigned surface distance error has improved from $1.38 \pm 1.08 \text{ mm}$ to $1.01 \pm 0.94 \text{ mm}$ for the prostate and from $1.04 \pm 1.00 \text{ mm}$ to $0.99 \pm 0.77 \text{ mm}$ for the bladder.

Table 1. Overall quantitative results. Mean \pm SD in *mm* for the unsigned surface distance error.

<i>Surface</i>	<i>DSC</i>	<i>Mean(mm)</i>	<i>Maximum(mm)</i>
Prostate	0.797	1.01 ± 0.94	5.46 ± 0.96
Bladder	0.900	0.99 ± 0.77	5.88 ± 1.29

For a visual performance assessment, the illustrative result with both computed contours and manual contours was displayed in Fig 4(a). The 3-D representation was shown in Fig 4(b).

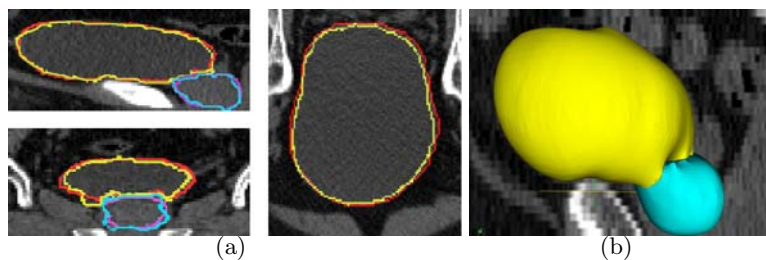


Fig. 4. (a) Sagittal, coronal and transverse slices with the computed result (blue for prostate and yellow for bladder) and the manual result (magenta for prostate and red for bladder). (b) 3-D representation of the prostate (blue) and the bladder (yellow).

The execution time on a WinXP PC (2.13×2 GHz, 2GB memory) was approximately 8 minutes per dataset.

4 Discussion and Conclusion

We have presented how to incorporate boundary, region and shape information into an optimal 3-D graph search framework for multiple objects segmentation. The general formulation has been adapted to the simultaneous segmentation of the prostate and the bladder. Both qualitative and quantitative validation were employed. The result shows the power of our algorithm.

References

1. Freedman, D., Radke, R.J., Zhang, T., Jeong, Y., Lovelock, D.M., Chen, G.T.Y.: Model-based segmentation of medical imagery by matching distributions. *IEEE Trans. Medical Imaging* 24(3), 281–292 (2005)
2. Vu, N., Manjunath, B.S.: Shape prior segmentation of multiple objects with graph cuts. In: *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2008)
3. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Trans. Pattern Anal. Machine Intell* 23(6), 681–685 (2001)
4. Rousson, M., Khamene, A., Diallo, M., Celi, J.C., Sauer, F.: Constrained surface evolutions for prostate and bladder segmentation in CT images. In: *Proc. of the First International Workshop on Computer Vision for Biomedical Image Applications*, pp. 251–260 (2005)
5. Costa, M.J., Delingette, H., Novellas, S., Ayache, N.: Automatic segmentation of bladder and prostate using coupled 3D deformable models. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 252–260. Springer, Heidelberg (2007)
6. Heimann, T., Munzing, S., Meinzer, H.P., Wolf, I.: A shape-guided deformable model with evolutionary algorithm initialization for 3d soft tissue segmentation. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007. LNCS*, vol. 4584, pp. 1–12. Springer, Heidelberg (2007)

7. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: Proc. IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 755–762 (2005)
8. Malcolm, J., Rathi, Y., Tannenbaum, A.: Graph cut segmentation with nonlinear shape priors. In: Proc. International Conference on Image Processing (2007)
9. Besbes, A., Komodakis, N., Langs, G., Paragios, N.: Shape priors and discrete mrfs for knowledge-based segmentation. In: Proc. IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1295–1302 (2009)
10. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images - a graph-theoretic approach. *IEEE Trans. Pattern Anal. Machine Intell.* 28(1), 119–134 (2006)
11. Li, K., Jolly, M.P.: Simultaneous detection of multiple elastic surfaces with application to tumor segmentation in CT images. In: Proc. SPIE Medical Imaging: Image Processing, vol. 6914 (2008)
12. Song, Q., Wu, X., Liu, Y., Smith, M., Buatti, J.M., Sonka, M.: Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 827–835. Springer, Heidelberg (2009)
13. Saha, P.K., Udupa, J.K.: Optimum threshold selection using class uncertainty and region homogeneity. *IEEE Trans. Pattern Anal. Machine Intell.* 23, 689–706 (2001)
14. Saha, P.K., Das, B., Wehrli, F.W.: An object class-uncertainty induced adaptive force and its application to a new hybrid snake. *Pattern Recognition* 40, 2656–2671 (2007)
15. Haeker, M., Abramoff, M.D., Wu, X., Russell, S.R., Burns, T.L., Sonka, M.: Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans. Medical Imaging* 28(9), 1436–1447 (2009)
16. Ishikawa, H.: Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Machine Intell.* 25(10), 1333–1336 (2003)
17. Song, Q., Wu, X., Liu, Y., Haeker, M., Sonka, M.: Simultaneous searching of globally optimal interacting surfaces with convex shape priors. In: Proc. IEEE International Conference on Computer Vision and Pattern Recognition (2010)
18. Wu, X., Chen, D.Z.: Optimal net surface problems with applications. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) ICALP 2002. LNCS, vol. 2380, pp. 1029–1042. Springer, Heidelberg (2002)

Segmentation of Cortical MS Lesions on MRI Using Automated Laminal Profile Shape Analysis

Christine L. Tardif¹, D. Louis Collins¹, Simon F. Eskildsen²,
John B. Richardson³, and G. Bruce Pike¹

¹ McConnell Brain Imaging Centre, Montreal Neurological Institute, Canada

² Dept. of Health Science and Technology, Aalborg University, Denmark

³ Dept. of Neuropathology, Montreal Neurological Institute/Hospital, Canada

Abstract. Cortical multiple sclerosis lesions are difficult to detect in magnetic resonance images due to poor contrast with surrounding grey matter, spatial variation in healthy grey matter and partial volume effects. We propose using an observer-independent laminal profile-based parcellation method to detect cortical lesions. Following cortical surface extraction, profiles are extended from the white matter surface to the grey matter surface. The cortex is parcellated according to profile intensity and shape features using a k-means classifier. The method is applied to a high-resolution quantitative magnetic resonance data set from a fixed *post mortem* multiple sclerosis brain, and validated using histology.

1 Introduction

Multiple sclerosis (MS) is classically defined as a white matter (WM) disease even though the involvement of grey matter (GM) in MS pathology has been recognized since the beginning of the 20th century. Recent immunohistochemistry (IHC) studies have shown that cortical GM lesions are common and widespread in MS brains [1,7], yet these lesions remain extremely difficult to detect *in vivo* using MRI. The characterization and segmentation of cortical lesions *in vivo* is essential to improve our understanding of the natural course of the disease and monitor its progression.

The lack of sensitivity of MRI is due to the different pathophysiology of GM in comparison to WM lesions, in particular the absence of inflammation and edema. The contrast between cortical lesions and surrounding normal appearing grey matter (NAGM) is further dampened by the lower myelin content of GM in comparison to WM (~10%), particularly in the superficial cortical layers. The morphology of subpial lesions, that extend from the pial surface and can span several gyri, may also cause them to be concealed by partial volume effects with cerebral spinal fluid (CSF). In addition to poor local contrast, intensity driven lesion segmentation is also hindered by the presence of biological variation across healthy GM and additional variation caused by diffuse MS pathology.

The most promising MRI sequence for *in vivo* lesion detection thus far is double inversion recovery (DIR) [6], yet according to histology reports cortical

lesions remain under-detected by this technique. Quantitative MR imaging techniques, such as relaxometry and magnetization transfer imaging, have reported abnormalities in MS patients *in vivo*; however, cortical lesions are not segmented from the NAGM in these techniques mainly due to low spatial resolution. A combined *post mortem* high resolution quantitative MRI and quantitative histology study of cortical lesions by Schmierer *et al* [9] clearly delineated cortical lesions and showed that T_1 is a predictor of neuronal density and T_2 of myelin content.

The cortical lesion segmentation technique proposed here is based on a quantitative architectural analysis method originally applied to histological data [8]. Under the assumption that the changes in myelo- and cyto-architecture laminar profiles will be reflected in the quantitative MR profiles, cortical parcellation based on laminar profile shape analysis is used to detect and delineate lesions. The methods consist of standard image processing tools adapted to analyze a unique very high-resolution 3D quantitative MRI data set of a fixed *post mortem* MS hemisphere, and validation using IHC.

2 Methods

2.1 *Post Mortem* Brain Tissue

The right hemisphere of an MS patient (79 year old female, 30 years disease duration, cause of death aspiration pneumonia) was provided by the Douglas Hospital Research Centre Brain Bank. The hemisphere was fixed in 10% buffered formalin after a *post mortem* delay of 41.25 hours, and had been fixed for approximately 4 years.

2.2 Magnetic Resonance Image Acquisition

All images were acquired on a Siemens TIM Trio 3T MRI scanner with a 32-channel receive-only head coil. The hemisphere was placed in an MR-compatible cylindrical container filled with formalin. 3D sagittal images were acquired with 0.35 mm isotropic resolution, $512 \times 512 \times 240$ matrix size and 6/8 partial Fourier phase encoding. The total acquisition time was of ~ 55 hours.

Relaxometry was performed using the variable flip angle method also known as DESPOT [3]. T_1 and relative proton density, M_0 , maps were calculated from two spoiled gradient echo (SPGR) images with a constant echo time (T_E) of 3.35 ms and repetition time (T_R) of 7.7 ms. The flip angles, optimized for the range of relaxation times of the fixed *post mortem* brain, were 4° and 22° . T_2 maps were derived from the measured T_1 times and two balanced steady state free precession (bSSFP) images with a fixed T_E of 3.84 ms and T_R of 7.7 ms, and optimal flip angles 20° and 70° . Each sequence was repeated 49 times, for a total scan time of ~ 38.5 hours.

The magnetization transfer ratio (MTR) was calculated from two proton density weighted SPGR acquisitions with α , T_E and T_R set to 25° , 4.09 ms and 25 ms respectively. The second acquisition included the MT saturation pulse provided on the Siemens Trio 3T scanner: a 500° Gaussian pulse of 10ms, 1200 Hz

off-resonance with a 100 Hz bandwidth. 13 repetitions were acquired for a total scan time of ~ 16.5 hours.

Due to the strong variations in the RF transmission field at 3T, we acquired a B_1 map [10] to correct the nominal flip angles for both DESPOT techniques. The B_1 map was derived from two magnetization prepared turbo-spin echo (TSE) images with α , $T_{EchoSpacing}$ and T_R were set to 20° , 15 ms and 2 s respectively, with a turbo factor of 7. The acquisition was 2D with 2 mm isotropic resolution and $128 \times 128 \times 50$ matrix size. The TSE readout was preceded by an α and 2α pulse for the first and second acquisition respectively, followed by a time delay equal to half the echo-spacing.

2.3 Histology and Immunohistochemistry

Coronal slices, 1 cm thick, were cut, photographed and scanned at a lower resolution ($0.35 \times 0.35 \times 1 \text{ mm}^3$) to facilitate the manual alignment of the histology to the whole hemisphere MRI. Tissue blocks representing cortical lesions and NAGM in the MR images were selected from several different cytoarchitectonic areas for paraffin embedding. Sections were cut at $5 \mu\text{m}$, and reacted with antibodies directed against myelin basic protein (MBP), and processed in a Ventana Benchmark XT with diaminobenzidine (DAB) as chromogen. Slides were digitized using a Zeiss MIRAX Scan automated slide scanner.

2.4 Image Processing

Preprocessing. The individual raw images were linearly aligned prior to averaging [2]. The average images were used to calculate the quantitative MR maps: T_1 , M_0 , T_2 and MTR. The M_0 and T_2 maps were non-uniformity corrected [11] for reception field inhomogeneity and banding artifacts respectively.

The MTR map, which most resembles *in vivo* T_1 -weighted contrast, was used for cortical surface extraction. The MTR image was linearly and then non-linearly aligned to the right hemisphere of the ICBM152 non-linear atlas and masked. The non-uniformity corrected MT-weighted SPGR image was used in combination with the MTR map for discrete tissue classification [14]. As shown in Figure 1, the formalin is clearly distinguishable from the surrounding GM in the sulci on the MT-weighted SPGR image due to the high contrast and resolution. The discrete tissue classification results were subsequently corrected for partial volume effects [12] and fuzzy tissue maps were created for WM, GM and formalin. Stereotaxic masks were applied to remove the brain stem and cerebellum from the tissue maps, and to label the subcortical GM and ventricles as WM. Small manual corrections of the masks were required due to the deformation caused by the fixation, otherwise the preprocessing was fully automated. An edge map of the GM/formalin interface was created by calculating the gradient of the sum of the WM and GM fuzzy tissue maps.

Cortical Boundary and Laminal Profile Extraction. The cortical boundary surfaces were extracted using FACE (Fast Accurate Cortex Extraction) [4,5],

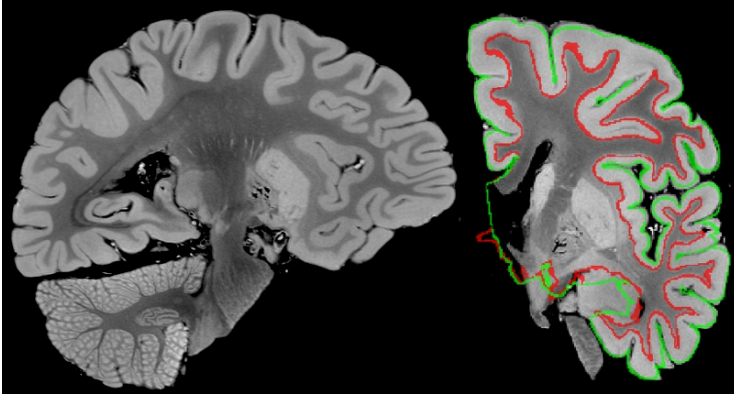


Fig. 1. Cortical surface extraction. Left: High-resolution MT-weighted SPGR image of the fixed MS hemisphere. Right: The WM and GM surfaces superimposed on the MT-weighted image in red and green respectively.

which uses deformable surfaces and a force balancing scheme. An initial surface was created by applying an iso-surface algorithm to the WM map to create a closed surface consisting of a triangulated mesh. This initial surface was then deformed iteratively to the WM/GM boundary under the influence of forces derived from the fuzzy tissue maps and gradient image. The WM surface was expanded under the influence of deformation forces derived from the surface normals, a gradient vector field, and the GM/formalin edge map. The resulting surfaces, shown superimposed on the MT-weighted image in Figure 1, consisted of approximately 240 thousand vertices uniformly distributed over the cortex.

Correspondence between vertices on the GM and WM surfaces was determined by a combination of the GM surface normals and the nearest point on the opposite WM surface. The nearest point Euclidean distances were used to constrain the results from the surface normals. The laminar profiles were extracted by sampling 20 points between corresponding vertices on the WM and GM surfaces using linear interpolation.

Laminar Profile Analysis. The laminar profile shape analysis is based on an observer-independent quantitative architecture analysis technique for cortical mapping of histological data [8]. This technique has been applied to high-resolution MRI to parcellate the visual cortex *in vivo* [13]. The first and last 10% of samples were removed to avoid contamination from neighbouring tissues. A 10-parameter feature vector was extracted from each laminar profile, including the mean amplitude and first four central moments of the profile and its absolute first derivative [8, 13]. The feature vectors derived from the 4 quantitative MR maps (T_1 , M_0 , T_2 and MTR) were concatenated, for a total of 40 features per profile. Each feature was z-scored to equally weigh the k-means classification. The profiles were classified into 4 clusters using a k-means classifier with a squared Euclidean distance metric. The seeds of the 4 clusters were sampled at random.

3 Results

The MBP IHC and segmentation results of a section of the superior frontal gyrus are shown in Figure 2. In the MBP section, the myelin fibers are stained brown such that the dark brown area corresponds to the densely myelinated WM and the lighter brown area corresponds to cortical GM. The light suprabial band along the sulcus corresponds to a demyelinated cortical MS lesion. The corresponding laminar profile-based parcellation results differentiate the lesion, labeled white, from the neighbouring NAGM, labeled green.

The cortical segmentation results are shown on the mid-cortical surface, i.e. the median of the WM and GM surfaces, in Figure 3. The cortex was classified

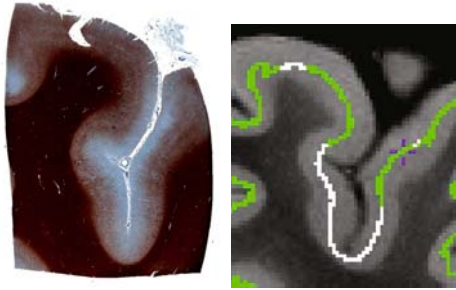


Fig. 2. Validation of cortical parcellation results with histology. Left: Tissue section of the superior frontal gyrus immuno-stained against MBP. Right: Corresponding parcellation results superimposed on an MR image. The NAGM is labeled green, and the demyelinated cortex is labeled white.

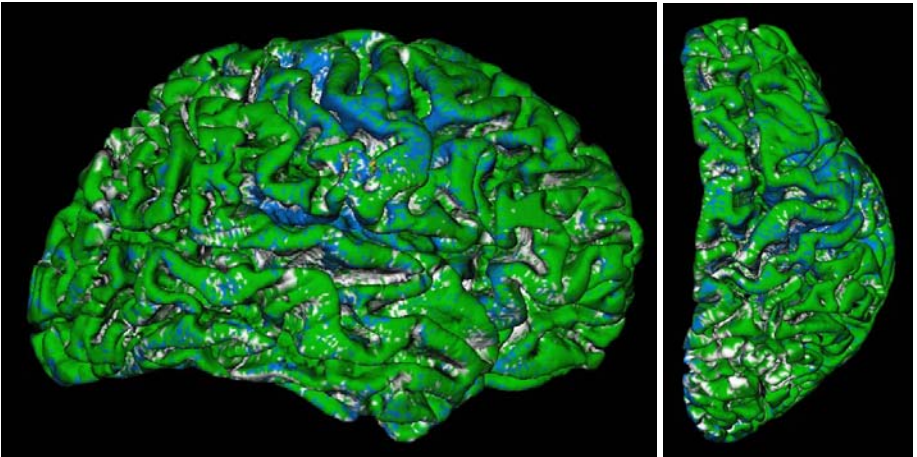


Fig. 3. Cortical parcellation results displayed on the mid-cortical surface

Table 1. Quantitative MRI results of the fixed MS hemisphere. Mean (standard deviation).

	Normal appearing GM Cortical lesion	
T ₁ (ms)	187 (13)	236 (12)
T ₂ (ms)	58 (6)	74 (2)
M ₀	3658 (81)	3914 (69)
MTR (%)	7.35 (1.06)	7.52 (1.2)

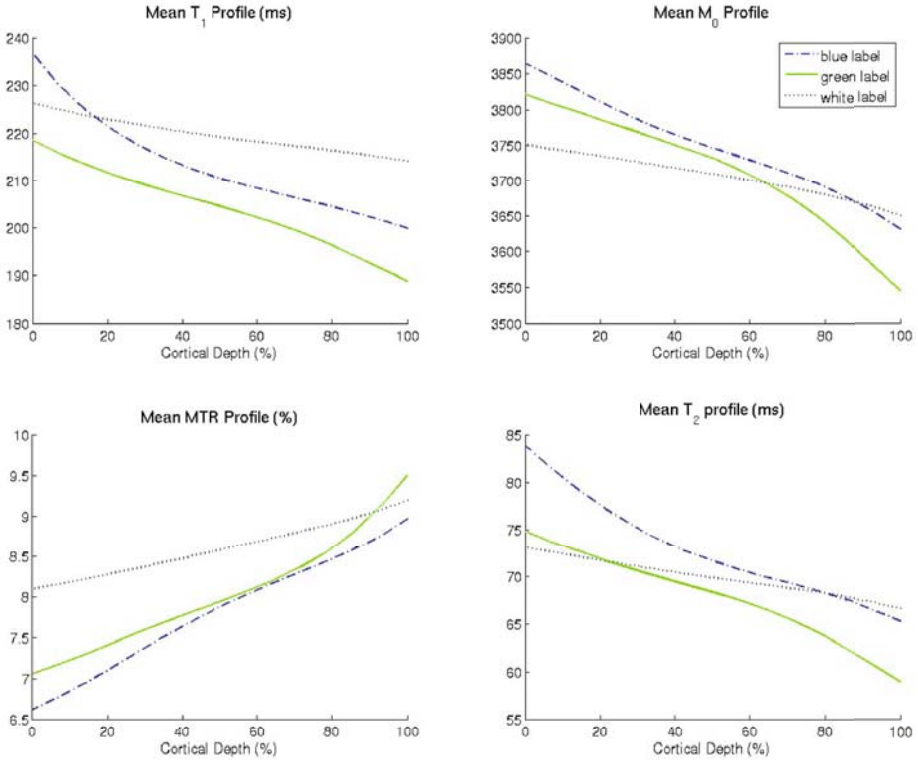


Fig. 4. Mean cortical profiles of the quantitative MR maps for each class of the cortex

into four different laminar patterns. The first class, labeled green, corresponds to NAGM. The classes labeled blue and white represent different types of cortical pathology. The fourth class represents non-cortical tissue.

Three cortical lesions were delineated in the MBP immuno-stained sections, and corresponding regions of interest (ROIs) drawn in the quantitative MRI maps. ROIs of NAGM were also chosen for comparison. Paired Student’s T-tests were performed to determine the significance of the differences observed. The quantitative MRI results for the NAGM and cortical lesions are listed in Table 1. The differences in T₁, T₂, M₀ and MTR between cortical lesions and NAGM were all statistically significant ($p < 0.0001$). The M₀ values were taken from the

non-uniformity corrected image. The ROIs were placed in regions that were not affected by the banding artifacts in the SSFP images. T_1 and T_2 relaxation times are higher in the cortical lesions in comparison to the NAGM. These results are in agreement with a previous fixed *post mortem* study by Schmierer *et al* [9]. The relative proton density M_0 is also increased in cortical lesions. We also observed a small but significant increase in MTR in the cortical lesions in comparison to the NAGM, whereas a decrease in MTR was reported by Schmierer *et al* in GM lesions.

The mean cortical profiles of the quantitative MR maps for each class are plotted in Figure 4, where 0% cortical depth corresponds to the GM/formalin boundary and 100% cortical depth to the WM/GM boundary. The areas labeled blue on the surfaces in Figure 3 show an increase in T_1 and T_2 times in comparison to the NAGM labeled green, in particular towards the superficial cortical layers. These regions are also characterized by an increase in M_0 and a decrease in MTR near the tissue boundaries. The areas labeled white (black dotted line in Figure 4) are characterized by a more uniform laminar profile and an increase in T_1 and MTR in comparison to the NAGM in green.

4 Discussion

We presented an automated laminar profile shape analysis technique for the segmentation of cortical MS lesions. The technique was demonstrated on a high-resolution 3D quantitative MR data set of a fixed MS hemisphere and validated with IHC. This unique combined *post mortem* quantitative MRI and IHC study is essential to improve our understanding of the relationship between MR parameters and the pathological substrates of cortical lesions.

This segmentation technique could also be applied to *in vivo* data with a lower SNR and CNR. Diffuse smoothing along the cortical mantle could be applied to improve sensitivity of profile-based morphometry to cortical lesions, in particular subpial lesions given their morphology. Furthermore, this data set can be used to guide the design of new *in vivo* acquisition techniques with optimal sensitivity to cortical pathology. We believe that this combined optimized image acquisition and analysis approach is very promising for *in vivo* lesion detection, which is essential to understand the natural course of the disease and monitor its progression.

A potential source of error in the cortical parcellation method is the inaccuracy in tissue classification and subsequent surface extraction. Cortical lesions could be misclassified as formalin/CSF, for instance, and thus remain undetected by this method.

References

1. Bo, L., Vedeler, C.A., Nyland, H.I., Trapp, B.D., Mork, S.J.: Subpial demyelination in the cerebral cortex of multiple sclerosis patients. *J. Neuropathol. Exp. Neurol.* 62(7), 723–732 (2003)

2. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3D intersubject registration of MR volumetric data in standardized talarach space. *Journal of Computer Assisted Tomography* 18(2), 192–205 (1994)
3. Deoni, S.C., Rutt, B.K., Peters, T.M.: Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magn. Reson. Med.* 49(3), 515–526 (2003)
4. Eskildsen, S.F., Ostergaard, L.R.: Active surface approach for extraction of the human cerebral cortex from MRI. In: *Int. Conf. Med. Image Comput. Assist. Interv.*, vol. 9(Pt. 2), pp. 823–830 (2006)
5. Eskildsen, S.F., Ostergaard, L.R., Rodell, A.B., Ostergaard, L., Nielsen, J.E., Isaacs, A.M., Johannsen, P.: Cortical volumes and atrophy rates in FTD-3 CHMP2B mutation carriers and related non-carriers. *NeuroImage* 45(3), 713–721 (2009)
6. Geurts, J.J., Pouwels, P.J., Uitdehaag, B.M., Polman, C.H., Barkhof, F., Castelijns, J.A.: Intracortical lesions in multiple sclerosis: improved detection with 3D double inversion-recovery MR imaging. *Radiology* 236(1), 254–260 (2005)
7. Kutzelnigg, A., Lucchinetti, C.F., Stadelmann, C., Bruck, W., Rauschka, H., Bergmann, M., Schmidbauer, M., Parisi, J.E., Lassmann, H.: Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain* 128(Pt. 11), 2705–2712 (2005)
8. Schleicher, A., Morosan, P., Amunts, K., Zilles, K.: Quantitative architectural analysis: A new approach to cortical mapping. *J. Autism Dev. Disord.* (2009)
9. Schmierer, K., Parkes, H.G., So, P.W., An, S.F., Brandner, S., Ordidge, R.J., Yousry, T.A., Miller, D.H.: High field (9.4 Tesla) magnetic resonance imaging of cortical grey matter lesions in multiple sclerosis. *Brain* 133(Pt. 3), 858–867 (2010)
10. Sled, J.G., Pike, G.B.: Correction for B1 and B0 variations in quantitative T2 measurements using MRI. *Magn. Reson. Med.* 43(4), 593 (2000)
11. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17(1), 87–97 (1998)
12. Tohka, J., Zijdenbos, A., Evans, A.: Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage* 23(1), 84–97 (2004)
13. Walters, N.B., Eickhoff, S.B., Schleicher, A., Zilles, K., Amunts, K., Egan, G.F., Watson, J.D.: Observer-independent analysis of high-resolution MR images of the human cerebral cortex: In vivo delineation of cortical areas. *Hum. Brain Mapp.* 28(1), 1–8 (2007)
14. Zijdenbos, A.P., Forghani, R., Evans, A.C.: Automatic ‘pipeline’ analysis of 3D MRI data for clinical trials: application to multiple sclerosis. *IEEE Transactions on Medical Imaging* 21(10), 1280–1291 (2002)

3D Knowledge-Based Segmentation Using Pose-Invariant Higher-Order Graphs

Chaohui Wang^{1,2}, Olivier Teboul^{1,3}, Fabrice Michel^{1,2},
Salma Essafi^{1,2}, and Nikos Paragios^{1,2}

¹ Laboratoire MAS, Ecole Centrale de Paris, France

² Equipe GALEN, INRIA Saclay - Île de France, Orsay, France

³ Microsoft France

chaohui.wang@ecp.fr

Abstract. Segmentation is a fundamental problem in medical image analysis. The use of prior knowledge is often considered to address the ill-posedness of the process. Such a process consists in bringing all training examples in the same reference pose, and then building statistics. During inference, pose parameters are usually estimated first, and then one seeks a compromise between data-attraction and model-fitness with the prior model. In this paper, we propose a novel higher-order Markov Random Field (MRF) model to encode pose-invariant priors and perform 3D segmentation of challenging data. The approach encodes data support in the singleton terms that are obtained using machine learning, and prior constraints in the higher-order terms. A dual-decomposition-based inference method is used to recover the optimal solution. Promising results on challenging data involving segmentation of tissue classes of the human skeletal muscle demonstrate the potentials of the method.

1 Introduction

Knowledge-based segmentation consists in recovering a region of anatomical interest in a new image. The process often combines data support with manifold learning on the space of adequate solutions. The data term is usually either edge-based or region driven. In the first case, one seeks to position the solution onto pixels exhibiting important intensity discontinuities, which is achieved through a weighted surface integral. Region-based methods assume that the object and the background have distinct statistical properties and seek to create a partition that maximizes the posterior probability density with respect to them. Both methods fail to address the case of anatomical regions of interest being part of the same tissue class, where either edges are not present or statistical separation is not feasible. The case of calf muscle segmentation is a typical example.

Prior knowledge is often encoded through constraining the solution space. This is achieved either through linear models describing the variation of the training set (ASMs) [1], or through projection and minimization of the distance to the learned manifold. The use of simple statistical means (atlas-based methods) [2], parametric or non-parametric priors considered in an explicit [3] or implicit

level set formulation [4] were considered. Such techniques exhibit two important limitations, the first is related to pose invariance and the second is related to their ability to capture statistics on high-dimensional spaces from a small number of training examples. The pose estimation problem arises both in the training and in the inference steps and introduces certain bias on the segmentation process since data are often to be registered in the reference space. The ratio samples versus dimensionality of representations is also a well-known problem in medical imaging.

In this paper, we propose a novel approach that is able to address segmentation for challenging data sets while being pose invariant and being able to capture local variations from small training sets. To this end, we employ a higher-order Markov Random Field (MRF) formulation. The representation of the shape is a point distribution model (PDM) that is used to provide the entire surface through conventional interpolation algorithms. Prior knowledge is modeled through the use of higher-order relative statistics of the PDM. These are invariant to rotation and scale while they can be learned from a small number of training examples. The entire manifold is described through the accumulation of such local constraints. This representation is associated with the randomized forest [5] learning approach that provides an efficient detection algorithm for points of interest exhibiting some statistical properties. These properties can be derived from Gabor-filter-based scale/rotation invariant features. In order to optimize the higher-order MRF’s energy, we propose to decompose the original problem, which is difficult to solve directly, into a series of sub-problems each of which corresponds to a factor tree [6,7]. The inference in a factor tree can be done exactly and very efficiently using max-product belief propagation algorithm [7]. In order to evaluate the performance of the method, we have considered the challenging application of segmentation of the calf muscle. The closest work refers to the segmentation algorithms in [8,9,10]. Opposite to these algorithms, our approach explores the 3D space and higher-order interactions between the model variables, inherits natural invariance with respect to the global pose (opposite to [9]), exploits invariant features with respect to this pose (opposite to [9,10]), builds pose invariant statistics (opposite to [8,9,10]) and provides an one-shot optimization result that does not depend on the initial conditions or the reference pose parameters.

The remainder of this paper is organized as follows: we present the higher-order shape prior, the MRF segmentation formulation and the MRF optimization method in section 2. The experimental validation composes section 3, while section 4 concludes the paper.

2 Knowledge-Based Segmentation

2.1 Shape Modeling

The shape model consists of a set \mathcal{V} of control points/landmarks that are located on the boundary of the object (Fig. 1(a)). Let x_i ($i \in \mathcal{V}$), a 3-dim vector, denote the 3d position of landmark i and $\mathbf{x} = (x_i)_{i \in \mathcal{V}}$ denote the position of

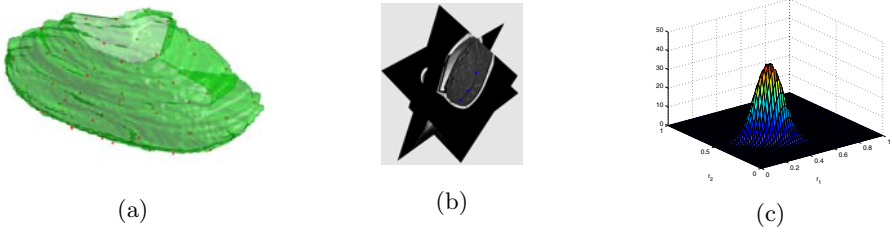


Fig. 1. Shape Model. (a) Distribution of the landmarks on the muscle boundary. (b) Two perpendicular slices with a triplet of landmarks (the blue asterisks). (c) The learned Gaussian distribution on \mathbf{r}_c for the triplet shown in (b).

all the landmarks. Given a training set \mathcal{T} composed of M aligned shapes, i.e., $\mathcal{T} = \{\mathbf{x}^{(m)}\}_{m \in \{1, 2, \dots, M\}}$ [11], we aim to learn a prior probability distribution on the different configuration of the shape model from the training set. Instead of learning the statistics on the pair of landmarks (e.g., [8]), which is not appropriate to get a scale-invariant modeling, we propose to learn the statistics of the measurements that are scale-invariant. To this end, let us consider a triplet of points, $c \in \mathcal{C} = \{(i, j, k) | i, j, k \in \mathcal{V} \text{ and } i \neq j, j \neq k, k \neq i\}$ and learn statistics on the relative lengths $(r_{icj_c}, r_{jck_c}, r_{kci_c})$ of the sides [12], which are defined as (take r_{icj_c} for example): $r_{icj_c} = l_{icj_c} / (l_{icj_c} + l_{jck_c} + l_{kci_c})$, where l_{icj_c} denotes the Euclidean distance between points i_c and j_c . An important advantage to use the relative lengths is that they can be computed much faster than the angle measurements which are also scale-invariant. For a triplet, it is sufficient to only consider the relative lengths of two sides since the third one is a linear combination of them (i.e., $r_{kci_c} = 1 - r_{icj_c} - r_{jck_c}$). Thus, without loss of generality, we use a Multivariate Gaussian Distribution (Fig. 1(c)) to model the distribution of $\mathbf{r}_c = (r_{icj_c}, r_{jck_c})^T$, i.e., $p_c(\mathbf{r}_c) = \mathcal{N}(\mathbf{r}_c | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where the mean $\boldsymbol{\mu}_c$ and the covariance matrix $\boldsymbol{\Sigma}_c$ can be learned from the training set by maximum likelihood:

$$\boldsymbol{\mu}_c = \frac{1}{M} \sum_{m=1}^M \mathbf{r}_c^{(m)} \quad (1)$$

$$\boldsymbol{\Sigma}_c = \frac{1}{M} \sum_{m=1}^M (\mathbf{r}_c^{(m)} - \boldsymbol{\mu}_c)(\mathbf{r}_c^{(m)} - \boldsymbol{\mu}_c)^T \quad (2)$$

Finally, we get the higher-order shape model $\mathcal{P} = (\mathcal{V}, \mathcal{C}, \{\mathcal{N}(\cdot | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)\}_{c \in \mathcal{C}})$, where \mathcal{V} and \mathcal{C} determine the topology of the model while $\{\mathcal{N}(\cdot | \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)\}_{c \in \mathcal{C}}$ characterizes the statistic geometry constraints between the triplets.

2.2 Landmark Candidate Detection

Given such a prior model, segmentation can be viewed as finding a mapping between the model points and a new volume. This can be expressed as a correspondence problem that requires detection for the model points in the image or

finding a set of correspondences for each point i ($i \in \mathcal{V}$). To this end, we first learn a classifier for each landmark, and then compute a score for each possible localization, and finally select the L positions that have the best scores. In the experiments, we adopt Randomized Forest [5] to perform the classification.

A randomized forest is composed of a set of T random decision trees. In the decision trees, an internal node consists of a random test on an input feature vector, the result (true or false) of which decides which (left or right) child node the feature vector goes to. And a leaf consists in a histogram $h = (h_1, \dots, h_W)$ (W is the number of classes), which is obtained during the training phase by counting the number of labeled feature vectors that arrive at this leaf. During the testing phase, an unlabeled feature vector is dropped in each decision tree τ and reaches the leaf l_τ , and the normalized histogram of l_τ provides a probability estimation for the feature vector belonging to each class w :

$$P(w|l_\tau) = \frac{h_w}{\sum_i h_i} \quad (3)$$

Finally, we average the probabilities of all the trees to obtain the probability over the forest:

$$P(w|(l_1, \dots, l_T)) = \frac{1}{T} \sum_{\tau} P(w|l_\tau) \quad (4)$$

We consider all the voxels in a volume as the possible localization of the landmarks. Each voxel is associated with a feature vector which is used as the input for the classifiers. Different features can be considered towards achieving a high quality detection. In order to well capture the local image structure information, we can use a series of 3d Gabor filters [13] with different scale, rotation parameters. We adopt the method proposed in [14] to sample these parameters such that scaling/rotation of the image becomes a translation of these parameters and then estimate the Fourier Transform Modulus (FTM) of the filter output to eliminate variations due to these translations (because the FTM is translation invariant). Due to the symmetry of FTM, it is enough to consider only half of the FTM domain by removing the redundant coefficients. In such a way, we get a feature vector for each voxel that is scale and rotation invariant.

Fig. 2 shows the detected candidate results for four landmarks at different locations on a testing muscle data.

2.3 Higher-Order MRF Segmentation Formulation

The shape model, together with the evidence from the image support, is formulated within a higher-order MRF towards image segmentation.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{C})$ denote a hypergraph with a node set \mathcal{V} and a clique set \mathcal{C} , $\mathbf{U}^{\mathcal{G}} = \{U_q(\cdot)\}_{q \in \mathcal{V}}$ the singleton potentials defined on the node set \mathcal{V} , and $\mathbf{H}^{\mathcal{G}} = \{H_c(\cdot)\}_{c \in \mathcal{C}}$ the clique potentials defined on the clique set \mathcal{C} . And then let $\text{MRF}^{\mathcal{G}}(\mathbf{U}^{\mathcal{G}}, \mathbf{H}^{\mathcal{G}})$ denote a higher-order MRF with topology \mathcal{G} as well as the potentials $\mathbf{U}^{\mathcal{G}}$ and $\mathbf{H}^{\mathcal{G}}$. In our problem, we associate a landmark to a node i

($i \in \mathcal{V}$), a triplet to a third-order clique c ($c \in \mathcal{C}$), with one-to-one mappings¹. The latent variable X_i corresponding to node i is a 3-dim vector denoting the position of node/point i . $\mathbf{x}^{\mathcal{G}}$ represents the configurations of all the nodes, i.e. $\mathbf{x}^{\mathcal{G}} = (x_i)_{i \in \mathcal{V}}$. The candidate set of each variable is denoted by \mathcal{X}_i ($i \in \mathcal{V}$). Thus the candidate set $\mathcal{X}^{\mathcal{G}}$ of all the variable of the MRF is defined as: $\mathcal{X}^{\mathcal{G}} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$, which composes all the possible configurations of the shape model. The candidate set \mathcal{X}_i for each node consists of the detected landmark candidates (section 2.2).

The segmentation problem is transformed into estimating the optimal positions of the landmarks, i.e., the optimal configuration $\mathbf{x}_{\text{opt}}^{\mathcal{G}}$ of $\text{MRF}^{\mathcal{G}}$, which is formulated as a minimization of the MRF's energy $E(\mathbf{x}^{\mathcal{G}})$:

$$\mathbf{x}_{\text{opt}}^{\mathcal{G}} = \arg \min_{\mathbf{x}^{\mathcal{G}} \in \mathcal{X}^{\mathcal{G}}} E(\mathbf{x}^{\mathcal{G}}) \quad (5)$$

where the energy of $\text{MRF}^{\mathcal{G}}$ is defined as:

$$E(\mathbf{x}^{\mathcal{G}}) = \sum_{q \in \mathcal{V}} U_q(x_q) + \sum_{c \in \mathcal{C}} H_c(\mathbf{x}_c) \quad (6)$$

where \mathbf{x}_c denotes the configuration $(x_q)_{q \in c}$ of clique c . The singleton potentials $U^{\mathcal{G}}$ and third-order clique potentials $\mathbf{H}^{\mathcal{G}}$ are presented as follows:

The singleton potential $U_q(x_q)$ ($q \in \mathcal{V}$) is the negative log-likelihood which imposes penalty for the landmark q being located at position x_q in image \mathbf{I} :

$$U_q(x_q) = -\log p(\mathbf{I}|x_q) \quad (7)$$

where we define $p(\mathbf{I}|x_q)$ using the classifier's output probability value for landmark q being located at x_q (Eq. 4).

The higher-order clique potential $U_c(\mathbf{x}_c)$ ($c \in \mathcal{C}$) encodes the statistic geometry constraints between the triplet c of points and is defined as:

$$U_c(\mathbf{x}_c) = -\alpha \cdot \log p_c(\mathbf{r}_c(\mathbf{x}_c)) \quad (8)$$

where $\alpha > 0$ is a weight coefficient, $\mathbf{r}_c(\mathbf{x}_c)$ denotes the mapping from the position of the triplet c to the 2-dim relative lengths of the sides, and the prior distribution p_c has been defined in section 2.1.

2.4 MRF Decomposition and Optimization

In order to perform inference in the proposed higher-order MRF, we adopt the well-known dual-decomposition optimization framework [15,16]. The key idea of such a framework is: instead of minimizing directly the energy of the original problem, we maximize a lower bound on it by solving the dual to the linear programming (LP) relaxation [15].

To this end, we first decompose the original problem (corresponding to $\text{MRF}^{\mathcal{G}}$) into a set of sub-problems (corresponding to $\{\text{MRF}^{\mathcal{G}_s}\}_{s \in \mathcal{S}}$, where \mathcal{S} is the set

¹ Due to such one-to-one mappings, in this section, we reuse the notation \mathcal{V} and \mathcal{C} to denote the node set and the clique set, respectively.

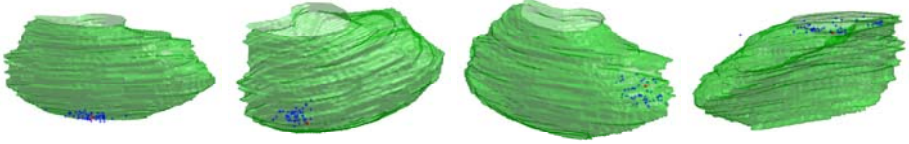


Fig. 2. Landmark detection results. The red hexagram represents the ground truth while the blue plus signs represent 50 detected candidates. The reference segmentation surface is provided to visually measure the distance between candidates and the ground truth.

of sub-problems), each of which is smaller and solvable. More concretely, we decompose the original hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{C})$ into a set of sub-hypergraphs $\{\mathcal{G}_s = (\mathcal{V}_s, \mathcal{C}_s)\}_{s \in \mathcal{S}}$ such that $\mathcal{V} = \cup_{s \in \mathcal{S}} \mathcal{V}_s$ and $\mathcal{C} = \cup_{s \in \mathcal{S}} \mathcal{C}_s$. In order to form $\{\text{MRF}^{\mathcal{G}_s}\}_{s \in \mathcal{S}}$ corresponding to the sub-problems, the original MRF potentials are decomposed into the sub-hypergraphs such that $\mathbf{U} = \sum_{s \in \mathcal{S}} \mathbf{U}^{\mathcal{G}_s}$ and $\mathbf{H} = \sum_{s \in \mathcal{S}} \mathbf{H}^{\mathcal{G}_s}$, which can be achieved simply by setting $U_q^{\mathcal{G}_s} = \frac{U_q^{\mathcal{G}}}{|\{s|q \in \mathcal{V}_s\}|}$ and $H_c^{\mathcal{G}_s} = \frac{H_c^{\mathcal{G}}}{|\{s|c \in \mathcal{C}_s\}|}$. The dual-decomposition [15] states that the sum of the minimum energies of the sub-problems provides a lower bound to the minimum energy of the original MRF. Furthermore, the problem of maximizing such a lower bound over its feasible set is then convex. Like [16], we adopt a projected subgradient method to perform this maximization so as to combine the solution of the sub-problems to get the solution of the original problem.

The most challenging component is how to define the sub-problems to decompose the original problem [2]. For the purpose of solving the inference, we adopt *factor graph* [6, 7] to represent the MRFs. To this end, we introduce additional nodes for the factors corresponding to the singleton potentials $U_q(x_q)$ ($q \in \mathcal{V}$) and the third-order clique potentials $H_c(\mathbf{x}_c)$ ($c \in \mathcal{C}$). Considering both the complexity and the quality of the optimum, we propose to decompose the original factor graph into a series of factor trees (i.e., factor graphs without loop) such that a higher-order clique factor appears in one and only one factor tree. The inference in a factor tree can be done exactly and very efficiently using max-product belief propagation algorithm [7] with complexity $O(NLK)$, where N , L and K denote the number of nodes, the number of candidates for each node, and the maximum order ($K = 3$) of the factors, respectively.

3 Experimental Results

We used the data set that was previously used in [17] to validate the proposed method. This data set consists of 25 3D MRI subjects whose calf part was imaged. The voxel spacing is of $0.7812 \times 0.7812 \times 4$ mm and each volume consists of 90 slices of 4mm thickness acquired with a 1.5 T Siemens scanner. Standard

² We cannot use the decomposition scheme proposed in [16], since the higher-order clique potentials are not pattern-based.

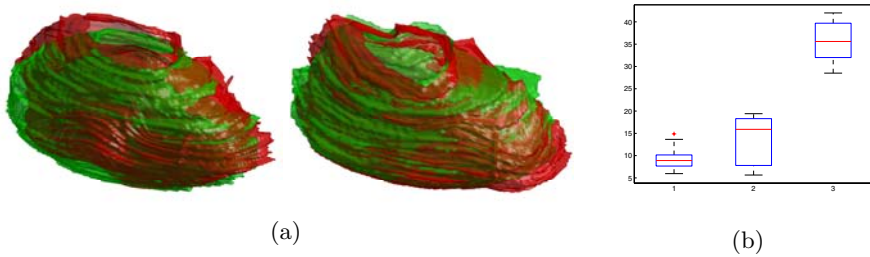


Fig. 3. Experimental Results. (a) Surface reconstruction results (green: reference segmentation, red: reconstruction result). (b) Boxplots of the average landmark error measure in voxel (1. our method, 2. method in [17], 3. standard ASM method.). On each box, the central mark in red is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points.

of reference was available, consisting of annotations provided by experts for the Medial Gastrocnemius (MG) muscle.

We performed a leave-one-out cross validation on the whole data set. For comparison purpose, we considered as alternative segmentation methods³ the ones presented in [17]. We present in Fig. 3(a) the surface reconstruction results using the estimated position of the landmarks and thin plate spline (TPS), while in Fig. 3(b) the average distance between the real landmark position and the one estimated from our algorithm, and the ones reported in [17] including the one obtained using standard active shape models. We reduce landmark localization error by an average of factor 2 in terms of voxel error compared to [17] that is considered to be the state of the art. The analysis of the results shows that the proposed prior and the inference using higher-order graphs globally perform well while the main limitation is introduced from the landmark candidate detection process. Since the method establishes correspondences between the model and the detected landmarks, in the absence of meaningful candidates the method fails to recover optimally the global shape. Regarding computational complexity, the method is linear with respect to the number of higher-order cliques and cubic with respect to the number of landmarks candidates per point.

4 Conclusion

In this paper, we have proposed a novel approach for 3D segmentation using pose invariant higher-order MRFs. Our method models the prior manifold through accumulation of local densities involving pose invariant combinations of points. Segmentation is expressed as a higher-order MRF optimization problem, where machine learning techniques and pose invariant features are considered to determine candidate positions for the model points. Promising results that clearly

³ Opposite to [17], we have considered a subset of 50 from the 895 model landmarks uniformly distributed in the model-space (Fig. 1).

outperform the prior art in very challenging data sets demonstrate the potentials of the method.

Extending the framework to deal with missing correspondences is the most promising direction to overcome the challenge of correctly estimating the position of all the landmarks of the model. Redundancy is a natural property inherited from the exhaustive construction of the higher-order model. The optimization of the graph connectivity towards reducing the computational complexity of the method is a straightforward direction as suggested in [8] through dimensionality reduction on the graph space. Last, but not least the case of spatio-temporal higher-order priors on anatomical structures with dynamic behavior is currently under investigation.

References

1. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. *Comput. Vis. Image Underst.* 61(1), 38–59 (1995), doi:10.1006/cviu.1995.1004
2. Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (icbm). *NeuroImage* 2(2, Part 1), 89–101 (1995)
3. Staib, L.H., Duncan, J.S.: Boundary finding with parametrically deformable models. *IEEE Trans. Pattern Anal. Mach. Intell.* 14(11), 1061–1075 (1992)
4. Leventon, M.E., Grimson, W.E.L., Faucher, O.D.: Statistical shape influence in geodesic active contours. In: *CVPR* (2000)
5. Breiman, L.: Random forests. *Machine Learning* 45, 5–32 (2001)
6. Frey, B.J.: *Graphical Models for Machine Learning and Digital Communication*. MIT Press, Cambridge (1998)
7. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Heidelberg (August 2006)
8. Besbes, A., Komodakis, N., Lings, G., Paragios, N.: Shape priors and discrete mrfs for knowledge-based segmentation. In: *CVPR* (2009)
9. Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal shape and intensity cost path segmentation. *IEEE TMI* 26(8), 1115–1129 (2007)
10. Donner, R., Lings, G., Bischof, H.: Sparse mrf appearance models for fast anatomical structure localisation. In: *British Machine Vision Conference* (2007)
11. Luo, B., Hancock, E.R.: Iterative procrustes alignment with the em algorithm. *Image Vision Comput.* 20(5-6), 377–396 (2002)
12. Lekadir, K., Merrifield, R., Yang, G.: Outlier detection and handling for robust 3-d active shape models search. *IEEE TMI* 26, 212–222 (2007)
13. Bernardino, A., Santos-Victor, J.: Fast iir isotropic 2-d complex gabor filters with boundary initialization. 15, 3338–3348 (2006)
14. Kokkinos, I., Yuille, A.: Scale invariance without scale selection. In: *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2008)
15. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
16. Komodakis, N., Paragios, N.: Beyond pairwise energies: Efficient optimization for higher-order mrfs. In: *CVPR* (2009)
17. Essafi, S., Lings, G., Deux, J.F., Rahmouni, A., Bassez, G., Paragios, N.: Wavelet-driven knowledge-based mri calf muscle segmentation. In: *Proceedings of the Sixth IEEE international conference on Symposium on Biomedical Imaging* (2009)

Markov Random Field driven Region-Based Active Contour Model (MaRACel): Application to Medical Image Segmentation

Jun Xu*, James P. Monaco*, and Anant Madabhushi

Department of Biomedical Engineering, Rutgers University, USA
{junxu, jpmonaco, anantm}@rci.rutgers.edu

Abstract. In this paper we present a Markov random field (MRF) driven region-based active contour model (MaRACel) for medical image segmentation. State-of-the-art region-based active contour (RAC) models assume that every spatial location in the image is statistically independent of the others, thereby ignoring valuable contextual information. To address this shortcoming we incorporate a MRF prior into the AC model, further generalizing Chan & Vese’s (CV) and Rousson and Deriche’s (RD) AC models. This incorporation requires a Markov prior that is consistent with the continuous variational framework characteristic of active contours; consequently, we introduce a continuous analogue to the discrete Potts model. To demonstrate the effectiveness of MaRACel, we compare its performance to those of the CV and RD AC models in the following scenarios: (1) the qualitative segmentation of a cancerous lesion in a breast DCE-MR image and (2) the qualitative and quantitative segmentations of prostatic acini (glands) in 200 histopathology images. Across the 200 prostate needle core biopsy histology images, MaRACel yielded an average sensitivity, specificity, and positive predictive value of 71%, 95%, 74% with respect to the segmented gland boundaries; the CV and RD models have corresponding values of 19%, 81%, 20% and 53%, 88%, 56%, respectively.

Keywords: Segmentation, medical imaging, prostate cancer, breast cancer, MRI, digital pathology.

1 Introduction

An active contour (AC) model performs segmentation by evolving a curve towards the minimum of an energy functional. Based on the type of image information driving the model, an AC may be categorized as either boundary-based [1] or region-based [2]. Region-based AC (RAC) models (the type of model considered in this work) employ statistical information derived from the segmented regions to drive the AC. Additionally, RACs employ level sets [3] instead of parameterized contours [2, 4]. The attraction of level sets lies in their ability to

* J. Xu and J. Monaco are joint first authors.

implicitly handle topological changes such as boundary splitting and merging. An early RAC model, called the region competition model [5], used generalized Bayes and minimum description length criteria to formulate the curve evolution function. Inspired by the Mumford-Shah functional, Chan & Vese’s (CV) [2] and Rousson and Deriche’s (RD) [4] RAC models used statistics derived from the intensities of each region to drive their ACs.

The shortcoming of most RAC models lies in their assumption that each spatial location in the image is statistically independent of the others. Unfortunately, this assumption does not hold for most (or perhaps any) real images. To address this assumption, we propose the integration of Markov random fields (MRFs) into the RACs. MRFs [6] provide an effective and tractable (via the Markov property) means for modeling spatial information within a Bayesian framework [6].

The contribution of this paper is an MRF-driven RAC model (MaRACel). Specifically, MaRACel incorporates a Markov prior into the RD model [4]. Intuitively, the effect of integrating this prior is similar to performing a probabilistic morphological close operation: edges randomly occurring (due to noise) in the regions are much less likely to result in spurious boundaries. For the Markov prior we introduce a continuous analogue of the discrete Potts model. This continuous extension is needed to ensure compatibility with variational calculus, the foundation of AC models. To demonstrate the superiority of MaRACel over the popular CV and RD models we qualitatively and quantitatively compare their performances in the task of segmenting prostatic acini (glands) in 200 digitized images of core needle biopsies.

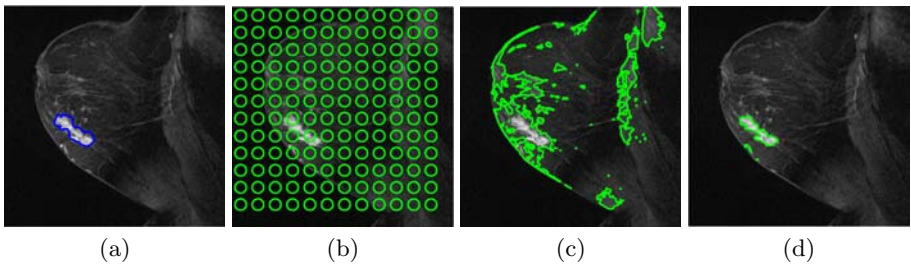


Fig. 1. Qualitative segmentation results for tumor boundary segmentation for breast DCE-MR image. (a) Single slice of DCE-MR image with manual segmentation of tumor region (in blue); segmentation results (in green) for (b) CV model, (c) RD model and (d) MaRACel.

To better highlight the need for MaRACel, we provide an illustrative example (Figure 1) of the automated detection of a lesion on a dynamic contrast enhanced (DCE) breast MR image [7]. The segmentation results for the RD and CV models, shown in columns (b) and (c), reveal the failure of these models to eliminate spurious noisy regions. (Note that CV model performs so poorly that the model does not evolve beyond its initial contours.) This is a result of the

RD and CV models' inability to account for the fact that spatially proximate pixels are more likely to belong to the same regions. As evinced in column (d), this issue can be rectified by incorporating a Markov prior — such as the Potts model — into the AC, thereby encouraging neighboring pixels to share the same class (i.e. be included in the same segmented region).

The remainder of the paper proceeds as follows: Section 2 establishes the statistical model for describing random images and introduces the MRF prior. Section 3 discusses how the curve evolution function can be derived from this statistical model. In Section 4 we apply MaRACel to the segmentation of prostatic glands, qualitatively and quantitatively evaluating its segmentation performance. Finally, we offer concluding remarks in Section 5.

2 Statistical Framework for the MaRACel Model

Let $\mathcal{C} = (C, \mathbf{f})$ define a color image, where $C \subset \mathbb{R}^2$ establishes the image region and $\mathbf{f}: C \rightarrow \mathbb{R}^3$ reflects the intensities of the three color channels. Let $x: C \rightarrow \{A_1, A_2\}$ be a function mapping each point in C to one of two discrete classes $\{A_1, A_2\}$. Thus x segments the image into regions C_1 and C_2 (e.g. the foreground and the background), where $C_1 \cup C_2 = C$ and $c \in C_i$ implies $x_c = A_i$.

The goal is to determine the function x that best segments the region C given the observed colors \mathbf{f} . Since we use a Bayesian framework, we consider the functions \mathbf{f} and x as observations of the random processes \mathbf{F} and X . Furthermore, the optimal segmentation (with respect to probability of error) is given by the maximum *a posteriori* (MAP) estimate [8]:

$$\max_x P(X = x | \mathbf{F} = \mathbf{f}) \propto \max_x P(\mathbf{F} = \mathbf{f} | X = x) P(X = x) \quad (1)$$

Thus, maximizing the *a posteriori* probability $P(X = x | \mathbf{F} = \mathbf{f})$ is equivalent to maximizing the product of the conditional $P(\mathbf{F} = \mathbf{f} | X = x)$ and prior $P(X = x)$ probabilities.

Note that $P(X = x)$, which signifies the probability of the event $\{X_c = x_c, \forall c \in C\}$, should not be confused with $P(X_c = x_c)$, which indicates the probability of the event $\{X_c = x_c\}$ at c . Additionally, when it does not cause ambiguity we will henceforth omit the random variables from the probability functions, e.g. $P(x) \equiv P(X = x)$.

2.1 Estimating the Prior Probability

To simplify the prior probability we invoke the Markov property: $P(x_c | x_{-c}) = P(x_c | x_{\eta_c})$, where x_{-c} abbreviates $\{x(r) : r \in C, r \neq c\}$, x_{η_c} abbreviates $\{x(r) : r \in \eta_c, r \neq c\}$, and the neighborhood η_c of c is any bounded region such that $s \in \eta_c \leftrightarrow c \in \eta_s$, $c \notin \eta_c$, and $c, s \in C$. Using these simplified conditional probabilities, the prior probability of the event $\{X_c = x_c, \forall c \in C\}$ can be approximated by the normalized pseudo-likelihood [9] as follows:

$$P(X = x) \approx \frac{1}{Z} \prod_{i=1}^2 \prod_{c \in C_i} P(X_c = A_i | x_{\eta_c})^{d_c} \quad (2)$$

where Z ensures summation to one and dc represents the bin volume which guarantees the correct continuum limit [10]. We choose to model each of the conditional distributions in (2) using a continuous analogue of the Potts model [11]:

$$P(X_c = A_i | x_{\eta_c}) = \frac{1}{Z_c} \exp \left\{ -\lambda \int_{\eta_c \cap C_i} dc \right\} = \frac{1}{Z_c} \exp \{ -\lambda A(\eta_c, A_i) \}, \quad (3)$$

where Z_c is a normalizing constant, $A(\eta_c, C_i)$ signifies the area of the region defined by $\eta_c \cap C_i$, and $\lambda \in \mathbb{R}$ is an appropriately selected constant. Note that the greater the value of λ , the greater the tendency for neighboring points in the MAP estimate to belong to the same class.

2.2 Estimating the Conditional Probability

We next estimate the conditional probability $P(\mathbf{f}|x)$. Since each vector \mathbf{F}_c given $\{X_c = x_c\}$ is assumed conditionally independent of all other \mathbf{F}_c , $P(\mathbf{f}|x)$ can be expressed as follows:

$$P(\mathbf{f}|x) = \prod_{i=1}^2 \prod_{c \in C_i} P(\mathbf{f}_c | X_c = A_i)^{dc}, \quad (4)$$

where dc again assures the correct continuum limit [10].

Similar to [5, 4], we assume each conditional probability $P(\mathbf{f}_c | X_c = A_i)$ is distributed normally with the following mean μ_i and covariance Σ_i :

$$\mu_i = \frac{1}{|C_i|} \int_{C_i} \mathbf{f}_c dc \quad \text{and} \quad \Sigma_i = \frac{1}{|C_i|} \int_{C_i} (\mathbf{f}_c - \mu_i)(\mathbf{f}_c - \mu_i)^T dc. \quad (5)$$

where $|C_i| = \int_{C_i} dc$ is the area of region C_i .

3 Variational Framework for the MaRACel Model

3.1 Energy Functional of MaRACel Model

We now wish to perform MAP estimation within the active contour framework. We begin by using Equations (1), (2), and (4) to express the log of the *a posteriori* probability — whose maximum will correspond to that of the *a posteriori* probability — in integral form:

$$\begin{aligned} \log P(x|\mathbf{f}) &= \log P(\mathbf{f}|x) + \log P(x) \\ &= \sum_{i=1}^2 \int_{C_i} \log P(\mathbf{f}_c | X_c = A_i) + \log P(X_c \approx A_i | x_{\eta_c}) dc - \log(Z). \end{aligned} \quad (6)$$

This formulation can be rewritten as an energy functional [2]:

$$\begin{aligned}
 E(\phi) = & - \int_C \{ \alpha [H(\phi) \log P(\mathbf{f}_c | X_c = A_1) + (1 - H(\phi)) \log P(\mathbf{f}_c | X_c = A_2)] \\
 & + \beta [H(\phi) \log P(X_c \approx A_1 | x_{\eta_c}) + (1 - H(\phi)) \log P(X_c \approx A_2 | x_{\eta_c})] \\
 & - \gamma |\nabla H(\phi)| \} dc,
 \end{aligned} \quad (7)$$

where $H(\cdot)$ is the Heaviside function and $\phi: C \rightarrow \mathbb{R}$ is a level set function such that $\phi(c) > 0$ if $c \in C_1$ and $\phi(c) < 0$ if $c \in C_2$. Thus, the ϕ that minimizes $E(\phi)$ establishes the boundary that corresponds to the MAP estimate (under the pseudo-likelihood approximation in (2)). Note that the final term $\gamma |\nabla H(\phi)|$ in (7) results from incorporating the boundary length for the purpose of regularization; this term was not derived from (6), and thus the minimizer of (7) will not precisely correspond to the MAP estimate.

3.2 Curve Evolution Functional of MaRACel Model

Applying the Euler-Lagrange equations to (7) and then substituting in (3) and (5) yields the curve evolution function for MaRACel

$$\begin{aligned}
 \frac{\partial \phi}{\partial t} = & \delta(\phi) \left\{ \alpha [(\mathbf{f}_c - \mu_2)^T \Sigma_2^{-1} (\mathbf{f}_c - \mu_2) - (\mathbf{f}_c - \mu_1)^T \Sigma_1^{-1} (\mathbf{f}_c - \mu_1)] \right. \\
 & \left. + \log \left[\frac{|\Sigma_2|}{|\Sigma_1|} \right] + \beta (A(\eta_c, C_1) - A(\eta_c, C_2)) + \gamma \mathbf{div} \left(\frac{\nabla \phi}{\|\nabla \phi\|} \right) \right\},
 \end{aligned} \quad (8)$$

where $\beta = \alpha \cdot \lambda$. Note that the relationship between MaRACel and the RD and CV models is as follows: Setting $\beta = 0$ in the MaRACel evolution function (8) reduces it to the RD model. If we further stipulate that Σ_1 and Σ_2 are both identity matrices, the RD model devolves into the CV model.

4 Experimental Results and Discussion

To illustrate the effectiveness of our proposed model, we now qualitatively and quantitatively compare the gland segmentation performance of MaRACel to those of the CV and RD models. Note that since cancerous and benign glands have unique morphologies, the ability to accurately delineate their boundaries is essential in both automated cancer detection [11] and Gleason grading (i.e. the stratification of prostate cancer by aggressiveness).

4.1 Model Parameters and Data Description

The initial contours for the CV, RD, and MaRACel models are given by an array of circles dispersed regularly across the image. (These circles were unintentionally depicted by the poor segmentation results in Figure 1(b).) Empirically, we found

that we achieve more robust performance if the MRF weighting term β in (8) begins at zero and then increases with each iteration (i.e. as the segmentations become more accurate) before leveling off. Thus β is determined as follows: $\beta = \frac{2\beta_0}{\pi} \arctan[0.2(t-1)]$, where t is the time in iterations. The values for α , β_0 and γ in (8) are 5, 0.1 and 15, respectively. The neighborhood η_c is a 21×21 window centered at pixel c (and excluding c). Thus the area $A(\eta_c, \Omega_i)$ in (8) is simply the number of pixels with class A_i that lie within the window η_c (excluding the center pixel c).

The dataset includes 200 images obtained from Hematoxylin & Eosin (H&E) stained prostate biopsy samples digitized at 20x optical magnification using an Aperio whole-slide digital scanner. Each image includes one or more prostatic glands. The glands in each of the 200 images were manually segmented (to provide ground truth for quantitative evaluation).

4.2 Qualitative Results

The goal is to accurately delineate the boundaries of the glandular lumens for each image in column (a) of Figure 2. The segmentation for the RD and CV models, shown in columns (c) and (d) in Figure 2, reveal the inability of these models to eliminate the small, spurious regions that appear due to noise. That is, the CV and RD models find background regions within the glands. The segmentation results for MaRACel are given in column (d). MaRACel removes these false regions, yielding a single segmented region for each glandular structure.

4.3 Quantitative Results

For each image the set of pixels lying within the manual delineations of the glands is denoted as $\mathcal{A}(G)$. The set of pixels lying within any boundary resulting from the specified AC model is denoted as $\mathcal{A}(S)$. The sensitivity (SN), specificity (SP), and positive predictive value (PPV) are then defined as follows: $SN = \frac{|\mathcal{A}(S) \cap \mathcal{A}(G)|}{|\mathcal{A}(G)|}$, $SP = \frac{|C| - |\mathcal{A}(S) \cup \mathcal{A}(G)|}{|C| - |\mathcal{A}(G)|}$, and $PPV = \frac{|\mathcal{A}(S) \cap \mathcal{A}(G)|}{|\mathcal{A}(S)|}$, where $|\cdot|$ denotes the cardinality of the set. We compute SN, SP and PPV for each image, and then determine the average and standard deviation across the 200 images. These statistics are reported in Table 1 for the CV, RD and MaRACel models.

Table 1. Quantitative evaluation of segmentation results for the CV, RD and MaRACel models. The average and standard deviation of the SN, SP and PPV over the 200 histopathology images are provided.

	Sensitivity (SN)	Specificity (SP)	Positive Predictive Value (PPV)
CV	0.19 \pm 0.06	0.81 \pm 0.01	0.20 \pm 0.06
RD	0.53 \pm 0.10	0.88 \pm 0.11	0.56 \pm 0.11
MaRACel	0.71 \pm 0.10	0.95 \pm 0.01	0.74 \pm 0.10

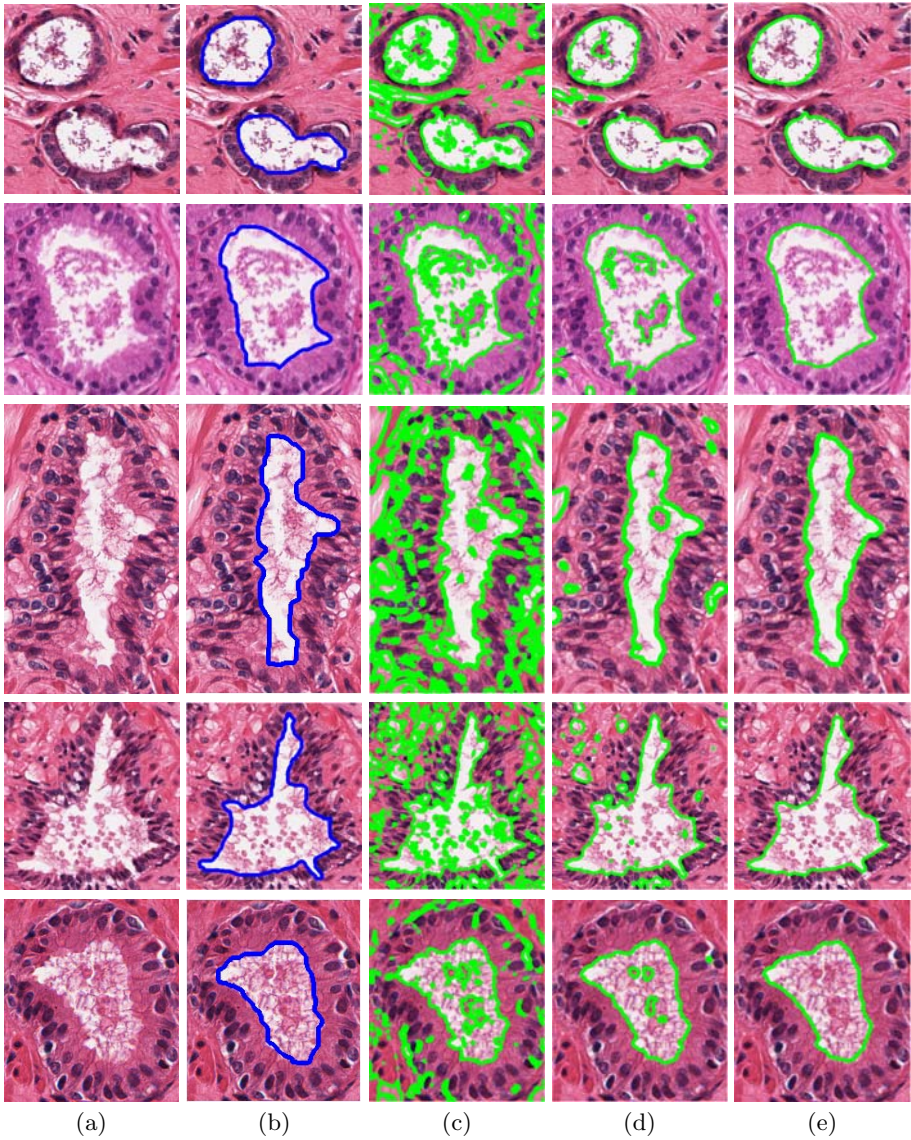


Fig. 2. Qualitative segmentation results for prostatic glands in digitized biopsy samples. (a) original images; (b) manual segmentations of the glandular boundaries (in blue); segmentation results (in green) for (c) CV model, (d) RD model and (e) MaRACel.

5 Concluding Remarks

In this paper we presented MaRACel, a generalization of Chan & Vese's and Rousson and Deriche's active contour models, that incorporates Markov random

fields into the AC. Currently, the region-based AC models only extract global statistical information and ignore valuable local contextual information. Incorporating an MRF prior into the AC provides a means for modeling this contextual information which is so essential in rejecting spurious edges and other forms of noise. To validate our model we quantitatively and qualitatively compared MaRACel with Chan & Vese's and Rousson and Deriche's AC models in the task of identifying glandular boundaries in prostate histopathology images. MaRACel significantly outperformed the other models in terms of average sensitivity, specificity, and positive predictive value.

Acknowledgments

This work is supported by NIH Grant Nos. R01-CA-136535-01, R21-CA-127186-01, and R03-CA-128081-01, the NJ Commission on Cancer Research, the Coulter Foundation, the Cancer Institute of NJ, and Bioimagine Inc.

References

1. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *IJCV* 22(1), 61–79 (1997)
2. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE TIP* 10(2), 266–277 (2001)
3. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences of the United States of America* 93(4), 1591–1595 (1996)
4. Rousson, M., Deriche, R.: A variational framework for active and adaptative segmentation of vector valued images, pp. 56–61 (2002)
5. Zhu, S., Yuille, A.: Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE TPAMI* 18(9), 884–900 (1996)
6. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE TPAMI* 6(6), 721–741 (1984)
7. Agner, S., Soman, S., Libfeld, E., McDonald, M., Thomas, K., Englander, S., Rosen, M., Chin, D., Noshier, J., Madabhushi, A.: Textural kinetics: A novel dynamic contrast-enhanced (dce)-mri feature for breast lesion classification. *Journal of Digital Imaging*, May 28 (2010) [Epub. ahead of print]
8. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. *IJCV* 46(3), 223–247 (2002)
9. Besag, J.: Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society, Series D (The Statistician)* 24(3), 179–195 (1975)
10. Cremers, D., Mikael, R., Rachid, D.: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. *IJCV* 72, 195–215 (2007)
11. Monaco, J.P., Tomaszewski, J.E., Feldman, M.D., Hagemann, I., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A.: High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *Medical Image Analysis* 14(4), 617 (2010)

Predicting Target Vessel Location for Improved Planning of Robot-Assisted CABG Procedures

Daniel S. Cho^{1,2}, Cristian A. Linte^{1,2}, Elvis Chen¹, Chris Wedlake¹, John Moore¹, John Barron⁴, Rajni Patel³, and Terry M. Peters^{1,2,3}

¹ Imaging Research Laboratories, Robarts Research Institute

² Biomedical Engineering Graduate Program, University of Western Ontario

³ Canadian Surgical Technologies and Advanced Robotics

⁴ Department of Computer Science, University of Western Ontario,
London, Ontario, Canada

Abstract. Prior to performing a robot-assisted coronary artery bypass grafting procedure, a pre-operative computed tomography scan is used to assess patient candidacy and to identify the location of the target vessel. The surgeon then determines the optimal port locations to ensure proper reach to the target with the robotic instruments, while assuming that the heart does not undergo any significant changes between the pre- and intra-operative stages. However, the peri-operative workflow itself leads to changes in heart position and consequently the intra-operative target vessel location. As such, the pre-operative plan must be adequately updated to adjust the target vessel location to better suit the intra-operative condition. Here we propose a technique to predict the position of the peri-operative target vessel location with ~ 3.5 mm RMS accuracy. We believe this technique will potentially reduce the rate of conversion of robot-assisted procedures to traditional open-chest surgery due to poor planning.

1 Introduction

The quest for alternate approaches to conventional cardiac therapy has increased the number of minimally invasive procedures being performed, leading to less trauma, shorter hospital stays, faster return to normal activities, and improved cosmesis [1,2,3,4]. Robot-assisted (RA) surgery represents a paradigm shift in the delivery of health care for both the patient and the surgeon [5] and it has been adopted as standard of care at many institutions worldwide [5]; one of the popular cardiac interventions performed under robot-assistance is the coronary artery bypass grafting (CABG) procedure.

In current clinical practice, a pre-operative computed tomography (CT) scan of the patient is used to assess his/her candidacy for undergoing a RA-CABG procedure. Based on the pre-operative scan, the surgeon identifies the location of the surgical target — the left anterior descending (LAD) coronary artery, examines whether there is sufficient workspace inside the chest wall for the robot arms, and ultimately estimates the optimal locations of the port incisions to

ensure proper reach of the surgical targets with the robotic instruments. However, it is not unusual that after setting up the patient for the robot-assisted procedure, the surgeons encounter difficulties due to the inability in reaching the target, robot arm collisions or reduced dexterity [6]. In fact, 20-30% of the RA-CABG interventions require conversion to traditional open-chest surgery [7], mainly due to the migration of the heart during the peri-operative workflow, not accounted for in the pre-operative plan.

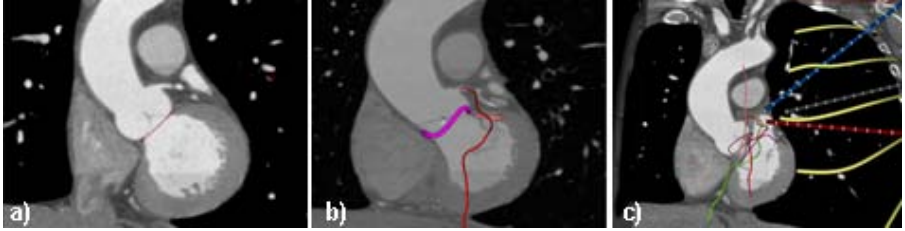


Fig. 1. Pre-operative planning stage showing patient's cardiac CT scan (a), the coronary vessel displayed relative to the valve annuli (b) and the port placement to ensure proper reach of the target vessel with the robotic instruments (c) *Note: the yellow lines represent intercostal spaces.*

It is believed that the workflow stages - lung collapse and CO₂ chest insufflation - will induce an overall shift of the heart inside the thoracic cavity, which may invalidate the surgical plan based solely on the pre-operative data. Therefore, it is critical to estimate the global heart displacement during the typical peri-operative workflow and improve the pre-operative plan by estimating the intra-operative location of the target vessel. However, the global migration of the heart has never been measured, nor has the change in location of the LAD between the pre- and intra-operative stages. The contribution of this paper is twofold. First, we provide an estimate of the heart migration measured from clinical patient data by registering the pre-operative cardiac CT (containing the LAD location identified from the CT dataset - **Fig. 1**) to images acquired peri-operatively using tracked US. Secondly, to overcome the clinical limitation arising due to the invisibility of the LAD in the US images, we conducted an *in vitro* study where we simulated the clinically observed heart shift and demonstrated the feasibility of our technique towards predicting the LAD location.

While no known accuracy constraints have been reported for this specific application, our collaborating cardiac surgeons have recommended that a target prediction accuracy on the order of one intercostal space (~ 10 - 15 mm, depending on patient size) is desired. From a clinical feasibility perspective, this constraint is valid: as long as the intra-operative LAD location is correctly predicted to within one intercostal space from its actual location, it can be reached by positioning the trocar on either side of the adjacent rib.

2 Materials and Methods

2.1 Clinical Procedure Workflow

In a typical RA-CABG procedure, the patient is first imaged pre-operatively (Stage₀) in the same position as during the intervention. Peri-operatively, following intubation and anesthesia delivery (Stage₁), the left lung is collapsed (Stage₂), and the chest is insufflated (Stage₃) to provide sufficient work space.

Image Acquisition. We employed real-time 2D trans-esophageal echocardiography (TEE) to monitor the heart during the interventional workflow. The images were acquired using a spatially tracked TEE probe modified by embedding a 6 DOF Aurora magnetic sensor coil (Northern Digital Inc., Waterloo, Canada) inside the encasing of the transducer [8]. Images were collected from three different views in order to capture all of the necessary cardiac features. For our work, mid-esophageal-4-chamber view images were captured at 20° increments from 0° to 180° for the mitral valve annulus (MVA) and the left ventricular apex (LVAp). Five long-axis view images with 10° increments and one short-axis view of the aorta at 30° were also acquired to visualize the aortic valve annulus (AVA) and the left coronary ostium (LCO). US image data were acquired by an expert anesthetist and repeated three times at each workflow stage.

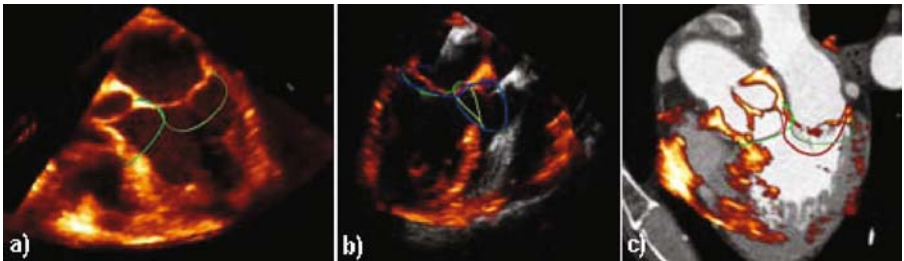


Fig. 2. a) Patient heart instance at Stage₁ acquired using tracked US and showing the valve annuli; b) Stage₁ (orange) and Stage₂ (gray) instances of the heart showing relative heart displacement and corresponding segmented valvular structures; c) Initial peri-operative US instance (Stage₁) registered to the pre-operative dataset and displayed within the CT coordinate space.

Estimating Global Heart Displacement. The peri-operatively acquired instances were then transferred into the CT coordinate system by aligning homologous features corresponding to the first peri-operative (Stage₁ US) and the pre-operative (Stage₀ CT) datasets (**Fig. 2**). These two stages are physiologically equivalent given the same patient position and dual-lung ventilation, and hence minimal anatomical variations are expected. As a result, the peri-operative displacements can be estimated with respect to the principal body axes.

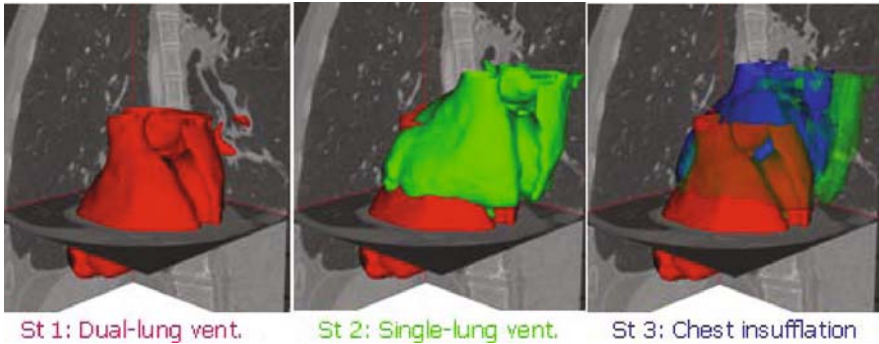


Fig. 3. Visual representation showing an automatically segmented epicardial model of a patient’s heart animated using the sequential peri-operative transforms based on the valvular structures. Note: Stage₁ is shown in red, Stage₂ in green and Stage₃ in blue.

We have shown the migration patterns in four patients undergoing RA-CABG procedure [9]. Our clinical data have suggested that the heart undergoes considerable displacement during the workflow, which should not be ignored during the planning process. As an example, we show the change in position of the epicardial surface of one patient’s heart segmented from the CT dataset and animated using the sequential peri-operative transforms (**Fig. 3**). Moreover, in spite of the observed displacements, the morphology of the identified features remains relatively consistent throughout the workflow [9], suggesting that no significant non-rigid deformations are occurring. Based on these clinical observations, we next conducted an *in vitro* validation study to assess the accuracy with which a newly-developed technique involving a rigid-body feature-based registration can predict the peri-operative location of the LAD vessel.

2.2 Predicting the LAD Location: Feature-Based Registration

Since the LAD can only be clearly seen in the pre-operative CT image and *not* in the peri-operative US images, its peri-/intra-operative location has to be deduced based on the rest of the data available peri-operatively. Therefore, we chose to predict its location via a registration algorithm that involves four features - LCO, MVA, AVA and LVAp - easily identifiable in both modalities and sufficiently close to the target vessel to provide improved accuracy in the region of interest. The LAD begins at the left coronary ostium and usually runs toward the apex, while the mitral and aortic valves are located on either side of the starting point of the LAD. A rigid-body registration driven by the above features was then applied to map the pre-operative dataset to the peri-operative datasets, therefore predicting the LAD location at each subsequent stage. The Euclidean distance between each set of homologous features was initially computed, followed by the optimization of the cost function shown in **Eq. 1** performed using the downhill simplex method [10].

$$cost = \alpha dist_{LCO} + \beta dist_{LVAp} + \gamma (dist_{MVA} + dist_{AVA}) \quad (1)$$

Since distances between the LCO and the LVAp were selected as the main constraints for the registration, the optimal values for the weighting factor parameters were empirically determined as follows: $\alpha = 1.0$, $\beta = 1.0$, and $\gamma = 0.6$, yielding an optimal alignment.

2.3 *In vitro* Experimental Validation

Experimental Apparatus: Considering that the LAD cannot be identified peri-operatively using US imaging, we conducted an *in vitro* validation study to assess the accuracy with which our technique can predict the LAD location. The experimental apparatus was set up in a configuration similar to that typically found in the OR: we simulated the migration patterns of the heart induced during RA-CABG procedures by altering the position and orientation of a heart phantom (The Chamberlain Group, Great Barrington, USA).

Sixteen CT-visible fiducials were attached to the surface of the phantom: ten were used to assist with the CT-to-phantom registration and the remaining six were used to “define” the path of the LAD vessel. The position of the heart phantom was tracked throughout the study using a 6-DOF NDI Aurora magnetic sensor rigidly attached onto the phantom. Two different modalities (CT and US) were employed for image acquisition: a pre-operative CT scan was acquired and a virtual surface model was constructed using automatic segmentation tools; peri-operative images at each workflow stage were acquired using a magnetically-tracked TEE probe similar to the one used in the OR (**Fig. 4**). The LAD vessel was initially identified from the CT image and its peri-operative location was predicted based on its pre-operative location using the proposed registration.

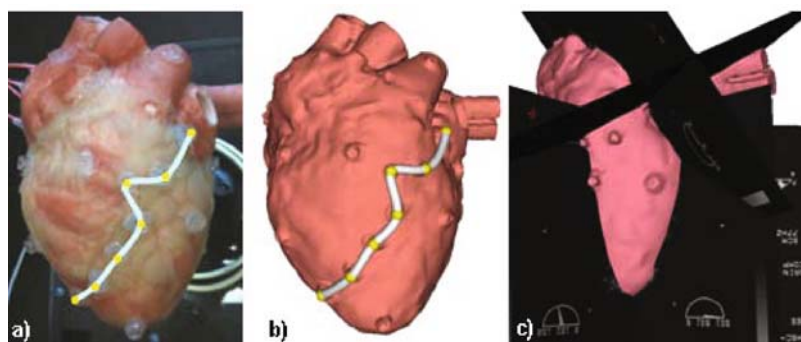


Fig. 4. a) Image and (b) virtual surface model of the heart phantom showing the LAD path; c) Peri-operative US image acquisition protocol showing imaging of the apex and coronary ostia using incrementally tracked 2D US images.

Intra-operative Image Acquisition: Since TEE is the standard of care for monitoring during cardiac procedures, in this study we collected the required tracked US images following the clinical workflow. The position of the heart

was altered twice, to mimic the actual intervention (i.e. lung collapse and chest insufflation). In each position, the image acquisition was repeated three times and the entire protocol was also repeated three times to minimize human errors. The four features were then extracted from the images using a custom-developed segmentation tool. All features were defined in the same 3D coordinate space; the mitral and aortic valves were represented as “rings”, while the ostium and the apex were represented as points.

Assessing Intra-operative Target Vessel Location: The six fiducials positioned along the LAD path were used to assess the target registration error (TRE) between the predicted LAD fiducial locations and their gold-standard locations, at each stage in the workflow. The gold-standard LAD fiducial locations were determined by recording the LAD fiducial locations at each stage using a magnetically tracked pointer and confirmed using the point-based registration transform corresponding to each peri-operative stage. The predicted LAD location was identified by mapping the pre-operative LAD fiducials using the feature-based registration transform described in section [2.2](#).

3 Evaluation and Results

We simulated three different RA-CABG-related workflows by altering the position and orientation of the heart phantom at three different stages. For each of the nine poses, we acquired three sets of tracked US images, defined the features of interest, and used the proposed registration algorithm to predict the location of the LAD vessel. [Table 1](#) summarizes the TRE between the actual locations of the LAD target fiducials (the ground truth) and their predicted locations from the registration.

Table 1. LAD TRE: Mean \pm SD and RMS (mm)

LAD Point	Stage ₀ to Stage ₁		Stage ₀ to Stage ₂	
	Mean \pm SD	RMS	Mean \pm SD	RMS
1	3.1 \pm 0.9	3.2	2.9 \pm 1.7	3.4
2	2.9 \pm 1.3	3.2	3.4 \pm 1.6	3.8
3	3.4 \pm 1.4	3.7	3.3 \pm 1.6	3.7
4	3.9 \pm 1.9	4.3	3.8 \pm 1.3	4.0
5	4.4 \pm 2.3	5.0	3.9 \pm 1.7	4.3
6	4.9 \pm 2.6	5.5	4.5 \pm 2.1	5.0
Overall	3.7 \pm 1.9	4.2	3.6 \pm 1.7	4.0

Table 2. RMS Feature Localization Error (mm)

Feature	Stage ₀	Stage ₁	Stage ₂
LVAp	2.3	4.7	4.0
AVA	1.4	1.3	1.2
MVA	0.6	1.9	1.5
CO	1.1	4.0	2.1

For a visual interpretation of the LAD TRE, [Fig. 5](#) shows the virtual model of the heart phantom along with the gold-standard and predicted LAD paths at both Stage₁ and Stage₂ in the peri-operative workflow, showing clinically-adequate alignment, well under the 10-15 mm intercostal space constraint.

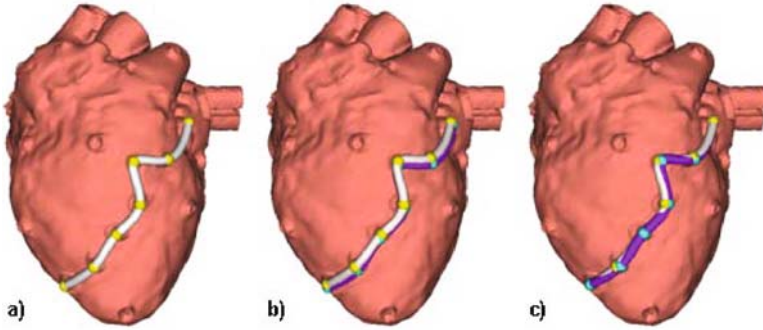


Fig. 5. a) Pre-operative heart phantom model at Stage₀ showing the LAD vessel; Visual display of the LAD TRE at Stage₁ (b) and Stage₂ (c), showing the gold-standard LAD (seashell white) obtained using the point-based registration transform and the predicted LAD (purple) determined using the proposed feature-based registration transform.

4 Discussion

This work constitutes the first steps towards optimizing pre-operative planning for RA-CABG procedures. Motivated by a recent clinical study that revealed substantial migration of the heart during the peri-operative procedure workflow, our goal is to predict the intra-operative location of the target vessel, to provide the surgeon with an optimized surgical plan that better reflects the intra-operative stage.

As a bridge to the *in vivo* validation, considering the limitations arising due to poor visualization and identification of the LAD coronary vessel in clinical US images, the *in vitro* phantom study was performed to assess the accuracy with which the proposed feature-based registration technique can predict the location of a target vessel. Our results have shown a root-mean-squared (RMS) TRE on the order of ~ 3.5 mm across the twenty-seven peri-operative poses simulated in our study. Moreover, considering that the target vessel location is predicted using a feature-based registration algorithm, we next assessed the error associated with the feature localization. **Table 2** includes a summary of the RMS localization error associated with the identification of each of the four features used to drive the registration: LCO, LVAp, MVA and AVA. As observed, the point-based localization of the LVAp and the LCO was consistently challenging, mainly due to the 2D nature of the US images used to identify a 3D structure.

These feature localization errors in fact explain the increasing TRE at the LAD fiducials closer to the apical region at both peri-operative stages. Considering that the anastomosis target site is typically located along the LAD path two thirds of the way from the ostia towards the apex, we have implemented a more robust apex localization approach from the US data which led to improved target registration error near the inferior end of the LAD. The new algorithm uses the apical region as a registration constraint as opposed to a single point, and include a robust estimator to reduce the TRE, as suggested by Ma *et al.* [11].

Nevertheless, in spite of these slight inaccuracies, our results are well within the 10–15 mm clinically-imposed constraint dictated by a typical intercostal space, allowing sufficient tolerance (over 10 mm) in the event that these errors amplify when using clinical data.

5 Conclusions

Driven by the clinical motivation to improve the pre-operative planning of RA-CABG procedures, here we have proposed and evaluated a technique used to predict the intra-operative target vessel location. Our technique was validated in an *in vitro* study simulating the clinically-observed RA-CABG procedure workflow and yielded 3.5 mm RMS accuracy in predicting the peri-operative LAD location. These results agree with the clinical constraints imposed by this application and provided an equally successful *in vivo* evaluation in our upcoming animal studies, currently under research ethics approval, employing DynaCT (Siemens, Erlangen, Germany) for intra-operative validation, we believe this technique has the potential to significantly improve the current pre-operative planning of RA-CABG procedures, and consequently lead to reduced rates of conversion to traditional open-chest surgery.

References

1. Modi, P., Hassan, A., Chitwood, W.R.J.: Minimally invasive mitral valve surgery: a systematic review and meta-analysis. *Eur. J. Cardiothorac. Surg.* 34, 943–952 (2008)
2. Doty, D.B., Flores, J.H., Doty, J.R.: Cardiac valve operations using a partial sternotomy technique. *J. Card. Surg.* 15, 35–42 (2000)
3. Vassiliades, T.A., Block, P.C., Cohn, L.H., et al.: The clinical development of percutaneous heart valve technology. *J. Thorac. Cardiovasc. Surg.* 129, 970 (2005)
4. Lutter, G., Ardehali, R., Cremer, J., et al.: Percutaneous valve replacement: current state and future prospects. *Ann. Thor. Surg.* 78, 2199–2206 (2004)
5. Modi, P., Rodriguez, E., Chitwood, W.R.J.: Robot-assisted cardiac surgery. *Interact. Cardiovasc. Thorac. Surg.* 9, 500–505 (2009)
6. Trejos, A.L., et al.: Optimizing port placement for robot-assisted minimally invasive cardiac surgery. *Int. J. Med. Robotics Comput. Assist. Surg.* 3, 355–364 (2007)
7. Damiano, R.Jr.: Robotics in cardiac surgery: The emperor’s new clothes. *The Journal of Thoracic and Cardiovascular Surgery* 134(3), 559–561 (2007)
8. Moore, J.T., Wiles, A.D., et al.: Integration of trans-esophageal echocardiography with magnetic tracking technology for cardiac interventions. In: *Medical Imaging 2010: Visualization, Image-Guided Procedures and Modeling. Proc. of SPIE*, vol. 7625, pp. 76252Y–1–10 (2010)
9. Linte, C.A., Carias, M., et al.: Estimating heart shift and morphological changes during minimally invasive cardiac interventions. In: *Medical Imaging 2010: Visualization, Image-Guided Procedures and Modeling. Proc. of SPIE*, vol. 7625, pp. 762509–1–11 (2010)
10. Press, W.H., Flannery, B.P., et al.: *Numerical recipes in C: The arts of scientific computing.* Cambridge University Press, Cambridge (1992)
11. Ma, B., Ellis, R.E., Fleet, D.J.: Spotlights: A robust method for surface-based registration in orthopedic surgery. In: Taylor, C., Colchester, A. (eds.) *MICCAI 1999. LNCS*, vol. 1679, pp. 936–945. Springer, Heidelberg (1999)

Quantification of Prostate Deformation due to Needle Insertion during TRUS-guided Biopsy

Tharindu De Silva^{1,2}, Aaron Fenster^{1,2,3},
Jagath Samarabandu², and Aaron D. Ward¹

¹ Imaging Research Laboratories, Robarts Research Institute

² Biomedical Engineering Graduate Program, University of Western Ontario

³ Department of Medical Biophysics, University of Western Ontario,
London, Ontario, Canada

tdesilva@imaging.robarts.ca, ward@robarts.ca

Abstract. Prostate biopsy is the clinical standard for the diagnosis of prostate cancer, and technologies for 3D guidance to targets and recording of biopsy locations are promising approaches to reducing the need for repeated biopsies. In this study, we use image-based non-rigid registration to quantify prostate deformation during needle insertion and biopsy gun firing, in order to provide information useful to the overall assessment of a TRUS-guided biopsy system's expected targeting error. We recorded mean tissue displacements of up to 0.4 mm, accounting for 16% of the clinically-motivated maximum desired RMS error of a guidance system.

1 Introduction

Prostate cancer is one of the most common cancers among men, second only to skin cancer [1]. Prostate biopsy is currently the standard clinical practice to obtain a definitive diagnosis of cancer. Two-dimensional (2D) transrectal ultrasound (TRUS) is the most common imaging modality used for image guidance in prostate biopsy due to its low cost and high frame rate. Recently, 3D magnetic resonance imaging (MRI) and TRUS-guided biopsy systems [2,3,4] have shown promise for the accurate guidance of biopsy needles to predefined targets. Such systems also provide a 3D record of biopsy locations, which is useful in planning targets for subsequent sessions if pathological analysis of biopsy samples reveals the need to rebiopsy the same or nearby targets in a later session. Biopsy target locations can also be determined based on other modalities such as MRI, requiring MRI-US registration. The need for accurate 3D guidance and recording of biopsies is clear; reports have shown that cancer detection rates during second and third biopsy sessions are only 20% and 7%, respectively, for 2D TRUS-guided biopsy [5].

Since the smallest tumours considered to be clinically significant have volumes of 0.5 cm³ or greater [6], corresponding to a spherical tumour with a radius of approximately 5 mm, the RMS error of a TRUS-guided biopsy system in delivering a needle to a target must be no more than 2.5 mm in order

to biopsy clinically significant tumours 95% of the time. There can be multiple potential sources of error that can cause the actual target biopsy location to be different from the expected target, including: (1) tolerances in the design and construction of mechanical needle guidance systems, (2) errors in imaging and calibration to the needle guidance systems, (3) patient and prostate motion and deformation during the procedure due to interaction with the TRUS probe and discomfort during biopsy, (4) prostate deformation due to slow biopsy needle insertion in preparation for biopsy gun firing, and (5) prostate deformation due to rapid biopsy needle insertion after firing the biopsy gun. There is previous research in measurement of and correction for the first three sources of error [4,7]. However, to the best of our knowledge, prostate deformation due to needle insertion through the rectal wall and biopsy gun firing has not yet been quantified in the context of prostate biopsy. Although the cumulative effect of all five of the above sources of error has been measured using MRI in the context of a specific MR-compatible robotic biopsy system [8], the specific contributions of needle insertion and biopsy gun firing to the overall error were not reported. In addition, we hypothesize that deformations in response to needle insertion and biopsy gun firing are different in the context of TRUS-guided biopsy due to several important differences in physical configuration. In contrast to the robotic procedure measured in [8], where an endorectal coil in a cylindrical housing is placed parallel to the rectal wall for imaging, TRUS-guided biopsy is typically conducted using an end-firing ultrasound transducer, where the transducer tip is manipulated against the anterior rectal wall in order to obtain images. The MR-guided robot in [8] inserts needles into the prostate through the rectal wall at an oblique angle to the endorectal coil housing, whereas in end-firing TRUS biopsy, the needles are parallel to the probe axis. It is reasonable to expect that these differences in physical configuration may lead to differing mechanical dynamics at the time of biopsy needle insertion and gun firing, resulting in different prostate deformation characteristics. The effect of needle insertion on prostate motion has been studied extensively in the context of brachytherapy procedures [9,10], where the patient is under general anesthesia and the brachytherapy needles are inserted slowly (relative to the firing speed of a biopsy gun) through the perineum. We hypothesize that the effect of the needle in the context of biopsy is different due to the reactions (e.g. in the form of pelvic floor muscle contractions) of an awake, uncomfortable patient and the high speed of needle insertion by the biopsy gun.

In this work, we apply intensity-based, non-rigid image registration to measure the deformation observed in the prostate during (1) the manual insertion of the biopsy needle into the prostate, and (2) the firing of the biopsy gun to acquire the tissue sample during hand-held, 2D TRUS-guided biopsy procedures. At a high-level, our procedure is as follows: (1) acquire video-rate sequences of 2D TRUS images during the acquisition of each prostate biopsy sample from a series of clinical biopsy sessions; (2) for each biopsy, select three frames, occurring before needle insertion, before biopsy gun firing, and after biopsy gun firing; (3) non-rigidly register these three frames together, validating the registration method

by measurement of its target registration error (TRE); and (4) quantify prostate deformation by taking clinically relevant measurements of the deformation vector fields given by the registration algorithm.

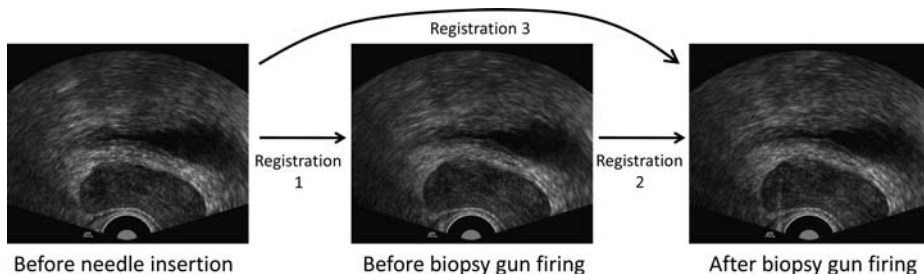


Fig. 1. Images in a biopsy sequence, and the names used for the three indicated registrations throughout this paper. In all of our images, the needle is to the left of the transducer, as shown in the rightmost image.

2 Method

Using a hand-held 2-dimensional (2D) transrectal ultrasound (TRUS) probe, we acquired TRUS images at 30 frames per second during clinical TRUS-guided prostate biopsy sessions of 9 patients (3 by one physician and 6 by another). From each biopsy sample taken, we extracted 3 frames; we refer to each of these frame triplets as a “biopsy sequence”. A biopsy sequence (figure 1) consists of: (1) the frame occurring immediately prior to the physician’s insertion of the needle through the rectal wall, (2) the frame occurring immediately prior to the firing of the biopsy gun, and (3) the frame occurring immediately after the firing of the biopsy gun. From our 9 patients, we obtained 96 biopsy sequences; an example of a sequence is given in figure 1.

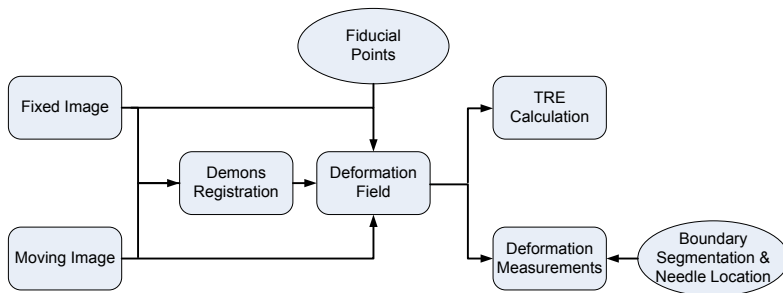


Fig. 2. High level method used in this paper

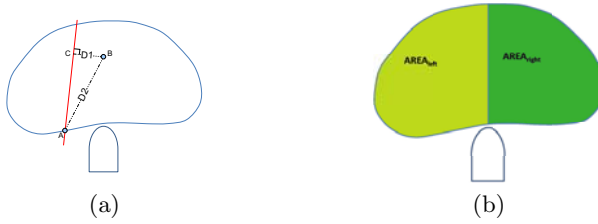


Fig. 3. (a) Distance from the needle axis (D1) and distance to the piercing point (D2). (b) Area to the left and right of the prostate from the probe axis.

Our method is given at a high level in the block diagram shown in figure 2. From each biopsy sequence, we choose a pair of images, with the image from the earlier time point designated as the moving image, and the image from the later time point designated as the fixed image. We then nonrigidly register the moving image to the fixed image, yielding a deformation field describing the motion and deformation of the prostate from the earlier time point to the later time point. The deformation field consists of vectors, each having two components that describe the magnitude and the direction of the displacement at each location. In order to validate the chosen registration algorithm, we measure the target registration error (TRE) of manually marked, intrinsic fiducials (calcifications), using the method described in 7. The spatial relationship between the needle and probe is fixed, and we oriented all the images in this study such that the needle lies at the same location in image space, to the *left* of the probe. We quantified the deformation fields given by registration using the following clinically-relevant calculations: (1) the mean deformation as a function of distance to the biopsy needle (distance D1 in figure 3(a)), (2) the mean deformation as a function of distance to the piercing point where the needle enters the prostate (distance D2 in figure 3(a)), and (3) the mean deformation as a function of the lateral position of the probe relative to the prostate, determined using the ratio of prostate areas to the left and right of the transducer (figure 3(b)). The latter two measurements require a prostate segmentation, which we perform manually in this study.

Our immediate use of non-rigid registration in figure 2 (i.e. without performing an initial rigid alignment) is due to our observation that rigid motion of the prostate within the biopsy sequences is minimal (mean rigid misalignment of 0.2 mm translation and 1 degree rotation). It is also intuitive that needle insertion and gun firing cause local deformations within the prostate. Therefore, we directly use a non-rigid registration algorithm having sufficient flexibility to measure these local deformations. A sample deformation vector field generated by the registration of an image prior to needle insertion to an image immediately after gun firing is shown in figure 4. Figure 4(a) represents the magnitudes of the deformation field vectors, including a manually contoured boundary of the prostate and location of the needle. Figure 4(b) and (c) represent the lateral and axial deformation fields respectively, to illustrate the anisotropy of the deformation field with respect to the orientation of the biopsy needle.

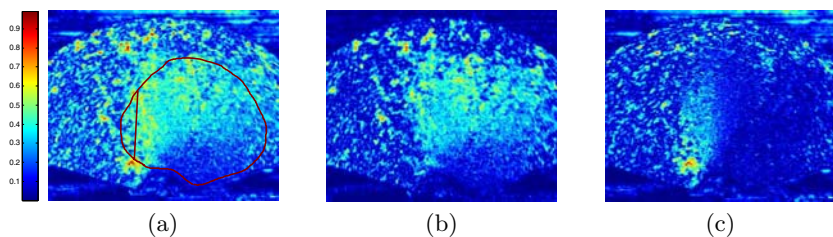


Fig. 4. (a) Deformation field vector magnitude with prostate boundary and needle location indicated. (b) Lateral (x) components of the deformation field (c) Axial (y) components of the deformation field.

3 Results

The TREs of our tested non-rigid registration algorithms are given in table 1. All tested algorithms improved the TRE, with the Demons algorithm providing the best overall results. However, it is worthy of note that the fiducial localization error (FLE) for trans-rectal ultrasound (TRUS) images has been determined in previous work to be 0.2 mm [7], which indicates that our reported TREs are approaching their lower bound. Nevertheless, the Demons algorithm was chosen for use in this study because (1) it provided the best overall TRE, (2) it is based on optical flow. Optical flow techniques have been shown to be useful in tracking fine-scale structure in ultrasound images in the presence of small tissue motion [11], which is expected in our study where tissue motion is localized in the vicinity of the needle.

Table 1. Comparison of TRE before and after registration

Registration Method	Registration 1 TRE(mm)	Registration 2 TRE(mm)
Before registration	0.31	0.21
Demons	0.29	0.14
Symmetric forces demons	0.32	0.16
Bspline	0.28	0.16

Figure 5(a) shows the relationship between the mean deformation magnitude and the perpendicular distance from the line of the needle insertion within the prostate for Registration 3. The curve plateaus to the right at a mean deformation value of approximately 0.28 mm. Figure 6 decomposes the deformation down into needle insertion (Registration 1) and gun firing (Registration 2) separately. Figure 5(b) shows the mean deformation magnitude as a function of the distance of the deformation vector from the insertion point of the needle into the prostate in Registration 3. Since we are measuring very small deformations, a valid concern is the potential effect of any noise present in the images on our measurements of the deformation fields. Figures 5 and 6 show the means of the *unsigned* vector magnitudes, i.e., $\frac{1}{N} \sum_{i=1}^N \sqrt{x_i^2 + y_i^2}$, which accumulate the

contributions of noise throughout the vector field, and may account for the unexpected positive plateau at approximately 0.28 mm. To address this issue, we show in figure 7(a) (corresponding to figure 5(a)) the magnitudes of the *signed* means of the vector components, i.e., $\sqrt{(\frac{1}{N} \sum_{i=1}^N x_i)^2 + (\frac{1}{N} \sum_{i=1}^N y_i)^2}$. Similarly, figure 7(b) gives the corresponding information to figure 5(b). We take the signed means of the vector components under the assumption that the true tissue motion is regular and any noise is irregular, with a tendency to self-cancellation when taking the signed mean. This effect is illustrated by the lower plateaus observed in figure 7. To explore the relationship between deformation and the relative positions of the probe and the prostate, we plotted the mean deformation magnitude of vectors within 5 mm of the needle (which is the clinical region of interest for biopsy targeting) versus the ratio of areas within the prostate to the left of the probe axis and to the right of the probe axis in figure 8. The point on the x axis at 0.5 depicts the situation where the probe is aligned with the central axis of the prostate. We used linear regression to fit a line to the resulting cluster of points; the correlation ratio was found to be -0.15.

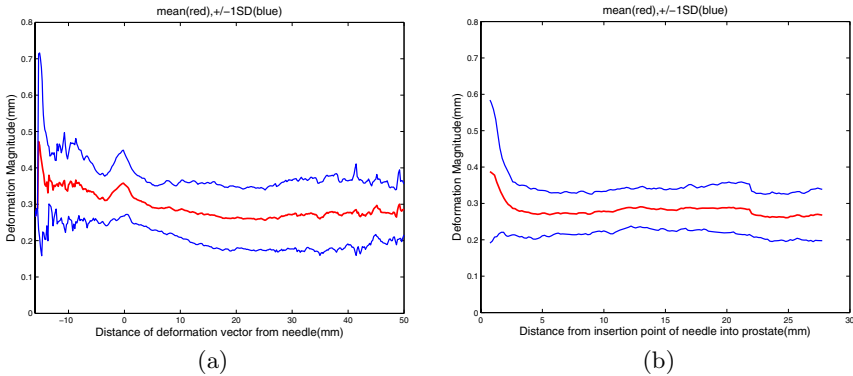


Fig. 5. Mean (red) and +/- 1 standard deviation (blue) of *unsigned* deformation vector field magnitudes for Registration 3, (a) as a function of distance from the needle, and (b) as a function of distance to the piercing point within the prostate

4 Discussion

In the graph in figure 5(a), the deformation to the left of the probe (nearer to the prostate edge) is approximately 0.1 mm higher than to the right. There is a slight trend of increased magnitude of deformation toward both edges of the prostate. Figure 6 decomposes this deformation into needle insertion (Registration 1) and gun firing (Registration 2) separately. The shapes of the two graphs in figure 6 are similar to that obtained for Registration 3 (figure 5(a)). However, note that the deformation after gun firing is approximately 0.1 mm higher than that due to needle insertion before gun firing. This shows that a significant portion of the deformation of Registration 3 (figure 5(a)) is due the firing of the gun.

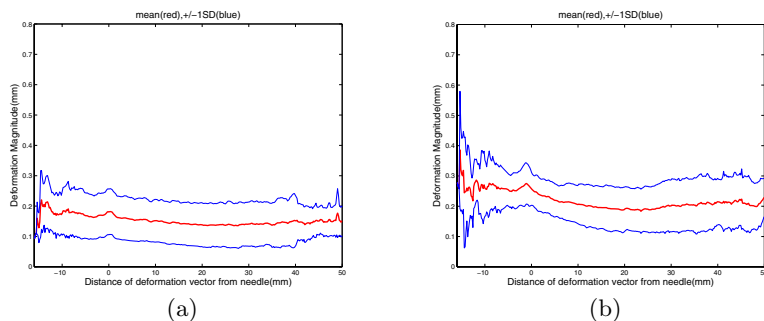


Fig. 6. Mean (red) and ± 1 standard deviation (blue) of *unsigned* deformation vector field magnitudes as a function of distance from the needle for (a) Registration 1, (b) Registration 2

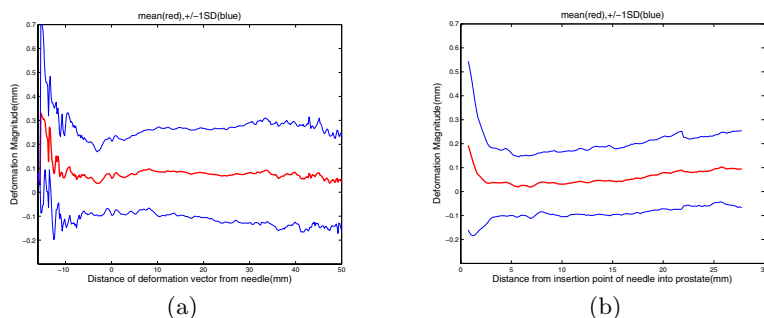


Fig. 7. Mean (red) and ± 1 standard deviation (blue) of the magnitudes of the *signed* vector component means for Registration 3, (a) as a function of distance from the needle, and (b) as a function of distance to the piercing point within the prostate

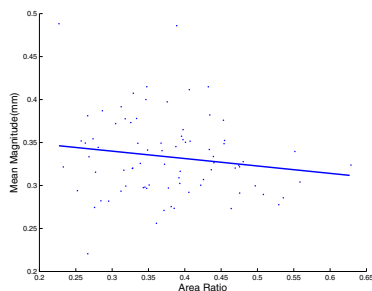


Fig. 8. Deformation magnitude as a function of the left:right (relative to the transducer centerline) ratio of prostate areas

Figure 5(b) shows a concentration of deformation within 5 mm of the insertion point of the needle into the prostate. At the piercing point, the deformation is approximately 0.4 mm. The deformation at the piercing point is about 0.2 mm higher than the deformation in the surrounding region. Registration 1 and

Registration 2 also showed a similar trend and the deformation was larger in Registration 2 (biopsy gun firing) than in Registration 1.

5 Conclusion

Observed mean tissue displacements as a consequence of needle insertion and gun firing are as high as 0.4 mm. These displacements could account for 10%–20% of the allowable 2.5 mm RMS error permitted in a biopsy system designed to successfully sample spherical targets with 5 mm radius 95% of the time. We observe that the majority of the deformation due to needle insertion and biopsy gun firing occurs lateral to the needle, away from the center of the prostate, and at the needle's piercing point at the prostate boundary.

References

1. Canadian Cancer Society's steering committee: Canadian cancer statistics 2009 (2009)
2. Cool, D., Sherebrin, S., Izawa, J., Chin, J., Fenster, A.: Design and evaluation of a 3D transrectal ultrasound prostate biopsy system. *Medical Physics* 35(10), 4695–4707 (2008)
3. Krieger, A., Susil, R.C., Ménard, C., Coleman, J.A., Fichtinger, G., Atalar, E., Whitcomb, L.L.: Design of a novel MRI compatible manipulator for image guided prostate interventions. *IEEE Transactions on Biomedical Engineering* 52(2), 306–313 (2005)
4. Bax, J., Cool, D., Gardi, L., Knight, K., Smith, D., Montreuil, J., Sherebrin, S., Romagnoli, C., Fenster, A.: Mechanically assisted 3D ultrasound guided prostate biopsy system. *Medical Physics* 35(12), 5397–5410 (2008)
5. Keetch, D.W., Catalona, W.J., Smith, D.S.: Serial prostatic biopsies in men with persistently elevated serum prostate specific antigen values. *The Journal of Urology* 151(6), 1571–1574 (1994); PMID: 7514690
6. Epstein, J.I., Sanderson, H., Carter, H.B., Scharfstein, D.O.: Utility of saturation biopsy to predict insignificant cancer at radical prostatectomy. *Urology* 66(2), 356–360 (2005); PMID: 16040085
7. Karnik, V.V., Fenster, A., Bax, J., Cool, D.W., Gardi, L., Gyacskov, I., Romagnoli, C., Ward, A.D.: Assessment of image registration accuracy in three-dimensional transrectal ultrasound guided prostate biopsy. *Medical Physics* 37(2), 802–813 (2010)
8. Xu, H., Lasso, A., Vikal, S., Guion, P., Krieger, A., Kaushal, A., Whitcomb, L.L., Fichtinger, G.: Accuracy validation for MRI-guided robotic prostate biopsy. In: *Medical Imaging, San Diego, California, USA*, vol. 7625, pp. 762517–8. SPIE, San Jose (2010)
9. Lagerburg, V., Moerland, M.A., Lagendijk, J.J., Battermann, J.J.: Measurement of prostate rotation during insertion of needles for brachytherapy. *Radiotherapy and Oncology* 77(3), 318–323 (2005)
10. Stone, N.N., Roy, J., Hong, S., Lo, Y., Stock, R.G.: Prostate gland motion and deformation caused by needle placement during brachytherapy. *Brachytherapy* 1(3), 154–160 (2002)
11. Meunier, J.: Tissue motion assessment from 3D echographic speckle tracking. *Physics in Medicine and Biology* 43, 1241–1254 (1998)

Optimized Anisotropic Rotational Invariant Diffusion Scheme on Cone-Beam CT

Dirk-Jan Kroon¹, Cornelis H. Slump¹, and Thomas J.J. Maal²

¹ Signals and System, University of Twente, The Netherlands

² Radboud University Nijmegen Medical Center, The Netherlands

Abstract. Cone-beam computed tomography (CBCT) is an important image modality for dental surgery planning, with high resolution images at a relative low radiation dose. In these scans the mandibular canal is hardly visible, this is a problem for implant surgery planning. We use anisotropic diffusion filtering to remove noise and enhance the mandibular canal in CBCT scans. For the diffusion tensor we use hybrid diffusion with a continuous switch (HDCS), suitable for filtering both tubular as planar image structures. We focus in this paper on the diffusion discretization schemes. The standard scheme shows good isotropic filtering behavior but is not rotational invariant, the diffusion scheme of Weickert is rotational invariant but suffers from checkerboard artifacts. We introduce a new scheme, in which we numerically optimize the image derivatives. This scheme is rotational invariant and shows good isotropic filtering properties on both synthetic as real CBCT data.

1 Introduction

Cone-beam computed tomography (CBCT) is an increasingly utilized imaging modality for dental surgery planning [1], due to the low hardware cost and high resolution images at a relative low radiation dose. For the surgical planning of implants, the mandibular nerve canals have to be segmented. In these scans the mandibular nerve canals are hardly visible. In implant placement, the segmentation is used to guard the safety margin around the canals during surgery. CBCT scanners have a relatively low radiation dose [1] thus the small mandibular canal is characterized by low contrast in a noisy image, see figure 4. The research goal of this paper, is to find a method to improve image contrast in CBCT scans for small structures.

Currently the best way to improve contrast in a CT image is to apply iterative reconstruction methods with regularization to suppress streak-artifacts and to improve smoothness in uniform regions [2]. In practice CBCT systems do not provide the required raw-scanner data for this approach. Therefore post reconstruction noise filtering is the practical method to improve image quality. A medical image is often assumed to have piecewise smooth regions with oscillatory noise, separated by sharp edges. There are many methods available in the literature to denoise such an image [3], in this paper we focus on edge enhancing diffusion filtering.

Linear diffusion equals Gaussian filtering in which the diffusion time controls the smoothing scale. To preserve the edges Perona-Malik introduced regularized non-linear diffusion (RPM) [4]. Edge preservation is achieved by lowering the scalar diffusion constant in the neighborhood of steep edges. This method results in piecewise smooth regions, however, image edges remain noisy. Instead of using a scalar diffusion constant, a tensor can be used to adapt the diffusion to the underlying image structure. So we smooth with small elongated kernels along edges, and Gaussian like kernels in uniform regions. The tensor can be constructed in two ways, as a coherence-enhancing diffusion (CED) [5] or as an edge-enhancing diffusion (EED). Recently the CED and EED algorithms were combined in an hybrid diffusion filter with a continuous switch (HDCS) [6]. If the local image structure is tubular HDCS switches to CED and if it is planar to EED.

The focus of this paper are the discretization schemes of the anisotropic diffusion tensor. We will evaluate the performance of the standard discretization scheme and the rotational invariant scheme of Weickert [7], and introduce a new scheme in which optimal filtering kernels are constructed using numerical optimization.

This paper is organized as follows, in the second section we introduce the diffusion filtering algorithm and discretization schemes. The new optimized scheme is introduced in the third section. Followed by evaluation of the diffusion schemes on synthetic and real images, and by the final section with discussion and conclusions.

2 Diffusion Filtering

Anisotropic diffusion filtering is an iterative edge preserving smoothing method. It describes the local image structure using a structure tensor also referred to as the "second-moment matrix", for details see [5]. This descriptor is transformed into a diffusion tensor \mathbf{D} . The diffusion equation is commonly written in an iterative forward difference approximation [7]:

$$\frac{\partial u}{\partial t} = \nabla \cdot (\mathbf{D} \nabla u) \quad \Rightarrow \quad u_{k+1} \cong u_k + (\nabla \cdot (\mathbf{D} \nabla u)) \quad (1)$$

Where u ($u = u(t, x, y, z)$) is the image, x, y, z the pixel coordinates and t the diffusion time. In the discrete function the continuous time is replaced by, τ the time step-size and k the number of the iterations. The eigenvectors of the diffusion tensor \mathbf{D} are set equal to the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ with $\mathbf{v}_1 = [v_{11}, v_{12}, v_{13}]$ of the structure tensor (note the symmetry):

$$\mathbf{D} = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{12} & D_{22} & D_{23} \\ D_{13} & D_{23} & D_{33} \end{bmatrix} \quad \text{with} \quad D_{ij} = \sum_{n=1..3} \lambda_n v_{ni} v_{nj} \quad (2)$$

The eigenvalues of the diffusion tensor are $\lambda_1, \lambda_2, \lambda_3$. Because our CBCT scans contain planar and tubular structures as well, we choose to use HDCS, with switches between CED and EED eigenvalues depending on the local image structure, for details see [6].

We can write the divergence operator equation [1](#) in 3D as:

$$\nabla \cdot (\mathbf{D}\nabla u) = \partial_x j_1 + \partial_y j_2 + \partial_z j_3 \tag{3}$$

With j_1, j_2, j_3 the flux components which are described by:

$$\begin{aligned} j_1 &= D_{11} (\partial_x u) + D_{12} (\partial_y u) + D_{13} (\partial_z u) \\ j_2 &= D_{12} (\partial_x u) + D_{22} (\partial_y u) + D_{23} (\partial_z u) \\ j_3 &= D_{13} (\partial_x u) + D_{23} (\partial_y u) + D_{33} (\partial_z u) \end{aligned} \tag{4}$$

For the standard discretization of the divergence operator central differences are used:

$$\begin{aligned} \partial_y (D_{12} (\partial_x u)) &= \frac{1}{2} \left(D_{12(i,j+1,k)} \frac{u_{(i+1,j+1,k)} - u_{(i-1,j+1,k)}}{2} \right. \\ &\quad \left. - D_{12(i,j-1,k)} \frac{u_{(i+1,j-1,k)} - u_{(i-1,j-1,k)}}{2} \right) \end{aligned} \tag{5}$$

The other terms are written in the same way [8](#), and are combined to a pixel-location dependent 3×3 or $3 \times 3 \times 3$ convolution stencil. Non-negative discretization makes the modification that stencil elements remain positive for various gray values. Rotation invariant anisotropic diffusion is important with curved like structures such as the mandibular canal. Weickert [7](#) showed that larger stencils than 3×3 (2D) are needed to fix the number of degrees of freedom to allow rotation invariance. This is achieved by implementing the equations [3](#) and [4](#), with Scharr’s rotational invariant 3×3 filters for the image derivatives ∂_x and ∂_y , resulting in an rotational invariant implicit 5×5 stencil.

3 Optimized Scheme

Another way to write the divergence operator using the product rule [9](#) is:

$$\nabla \cdot (\mathbf{D}\nabla u) = \text{div}(\mathbf{D})\nabla u + \text{trace}(\mathbf{D}(\nabla\nabla^T u)) \tag{6}$$

We obtain for the divergence part of the equation:

$$\begin{aligned} \text{div}(\mathbf{D})\nabla u &= (\partial_x u)(\partial_x D_{11} + \partial_y D_{12} + \partial_z D_{13}) \\ &\quad + (\partial_y u)(\partial_x D_{12} + \partial_y D_{22} + \partial_z D_{23}) \\ &\quad + (\partial_z u)(\partial_x D_{13} + \partial_y D_{23} + \partial_z D_{33}) \end{aligned} \tag{7}$$

We write the Hessian part of the equation as:

$$\begin{aligned} \text{trace}(\mathbf{D}(\nabla\nabla^T u)) &= (\partial_{xx} u)D_{11} + (\partial_{yy} u)D_{22} + (\partial_{zz} u)D_{33} \\ &\quad + 2(\partial_{xy} u)D_{12} + 2(\partial_{xz} u)D_{13} + 2(\partial_{yz} u)D_{23} \end{aligned} \tag{8}$$

Equation [7](#) is discretized using $3 \times 3 \times 3$ derivative kernels, and the Hessian of equation [8](#) with a $5 \times 5 \times 5$ second derivative kernel. In 2D the spatial kernels can be written as:

$$M_{xx} = \begin{bmatrix} p_1 & p_2 & p_3 & p_2 & p_1 \\ p_4 & p_5 & p_6 & p_5 & p_4 \\ -p_7 & -p_8 & -p_9 & -p_8 & -p_7 \\ p_4 & p_5 & p_6 & p_5 & p_4 \\ p_1 & p_2 & p_3 & p_2 & p_1 \end{bmatrix} M_{xy} = \begin{bmatrix} p_{10} & p_{11} & 0 & -p_{11} & -p_{10} \\ p_{11} & p_{12} & 0 & -p_{12} & -p_{11} \\ 0 & 0 & 0 & 0 & 0 \\ -p_{11} & -p_{12} & 0 & p_{12} & p_{11} \\ -p_{10} & -p_{11} & 0 & p_{11} & p_{10} \end{bmatrix} \quad (9)$$

$$M_x = \begin{bmatrix} p_{13} & p_{14} & p_{13} \\ 0 & 0 & 0 \\ -p_{13} & -p_{14} & -p_{13} \end{bmatrix} \quad (10)$$

The kernel values $\mathbf{p} = [p_1, p_2, \dots, p_{14}]$ can be found analytically or by numerical optimization. We choose numerical optimization, because it can optimize the whole process, while analytical derivation is only feasible for separate parts of the process, with simplifications such as ignoring numerical round of effects. We optimize the diffusion kernel using the following cost function:

$$\mathbf{p} = \arg \min_{\mathbf{p}} (e_f(\mathbf{p}) + \alpha e_g(\mathbf{p})) \quad (11)$$

This function finds a balance between the edge orientation invariant filtering performance e_f , and isotropic diffusion performance e_g , with weight constant α . With the first term e_f we want to find the best edge enhancement for edges with several orientations and spatial frequencies. Therefore we use the difference between an image with circles of varying spatial frequencies without noise I , and an image with Gaussian noise added I_{noise} , which is diffusion filtered. With $F(I_{noise}, \mathbf{p})$ the diffusion filtering of the image with noise using kernel values \mathbf{p} :

$$e_f(\mathbf{p}) = \sum_{\mathbf{x}} |F(I_{noise}, \mathbf{p}) - I|, \quad \text{with } I = \sin(x^2 + y^2) \quad (12)$$

With the second term e_g we want to achieve Gaussian like diffusion in uniform regions. We use an image I_{point} which is zero except the center pixel equal to one. The term e_g is set to the difference between the isotropic noise filtered image I_{point} and a least squares fitted Gaussian kernel. We set both diffusion tensor eigenvalues to one, corresponding to a uniform region.

$$e_g(\mathbf{p}) = \arg \min_a \sum_{\mathbf{x}} \left(F(I_{point}, \mathbf{p}) - \frac{1}{\pi\sqrt{a}} \exp(-|x|^2/a) \right)^2 \quad (13)$$

We use the Matlab Nelder-Mead Simplex minimizer [10] because it is robust against local minima. Also a quasi Newton minimizer is used [11], because the minimizer has a high convergence speed. We use 10 iterations of the Simplex Method followed by minimizing until convergence with the quasi Newton optimizer. This is done iteratively until the simplex method also converges. Parameters used for the circle image are, size 255×255 , $\tau = 0.1$, iterations 5, $\sigma = 1$, $\rho = 10$, CED eigenvalues, Gaussian noise variance 0.1, and x and y coordinates in the range $[-10, 10]$. The parameters of I_{point} are image size 51×51 and 5 iterations, constant $\alpha = 200$. The computed kernel values \mathbf{p} are:

0.008 0.049 0.032 0.038 0.111 0.448 0.081
 0.334 0.937 0.001 0.028 0.194 0.006 0.948

It is important to note that the scheme is optimized for rotational invariance, but that the derivative kernels are not rotational invariant, for instance M_x approximates a central difference instead of a Scharr like kernel.

In 3D the approach is the same, a spherical function in an image volume is used, with 33 instead of 14 unknown kernel variables. The optimized kernels are available in our open source diffusion toolbox¹.

4 Evaluation

We evaluate the properties of the standard, rotation and optimized diffusion scheme with respect to three image based criteria. The first is noise removal in uniform regions, the second preservation and enhancement of image edges independent of rotation and size. The final test is the combined filtering performance on a real CBCT dataset.

In this first test we look at noise smoothing in uniform regions. To do this we use the image I_{point} introduced in the optimization section, with the same filtering parameters and 100 iterations. Figure 1 shows the image results and difference between a least squares fitted Gaussian 2D function and the diffusion result. Ideal uniform diffusion is equal to Gaussian filtering, thus the standard diffusion and the optimized scheme perform well. The rotation invariant result does not look like a Gaussian, this is because the scheme is based on Sobel like derivative kernels, which do not use the local pixel value but only the neighboring values.

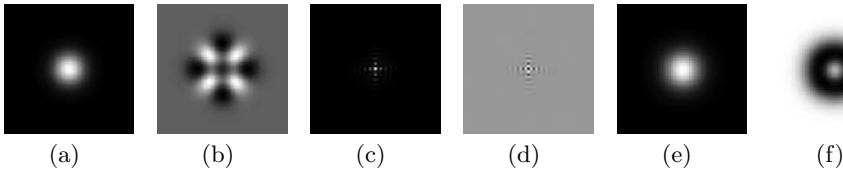


Fig. 1. Uniform Diffusion of a pixel with standard discretization (a), rotation invariant (c), optimized scheme (e). Sub figures (b), (d) and (f) show the difference between the image result and least squares fitted 2D Gaussian function. The values in (b) are in the order of $1 \cdot 10^{-5}$, (d) in the order of $1 \cdot 10^{-2}$ and (f) in the order of $1 \cdot 10^{-4}$.

In the second test we look at rotation invariant edge enhancement, using the circle image with Gaussian noise I_{noise} , the same parameters as in the optimization section and 100 iterations.

Figure 2 shows that only the rotational invariant and optimized scheme are edge orientation independent. The rotational invariant scheme suffers from checkerboard artifacts due to the Scharr derivative kernels which only uses neighbor pixels and not the current pixel.

¹ Source code, <http://www.mathworks.com/matlabcentral/fileexchange/25449>

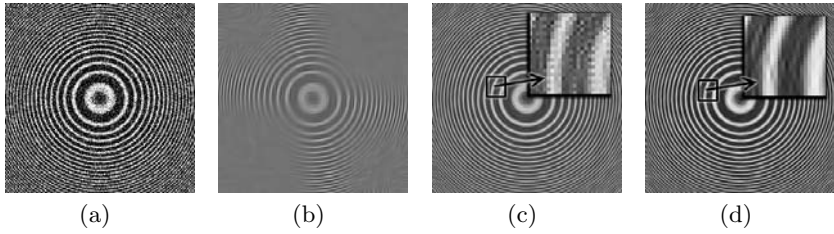


Fig. 2. The sub figures show the test image (a), after diffusion with the standard scheme (b), with the rotational invariant scheme (c), and the optimized scheme (d)

The final test is performed on 8 CBCT preprocessed human-head datasets of $400 \times 400 \times 551$ voxels. The preprocessing consist of clustering the data sets in to three intensity classes background, tissue and bone, using bias field corrected fuzzy clustering [12], which is robust to streak artifacts. The resulting image data serves as ground truth for the edges. The edges are detected by applying a threshold on the gradient magnitude. Uniform regions are defined as the pixels which are at least six voxels away from an edge. Finally Gaussian noise of variance 0.01 is added to the image data. The image data is filtered with the standard and the optimized scheme using HDCS eigenvalues, with parameters $\sigma = 0.5$, $\rho = 2$, $\tau = 0.15$, HDCS parameters $\lambda_e = 30$, $\lambda_h = 30$, $\lambda_c = 15$ and 26 iterations, see figure 3. Time to filter one dataset on an Intel Core 2 Duo desktop PC is approximately 25 minutes for the diffusion, and about 2.5 hours for the NLM filter.

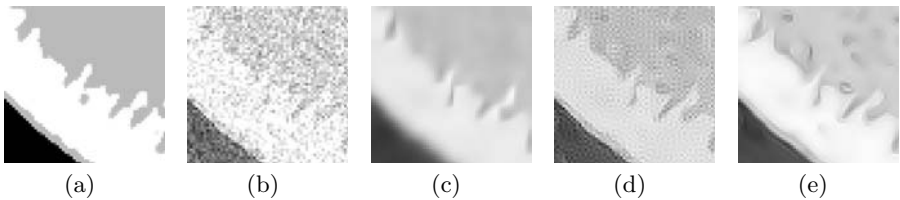


Fig. 3. Small part of HDCS filtered bone structure, ground truth (a), Gaussian noise added (b), standard scheme (c), rotation invariant (d) and optimized scheme (e)

We compare the performance between the standard, the optimized diffusion scheme, and the original non-local means (NLM) [13]. The summed squared pixel distance between Gaussian low pass filtered and original diffusion results is used as a performance value. A step edge contains high frequencies which will be removed by the low pass filter, resulting in a large pixel distance. In uniform regions high frequency noise will also be removed, thus a large pixel distance is a sign of noise which is not removed by the diffusion filtering. We calculate the smoothing pixel distance values for the edge pixels and for the uniform regions. The results are shown in table 1.

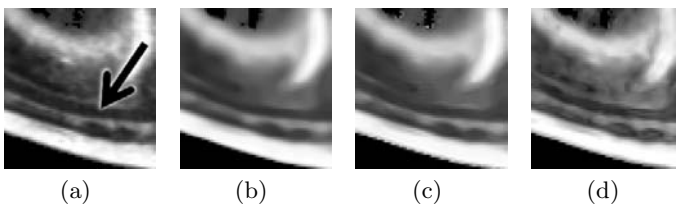
Table 1. Pixel distance between Gaussian smoothed and raw edge preserving diffusion filtering results of standard, optimized diffusion scheme and NLM

dataset	Edge				Uniform region			
	raw	optimized	standard	NLM	raw	optimized	standard	NLM
1	$6.2 \cdot 10^4$	$2.0 \cdot 10^4$	$1.1 \cdot 10^3$	$1.4 \cdot 10^3$	$4.6 \cdot 10^4$	$1.6 \cdot 10^3$	7.6	1.9
2	$6.6 \cdot 10^4$	$2.1 \cdot 10^4$	$1.5 \cdot 10^3$	$2.1 \cdot 10^3$	$4.7 \cdot 10^4$	$1.2 \cdot 10^3$	6.7	1.5
3	$6.4 \cdot 10^4$	$2.1 \cdot 10^4$	$1.2 \cdot 10^3$	$1.4 \cdot 10^3$	$4.6 \cdot 10^4$	$1.4 \cdot 10^3$	6.5	0.7
4	$6.5 \cdot 10^4$	$2.0 \cdot 10^4$	$1.5 \cdot 10^3$	$2.2 \cdot 10^3$	$4.7 \cdot 10^4$	$1.2 \cdot 10^3$	6.2	3.1
5	$6.9 \cdot 10^4$	$2.3 \cdot 10^4$	$1.3 \cdot 10^3$	$2.3 \cdot 10^3$	$4.6 \cdot 10^4$	$1.5 \cdot 10^3$	6.9	0.6
6	$6.7 \cdot 10^4$	$2.6 \cdot 10^4$	$2.0 \cdot 10^3$	$1.6 \cdot 10^3$	$4.5 \cdot 10^4$	$2.2 \cdot 10^3$	9.9	0.3
7	$7.0 \cdot 10^4$	$2.2 \cdot 10^4$	$1.5 \cdot 10^3$	$1.9 \cdot 10^3$	$4.7 \cdot 10^4$	$1.3 \cdot 10^3$	6.8	2.6
8	$6.4 \cdot 10^4$	$2.2 \cdot 10^4$	$1.3 \cdot 10^3$	$1.8 \cdot 10^3$	$4.6 \cdot 10^4$	$1.4 \cdot 10^3$	6.7	1.6

The NLM algorithm and standard scheme gives the best smoothing performance for uniform regions, with a 200 times smaller distance compared to the optimized scheme. This is because the optimized scheme preserved the edges of some random noise structures. The same noise structures are also visible in the rotation invariant scheme in image [3](#). In the HDCS eigenvalues there is a threshold value λ_h to separate between noise and a image structures. But in this case the signal to noise ratio is too low to allow a good separation between noise and real image structures. Also on the real object edges the optimized scheme gives the highest pixel distance. This can be due to remaining noise on the edges or due to a steeper image edge than with standard scheme. Figure [3](#) shows it is because the image edges are more pronounced.

The original 8 CBCT datasets were also filtered with the three methods, and slices were shown to three medical experts which use cone-beam CT. They preferred the optimized filtering despited the fact it sometimes enhances noise structures. They explained that the other methods lose small important details, while the optimized filtering enhanced some hardly visible structures. Noise structures are not a major problem because the anatomy is known.

Finally we show the filtering results of all schemes on an CBCT scan which is geometric transformed to make the jaw flat, see figure [4](#). The optimized scheme gives the best enhancement and preservation of the mandibular canal.

**Fig. 4.** Small part of HDCS filtered scan (a), mandibular canal (arrow), standard scheme (b), rotation invariant (c) and optimized scheme (d). The optimized scheme better preserves the original edges and image structure.

5 Conclusion

The introduced 2D/3D anisotropic diffusion scheme, shows better edge enhancement in our synthetic and CBCT data, compared to the standard, rotation invariant scheme and NLM. Filtering is Gaussian in uniform image regions without checkerboard artifacts. The results show that the better edge preservation also causes high noise structures to be preserved. Despite this artifact the medical experts preferred the introduced method because it enhanced also hardly visible anatomical structures. The cause of the problem is not the optimal scheme, but has to be solved by a better separation between noise edges and real edges in the diffusion tensor construction part.

References

1. Roberts, J., Drage, N., Davies, J., Thomas, D.: Effective dose from Cone-beam CT examinations in dentistry. *Br. J. Radiol.* 82, 35–40 (2009)
2. Sunnegardh, J.: Iterative Filtered Backprojection Methods for Helical Cone-Beam CT. PhD thesis, Computer Vision, The Institute of Technology (2009)
3. Awate, S., Whitaker, R.: Unsupervised, Information-Theoretic, Adaptive Image Filtering for Image Restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 364–376 (2006)
4. Perona, P., Malik, J.: Scale-space and Edge Detection using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 629–639 (1990)
5. Weickert, J.: Anisotropic Diffusion in Image Processing. PhD thesis, University of Copenhagen, Department of Computer Science (1998)
6. Mendrik, A., Vonken, E., Rutten, A., Viergever, M., van Ginneken, B.: Noise Reduction in Computed Tomography Scans using 3D Anisotropic Hybrid Diffusion with Continuous Switch. *IEEE Trans. Med. Imaging* 28(10), 1585–1594 (2009)
7. Weickert, J., Scharr, H.: A Scheme for Coherence-Enhancing Diffusion Filtering with Optimized Rotation Invariance. *J. Vis. Commun. Image Represent.* 13(1), 103–118 (2002)
8. Frangakis, A., Hegerl, R.: Noise Reduction in Electron Tomographic Reconstructions using Nonlinear Anisotropic Diffusion. *J. Struct. Biol.* (135), 239–250 (2001)
9. Felsberg, M.: On the Relation between Anisotropic Diffusion and Iterated Adaptive Filtering. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 436–445. Springer, Heidelberg (2008)
10. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM J. Opt.* 9(1), 112–147 (1998)
11. Shanno, D.: Conditioning of Quasi-Newton Methods for Function Minimization. *Math. Comput.* 24(111), 647–656 (1970)
12. Ahmed, M., Yamany, S., Farag, A., Moriarty, T.: Bias Field Estimation and Adaptive Segmentation of MRI Data Using a Modified Fuzzy C-Means Algorithm. In: Proc. IEEE Int. Conf. Comput. Vis. and Pattern Recognit., pp. 250–255 (1999)
13. Buades, A., Coll, B., Morel, J.: A Non-Local Algorithm for Image Denoising. In: Proc. of IEEE CVPR 2005, Washington, DC, USA, vol. 2, pp. 60–65. IEEE Computer Society, Los Alamitos (2005)

Control of Articulated Snake Robot under Dynamic Active Constraints

Ka-Wai Kwok, Valentina Vitiello, and Guang-Zhong Yang

The Hamlyn Centre for Robotic Surgery
Imperial College London, London, United Kingdom
{k.kwok07,v.vitiello07,g.z.yang}@imperial.ac.uk

Abstract. Flexible, ergonomically enhanced surgical robots have important applications to transluminal endoscopic surgery, for which path-following and dynamic shape conformance are essential. In this paper, kinematic control of a snake robot for motion stabilisation under dynamic active constraints is addressed. The main objective is to enable the robot to track the visual target accurately and steadily on deforming tissue whilst conforming to pre-defined anatomical constraints. The motion tracking can also be augmented with manual control. By taking into account the physical limits in terms of maximum frequency response of the system (manifested as a delay between the input of the manipulator and the movement of the end-effector), we show the importance of visual-motor synchronisation for performing accurate smooth pursuit movements. Detailed user experiments are performed to demonstrate the practical value of the proposed control mechanism.

1 Introduction

Recent technological advances in surgery are driven by early intervention, consistent surgical outcome and accelerated patient recovery. This requires surgical procedures with improved quality, accuracy and minimally invasive access, made possible by preoperative and intra-operative imaging combined with flexible, ergonomically enhanced surgical robots [1]. Technically, the development of snake robot is motivated by the recent investigation of natural orifice or single port transluminal endoscopic surgery [2], for which path-following and dynamic shape conformance are important. Procedures that are clinically relevant to such devices include, for example, drug delivery systems for embryonic stem cell transplant into scarred myocardium, minimally invasive bilateral pulmonary vein isolation, and transmural epicardial ablation. The increased flexibility of these hyper-redundant robots also imposes significant challenges on kinematic control. For surgical navigation, the relative pose of the camera at the instrument tip during articulated movement is also important to consider ensuring stable vision and the avoidance of disorientation. Further challenges include the singularity problem while solving the inverse kinematics as the resultant inverse kinematic solutions may exceed the allowable workspace or the physical constraints of the robot actuation. For *in vivo* applications, the internal organs are in constant motion, so how to ensure path-following whilst maintaining dynamic shape conformance is a significant technical challenge.

Previous work [3] has introduced a modelling scheme of dynamic active constraints with a volumetric pathway. It is able to adapt to tissue deformation, thus providing an explicit safety manipulation margin for the entire articulated device, rather than only the tip of the robot. Such a constraint is also used in this study, which has two main purposes. First, we will introduce a control scheme for stabilising the camera reference frame at the end-effector relative to the deforming tissue so that the relative pose and distance of the camera to the tissue are maintained. The search for the optimal joint configuration of the robot is formalised as a minimization problem, in which the measures of the visual stability and the conformation to the allowable spatial constraint are introduced into the objective function. Secondly, the dynamic frequency response due to physical actuation limit is also considered. We hypothesise that for manipulation of such a robot when performing smooth pursuit movement, it is important to match the kinematic response of the manipulator controlled by the operator to that of the end-effector to ensure natural, stable user interaction. Detailed quantitative performance and usability assessment was carried out on a group of subjects using the proposed control scheme in a virtual environment with fully controllable joint parameters to demonstrate the practical value of the method.

2 Methods

2.1 Kinematics of Articulated Robot under Dynamic Active Constraints

In this paper, the robot model consists of a series of rigid links connected by universal joints. Each universal joint is operated by two actuators providing two rotational DoFs along perpendicular axes. For the prescription of active constraints, it is assumed that the anatomical pathway is pre-defined through the use of pre- and intra-operative data. The spatial constraint follows the tissue surface and dynamically adapts to tissue deformation in real-time.

A 4×4 matrix ${}^{i-1}T(\alpha, \beta)$ describing the homogenous transformation between the frame of link i and the frame of link $i-1$ can be expressed in terms of two angular values α and β of the universal joint actuating link i . The transformation between the world coordinate system W and the link l (with L links in total) is expressed in Eq. (1), where wT is a constant matrix describing the pose of the first link with respect to the world reference frame.

$${}^wT(\mathbf{q}_l) = {}^wT \left\{ \prod_{i=1}^l {}^{i-1}T(q_{2i-1}, q_{2i}) \right\} : l = 1, \dots, L \quad (1)$$

The vector $\mathbf{q}_l = [q_1, \dots, q_{2l}]$ represents the values of the joint angles affecting the pose of the link l . The pose of the end-effector on the distal link (i.e. the camera) is therefore determined by the set of angular joint values $\mathbf{q} = [q_1, \dots, q_{2L}]$.

2.2 Optimal Robot Configuration for Visual Stabilization

For visual stabilisation and optimal joint control, the objective function used measures how well the joint configuration \mathbf{q} tracks and visualizes a moving target. This includes the following two sub-objective functions.

Sub-objective 1: To minimize the collision depth to the constraint pathway enclosing the robot, *i.e.*,

$$ObjV_1(\mathbf{q}) = w_1 \sum_{i=1}^L \left\{ \exp(k_i \cdot p_i(\mathbf{q})) - 1 \right\} : 0 < k_i \leq k_{i-1} \quad (2)$$

To simplify proximity queries, articulated devices are usually modelled as a series of primitives such as cylinders and spheres to represent the links and rotational joints [4]. However, to compute the deviation outside of the constraint, the robot is represented by a set of surface vertices. Assuming the vertex coordinates of link l are $\mathbf{x}_{j=1\dots V}^l$, $\mathbf{x} = (x, y, z, 1)^T$, the deviation of the robot link l due to joint configuration \mathbf{q} is calculated as:

$$p_l(\mathbf{q}) = \max_j \left\{ I \left({}^w T(\mathbf{q}) \mathbf{x}_{j=1\dots V}^l \right) \right\} \quad (3)$$

where $I(\mathbf{x})$ in Eq. (3) is an implicit function of a 3D Cartesian point and its iso-value is the shortest Euclidean distance from the input coordinate to the constraint surface $\partial\Omega_t$ varying with time t , where Ω_t denotes the forbidden region outside of the constraint zone. If a point is within the pathway, the output value is zero, as defined in Eq. (4).

$$I(\mathbf{x}) = \begin{cases} \min \left\| \mathbf{x} - \mathbf{x}_{\partial\Omega_t} \right\| : \mathbf{x}_{\partial\Omega_t} \in \partial\Omega_t, & \text{if } \mathbf{x} \in \Omega_t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The objective function in Eq. (2) weighted by w_1 is devised as a sum of exponential functions mapped from the deviations caused by each robot link. The exponential constants $k_i \leq k_{i-1}$ are set so that the constraint is reinforced on proximal links closer to the base of the robot. This ensures that the major portion of the robot adapts to the deformation without deviating outside of the pathway during the motion tracking.

Sub-objective 2: To maintain an optimal viewing angle at a given distance to the target, *i.e.*,

$$ObjV_2(\mathbf{q}) = \frac{w_2}{\pi} \cos^{-1} \left\{ -\hat{\mathbf{d}}_{tar} \cdot \mathbf{n}_{tar} \right\} \quad (5)$$

The joint configuration is updated frame-by-frame so as to keep steady visualization of the moving target by compensating for motion disparity. Large motion disparity relative to the camera coordinate frame causes residual images which often adversely affect the surgical performance. It can be reduced by aligning the relative coordinate frame along the tissue surface normal. We assume that the motion of the visual target $\mathbf{x}_{tar}(t)$ and its surface normal $\mathbf{n}_{tar}(t)$ are estimated accurately by online tissue tracking (e.g. [5]). By aligning the camera coordinate frame to the robot end-effector, we can express the vector \mathbf{d}_{tar} pointing from the camera to the target, and the camera viewing direction \mathbf{d}_{cam} as follows:

$$\mathbf{d}_{tar} = \mathbf{x}_{tar} - {}^w T(\mathbf{q})[0, 0, 0, 1]^T, \quad \mathbf{d}_{cam} = {}^w R(\mathbf{q})[0, 0, 1, 1]^T \quad (6)$$

where ${}^w R(\mathbf{q})$ is also a homogenous matrix given by the rotation elements of ${}^w T(\mathbf{q})$. To eliminate viewing disparity between time frames, the problem can be reformulated so as to satisfy the following non-linear constraints:

$$d_{diff} = \left(\|\mathbf{d}_{tar}\| - d_c \right)^2 \leq \Delta d^2 \quad \text{and} \quad \theta_v = \cos^{-1} \left\{ \hat{\mathbf{d}}_{tar} \cdot \mathbf{d}_{cam} \right\} \leq \Delta \theta \quad (7)$$

where d_c is the optimal distance, Δd and $\Delta\theta$ are respectively the tolerances of the viewing distance and angle to the target and the caret \wedge indicates the vector is in unit length. By constraining the target movement on the image plane, we devise the cost function as in Eq. (5) such that the search converges to the configuration in which the visual direction is in parallel to the surface normal \mathbf{n}_{tar} of the target.

2.3 Joint Space Motion Planning

To avoid local minima and enable the solution to converge towards an optimum, the initial configuration \mathbf{q}^0 of joint space motion is determined based on the profile of the constraint pathway. We assume that the target moves periodically, s.t. $\mathbf{x}_{tar}(t) = \mathbf{x}_{tar}(t+T)$, and there exists a time moment t_0 at which the target is temporarily at rest s.t. $\dot{\mathbf{x}}_{tar}(t_0) = 0$. Having set the robot initially in a straight configuration $\mathbf{q} = \bar{0}$ with the camera pointing at target $\mathbf{x}_{tar}(t_0)$ at distance d_c , each joint value $\mathbf{q}^0 = [q_1^0, \dots, q_{2L}^0]^T$ is then computed in descending order from joint L to 1 by imposing the condition of minimal deviation outside of the constraint. The solution \mathbf{q}^0 is within a physically allowed range of angular displacements $[\mathbf{q}_{lb}, \mathbf{q}_{ub}]$.

In addition to the optimization constraints introduced in Eq. (7), the physical limit of robotic actuation is also addressed by bounding the joint velocities $[\dot{\mathbf{q}}_{lb}, \dot{\mathbf{q}}_{ub}]$ according to the maximum motor speed and acceleration. Assuming a very short time interval Δt between frames, the velocity limit can be converted to an angular displacement range referring to the \mathbf{q}^{i-1} optimised at the previous frame $i-1$. Thus, the overall constraint is expressed as:

$$\begin{cases} \mathbf{q}_{lb} \leq \mathbf{q}^i \leq \mathbf{q}_{ub} \\ (\Delta t \cdot \dot{\mathbf{q}}_{lb} + \mathbf{q}^{i-1}) \leq \mathbf{q}^i \leq (\Delta t \cdot \dot{\mathbf{q}}_{ub} + \mathbf{q}^{i-1}) \end{cases} \quad (8)$$

Once the optimal configuration of $\mathbf{q}(t)$ is obtained frame-by-frame and interpolated to obtain a periodic robot motion, the motion is validated for implementation under the physical torque allowance τ_{max} of each joint. Given the motion in terms of $[\mathbf{q}(t), \dot{\mathbf{q}}(t), \ddot{\mathbf{q}}(t)]$ and the mass of each link, the required dynamic torque can be computed over the entire motion by using an iterative Newton-Euler dynamics algorithm [6]. In case of unfeasible motion, the corresponding configurations are re-optimised under a stricter velocity constraint $[\dot{\mathbf{q}}'_{lb}, \dot{\mathbf{q}}'_{ub}]$. The valid motion $\tilde{\mathbf{q}}(t)$ is then generated by combining all the feasible optimal configurations.

2.4 Human-Robot Interaction

To facilitate area, rather than single feature exploration, manual control is provided for manipulating the robot while dynamically tracking the tissue motion. Having computed the robot tracking motion for points surrounding the target, each motion can be parameterised by Hermite interpolation using only two parameters (u, v) . The parameterised motion function becomes $\mathbf{Q}(u, v, t)$, s.t. $\mathbf{Q}(0, 0, t) = {}^1\tilde{\mathbf{q}}(t)$, $\mathbf{Q}(1, 0, t) = {}^2\tilde{\mathbf{q}}(t)$, $\mathbf{Q}(0, 1, t) = {}^3\tilde{\mathbf{q}}(t)$ and $\mathbf{Q}(1, 1, t) = {}^4\tilde{\mathbf{q}}(t)$ by assuming a four-point neighbourhood, where $u, v \in [0, 1]$ and ${}^i\tilde{\mathbf{q}}(t): i = 1..4$ are the optimised motions generated by tracking the neighbouring targets. After parameterization, the resultant joint

velocity $\dot{\mathbf{q}}$ not only depends on the tracking motion itself, but also on the varying rate of the two parameters, (\dot{u}, \dot{v})

$$\dot{\mathbf{q}} = \dot{\mathbf{q}}_{uv}(\dot{u}, \dot{v}, t) + \frac{\partial \mathbf{Q}(u, v, t)}{\partial t} : \dot{\mathbf{q}}_{uv}(\dot{u}, \dot{v}, t) = \frac{\partial \mathbf{Q}(u, v, t)}{\partial u} \dot{u} + \frac{\partial \mathbf{Q}(u, v, t)}{\partial v} \dot{v} \quad (9)$$

2-DoF operator control is obtained by manual variation of the parameters using a haptic device for exploring the surrounding region of the target. In this case, it is necessary to provide a damping force to hinder the excessive robot dynamics generated by manual control. A threshold $\Delta\dot{q}$ is set to limit the joint velocities $\dot{\mathbf{q}}_{uv}(\dot{u}, \dot{v}, t)$ due to the manual control and added on the existing tracking motion. This can be considered as a safety margin for defining the damping coefficient as:

$$f_b = \frac{c}{b(\dot{u}, \dot{v}, t)} : b(\dot{u}, \dot{v}, t) = \min_i \{ \Delta\dot{q} - |\dot{q}_i(\dot{u}, \dot{v}, t)| \}, \quad i = 1 \dots 2L \quad (10)$$

where b is the minimal difference between $\Delta\dot{q}$ and $|\dot{q}_i|$ among all the joints. When one of the joint velocities is reaching the safety margin, stronger force is applied to stop the hand motion before $b \rightarrow 0$. Because of the physical limitations of the haptic device an elastic force is implemented instead, but the avoidance of excessive robot dynamics is still ensured.

3 Experiments and Results

The proposed tracking method is generic for snake type robots. To evaluate its performance under realistic conditions, epicardial ablation for modified MAZE procedure is considered. The target motion with sequence of 50 frames is first extracted from a 4D model reconstructed by the CT data. The target motion relative to the static camera coordinate frame is depicted in Fig. 1(a), which is highly correlated to the cardiac motion at 1Hz and the peak-to-peak displacement is about 6.1mm. The simulated 5-link ($L=5$) articulated robot features 10 rotational DoFs limited by $[q_{i,lb}, q_{i,ub}] = [-40^\circ, 40^\circ]$, $[\dot{q}_{i,lb}, \dot{q}_{i,ub}] = [-25^\circ, 25^\circ]s^{-1}$ with $i = 1, \dots, 2L$ and $\tau_{\max} = 16.6$ mNm. The mass, radius and length of each link are 33g, 6mm and 20mm, respectively. Manipulating the robot not exceeding the dynamic active constraints is always considered a higher priority during the minimization, s.t. $w_1 > w_2$. The processing time required mainly depends on the deviation distance calculation in Eq. (2). The computational complexity is approximately proportional to the number of vertices in the model. The minimization was performed off-line and preoperatively (we used MatLab on a PC with Intel Core2 Duo CPU T7100 and 2GB RAM). The time for proximity queries on 767 vertices of a 5-link robot is below 15ms.

The robot motion aiming to track the moving target is planned smoothly without exceeding the actuation limits, and the quantitative performances of the visual stabilization are shown in Fig. 1(b-e). Compared to the stationary robot without tracking enabled, the viewing angle relative to the target surface normal is slightly reduced to ensure a perpendicular view of the surface (Fig. 1(b)). However, the parameters characterizing the stability of the visual target along the camera are the viewing distance $\|\mathbf{d}_{tar}\|$ and angular accuracy θ_v (Eq. (7)) as shown in Fig. 1(c-d).

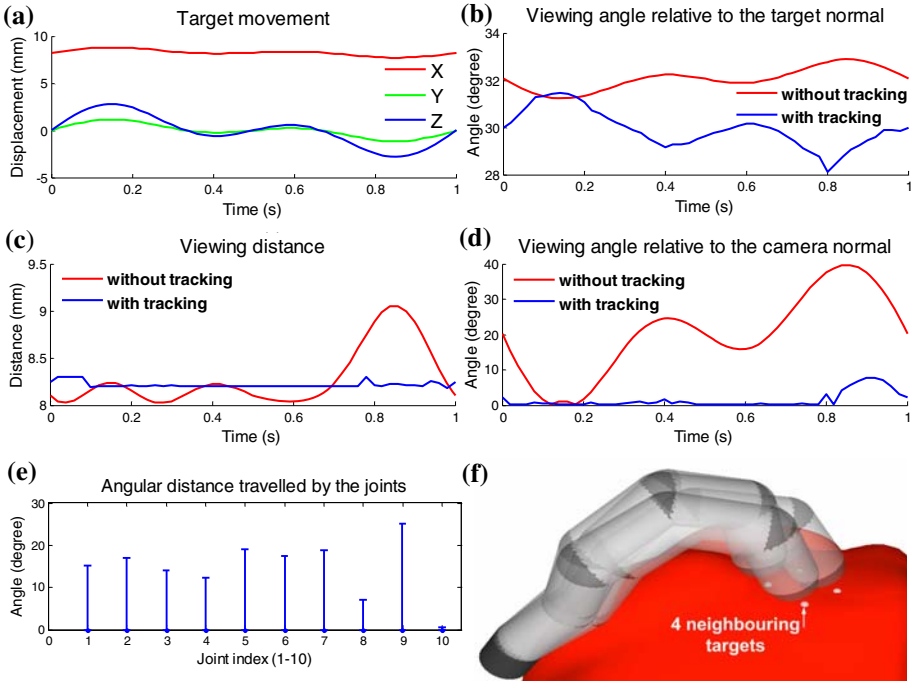


Fig. 1. (a) 3D time-varying position of the visual target; (b) viewing angle between camera direction and target surface normal; (c) viewing distance between the camera and the target; (d) angle reflecting how accurately the camera is pointing at the target during motion tracking; (e) angular travelled distance of each joint; (f) robot configurations aiming at different neighbouring targets

Compared to existing methods (e.g. [7]), the proposed approach eliminates the need of expressing the inverse kinematics of the end-effector in a closed-form, ensuring that a solution is always obtainable without having to deal with the presence of reduced-rank Jacobian singularities. Our method aims to determine an optimal configuration (i.e. minimum deviation from constraint and stable visualization) among a wide range of feasible solutions without having to specify the pose of the end-effector. As shown in the results, the camera can be stabilised at a predefined distance to the target $d_c = 8.2$ mm. Also, the angular deviation can be highly reduced below 8° , ensuring accurate target visualization at the centre of the image by compensating any rapid movement. The improved stability is attributed to the satisfaction of the nonlinear constraints in Eq. (7). The advantage of using an articulated robot with high kinematic redundancy is to provide adequate flexibility to satisfy the strict criteria in the complex nonlinear constraints (i.e. $\Delta d = 0.1$ mm and $\Delta\theta = 8^\circ$) under the actuation limits. Greater tolerances might be needed in case of low angular resolution of the joint actuator. Fig. 1(e) shows the angular travelled distance of the joints, demonstrating the movement burden is shared between joints. This ensures the tracking capability can be resumed by re-planning the joint motion in case of limited performance, such as motion

lock of a particular joint due to malfunction. In addition, singularities due to solutions with large joint angle differences between two consecutive timeframes are also avoided.

Once the joint motion is planned for tracking the neighbouring targets (Fig. 1(f)), visual exploration of the surrounding region can be performed by the operator. Curved reference paths and spherical target objects conforming to the deformation are pre-defined on the tissue surface within the region. With the virtual camera mounted on the simulated end-effector, the operator can see those objects being stabilised visually within the central part of the screen. The performance of the visual exploration can be recorded during the task.

Table 1. Measured performance indices averaged across the ten subjects studied

	Feature targeting					Path following				
	W/O haptic		With haptic		Imp%	W/O haptic		With haptic		Imp%
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Compleat. Time (Sec)	135.3	15.9	133.6	14.9	1.3	110.1	78.5	65.6	15.5	40.4
Dis. Travelled (mm)	937.4	92.0	925.2	72.1	1.3	668.8	462.0	416.3	91.1	37.8
Deviation (deg)	12.8	2.9	10.4	1.7	18.3	9.7	6.1	5.8	1.0	40.4

Ten subjects were recruited to assess the performance of the method, for which haptic feedback is provided to limit the hand manipulation so that its movement is in-sync with the frequency response of the end-effector. Subjects were asked to operate the haptic device (Omni Phantom, SensAble Tech. Inc., USA) for targeting multiple spherical targets in sequence or on a curved path. This allows for a comparison of the performance for either individual feature targeting or smooth pursuit for path following. Each task was performed twice, with and without the haptic force for maintaining the manual control within the safety velocity margin. The task order was randomized to minimize the effect of learning. Subjects were allowed to familiarise with the control interface before performing the tasks. Note that all the tests were held under motion tracking and stabilisation. When target tracking is disabled, the operator can fully utilise the actuation and have fast-response control in varying (u, v) . However, not only the deformation is rapid, but also the variation of (u, v) is not designed for tissue tracking. Consequently, the operator is not able to track the target objects. A conservative velocity limit $\Delta\dot{q} = 4.5^\circ\text{s}^{-1}$ was set to constrain $\dot{\mathbf{q}}_{uv}$ (Eq. (9)) for preventing the joints from exceeding their angular velocity limits while performing tissue tracking. Without haptic feedback, the subject introduces fast and wide movements, but a severe delay is imposed depending on the physical limit of the end-effector. Such a delay is common (and important to consider) in human-robot interaction especially when the user pushes the robot to its performance limit. The performance indices recorded in this study include task completion time, the total displacement of the robot end-effector and the path/target deviation, computed as the average of the total angular deviation between the camera viewing direction of the target or path and the optimal angle θ_v deduced from Eq. (7). The mean and standard deviation of these indices measured with and without the use of haptic force for the tasks of feature targeting and path following are summarized in Table 1. The results clearly show a significant

improvement in performing the path following task when visual-motor perception is synchronized. This implies that smooth pursuit movement requires matching of the kinematic response of the robot end-effector to the manipulation input of the operator.

4 Discussion and Conclusions

In this paper, we have proposed a control method to stabilise the visualisation of a deforming tissue using an articulated robotic device. Results have shown that visual stabilisation is particularly important when using a flexible redundant robot since the pose of the camera at the distal tip is affected by the joint configuration. To avoid disorientation, it is necessary to ensure that the area of interest is always within the field-of-view of the camera and at a stable relative viewing angle. Residual motion of the image due to tissue deformation can be minimised by aligning the direction of camera along the surface normal. The experimental results of the proposed method demonstrate the capability of a redundantly articulated robot for compensating dynamic tissue deformation. Further investigation of the proposed method is also conducted when integrating manual control with haptic interaction. In this case, the dynamic frequency response of the robot is affected by physical actuation limits. Such a delay causes a mismatching between the visual feedback and the motion input of the operator and we have shown that for pursuit movement it can significantly influence the task performance. A damping haptic force is therefore used to matching the frequency response of the user-controlled manipulator with that of end-effector. Such knowledge should be of importance to the actual deployment of snake robots for robotic-assisted transluminal endoscopic procedures.

References

1. Taylor, R.H., Stoianovici, D.: Medical robotics in computer-integrated surgery. *IEEE Transactions on Robotics and Automation* 19, 765–781 (2003)
2. Swanstrom, L.L., Whiteford, M., Khajanchee, Y.: Developing essential tools to enable trans-gastric surgery. *Surg. Endosc.* 22, 600–604 (2008)
3. Kwok, K.W., Mylonas, G.P., Sun, L.W., Lerotic, M., Clark, J., Athanasiou, T., Darzi, A., Yang, G.Z.: Dynamic Active Constraints for Hyper-Redundant Flexible Robot. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5761, pp. 410–417. Springer, Heidelberg (2009)
4. Li, M., Ishii, M., Taylor, R.H.: Spatial motion constraints using virtual fixtures generated by anatomy. *IEEE Trans. Robotics* 23, 4–19 (2007)
5. 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)
6. Luca, A.D., Ferrajoli, L.: A Modified Newton-Euler Method for Dynamic Computations in Robot Fault Detection and Control. In: *IEEE International Conference on Robotics and Automation*, pp. 3359–3364 (2009)
7. Fahimi, F., Ashrafiuon, H., Nataraj, C.: An Improved Inverse Kinematic and Velocity Solution for Spatial Hyper-Redundant Robots. *IEEE Trans. on Robotics and Automation* 18, 103–107 (2002)

Estimating Radiation Exposure in Interventional Environments

Alexander Ladikos¹, Cedric Cagniard¹, Reza Ghotbi²,
Maximilian Reiser³, and Nassir Navab¹

¹ Chair for Computer Aided Medical Procedures, Technische Universität München

² Center for Vascular and Endovascular Surgery, Klinikum Pasing, München

³ Department of Clinical Radiology, Ludwig-Maximilians-Universität München

Abstract. In the last decade the use of interventional X-ray imaging, especially for fluoroscopy-guided procedures, has increased dramatically. Due to this the radiation exposure of the medical staff has also increased. Although radiation protection measures such as lead vests are used there are still unprotected regions, most notably the hands and the head. Over time these regions can receive significant amounts of radiation. In this paper we propose a system for approximating the radiation exposure of a physician during surgery. The goal is to sensitize physicians to their radiation exposure and to give them a tool to quickly check it. To this end we use a real-time 3D reconstruction system which builds a 3D-representation of all the objects in the room. The reconstructed 3D-representation of the physician is then tracked over time and at each time step in which the X-Ray source is used the radiation received by each body part is accumulated. To simulate the radiation we use a physics-based simulation package. The physician can review his radiation exposure after the intervention and use the collected radiation information over a longer time period in order to minimize his radiation exposure by adjusting his positioning relative to the X-ray source. The system can also be used as an awareness tool for less experienced physicians.

1 Introduction

In contrast to even 10 years ago the use of interventional X-ray imaging and therefore the dosage received by the physician has increased dramatically, especially for fluoroscopy-guided procedures. In fluoroscopy the X-ray source may be turned on for more than 30-60 minutes [10] during which the physician is standing very close to the patient and therefore the X-ray source. Due to the increase in the use of fluoroscopy-guided procedures, interventional radiologists have nowadays a much greater risk of radiation damage. Although the dose received in a single intervention is not large, it is well known, that long-term radiation exposure even at low doses can lead to negative effects on the body, which in the extreme can lead to cell mutations and cancer. The physician is usually protected by a lead vest and sometimes a lead collar around his neck. However, this leaves some body parts still exposed to the radiation, most notably

the head, the arms and the hands. Especially the hands are at danger since they are used to work on the patient and therefore closest to the radiation source [10,15]. To control the radiation received a dosimeter is typically worn under the lead-vest. However, since the dosimeter is worn under the protective vest, its readings are not representative of the radiation received by the unprotected body parts. Moreover since it is placed at the level of the chest it is further away from the X-ray source than for instance the hands and therefore only of limited use. We therefore want to present a solution to physicians, especially the less experienced ones, which allows them to quickly gauge their radiation exposure during an intervention and over the course of a longer time frame, in order to sensitize them for the dangers associated with X-ray radiation and the options of limiting their exposure. To this end we propose a real-time 3D-reconstruction system combined with a geometric tracking algorithm which allows us to reconstruct and track the physician in the interventional room during the procedure and - with the knowledge about the position and movement of the X-ray source - to approximate his radiation exposure.

Contributions. The contributions of this paper are twofold: On the technical side we present a powerful methodology for tracking the physician from his real-time 3D-reconstruction and for modeling the scatter radiation created by the X-ray source. The tracking is a non-trivial problem since the physician has to be tracked based only on his geometry and correspondences have to be established between corresponding body parts in different reconstructions. The simulation of the scatter radiation is done using a physics framework for simulating particle physics [5]. In the simulation we model the C-arm and the part of the patient irradiated considering physical interaction effects such as Compton Scattering, Rayleigh Scattering and the Photoelectric Effect. The simulation results are used to determine the radiation deposited in an object at a given position relative to the C-arm, such as the physician.

On the medical side we show a concept for a system which sensitizes the physician to his radiation exposure and which helps him to take more informed decisions when using X-ray devices to minimize his risk for radiation related dangers. The system displays a color-coded map of the physician after each intervention which shows his radiation exposure during the procedure and if desired the accumulated radiation for a given time frame. This allows the physician to keep track of his long-term radiation exposure. The system can also be used for making inexperienced physicians more aware of the dangers of radiation exposure.

2 Related Work

There have been many medical studies concerning the radiation exposure of patients and surgeons during different interventions [13,11,5,6]. In [15] for instance the radiation exposure of the patient and the surgeon is modeled mathematically. However, only estimates about the average distance of the surgeon to X-ray source are used. Many studies conclude that the radiation exposure of the

physician during fluoroscopy-guided interventions is very high in the area of the hands [10] and in the lower extremities [13]. This has sparked some discussion about the possibility of protecting the surgeon and making him more aware of the dose he has already received. Our system attacks at this point since it allows the physician to check his exposure easily in an everyday setting.

There has also been work on simulating the radiation exposure for training novel physicians [16,2]. In this work the C-arm was modeled in a simulation and the radiation dose was measured at several spherical detectors around the C-arm. However, the position of the physician was not taken into account and the system was not meant for use in interventional rooms but for training. We are not aware of any work which tracks the physician during the intervention and which accumulates his radiation exposure over the course of an intervention.

On the technical side, which builds the foundation of our system, there has been a lot of work which was mainly published in the computer vision literature. For the real-time 3D-reconstruction of the scene we use the visual hull algorithm described in [7]. The visual hull [9] is an approximation of the true shape of an object which is computed from silhouette images by intersecting the silhouette-backprojections. The recovered shape is the shape maximally consistent with the input silhouette images.

The tracking stage of our system is responsible for establishing dense correspondences between the independent 3D-reconstructions of the scene computed at each time frame. As we are interested in the movement of the physician, our work relates directly to the vast body of literature addressing markerless human motion-capture in multi-camera environments. Most of these vision-based tracking algorithms use kinematic models in the form of a skeleton as discussed in the survey by Moeslund et al. [11]. Such skeletal models allow to effectively constrain the problem and reduce its dimensionality but they usually require manual initialization of the joint positions, as well as lengthy parametrization of the tracked person's body characteristics from generic mesh models. In contrast to these approaches, we chose to use the framework recently presented by [3]. This algorithm doesn't rely on strong priors on the nature of the tracked object and instead deforms the first reconstructed mesh to fit the geometry in the rest of the sequence.

3 Reconstruction and Tracking System

We use a multi-camera system [8] to reconstruct the shape of the objects inside the interventional room in real-time. The system consists of 16 optical cameras mounted on the ceiling as well as a few PCs to perform the reconstruction and the radiation modeling (see Figure 1). In an offline phase, the initial position of the C-arm inside the reconstruction volume is determined and the cameras are calibrated. In addition, background images of the room are acquired. In the online phase, the background images are used to segment the foreground objects. The segmented images are subsequently used to perform the 3D-reconstruction by intersecting the backprojections of the object silhouettes in space. The reconstruction of the physician is then tracked using a mesh deformation framework.



Fig. 1. View of our reconstruction studio. The cameras are marked by red circles.

In a final step, the scatter radiation created by the C-arm is modeled and accumulated using the reconstruction of the physician. Once the system has been set up, very little maintenance work is required. In the following sections we will describe each component of the system in more detail.

3.1 System Architecture

We use a distributed system for performing the 3D-reconstruction. A total of four PCs with four cameras each is connected through a Gigabit-Ethernet network. The PCs acquire the camera images, perform the background subtraction using the method presented in [4] and then perform a partial 3D-reconstruction using their locally available silhouette images [7]. The results are sent to a master PC which combines the partial reconstructions and performs the radiation accumulation. The results are then visualized.

We decided to place the cameras on the ceiling so that they do not take away any workspace from the operating room and to protect them from accidental collisions with other objects, which would require a recalibration of the system. The cameras are synchronized in order to perform a consistent reconstruction.

3.2 Calibration

Before the system can be used the cameras have to be calibrated extrinsically and intrinsically. To this end we use a calibration toolbox designed for multi-camera environments [14]. It is based on a calibration point which is moved around in the room and detected in the cameras. The method is easy to use and delivers an accurate calibration. Typical calibration times are in the range of 15 minutes. The recovered camera coordinate system is registered to the room coordinate system using target points at known positions in the room.

3.3 Segmentation

It is necessary to extract the foreground objects in the camera images in order to perform the 3D-reconstruction. This is done using a robust background subtraction method [4]. Reference images are acquired before the intervention

and are used during runtime in order to detect the changed parts in the scene. This results in the silhouette images which are the input for the reconstruction. In the case of static foreground objects, such as a surgical table, problems can arise, because these objects will occlude other objects which in turn cannot be fully reconstructed since they are not fully seen in the silhouette images. We solve this problem by masking out static objects in the silhouette images, effectively declaring them to foreground. This automatically adds them to the reconstruction.

3.4 Reconstruction

The silhouette images are used in a GPU-based visual hull computation method [7]. First the reconstruction volume is discretized into voxels. Each voxel is then checked for its occupancy by projecting it into all available silhouette images. If it is empty in at least one of the silhouette images it is marked as empty, otherwise it is marked as occupied. In order to speed this up an octree model of the reconstruction volume is used and the computations are streamlined on the GPU.

3.5 Tracking

The result of the 3D-reconstruction stage is a sequence of meshes independently reconstructed at each time frame. We use the non-rigid mesh registration framework presented by [3] to iteratively deform a reference mesh to fit the geometry in the rest of the sequence. The reference mesh is taken from the first frame of the sequence. We prefer this tracking method to skeletal-based alternatives as it does not need an initialization of the pose and of the body characteristics of the tracked person. The only requirement is that the tracked physician must be clearly distinct from the rest of the reconstructed geometry in the reference frame. The output of the tracking algorithm is a dense trajectory for each vertex of the reference mesh across the sequence, which is necessary to accumulate the radiation exposure over time.

4 Radiation Modeling

We model the scatter radiation created by the interaction of the X-rays with the patient's body using the particle-physics simulation system GEANT4 [5]. Using this system it is possible to model the different physical effects, which lead to the production of scatter radiation, namely Compton Scattering, Rayleigh Scattering and the Photoelectric Effect. We model the scene similar to [16]. In that work the authors also use the GEANT4 system to model X-ray scatter radiation. We model the C-arm as an intensifier tube and a detector both made out of iron. This has the effect that radiation is blocked above the detector and below the source. The patient is modeled as a cylinder containing water. As has been shown in [16,2] the model, although simplified from reality, still provides

reasonable simulation results which make it useful for educating physicians about their radiation exposure. After having modeled the C-arm and the patient we create a X-ray beam originating from the radiation source and direct it in a narrow cone at the patient body. In order to obtain the energy deposited by the X-rays in the environment we place a so-called detector sphere around the scene in our simulation and register the energy of the particles which fall onto it. By placing several such spheres with different radii and all centered on the source of the scatter radiation, i.e. the patient, we can estimate the distribution of the radiation inside the room and thereby the radiation received by the physician. We precompute an X-ray irradiation volume using the settings of the C-arm and compose it with the tracked mesh in the previous step. This allows us to accumulate the radiation received by each vertex and by interpolation the radiation received by the whole mesh.

5 Results

To validate our system we recorded an interventional scenario in our lab consisting of a C-arm, a patient and a physician. The physician is moving around

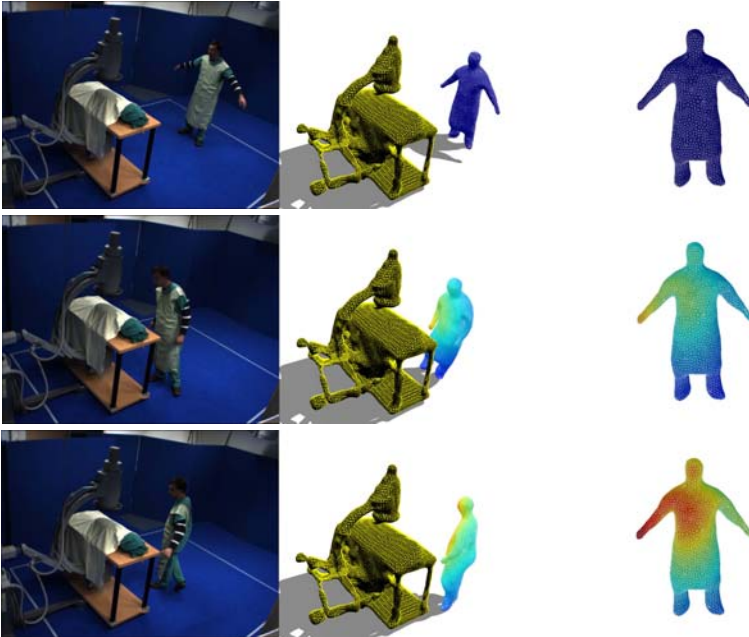


Fig. 2. Results on a sequence recorded in our lab. The three rows show the radiation exposure at the beginning, the middle and the end of the intervention. The first column shows one of the input images, the second column shows the tracked 3D-scene and the final column shows the physician in his reference pose with the color-coded radiation exposure. Red represents a high radiation exposure while blue represents a low radiation exposure.

the C-arm while the C-arm is constantly radiating. This is a typical scenario for a fluoroscopy-guided intervention, since in fluoroscopy the C-arm is also almost constantly radiating. The frame rate at which the reconstruction system is running is 20 fps. Once we obtain the reconstruction of the physician, we start the tracking algorithm. The physician is then tracked during the entire sequence. With the knowledge of the position of the X-ray source and the patient we can compute the source of the scatter radiation which is subsequently computed using the radiation simulation framework. Since we have the tracked mesh of the surgeon with corresponding vertices over time, we can simply add up the radiation at each vertex position for each frame. By performing this addition over the whole sequence we obtain the final radiation dose collected by the physician for each vertex of his reconstruction. These values are then interpolated to obtain the radiation on the whole mesh. This is reasonable since the mesh consists of only small triangles. Finally, we visualize the accumulated radiation using a heat map (see Figure 2 and the videos included as additional material). The scaling on the heat map is set so that the maximum amount of radiation received by the physician is marked as bright red. This makes it easier to visually gauge the exposure. It can be seen, as also observed in [15], that the hands receive most of the radiation.

6 Discussion

We understand our system as a proof-of-concept. Our goal was not to develop a system which can be directly used in an operating room, but to show what problems have to be addressed, which methods are available to solve them and how they can be combined in a sensible way. In particular the exposure computed by our system needs to be validated by experimental measurements and the dynamic and static environment of the interventional room needs to be taken into account in more detail. The use of our system is not limited to the estimation of the radiation exposure. The availability of a real-time 3D-reconstruction of the interventional environment can also be used for workflow analysis [12], collision detection [8] and documentation purposes.

7 Conclusion

We presented a system for modeling the radiation received by a physician during an intervention. Our system builds on a real-time 3D-reconstruction of the interventional room which is used for tracking the 3D-mesh of the physician using a mesh deformation framework. The radiation is modeled using a particle-physics simulation package. The contribution of the paper is the combination of real-time 3D-reconstruction and mesh-based tracking to help the physician estimate his radiation exposure and to allow him to collect statistics about his long-term exposure. Our system can also be used for making novice physicians more aware of their radiation exposure. Future work includes validating the radiation estimation by using sensors attached to the physician, creating profiles for each physician and bringing the system to a real operating room.

References

1. Bogaert, E., Bacher, K., Thierens, H.: A large-scale multicentre study in Belgium of dose area product values and effective doses in interventional cardiology using contemporary x-ray equipment. *Radiation Protection Dosimetry* 128(3), 312–323 (2008)
2. Bott, O., Wagner, M., Duwenkamp, C., Hellrung, N., Dressing, K.: Improving education on c-arm operation and radiation protection with a computer-based training and simulation system. *International Journal of Computer Assisted Radiology and Surgery* 4(4), 399–407 (2009)
3. Cagniard, C., Boyer, E., Ilic, S.: Probabilistic deformable surface tracking from multiple videos. In: *ECCV* (2010)
4. Fukui, S., Iwahori, Y., Itoh, H.i., Kawanaka, H., Woodham, R.: Robust background subtraction for quick illumination changes. In: Chang, L.-W., Lie, W.-N. (eds.) *PSIVT 2006*. LNCS, vol. 4319, pp. 1244–1253. Springer, Heidelberg (2006)
5. GEANT4, <http://www.cern.ch/geant4>
6. Giordano, B.D., Baumhauer, J.F., Morgan, T.L., Rehtine, G.R.: Patient and surgeon radiation exposure: Comparison of standard and mini-c-arm fluoroscopy. *Journal of Bone & Joint Surgery* 91(2), 297–304 (2009)
7. Ladikos, A., Benhimane, S., Navab, N.: Efficient visual hull computation for real-time 3d-reconstruction using cuda. In: *Proceedings of the 2008 Conference on Computer Vision and Pattern Recognition Workshops* (2008)
8. Ladikos, A., Benhimane, S., Navab, N.: Real-time 3d reconstruction for collision avoidance in interventional environments. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II*. LNCS, vol. 5242, pp. 526–534. Springer, Heidelberg (2008)
9. Laurentini, A.: The visual hull concept for silhouette-based image understanding. *PAMI* 16(2), 150–162 (1994)
10. Marx, M., Ellis, J.: Radiation protection of the hand in interventional radiology: should it fit like a glove? *Radiology* 200(1), 24–25 (1996)
11. Moeslund, T.B., Hilton, A., Krueger, V.: A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* 104(2), 90–126 (2006)
12. Padoy, N., Blum, T., Essa, I., Feussner, H., Berger, M.-O., Navab, N.: A boosted segmentation method for surgical workflow analysis. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 102–109. Springer, Heidelberg (2007)
13. Shortt, C.P., Al-Hashimi, H., Malone, L., Lee, M.J.: Staff radiation doses to the lower extremities in interventional radiology. *Cardiovascular Interventional Radiology* 30, 1206–1209 (2007)
14. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multi-camera self-calibration for virtual environments. *Presence: Teleoperators and Virtual Environments* 14(4), 407–422 (2005)
15. Tsalafoutas, I., Tsapaki, V., Kaliakmanis, A., Pneumáticos, S., Tsoronis, F., Koulentianos, E., Papachristou, G.: Estimation of radiation doses to patients and surgeons from various fluoroscopically guided orthopedic surgeries. *Radiation Protection Dosimetry* 128(1), 112–119 (2008)
16. Wagner, M., Duwenkamp, C., Dressing, K., Bott, O.J.: An approach to calculate and visualize intraoperative scattered radiation exposure. In: *MIE* (2009)

Force Adaptive Multi-spectral Imaging with an Articulated Robotic Endoscope

David P. Noonan^{1,2}, Christopher J. Payne^{2,3}, Jianzhong Shang², Vincent Sauvage^{1,2},
Richard Newton^{1,2}, Daniel Elson^{1,2}, Ara Darzi¹, and Guang-Zhong Yang^{2,3}

¹ Department of Surgery and Cancer, Imperial College London

² Institute of Biomedical Engineering, Imperial College London

³ Department of Computing, Imperial College London

{dnoonan, cjp04, j.shang, v.sauvage, r.newton}@imperial.ac.uk
{ds.elson, a.darzi, g.z.yang}@imperial.ac.uk

Abstract. Recent developments in optical spectroscopic techniques have permitted in vivo, in situ cellular and molecular sensing and imaging to allow for real-time tissue characterization, functional assessment, and intraoperative guidance. The small area sensed by these probes, however, presents unique challenges when attempting to obtain useful tissue information in-vivo due to the need to maintain constant distance or contact with the target, and tissue deformation. In practice, the effective area can be increased by translating the tip of the probe over the tissue surface and generating functional maps of the underlying tissue response. However, achieving such controlled motions under manual guidance is very difficult, particularly since the probe is typically passed down the instrument channel of a flexible endoscope. This paper describes a force adaptive multi-spectral imaging system integrated with an articulated robotic endoscope that allows a constant contact force to be maintained between the probe and the tissue as the robot tip is actuated across complex tissue profiles. Detailed phantom and ex-vivo tissue validation is provided.

Keywords: fluorescence spectroscopy, articulated robot, force control, in vivo tissue characterization.

1 Introduction

Point based spectroscopic approaches such as diffuse reflectance, Raman and fluorescence spectroscopy have been shown to be clinically useful for in vivo assessment of cancer, atherosclerosis and ischemia. Current techniques under development include visible diffuse light spectroscopy applied in the colon to detect colonic ischemia [1], the use of fluorescence spectroscopy to detect atherosclerotic lesions [2] and the use of Raman sensing to detect cervical cancer [3]. There has been increasing focus in recent years on simultaneous large area surveillance with high spectral resolution [4], (and therefore high sensitivity and specificity), potentially with multiple labels and multimodal spectroscopy. Simultaneously the continued development of targeted fluorescence probes that provide a specific spectral profile offer a further example of

the importance of being able to trace a path across the tissue surface in order to investigate the localization of the marker whilst still achieving a good detection of the marker fluorescence on top of the optical background signal.

In practice, due to the small sensing area of these techniques and the common requirement of tissue contact, performing large area surveillance and repeated examination is difficult due to in vivo tissue deformation and the difficulty in manually maintaining the probe-tissue distance. Previous results have shown that these techniques are affected by the pressure exerted between the probe and the tissue surface [5] due to the following effects: (a) the coupling efficiency between the fibres and the tissue, which depends on the refractive index mismatch between the two; (b) the change in the optical scattering properties of the tissue due to compression; and (c) the increase in signal from deeper tissue layers with increasing pressure due to the displacement of the upper layers, (which can be a particularly important consideration when sensing layered tissues such as epithelial structures).

The purpose of this paper is to demonstrate a force adaptive multi-spectral imaging system integrated with an articulated robotic endoscope. This allows a constant contact force to be maintained between the probe and the tissue as the robot tip is translated across complex tissue profiles. A similar approach for measuring contact forces remotely has been employed for both fixed transformations between the sensor and contact site [6] [7] and flexible transformations [8]. An additional technical advantage of the system described in this paper is that the estimation of distal tip position allows for localization of specific spectral responses within the field of view. Detailed validation is provided using a multi-spectral phantom as well as results demonstrating successful force adaptive acquisition of tissue autofluorescence on ex-vivo liver tissue samples.

2 Methods

2.1 Articulated Robotic Endoscope Design and Operation

The articulated robotic endoscope used in this paper features five controllable degrees of freedom arranged into three serially connected link segments and attached to a 500mm long rigid aluminum shaft. The degrees of freedom are arranged into two universal joints (intersecting pitch and yaw) at the distal and medial joints and a single yaw DoF at the proximal joint. Each DoF is independently addressable and capable of actuating $\pm 45^\circ$, enabling the device to servo its tip over a large area ($\pm 90^\circ$ vertically and $\pm 135^\circ$ horizontally). An advantage of the serially connected joint architecture is that the proximal/medial joint(s) can be used to avoid obstacles, such as anatomical structures or other instrumentation, to safely position the distal joint at a more advantageous position. As can be seen in Fig. 1(a), the distal joint alone can then continue scanning to provide the operator with both visual feedback and the ability to interact at the target site.

The diameter of the shaft and articulated segments is 12.5mm so as to be compatible with existing trocar ports. Each DoF is actuated by an embedded 4mm diameter Brushless DC micromotor (Namiki Precision Jewel Ltd, SBL04-0829) with a 337:1 planetary gearbox. The rotary motion of the motor is converted to joint rotation via an

internal mechanism (patent pending). Joint position is estimated by counting a single pulse output from the motor drive electronics for each full rotation of the motor. The length of the proximal and medial segments is 44mm while the distal segment is 36mm long. The device also features three internal channels ($2 \times 3\text{mm}$, $1 \times 1.8\text{mm}$). The latter houses power and signal lines for the motors while the two other channels can be used for various additional instrumentation. An IntroSpicioTM 115 Micro CCD video camera (Medigus Ltd, Israel) and a custom fluorescent imaging probe, which will be further described in Section 2.2, are used for this study.

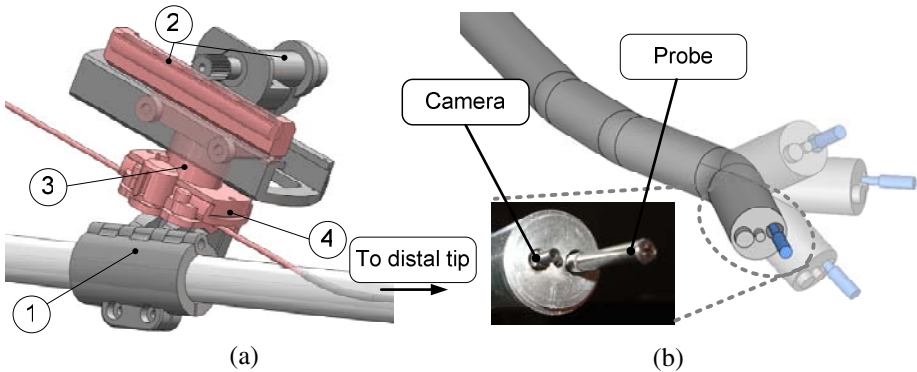


Fig. 1. (a) Schematic illustrating the force adaptive linear servoing mechanism. Moving parts are marked in red. It consists of four key components: a clamp (1) which attaches to the robot shaft, a Maxon RE10 Brushed DC motor driving a rack and pinion linear slider (2), a Nano17 F/T sensor (3) and a housing to clamp and thus translate the imaging probe (4) (b) A CAD schematic illustrating the concept of the robotic scanning device using controlled actuation of the distal tip. The 1.8mm Medigus Camera and fluorescent imaging probe are also marked.

2.2 Force Sensitive Probe Operation

One key component of the articulated robotic device is a force sensitive linear servoing mechanism, into which the proximal end of the fluorescent imaging probe is clamped. The mechanism is located approximately 500mm from the distal opening of the robot's 3mm biopsy channel at the tip of the device. An embedded ATI Nano17 6-axis Force/Torque sensor (SI-12-0.12) measures the axial load along the central axis of the imaging probe. Naturally, inaccuracies in the absolute force measurement will be introduced due to friction in the instrument channel, which will increase with the tortuosity of the robot (due to frictional interface effects along the length of the probe). However, the purpose of the system is not to provide the operator with an accurate absolute force measurement between the probe and the tissue, but rather to maintain a constant force during image acquisition. It is assumed that the force measured remotely from the tip and transmitted through the probe is sufficiently representative of the tip interaction force to drive the controller. Additionally, the probe is only expected to undergo a linear translation of less than 3–4mm during operation which significantly reduces the negative effects of the long transmission distance. These issues will be further discussed in Section 3.

Fig. 1(a) illustrates the F/T sensor clamped into the rack and pinion based linear servo mechanism. The pinion (24t, MOD 0.25) is actuated via a brushed DC geared motor (Maxon RE10 6V, 16:1 gearhead, MR Type Integrated 256cpr Quadrature Encoder). The rack is seated in a groove which constrains all but a translational degree of freedom along the centerline of the imaging probe. A rack and pinion system was chosen for back-drivability, ease-of-use and so as to provide a fast response time. The probe is fed directly into the internal biopsy channel to prevent undesirable deflections or buckling as it is loaded axially.

A closed-loop force control scheme was created in the *LabView* development environment. The force along the centerline of the probe is measured by the force-torque sensor and acquired using a NI-6221 acquisition card at 500Hz. This signal is then fed to a Proportional-Integral controller and the output signal is processed by a differential operational amplifier (so as to allow bipolar control of the motor) which in turn forms the input of a pulse-width modulation-based speed controller for the DC motor (Maxon, LSC 30/2 linear 4Q Servoamplifier). The speed controller has an internal feedback loop which uses the integrated quadrature encoder to regulate the motor speed according to the input signal described above.

Fluorescence emission measurements were taken using a 2 mm diameter flexible fibre-optic probe (Romack, V.A., USA) consisting of 7 hexagonally assembled fibres, each having a 200 μ m core diameter and a 0.22 numerical aperture. Excitation light was provided by a 375nm laser diode collimated with a 4.6mm focal length lens and focused into a central excitation fibre with an 8mm focal length lens. The fluorescence emission light collected by the six fibers surrounding the central fibre was arranged into a linear array which was imaged onto the 80 μ m wide slit of a spectrograph (Specim Ltd, Finland). A 405nm long-pass filter was inserted in to block the reflected excitation light and the light was dispersed by a prism-grating-prism element of the spectrograph and detected with a cooled CCD camera (Retiga EXI, QImaging, 1392 \times 1040 pixels). The camera and the beam shutter were also controlled through *LabView* (exposure time phantom, 1ms; exposure time liver, 10ms; region of interest 1392 \times 110) and the autofluorescence spectrum was acquired in a single shot with a frame-rate of 130ms. Image acquisition from the onboard Medigus camera is also synchronized with the autofluorescence signal acquisition, which allows the immediate localization of the probed site within the camera field of view.

3 Experimental Design and Results

Three experiments were performed in order to access the efficacy of system. The first was a system characterization to evaluate what contact force should be used, the second employed a fluorescent stained phantom to validate the ability of the system to accurately reconstruct a multi-spectral map and finally, the third experiment was performed on ex-vivo fowl liver to demonstrate the ability of the system to maintain sufficient contact with the tissue to acquire autofluorescence spectra. The first set of experiments were performed to characterize: a) the effect of different angular displacements of the distal tip on the minimum force required to advance the imaging probe; and b) the effect of different contact forces on the intensity of the resulting spectroscopy image.

To characterize the effect of different angular displacements, the distal tip was servoed to five different positions (straight and locations in each of the four quadrants of the Cartesian co-ordinate system). At each position, the setpoint of the force controller was increased in increments of 10mN until the probe advanced and made contact with a fluorescein stained silicone phantom. The results show that larger forces are required to move the imaging probe when the tip is deflected than when it is straight (20mN straight, 30 – 40mN at deflections of $\pm 12^\circ$ on each axis of the universal joint). This is due to friction interference effects between the imaging probe and the channel inside the robot as it deflects. All subsequent experiments were carried out with a minimum setpoint of 40mN and for all experiments described in this paper, only the distal tip of the robot was actuated.

Further experiments were performed to characterize the spectral response of the probe for different contact forces with ex-vivo fowl liver tissue (Fig. 2). The probe was advanced into contact with the tissue and the contact force was increased in 10mN increments up to a limit of 150mN as spectroscopic images were acquired. The results indicated that while the magnitude of the tissue autofluorescence remains constant for all contact forces, the effect of ambient light on the spectral trace decreases in proportion with increasing contact force. The presence of this signal is due to diffusion of the illumination light through the tissue to the detection fibers even though the probe is in full contact with the tissue. Based on feedback from the on-board camera, it can be seen that as the tissue undergoes higher deformation due to high contact forces, less ambient light is detected. However, these same high contact forces do not affect the intensity of the autofluorescence signal from the tissue. These results indicate that with this system we can maintain a low but constant force that achieves a sufficient optical signal without applying excessive load to the tissue.

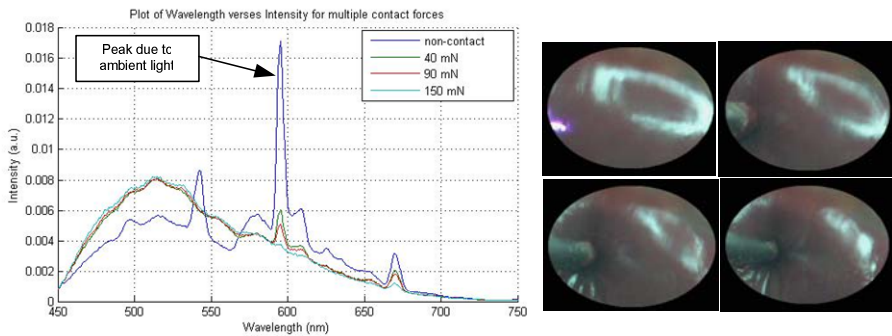


Fig. 2. Plot of Intensity versus Wavelength for contact forces of 0mN, 40mN, 90mN and 150mN. The synchronized images from the onboard camera illustrate the probe-tissue interaction for the four contact forces (*clockwise from top left*) 0mN, 40mN, 90mN and 150mN.

In order to validate the ability of the system to accurately trace a trajectory and reconstruct the underlying multi-spectral fluorescence images, a double stained phantom was developed. The silicone phantom featured four horizontal lines mounted on a flat surface. The first and third lines consisted of yellow fluorescent dye while the second and fourth lines consisted of green fluorescent dye. The lines were approximately 1.5mm thick and were separated by a gap of approximately 1.5mm.

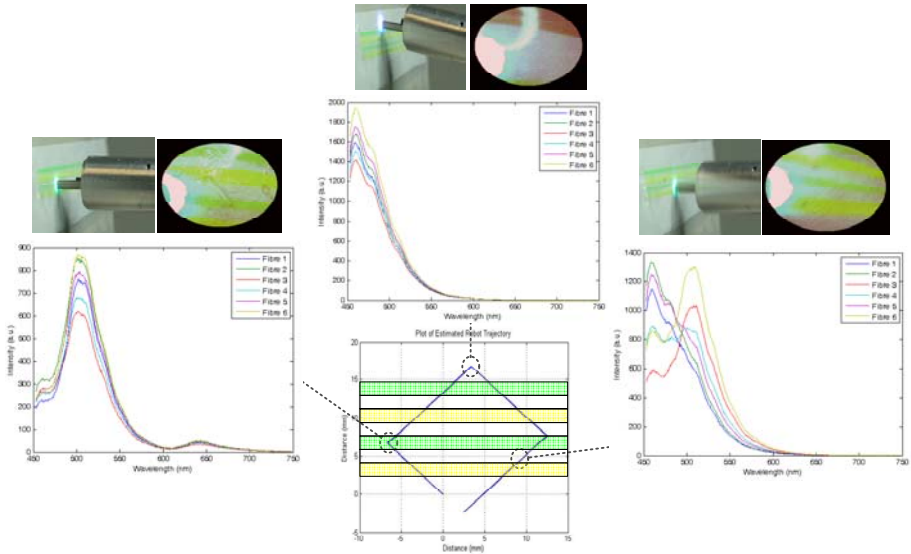


Fig. 3. Plot of the trajectory of the robot tip as it passes over the green and yellow fluorescence dye. Three results sequences displaying the spectral information (*bottom*) the onboard camera (*top-right*) and an external camera for visual confirmation of the spectral result (*top-left*) are also shown.

In order to obtain continuous multi-spectral measurements along 2D scan patterns, each of the separate components (robot control, linear actuator control, onboard camera acquisition and fluorescence image acquisition) were synchronized and the robot was driven under joystick control along a trajectory. This synchronization allows the spectral information at each acquisition to be localized along the tip trajectory using the estimation of tip position recorded from the motor pulses. The location of each image can be confirmed by the use of the onboard camera and a second, external camera which was used solely for validation purposes. Additionally, the cameras were used to confirm that the probe maintained contact with the surface at all times. Since the probe tip moves in an arc during robotic joint actuation the linear servoing mechanism must actuate the probe in order to maintain contact during all stages of the image acquisition. A similar situation would arise with biological tissue where the surface to be probed would have a different profile to the arc traced by the robot tip.

Fig. 3 shows the robot tip trajectory projected onto a 2D plane (located at the robot distal face when the robot is straight) and a sample of three result sets from three example points along the trajectory. Each result sequence shows the spectral image acquired at that location (*bottom*), the onboard camera view (*top-right*) and the external camera view (*top-left*). The location of the spectral peak at different locations (paper: 460nm, yellow dye: 515nm, green dye: 505nm and 640nm) clearly illustrate the ability of the system to differentiate between the different wavelengths recorded at the different locations. The results on the right hand side of Fig. 3 are particularly interesting as different fibers within the imaging probe display different excitation wavelengths. During these acquisitions, the probe was straddling the boundary of

more than one excitation. This indicates that higher resolution multi-spectral information can be obtained by incorporating the relevant spatial distribution of the fibers in the probe.

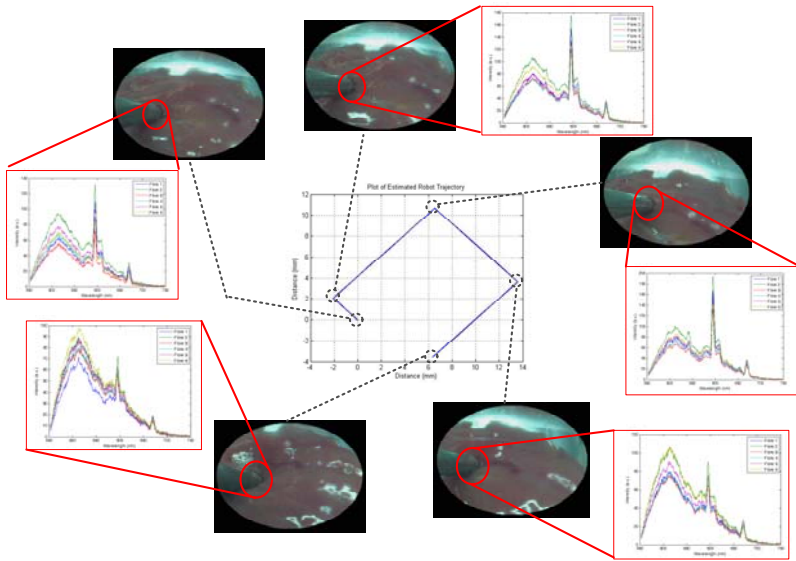


Fig. 4. (middle) 2D plot of the trajectory of the probe tip with five results sets indicated. Each results set displays the spectral information and the synchronized onboard camera image.

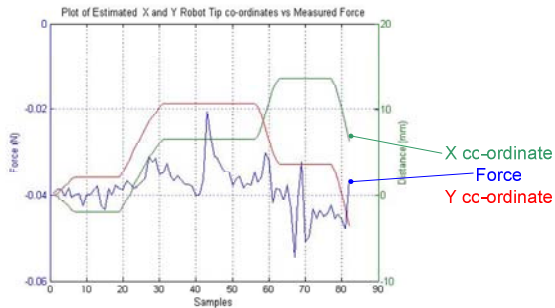


Fig. 5. Plot of the X and Y co-ordinates of the probe tip (green and red respectively) versus the force measured by the F/T sensor (blue)

To demonstrate the ability of the system to employ the protocol described above to record autofluorescence from biological soft tissue a similar experiment was performed on ex-vivo fowl liver. These results are shown in Fig. 4. Images from the onboard camera are highlighted as well as the spectral information captured by the probe. The magnitude of the autofluorescence peak in the spectra (peak approx. 510nm) remains constant for all acquisitions, indicating that good contact is maintained throughout the experiment. Fig. 5 illustrates the relationship between the

movement of the tip and the measured force during this acquisition sequence. As can be seen from the plot the force controller maintains a constant contact force of 40mN for the duration of the run. Deviations from the constant 40mN force occur after approx. 43 and 66 samples. These deviations are due to changes in the tissue surface profile which cause the probe to almost lose contact with the tissue or become buried deeper into the tissue. In each of these cases the controller servos the probe to maintain the 40mN force. This motion can be observed in the onboard camera.

4 Discussion and Conclusion

The results presented in this paper demonstrate the ability of a robotically-assisted approach for consistent spectroscopic sensing by overcoming the challenges of manual probe manipulation. By decoupling the operator from the complex manipulations required to obtain such images, the focus of the clinical investigation can be shifted towards identifying potential target sites and interpreting the spectroscopic information. The ability of the proposed system to obtain continuous multi-spectral information along 2D scan patterns across complex surface profiles, while maintaining a constant contact force, has clear potential clinical applications in areas such as tumour margin identification and localization of fluorescence markers.

References

1. Friedland, S., Benaron, D., Coogan, S., Sze, D.Y., Soetikno, R.: Diagnosis of chronic mesenteric ischemia by visible light spectroscopy during endoscopy. *Gastrointestinal Endoscopy* 65, 294–300 (2007)
2. Robichaux-Viehoever, A., Kanter, E., Shappell, H., Billheimer, D., Jones, H., Mahadevan-Jansen, A.: Characterization of Raman spectra measured in vivo for the detection of cervical dysplasia. *Applied Spectroscopy* 61, 986–993 (2007)
3. Jo, J.A., Fang, Q., Papaioannou, T., Baker, J.D., Dorafshar, A.H., Reil, T., Qiao, J.H., Fishbein, M.C., Freischlag, J.A., Marcu, L.: Laguerre-based method for analysis of time-resolved fluorescence data: application to in-vivo characterization and diagnosis of atherosclerotic lesions. *Journal of Biomedical Optics* 11 (2006)
4. Noonan, D.P., Elson, D., Mylonas, G., Darzi, A., Yang, G.-Z.: Laser Induced Fluorescence and Reflected White Light Imaging for Robot-Assisted Minimally Invasive Surgery. *IEEE Transactions on Biomedical Engineering* 56, 889–892 (2009)
5. Ti, Y., Lin, W.-C.: Effects of probe contact pressure on in vivo optical spectroscopy. *Optics Express* 16, 4250–4262 (2008)
6. Brouwer, I., Ustin, J., Bentley, L., Sherman, A., Dhruv, N., Tendick, F.: Measuring In Vivo Animal Soft Tissue Properties for Haptic Modeling in Surgical Simulation. In: *Medicine Meets Virtual Reality*, pp. 69–74 (2001)
7. Wagner, C.R., Stylopoulos, N., Howe, R.D.: The Role of Feedback in Surgery: Analysis of Blunt Dissection. In: *10th Symp. on Haptic Interfaces For Virtual Envir. & Teleoperator Systems*, pp. 68–74 (2002)
8. Misra, S., Reed, K.B., Douglas, A.S., Ramesh, K.T., Okamura, A.M.: Needle-Tissue Interaction Forces for Bevel-Tip Steerable Needles. In: *IEEE International Conference on Biomedical Robotics and Biomechanics*, pp. 224–231 (2008)

Motion Tracking in Narrow Spaces: A Structured Light Approach

Oline Vinter Olesen^{1,2,3}, Rasmus R. Paulsen¹, Liselotte Højgaard²,
Bjarne Roed³, and Rasmus Larsen¹

¹ Informatics and Mathematical Modelling, Technical University of Denmark
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark
ovol@imm.dtu.dk

<http://www.imm.dtu.dk/>

² Department of Clinical Physiology, Nuclear Medicine & PET, Rigshospitalet,
Copenhagen University Hospital, University of Copenhagen

³ Siemens Healthcare, Siemens A/S, Denmark

Abstract. We present a novel tracking system for patient head motion inside 3D medical scanners. Currently, the system is targeted at the Siemens High Resolution Research Tomograph (HRRT) PET scanner. Partial face surfaces are reconstructed using a miniaturized structured light system. The reconstructed 3D point clouds are matched to a reference surface using a robust iterative closest point algorithm. A main challenge is the narrow geometry requiring a compact structured light system and an oblique angle of observation. The system is validated using a mannequin head mounted on a rotary stage. We compare the system to a standard optical motion tracker based on a rigid tracking tool. Our system achieves an angular RMSE of 0.11° demonstrating its relevance for motion compensated 3D scan image reconstructions as well as its competitiveness against the standard optical system with an RMSE of 0.08° . Finally, we demonstrate qualitative result on real face motion estimation.

1 Introduction

The tomographic reconstruction of 3D and time varying 3D medical images from a series of scanning modalities including X-Ray computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) requires sequential data recording over time. Patient motion during this time will result in a lower image quality or even render the examination useless, cf. [1] for PET imaging. The probability of patient motion occurring grows with increasing acquisition time. For structural or anatomical imagery, patient motion can sometimes be estimated and compensated directly from the scan recordings, e.g. in cardiac MRI [2] and lung CT [3]. For functional 3D scans such as PET and fMRI lower contrast and more spatially sparse events hampers the direct estimation of motion from the recordings themselves. Our focus is on the Siemens High Resolution Research Tomograph (HRRT) PET brain scanner, which is a

brain dedicated scanner with a spatial resolution down to 1.4 mm [4]. Movement induced image degradation increases with increasing scanner resolution and thus head motions end up counteracting the technological advances of high resolution scanners. Due to the low count rate and resulting low contrast information the patient motion is assumed known for most of the suggested motion correction methods in PET imaging [5,6,7,8]. An optical real-time motion tracking system has been preferred (Polaris System, Northern Digital Inc.) [9]. This system registers a rigid tracking tool with 3-6 infrared reflecting markers. The tracking tool is fixed to the patient's head using different types of band-aid, helmets, wet-caps, or goggles. However, it has been reported that, using fixation, these methods can cause artifacts on the PET images [10]. We have previously described a structured light (SL) based system for 3D face surface reconstruction that: (1) does not need any markers; (2) fits to the narrow geometry of the Siemens HRRT PET scanner; and (3) can potentially be built into future PET scanners [11].

In this paper we will show how 3D point clouds captured using this system by use of resistant and robust iterative closest point (ICP) registration to a template surface can be used to estimate rigid body motion of the head inside the scanner patient tunnel. In the method proposed, the pose estimation is based on the rigid alignment of scans to either a pre-computed reference scan or the previous scan. The alignment is a special case of the 3D point cloud registration. Point correspondence cannot be assumed and since the scanner output are patches, there will only be partial overlap between reference/previous scans and the current scan. Our method is based on the classical ICP algorithm [12] that aligns two point clouds with no prior correspondence. Several efficient variants of the ICP algorithm have been published [13]. The possible enhancements of the normal ICP includes matching based on differential properties of point sets and rejection of invalid point matches. Due to the partial overlap of our scans, we employ a point rejection approach, where points falling on the border of the target are rejected. It is therefore necessary to detect borders of the 3D scan. This is non-trivial to do with raw points clouds. To overcome this, we represent our target as a triangulated surface that has been computed using a state-of-the-art surface reconstruction algorithm [14]. Compared, to the popular Poisson surface reconstruction algorithm [15], the Markov Random Field surface reconstruction algorithm [14] deals particularly well with human body scans. We provide a quantitative evaluation of the performance of the structured light system for pose estimation.

2 Experiments and Methods

The SL system consists of a DLP projector (DLP Pico Projector, Texas Instruments) with HVGA resolution (480×320) and two grayscale CCD cameras (Chameleon, Point Grey Research) with a resolution of 1280×960 as described in [11]. The SL system was mounted on the gantry of the HRRT PET scanner as shown on Fig. 1 just above the patient tunnel. The performance of the new tracking approach is evaluated by a set of experiments on the HRRT PET scanner with simultaneous tracking using the Polaris Vicra system. A mannequin

head was placed inside the patient tunnel as a patient would be. It was mounted onto a nano rotary motor stage from Thorlabs. The stage made it possible to perform highly controllable rotation of the head. The stage was programmed to rotate in steps of 5 degrees from -15 to 15 degrees and the movements were repeated four times. At each stationary position a set of images were captured with the SL system.

The Polaris tracking tool was fixed to the forehead of the mannequin head using a band-aid as used for patients to track the head motions during the PET acquisition. While the Thorlabs stage provides baseline rotation data, the Polaris system recorded the motions of the head simultaneously with the image capturing of the SL system. Fig. 1 shows the set up of the experiments where the SL system is seen in the front and the Polaris sensor is seen in the back behind the patient tunnel. Fig. 1(a) shows the head in the reference position at 0 degrees where the region of interest (ROI) is seen as the bright region around the bridge of the nose. This ROI is chosen due to the limited facial movements of the bridge of the nose. At each of the seven positions of the four runs of experiments 3D point clouds are reconstructed using phase-shifting interferometry (PSI) [16]. PSI is used to determine the correspondence between the two image planes; the projector image plane and the image plane of one of the cameras. From a series of three captured interferograms (2D images) the wavefront phase is computed and converted to line positions on the projector image plane. Thus a given phase of the cosine patterns on the captured images correspond to a position on the projector image plane after phase unwrapping. Since the phase is periodic, the phase has to be unwrapped to achieve a continuous phase image. Several methods to perform phase unwrapping exist. Experiments showed that the method described in [17] perform well with our data.

The points on the image planes are converted into 3D coordinates using a simple pinhole model for both the cameras and the projector and assuming

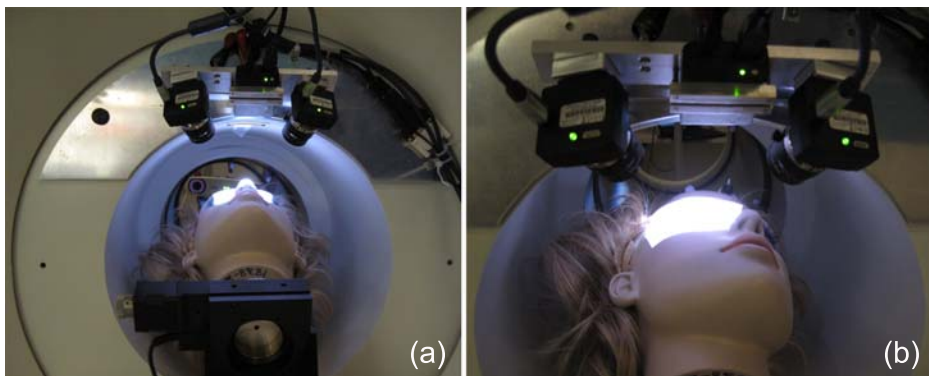


Fig. 1. Photographs of the mannequin head inside the HRRT PET scanner with the SL system in the front mounted to the HRRT PET gantry. (a) The motor stage is seen in the bottom and the Polaris sensor in the back. (b) Mannequin head rotated to the right (the tracking tool can just be discerned above the forehead).

the calibrations parameters for all three components are known. The details of system calibration and 3D coordinate computations can be found in [11].

We use the scan acquired at 0° as the reference and in the following, the pose of the head is estimated relatively to this reference in the following. We want to estimate the rigid body transformation from the current 3D scan to the reference scan. The scans are unstructured point clouds where approximate estimates of the point normals exist. We are using a specialised version of the ICP [12] algorithm. Initially, a surface is created based on the reference scan. Both cameras produce a point cloud representation of the part the head in its field of view. These two point clouds are aligned and merged to create a reference scan that covers the field of view of both cameras. The surface is created using the novel Markov Random Field surface reconstruction algorithm [14]. It is based on an implicit description of the surface combined with a regularization step that makes it well suited for human body scans. Since the surface reconstruction algorithm by default computes surfaces that extend beyond the point cloud, a post-processing step is needed where the surface is cropped to fit the point cloud. This is done by removing parts of the surface that are not supported by reliable input points. Support is defined as being within a distance, d , of an input point. d is estimated as the average neighbour distance in the input point cloud. The result is a polygonised surface patch, where the edge vertices are defined by having only one adjacent triangle. For each point in the current scan, the closest point on the triangulated surface is found using a k D-tree based approach. If the point falls on an edge vertex, the point match is discarded. The remaining point matches are used to compute the rigid body transformation using the solution found in [18]. Using this method the transformation bringing the current scan into alignment with the reference surface is computed. The transformation consists of an estimated 3×3 rotation matrix \mathbf{R} and a translation vector \mathbf{t} .

To be able to compare the rotation estimate from our method and the Polaris system with the baseline rotation provided by the Thorlabs stage, the rotation angle θ , direction of rotation axis \mathbf{v} , and a point on the line \mathbf{c} is determined as [18]:

$$\begin{aligned}\theta &= \arccos((\text{trace}(\mathbf{R}) - 1)/2) \\ \mathbf{v} &= 1/(2 * \sin(\theta)) * [R_{32} - R_{23} \ R_{13} - R_{31} \ R_{21} - R_{12}]^T \\ \mathbf{c} &= (\mathbf{I} - \mathbf{R})^{-1}\mathbf{t}\end{aligned}$$

3 Data and Results

Fig. 2 shows the 3D point clouds at the different positions for one of the four experiment runs with the right (red) and left (blue) camera respectively. The image to the left represent the position at -15 degrees and the image to the right at a position of 15 degrees. These are aligned into the reference surface seen in the center. As seen, the point clouds are highly detailed with little noise and outliers, supporting a high spatial resolution of the system. Two results of the ICP alignment are shown in Fig. 3 at ± 10 degrees from the right camera. The



Fig. 2. 3D point clouds of the mannequin head at the seven positions. Left to right from -15 degrees to 15 degrees. Blue images represent the left camera and red images represent the right camera. The center images at 0 degrees are the reference surfaces.

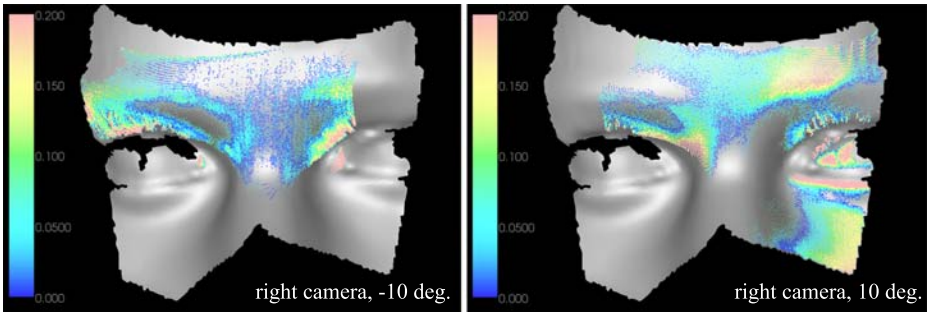


Fig. 3. Results of the ICP alignment of the right camera at two positions ± 10 degrees. Alignments into the reference position are shown on top of the reference surfaces. The colors represent the errors as the distance to target in mm.

errors between the target and the aligned points are in the order of 0-0.2mm with the largest errors around the eyes. As previously mentioned, the motion of the Thorlabs stage is considered the ground truth motion. The errors of the estimated motion are plotted as a function of the ground truth motion in Fig. 4. The red and blue points represent the right and left camera respectively and the black points are the results from the Polaris system. The errors of the SL system are less than 0.2 degrees from the performed rotation when using the right camera (red) for negative rotations and the left camera (blue) for the positive rotation with a RMS error of 0.11 degrees. This is a similar result as the Polaris system, which has a RMS error of 0.08 degrees. There are two main reasons why the results are less accurate when the head is rotated toward the camera; the overlap with the target around the bridge of the nose becomes less and the overlap around the eye becomes larger. The eye is a non-robust region due to edges and hair. This is an even larger problem on humans. In the future, this region will be excluded prior to alignment.

The system has been tested on a human test subject in the HRRT setup. The purpose is to demonstrate the clinical usability of the system. Currently,

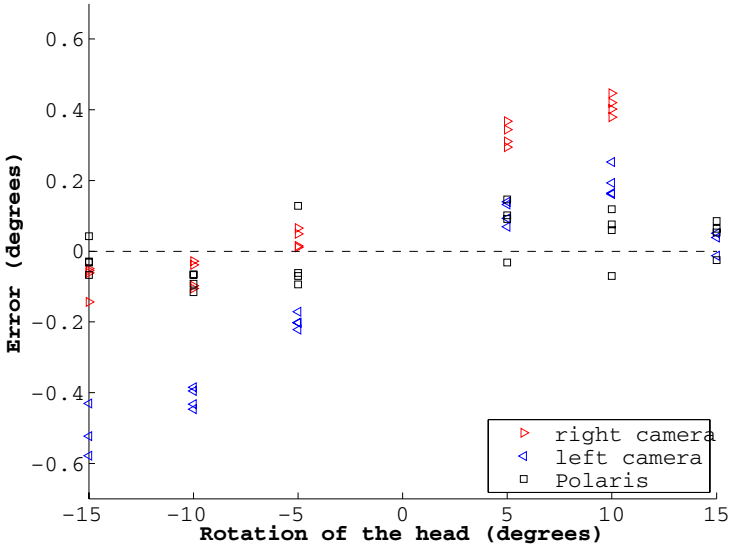


Fig. 4. Comparison of the SL system and the Polaris system; differences between the estimated rotations and the performed ones as a function of the performed rotations.

we are only aiming for a quantitative evaluation and therefore no ground truth data was recorded and no PET acquisition was done. The subject moved 1-3 cm between each scan. The results can be seen in Fig. 5, where the reference scan is represented as a grey reconstructed surface. Two following scans are aligned to the reference scans and the aligned point clouds are seen. It is clearly seen, that the reference scans are less complete than for the mannequin. This is due to problems with shadows around the nose. We are currently optimising the hardware configuration of the system with respect to real humans in the clinical setting. In Fig. 5, the colour coding of the aligned scans represents the individual

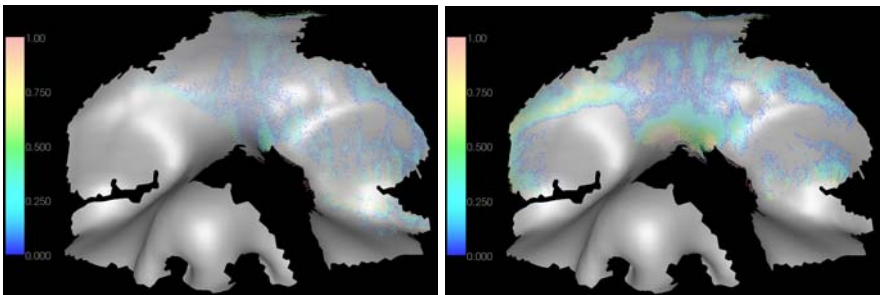


Fig. 5. Test with real human scanings. Point clouds aligned to the reconstructed surface of the reference scan (grey surface) and the point errors are colourcoded. The medians of the point errors are 0.13 and 0.17 mm.

per-point alignment error. It is computed as the distance from the point to the closest point on the reference scan. The median values of these point errors are 0.13 and 0.17 mm for the two scans in Fig. 5. These results are representative for the experiments we have performed on human volunteers. While not optimal, the magnitude of this error indicates that the system will be able to accurately determine the pose changes for real humans in the clinical environment.

4 Summary and Conclusions

We have presented a structured light system adapted for motion compensation in high-resolution PET scanners. While the systems accuracy is comparable with the current state-of-the-art optical trackers, it is more flexible and easier to adapt to the narrow patient tunnels of PET scanners. Furthermore, the system is fully automatic and does not rely on markers that are notoriously difficult to use in a clinical setting. The system was tested on a setup, where a mannequin head mounted on a robot system created baseline data. The results show, that the proposed framework was able to estimate the head rotation with an accuracy better than 0.2° for a head movement between -15° and 15° . Furthermore, preliminary experiments were performed on human test subjects. Quantitative analysis shows that the system is able to robustly estimate the pose changes for human subjects, strongly indicating that the method will be useable in the clinical practise.

Acknowledgements. The authors would like to thank the staff at The Mechanical Workshop at Rigshospitalet, Copenhagen, Denmark who helped us produce the mechanical devices. Ms. Olesen also thanks the Danish Industrial PhD programme for supporting this research.

References

1. Green, M.V., Seidel, J., Steina, S.D., Tedder, T.E., Kempner, K.M., Kertzman, C.: Head movement in normal subjects during simulated PET brain imaging with and without head restraint. *J. Nuclear Medicine* 35(9), 1538–1546 (1994)
2. Stegmann, M.B., Larsson, H.B.W.: Motion-compensation of cardiac perfusion MRI using a statistical texture ensemble. In: Magnin, I.E., Montagnat, J., Clarysse, P., Nenonen, J., Katila, T. (eds.) *FIMH 2003*. LNCS, vol. 2674. Springer, Heidelberg (2003)
3. Ehrhardt, J., Werner, R., Frenzel, T., Säring, D., Lu, W., Low, D., Handels, H.: Reconstruction of 4D-CT data sets acquired during free breathing for the analysis of respiratory motion. In: *Proc. SPIE Medical Imaging*, vol. 6144, pp. 614414–1–8 (2006)
4. Olesen, O.V., Sibomana, M., Keller, S.H., Andersen, F., Jensen, J.A., Holm, S., Svarer, C., Højgaard, L.: Spatial resolution of the HRRT PET scanner using 3D-OSEM PSF reconstruction. In: *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. M13-225, pp. 3789–3790 (2009)
5. Picard, Y., Thompson, C.J.: Motion correction of PET images using multiple acquisition frames. *IEEE Trans. on Medical Imaging* 16(2), 137–144 (1997)

6. Woo, S.K., Watabe, H., Choi, Y., Kim, K.M., Park, C.C., Bloomfield, P.M.: Sinogram-based motion correction of PET images using optical motion tracking system and list-mode data acquisition. *IEEE Trans. on Nuclear Science* 51(3), 782–788 (2004)
7. Rahmim, A., Dinelle, K., Cheng, J.C., Shilov, M.A., Segars, W.P., Lidstone, S.C., Blinder, S., Rousset, O.G., Vajihollahi, H., Tsui, B., Wong, D.F., Sossi, V.: Accurate event-driven motion compensation in high-resolution PET incorporating scattered and random events. *IEEE Trans. on Medical Imaging* 27(8), 1018–1033 (2008)
8. Raghunath, N., Faber, T.L., Suryanarayanan, S., Votaw, J.R.: Motion correction of PET brain images through deconvolution: II. Practical implementation and algorithm optimization. *Physics in Medicine and Biology* 54(3), 813 (2009)
9. Lopresti, B.J., Russo, A., Jones, W.F., Fisher, T., Crouch, D.G., Altenburger, D.E.: Implementation and performance of an optical motion tracking system for high resolution brain PET imaging. *IEEE Trans. on Nuclear Science* 46(6), 2059–2067 (1999)
10. Herzog, H., Tellman, L., Fulton, R., Pietrzyk, U.: Motion correction in PET brain studies. In: *IEEE Proc. The Fourth International Workshop on Multidimensional Systems*, pp. 178–181 (2005)
11. Olesen, O.V., Jørgensen, M.R., Paulsen, R.R., Højgaard, L., Roed, B., Larsen, R.: Structured light 3D tracking system for measuring motions in PET brain imaging. In: *Proc. SPIE Medical Imaging*, vol. 7625, p. 76250X (2010)
12. Besl, P.J., McKay, N.: A method of registration of 3D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(2), 239–256 (1992)
13. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: *Proc. Int. Conf. 3-D Digital Imaging and Modeling*, pp. 145–152 (2001)
14. Paulsen, R., Bærentzen, J., Larsen, R.: Markov Random Field Surface Reconstruction. *IEEE Transactions on Visualization and Computer Graphics* (2009)
15. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson Surface Reconstruction. In: *Proc. Symposium on Geometry Processing*, pp. 61–70 (2006)
16. Huang, P., Hu, Q., Jin, F., Chiang, F.: Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering* 38, 1065 (1999)
17. Herráez, M., Burton, D., Lalor, M., Gdeisat, M.: Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. *Appl. Opt.* 41, 7437–7444 (2002)
18. Horn, B.K.P.: Closed form solution of absolute orientation using unit quaternions. *Journal of the Optical Society A* 4(4), 629–642 (1987)

Tracking of Irregular Graphical Structures for Tissue Deformation Recovery in Minimally Invasive Surgery

Marco Visentini-Scarzanella¹, Robert Merrifield², Danail Stoyanov²,
and Guang-Zhong Yang^{1,2}

¹ Royal Society/Wolfson Foundation MIC Laboratory,

² Institute of Biomedical Engineering,

Imperial College London, London SW7 2AZ, United Kingdom

{marcovs, robert.merrifield, danail.stoyanov}@imperial.ac.uk
g.z.yang@imperial.ac.uk

Abstract. Tissue deformation tracking is an important topic of minimally invasive surgery with applications ranging from intra-operative guidance to augmented reality visualisation. In this paper, we present a technique for visual tracking of irregular structures with an arbitrary degree of connectivity in space. The variational formulation of the proposed method ensures that correlation is maximised between tracked points and their computed new positions while the overall structure shape variation is minimised, thus maintaining spatial coherence of the tracked structure. The proposed method is applied to surgical annotation and tracking in 3D for telementoring and path-planning. The results are validated both on a CT-scanned phantom model and *in vivo*, showing an average alignment error of 1.79 mm (± 0.72 mm).

1 Introduction

Tissue deformation tracking during MIS has been the focus of considerable research efforts with the goal of enhancing the user experience particularly during robotically assisted procedures. For MIS, issues such as limited field-of-view, large-scale tissue deformation and disorientation can be addressed by image mosaicing, virtual motion compensation, ablation guidance [1], imposition of dynamic active constraints and motor channelling [2], all of which hinge on reliable image feature tracking techniques [3]. While these techniques could be used in conjunction with 2D feature tracking plus stereo matching, a 2D approach to tracking does not reflect the physical structure of the object and is unsuitable for integrating *a priori* motion information.

In this paper, a *structure tracking* approach is proposed. It differs from feature or point based tracking methods in that its optimality criteria are global to the structure tracked and it can operate in any dimensional space, in contrast to unconstrained feature tracking where the optimality is local to each feature tracked. While existing methods for structure tracking operate on regular grids in 2D, the proposed method relies on the construction and tracking of networks from sparse point clouds with arbitrary degrees of connectivity in d dimensions. A novel feature correlation function tailored to MIS data is formulated and the proposed method is applied to MIS 3D

spatio-temporal telestration for path planning and telementoring. The method is evaluated with both a synthetic heart model with known ground-truth 3D geometry and deformation, as well as *in vivo* data acquired during a Totally Endoscopic Coronary Artery Bypass (TECAB) procedure.

2 Method

The proposed method consists of three distinct steps. First, a network is built from a collection of points given the parameters about their dimensionality and connectivity level. Then, as their motion fields are updated, a point-wise correlation is computed for each possible position in the search space for all points in the cloud. Finally, a variational technique is used to maximise correlation between the points and their proposed updated positions whilst minimising the overall network shape variation.

2.1 Network Building

Given a sparse collection of points in d dimensions, a connected network can be built given a scoring function and the maximum degree of connectivity k allowed. The maximum degree of connectivity is determined by the highest total number of links a point in the network has, divided by two: for instance, a regular 2D mesh would have a degree of connectivity of 2. Determining the degree of connectivity required implies a trade-off between strong geometric regularisation constraints for high degrees due to the larger number of dependencies and the inability of fully capturing the characteristics of the underlying structure being represented for low degrees.

To construct a network given a set of points and a degree k , the d -dimensional space is first partitioned into $2k$ equal sections. For each space partition and each pair of points that could be connected by a line within the partition under consideration, their L_d norm is computed, and the pair separated by the minimum distance is connected.

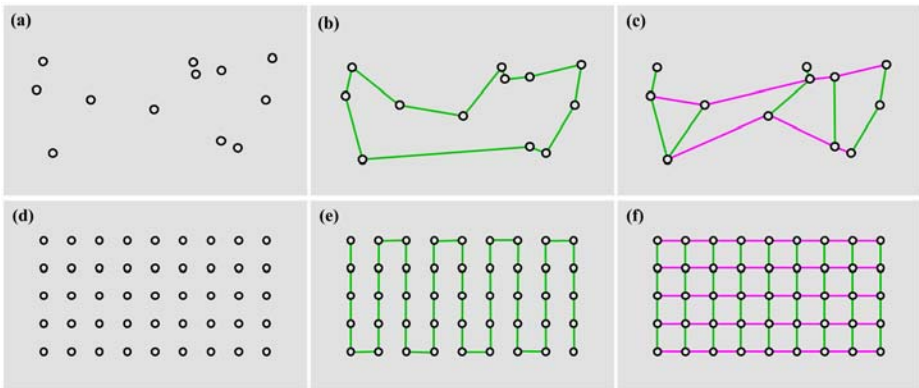


Fig. 1. (a) Sparse set of 2D points. (b) Result of the algorithm with degree of connectivity $k=1$ and (c) $k=2$. (d) Regular 2D point grid. (e) Result of the algorithm with $k=1$ and (f) $k=2$.

The process is iterated until no more points can be linked without breaking the limit of two connections for every k . For regular grids and $k=d$, the process converges towards a conventional lattice structure, as shown in Fig. 1 above for $d=2$:

2.2 Correlation Estimation

Variational approaches to optical flow estimation commonly involve solving the following minimisation problem [4]:

$$\min_u \left\{ \int_{\Omega} \lambda g(x, u(x)) + \psi(\nabla u(x)) d\Omega \right\} \quad (1)$$

In the expression above, $g(\mathbf{x})$ is the data term measuring the correlation between the point \mathbf{x} and the displaced candidate $\mathbf{x}+u(\mathbf{x})$, the second term is a regularisation constraint enforcing a smooth displacement field, and λ is a factor weighing the relative contribution of data and regularisation terms.

The nature of the data term depends on the optimisation procedure adopted: faster techniques involve the use of differentiable data terms [5], whereas more sophisticated non-differentiable penalty functions are coupled with complete searches [6]. Recently, Giannarou et al. [7] proposed an affine-invariant anisotropic corner detector tailored for MIS applications and robust to lighting variations and image blur. Detected features are described as ellipses whose size, orientation and radii ratio summarise the structural information of the feature patch examined.

In this study, a non-differentiable penalty function based on the intersection area between the ellipses representing the original and candidate points is used. The measure is then weighed by the difference in colour intensity distribution between the two patches, in order to include both structural and intensity information:

$$g(\mathbf{x}, u(\mathbf{x})) = 1 - \varphi(I(\mathbf{x}), I'(\mathbf{x} + u(\mathbf{x}))) \cdot \sum_{\Omega'} \sqrt{P(I(\mathbf{x}))P(I'(\mathbf{x} + u(\mathbf{x})))} \quad (2)$$

In (2), $I(\mathbf{x})$ and $I'(\mathbf{x})$ are the previous and current intensity images for which the flow needs to be calculated, $\varphi(I(\mathbf{x}), I'(\mathbf{y}))$ is the ratio of overlapping area to total area between the ellipses representing points \mathbf{x} and \mathbf{y} in the previous and current frames, and the summation term is the Bhattacharyya coefficient of the channel intensities distributions in the two patches Ω' . The penalty function is bounded between 0 and 1.

The overlapping area between two generic non-axis aligned ellipses can be calculated first by finding their intersections: two polynomials share a common root if and only if their Bezout determinant is zero, which in the elliptical case is a quartic whose zeros can be efficiently found with a polynomial solver. Given a set of intersections Φ , the area A_{int} can be found by integrating anticlockwise along the elliptical arcs between intersection points according to Green's theorem:

$$A_{int} = \frac{1}{2} \sum_{\varphi_i \in \Phi} R \left(-x_c \sin(\theta) + y_c \cos(\theta) - \cos(t) \right)_{\varphi_i}^{\varphi_{i+1}} + r \left(x_c \cos(\theta) + y_c \sin(\theta) + \sin(t) \right)_{\varphi_i}^{\varphi_{i+1}} + Rr t_{\varphi_i}^{\varphi_{i+1}} \quad (3)$$

In Eq. (3), R , r and θ are the major radius, minor radius and orientation of the ellipse whose arc two consecutive intersections lie on. The integrating variable t is the angle between the ellipse centre and a point. The overlapping area to total area ratio is then:

$$\varphi(\mathbf{x}, \mathbf{y}) = \frac{A_{\text{int}}}{\pi(R_x r_x + R_y r_y)} \tag{4}$$

2.3 Variational Optimisation

The problem formulated by Eq. (1) can be solved with differentiable data terms in the case where ψ is the TV- L_1 operator [8]. The general principle can be extended to cases when non-differentiable penalty functions are used [9]. Here we further extend the method to irregular graphical structures with an arbitrary degree of connectivity. As Eq. (1) is nonconvex in u , an auxiliary vector field v is introduced to decouple the pointwise penalty term from the regularity term, leading to the convex approximation:

$$\min_{u,v} \left\{ \int_{\Omega} \lambda g(\mathbf{x}, v(\mathbf{x})) + \frac{1}{2\theta} (v(\mathbf{x}) - u(\mathbf{x}))^2 + \psi(\nabla u(\mathbf{x})) d\Omega \right\} \tag{5}$$

Such an approximation is convex in u and nonconvex in v , and it can be shown that $u \rightarrow v$ as $\theta \rightarrow 0$. Its solution involves alternating two optimisation steps while decreasing θ between iterations: for non-differentiable data terms, the optimal v (with constant u) is found by a complete search. For each iteration i , the optimal element of u (with constant v) for each dimension d can be found by the following scheme:

$$\begin{cases} u_d^{i+1} = v_d^{i+1} - \theta \text{div} p_d^i \\ p_d^{i+1} = \frac{p_d^i + \tau \nabla (\text{div} p_d^i - v/\theta)}{1 + |\tau \nabla (\text{div} p_d^i - v/\theta)|} = \frac{p_d^i + \tau \nabla u_d^{i+1}}{1 + |\tau \nabla u_d^{i+1}|} \end{cases} \tag{6}$$

Above, p is a collection of d k -dimensional auxiliary vectors each initialised at $p^0 = 0$. The irregular element spacing throughout the graphical model is reflected by the gradient operator which weighs the variation of neighbouring elements proportionally to the reciprocal of their Euclidean L_d norm, constraining closer elements more than farther away ones.

2.4 Applications to MIS

The above theoretical framework is applied to stereo MIS video for the specific case of $d=3$ and $k=1$. This corresponds to an arbitrary connected trace (open or closed) in 3D space, and can be applied to, for example, tracking custom freehand inputs overlaid on the visualised 3D scene for preoperative path-planning or surgical training. Such interfaces have been investigated for natural scenes [10] with the additional requirement of prior offline processing, and for telementoring in MIS [11] using conventional algorithms on static scenes. The proposed algorithm allows using directly prior 3D velocity information to restrict the search space for the optimal motion vector and it adds robustness to the process due to the penalty term tailored to

MIS data. Also, it yields subpixel accuracy if a fine discretisation of the 3D space is adopted. More importantly, contrarily to unconstrained point tracking it strives to preserve the shape of the original annotation due to its regularisation term.

Under this scheme, the user is prompted to draw an annotation on a still frame from one of the stereo channels. Stereo matches for the annotation points are found on the remaining channel with the algorithm from [12], the matching pairs are then triangulated yielding a 3D representation of the original annotation. The annotation is then tracked across the incoming frames with the algorithm presented. The penalty function is modified to be an average of the two penalties from the projection of the 3D candidate point on the left and right channel and their comparisons with the previous frames, thus enforcing inter-channel consistency. Fig. 2 below schematically illustrates the proposed technique.

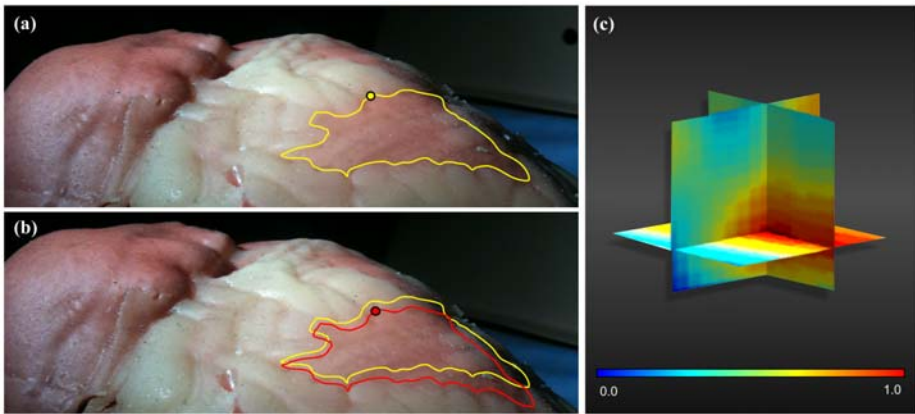


Fig. 2. An illustration of the proposed method. A 3D point (a) can be moved to a new position (b) within a rectangular parallelepipedal window depending on the correlation expressed by the penalty function (c). Uneven displacement between connected points results in stronger constraints from the regularisation term.

3 Experimental Setup and Results

The proposed technique has been tested both with a synthetic phantom model and *in vivo* to evaluate its performance. Both sequences were recorded in standard definition at 25fps from a da Vinci surgical platform with known stereo camera parameters, and input to an Intel Core 2 Duo 2.4GHz system with 2GB of RAM for processing. With the above setup, the algorithm operates at ~ 1.5 fps. For all sequences, the algorithm considers a velocity search space of 2mm^3 per point per frame sampled at intervals of 0.025mm , with values for λ and τ set to 1.1 and 0.8 respectively.

Phantom experiment. The method was first applied to the dynamic sequence of a silicone phantom heart (Chamberlain Group, MA, USA) beating at 90bpm. A closed contour with 39 control points was defined in correspondence of a section of the left anterior descending artery (LAD). Fig. 3 (a) and (b) below show the reconstructed annotation volume over a section of the sequence with highlighted diastolic frames.

The tracked annotation shows a small drift of its left and bottom segment, while its centroid remains focused on the region of interest. The upper and right sections remain stable throughout. The drift is due to the fact that the annotation is crossing two completely uniform areas without feature information. The regularisation constraint however limits the amount of drift by driving the whole volume in a uniform direction determined by the annotation segments crossing feature points. Tracking regularity over the whole sequence is quantitatively shown in Fig. 3 (c):

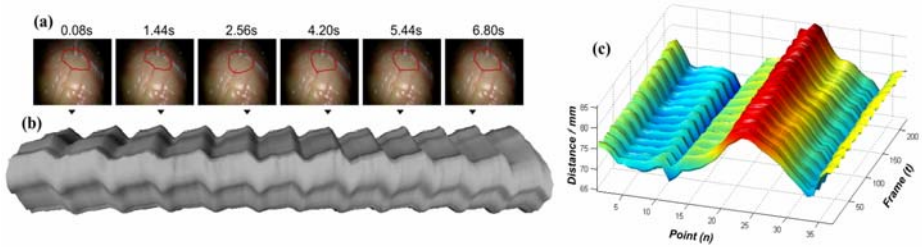


Fig. 3. (a) Selected diastolic frames with the tracked annotation corresponding to ridges in (b). (b) Reconstructed annotation volume over time. (c) Distance of annotation points from the camera over the sequence.

It can be seen that the overall drift is less than 2mm, while it stabilises in the latter half of the sequence. All other points are stable throughout. The closed-loop configuration is maintained during the sequence, as shown by the surface’s sinusoidal characteristics, and alternating cardiac phases for all points are evidenced by the ridges in the plot.

In vivo experiment. The algorithm was further evaluated with a TECAB sequence by delineating an open segment on the pericardium. The sequence presents some significant deformation, with tissue sections becoming occluded from the heart contractions. Fig. 4 below shows the performance of the technique together with selected frames:

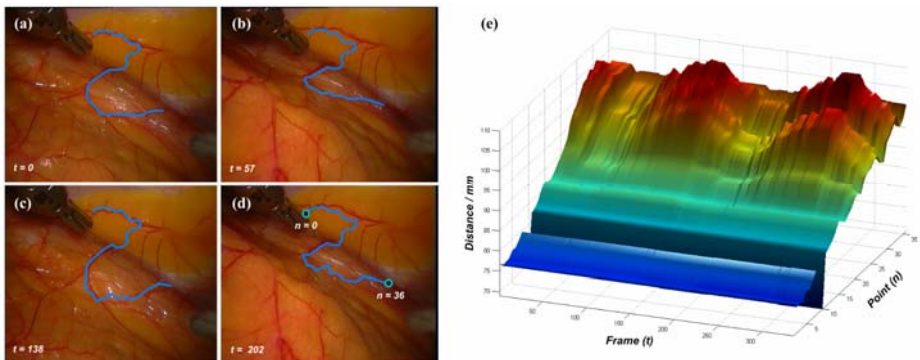


Fig. 4. (a) Initial contour. (b) First diastole. (c) First systole. (d) Second diastole. (e) Distance from camera centre of the tracked points along the annotation. over time.

Despite the degree of tissue deformation shown in Fig. 4 (b), (d) and the lack of explicit occlusion handling, the proposed method was able to maintain the annotation regularity and recover the tissue motion as shown in Fig. 4 (e), where the diastoles and systoles corresponds to the troughs and peaks in the surface plot respectively.

Furthermore, the top half of the images containing the first 15 points of the annotation was motionless throughout the sequence. This is reflected by the surface plot in Fig. 4 (e), where motion characteristics are clearly discernible between the first and last portion of the annotation.

Validation results. To facilitate validation, the phantom heart was augmented with 15 silicone fiducials for registration visible on its surface and scanned with a Siemens Somatom Sensation 64 CT scanner. The resulting volumes had a spatial resolution of $0.41 \times 0.41 \times 0.5$ mm and a temporal resolution of 0.33s. During the scan the heart rate was set to 90bpm, yielding twenty 3D volumes covering the entire cardiac cycle.

The cardiac volumes were then manually registered with the video sequence, so that the correspondence between frame number and cardiac phase was known. For quantitative validation, two annotations have been considered: the open contour shown in the phantom experiment section, and a closed contour delineating the LAD. For both sequences, the initial 3D annotation was stored internally together with its corresponding cardiac phase; the distance between the initial 3D points and the estimated annotation position was then computed whenever the frame under consideration mapped to the initial cardiac phase. Table 1 below quantifies the overall performance of the algorithm:

Table 1. Numerical evaluation of the performance from the proposed algorithm

Error statistics (mm)	Mean	Standard deviation	Maximum	Minimum
Open contour	2.133	0.810	3.695	0.912
Closed contour	1.792	0.721	3.561	0.433

Overall, the open and closed contour sequences show similar performance. The relatively small error in the case of closed contours is due to stronger constraints imposed. Also, given two segments of equal lengths the closed segment would cover a more localised area of the heart surface which could exhibit a higher degree of motion cross-correlation.

4 Discussion and Conclusions

In this paper, we have presented a novel technique for 3D tracking of arbitrary structures with applications to annotation tracking in MIS. Robustness is provided by a novel penalty function while motion consistency is enforced by a TV- L_1 regularisation constraint. Tracking directly in 3D space enforces inter-channel consistency and allows for the reduction of the search space with prior motion information without explicit motion models. The technique has been tested on phantom and *in vivo* with an average discrepancy during validation of 1.79 mm (± 0.72 mm). The regularity of the algorithm

during phantom experiments shows its validity, and future work will focus on a real-time implementation exploiting the method's suitability for parallel processing, a reinitialisation procedure for cyclic motion and occlusion detection.

References

1. Stoyanov, D., Mylonas, G., Yang, G.-Z.: Gaze-Contingent 3D Control for Focused Energy Ablation in Robotic Assisted Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 347–355. Springer, Heidelberg (2008)
2. Mylonas, G., Kwok, K.W., Darzi, A., Yang, G.-Z.: Gaze-Contingent Motor Channelling and Haptic Constraints for Minimally Invasive Robotic Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 676–683. Springer, Heidelberg (2008)
3. Mountney, P., Stoyanov, D., 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)
4. Horn, B.K.P., Schunck, B.G.: Determining optical flow. *Artificial Intelligence* 17, 18 (1981)
5. Zach, C., Pock, T., Bischof, H.: A Duality Based Approach for Realtime TV-L1 Optical Flow. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 214–223. Springer, Heidelberg (2007)
6. Steinbruecker, F., Pock, T., Cremers, D.: Advanced Data Terms for Variational Optic Flow Estimation. In: Vision, Modeling, and Visualization Workshop, Braunschweig, Germany (2009)
7. Giannarou, S., Visentini-Scarzanella, M., Guang-Zhong, Y.: Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp. 1059–1062 (2009)
8. Chambolle, A.: An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision* 20(1-2), 89–97 (2004)
9. Steinbruecker, F., Pock, T., Cremers, D.: Large Displacement Optical Flow Computation Without Warping. In: IEEE International Conference on Computer Vision, Kyoto, Japan (2009)
10. Rav-Acha, A., Kohli, P., Rother, C., Fitzgibbon, A.: Unwrap Mosaics: A new representation for video editing. In: ACM Transactions on Graphics (SIGGRAPH 2008) (2008)
11. Ali, M.R., et al.: 3-D telestration: a teaching tool for robotic surgery. *J. Laparoendosc. Adv. Surg. Tech.* 18(1), 107–112 (2008)
12. Stoyanov, D., Darzi, A., Yang, G.Z.: Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 41–48. Springer, Heidelberg (2004)

Graph Based Interactive Detection of Curve Structures in 2D Fluoroscopy

Peng Wang, Wei-shing Liao, Terrence Chen, Shaohua K. Zhou,
and Dorin Comaniciu

Siemens Corporate Research, 755 College Road East, Princeton NJ, U.S.A.

Abstract. An accurate and robust method to detect curve structures, such as a vessel branch or a guidewire, is essential for many medical imaging applications. A fully automatic method, although highly desired, is prone to detection errors that are caused by image noise and curve-like artifacts. In this paper, we present a novel method to interactively detect a curve structure in a 2D fluoroscopy image with a minimum requirement of human corrections. In this work, a learning based method is used to detect curve segments. Based on the detected segment candidates, a graph is built to search a curve structure as the best path passing through user interactions. Furthermore, our method introduces a novel hyper-graph based optimization method to allow for imposing geometric constraints during the path searching, and to provide a smooth and quickly converged result. With minimum human interactions involved, the method can provide accurate detection results, and has been used in different applications for guidewire and vessel detections.

1 Introduction

In medical imaging, an accurate and robust method to detect curve structures, such as a vessel branch in angiogram, or a guidewire or a catheter in fluoroscopy, is required in many practical applications. Although a fully automatic method is always highly desired, current automatic methods are prone to the errors that are caused by image noise, and curve-like artifacts. In some applications where a large error is not tolerated, human interactions are introduced to make corrections. There are also some cases where human interactions are needed to specify a curve structure of particular interests. For examples, clinicians need to specify a vessel branch where interventions will take place, or users are more interested in a guidewire instead of a catheter, as shown in Figure 1. Such situations highlight the necessity to have interactions in practical systems. In this paper, we present a novel method based on a graph optimization to interactively detect a curve structure in a 2D fluoroscopy image with a minimum requirement of human corrections.

There are existing methods to automatically detect curve structures, such as guidewire [1], and vessels [2], while less research has been done on the interactive curve detection. Some existing interactive detection and segmentation methods [3,4] apply a graph cut algorithm, which needs user inputs to initial positive

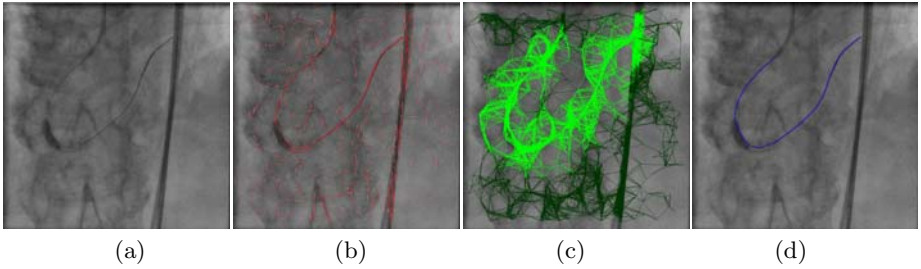


Fig. 1. Graph based interactive detection. (a) an original frame; (b) detected curve segments shown on red; (c) an original graph shown in dark green, and the second stage graph shown in bright green; (d) an interactively detected guidewire in blue.

and negative samples. However, the graph cut based methods mainly aim at the region segmentation, while in 2D fluoroscopy, the curve structure such as guidewire can be thin. The work [5] by Mazouer and his colleagues is close to the work in this paper. It detects guidewire segments in a hierarchical manner, and then finds the guidewire by minimizing the connectivity cost between segments using Dynamic Programming. However, the method only considers the pair-wise segment connectivity, and ignores the curve geometry constraints. When the image quality of 2D fluoroscopy is poor, the method is prone to image noises, and cannot converge to a satisfactory accuracy. It is shown in [5] that false detection rate drops to 4% after 7 user clicks, but there are no further improvements even given more user clicks.

To address the aforementioned problems, this paper presents a novel graph based framework and a hyper-graph based extension for interactive detection. The method automatically detects curve segments, which, together with user interactions, are used to construct a graph. In the graph, the nodes represent the curve segments, and the edges represent the curve connectivity. The method then formalizes the curve detection as to find the shortest path in the graph given user inputs. To impose curve geometric constraints, our method introduces a hyper-graph, where the geometric constraints can be imposed as an inherent property. The presented work integrates the automatically detected curve segments, user interactions, and generic geometric constraints into one framework, and provides a smooth and quickly converged detection result. Extensive experiments demonstrate that with minimum human interactions involved, the method can detect various curve structures, such as guidewire and vessels, in 2D fluoroscopic images.

2 Interactive Detection Based on Graph Optimization

The presented interactive method consists of three components: automatically detecting curve segments, constructing a graph based on automatic detections and user interactions, and detecting a curve based on graph optimization. In this section, we first introduce the framework based on a conventional graph, and then extend it based on a hyper-graph in the next section.

2.1 Hierarchical Learning-Based Curve Segment Detection

The method adapts a hierarchical learning based method that is similar to the method presented in the work [1]. One difference is that the detectors used in this method only aim at detecting curve segments, not a whole curve. The detection includes two steps: first to detect piece-wise curve segments, and then to detect pair-wise connections between two curve segments. In the first step, the piece-wise segment detector detects a single piece of curves, namely a short line segment with a constant length. Such a curve segment has three parameters (x, y, θ) , where (x, y) is the segment center location and $\theta \in [-90, 90]$ is the segment orientation. Such a learning based detector is trained from offline collected data. All the points on the annotated curves are considered positive training samples, while negative samples are randomly obtained from the regions containing no curve structures. The piece-wise segment detector uses Haar features [6] and the PBT classifier [7]. To detect segments of different orientations, an image is rotated at discrete angles to search the rotation angles of curve segments. Figure 1(b) shows detected piece-wise curve segments of a guidewire.

Since the piece-wise detector usually produces a lot of false alarms, a pair-wise curve segment detector is used to prune false detections. In the pair-wise segment detection, every pair of two curve segments are classified if they belong to a curve, based on the image intensity and geometric features. Such geometric features include the center distance and angle difference between two segments. The same as the piece-wire segment detector, a PBT classifier is used to train the pair-wise curve segment detector. The probabilistic output from the pair-wise segment detector, noted as $p_{i,j}$, quantifies the connectivity between the i -th and j -th curve segments.

2.2 Graph Representation for Interactive Detection

In this method, a graph is used to organize all the detected curve segments, and the curve structure detection is then formalized as an optimization problem based on the graph. As shown in Figure 2 (a), each node in a graph represents a detected segment, and an edge between two nodes represents a possible connection between them. We denote the node as n_k for the k -th segment, and denote

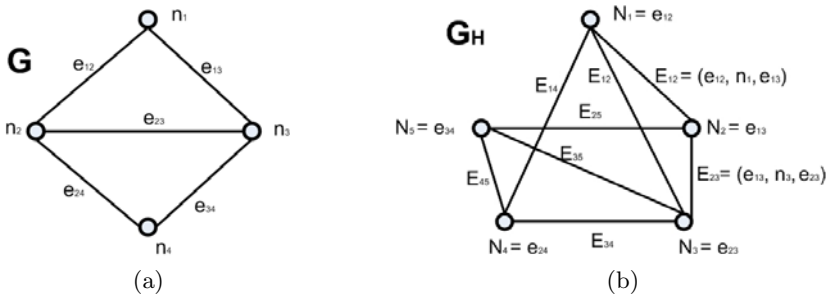


Fig. 2. Graph representation for curve detection (a) an original graph; (b) a hypergraph

the edge as e_{ij} for the detected connection between i -th and j -th node. Each edge e_{ij} is associated with a cost c_{ij} , while the cost c_k associated with the k -th node is usually ignored in the original graph.

The edge cost in our method is defined by the probabilistic outputs of the pair-wise detector, because it quantifies the connection between two nodes. Here we define the cost of an edge as in Eqn. (1):

$$c_{ij} = -\log(p_{i,j}). \quad (1)$$

By this definition, the higher the probability is, the lower the cost is. Such a definition allows for formalizing the interactive detection as an optimization problem in the graph, as in Eqn. (2):

$$\Gamma = \arg \min_L \sum_{(i,j) \in L} c_{i,j} = \arg \max_L \sum_{(i,j) \in L} \log(p_{i,j}), \quad (2)$$

where Γ is the detected curve, and L is a valid path in the graph. Therefore, the curve detection is to find a best path in the graph, given the edge cost defined as in Eqn. (1). Given two end points on the curve, one as the source and the other as the destination node, the curve detection can be solved by a classical shortest path algorithm, such as the Dijkstra's algorithm [8,9] used in this method.

In the interactive detection, the source and destination nodes are specified by the first two user clicks, e.g., the first input point as the source node, and the second as the destination node. Each user click will create a new curve segment, whose position is defined by the user click, and whose orientation is interpolated from neighboring detected segments using the extension field [10]. The source and destination segments generated from user clicks are then used in the shortest path computations. After the first two clicks, additional user clicks may be needed to improve an interactive detection result. The additional user interactions are used to generate new curve segments in the same way as mentioned above. The user interaction continues until a satisfactory result is obtained.

3 Extension to Hyper-Graph Based Optimization

The graph optimization provides a solid framework for interactive detections. However, the best path solved in Eqn. (2) only involves the minimization of costs based on the classification scores from the pair-wise segment detector. Geometric properties of the curves, such as curve smoothness, have not been taken into considerations. Without geometric constraints, the detected curves can easily be distracted by false detections, and could generate unwanted results. A problematic detection result is shown in Figure 3(b). Such path-related curve smoothness constraints are difficult to be incorporated in the conventional graph optimization algorithms, such as Dijkstra's. To address this issue, our method further extends an original graph to a hyper-graph which can accommodate geometric constraints in the detection framework.

The basic idea of constructing a hyper-graph is to model a segment of curve path, which involves more than a pair of nodes, as a hyper-edge (an edge in the

hyper-graph, denoted as $E_{i,j}$), and also to model an edge in the original graph as a hyper-node (a node in the hyper-graph, denoted as N_k) in the hyper-graph. A simple example is shown as in Figure 2. An original graph G is shown in Figure 2(a), and its corresponding hyper-graph G_H is shown in Figure 2(b). The hyper-node N_i is transformed from one of original graph's edges $e_{j,k}$, and a hyper-edge $E_{i,j}$ in G_H is transformed from a path ($e_{m,n} \rightarrow n_m \rightarrow e_{m,l}$) in G , where j, k, m, n, l are indices. In this way, a hyper graph is constructed from an original graph, and such a graph transformation is unique and reversible.

In the hyper-graph, a hyper-node inherits the edge cost in the original graph, i.e. $C_{N_k} = c_{ij}$ if N_k corresponds to e_{ij} . For the cost of a hype-edge, since a hype-edge corresponds to a part of path on a curve, the cost of the hype-edge is then defined to impose the desired geometric constraints. In this method, two types of geometric constraints, shortness and the smoothness, are used to define the edge cost as in Eqn. (3):

$$C_{E_{(i,j)}} = \alpha(1 + \cos(\angle(E_{(i,j)}))) + \beta \log\left(\frac{1}{1 + \exp(-|E_{(i,j)}|/B)}\right), \quad (3)$$

where $\angle(E_{(i,j)})$ is the angle at the center of curve path that corresponds to the hyper-node, and $|E_{(i,j)}|$ is the arc length of the path. The first term $1 + \cos(\angle(E_{(i,j)}))$ imposes the non-negative smoothness measurements of a curve path, and the second term $\log\left(\frac{1}{1 + \exp(-|E_{(i,j)}|/B)}\right)$ favors a shorter curve. The parameters α and β balance the two types of geometric constraints, and B is a parameter in the sigmoid function. These parameters are empirically set in this method.

Given the constructed hype-graph and the redefined costs on edges and nodes, a graph shortest path algorithm finds the optimal curve, based on the criteria form both pair-wise connectivity and geometric constraints, as in Eqn. (4):

$$\Gamma = \arg \min_L \sum_{N_k \in L \& E_{(i,j)} \in L} C_{N_k} + C_{E_{(i,j)}} \quad (4)$$

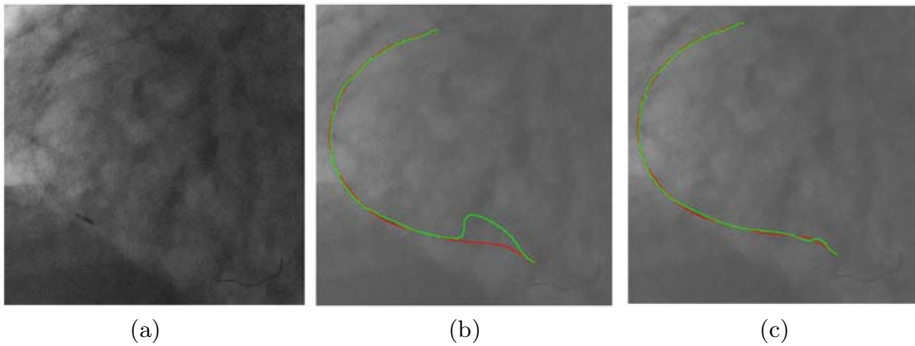


Fig. 3. Using hyper-graph to impose geometric constraints. (a) a frame; (b) the result based on the original graph; (c) the result based on the hyper-graph. The green lines represent interactive detections, and red lines are annotations.

-
- Given the original graph \mathbf{G} , the source node n_s , and the destination node n_d
 - Construct a hyper-graph \mathbf{G}_H
 1. Set each edge in \mathbf{G} as a hyper-node in \mathbf{G}_H , and copy the edge cost in \mathbf{G} as a node cost in \mathbf{G}_H
 2. Connect two hyper-nodes if they share a node in \mathbf{G} , and compute the hyper-edge cost as in Eqn. (3)
 - Find the shortest path in the hyper-graph \mathbf{G}_H
 1. Get a set of edges $\{e_{i,s}\}$ that connect to n_s in \mathbf{G} , and a set of edges $\{e_{j,d}\}$ that connect to n_d in \mathbf{G} .
 2. Compute a bundle of shortest distance. Each shortest distance is for a pair of hyper-nodes in \mathbf{G}_H . In each pair, one hyper-node corresponds to one of $\{e_{i,s}\}$, and the other corresponds to one of $\{e_{j,d}\}$
 3. Select the optimal path from the bundle of shortest paths
 4. Map the shortest path in \mathbf{G}_H back to the original path in \mathbf{G}
-

Fig. 4. Algorithm of hyper-graph based curve detection

Slightly different from the original graph, the single-source shortest path algorithm needs to be run multiple times to find a bundle of shortest paths in the hyper-graph, because each source node in the original graph corresponds to multiple edges in the hyper-graph, as shown in Figure 2. From the bundle of shortest paths, an optimal curve with the least cost is selected as the final detection result. Figure 4 summarizes the hyper-graph based detection algorithm. Figure 3 shows an example of the improved detection of using hyper-graph. The numerical evaluations are further conducted in Section 4.

When constructing a hyper-graph from an original graph, the size of hyper-graph grows linearly with the number of edges and nodes in the original graph. With the increased graph size, the computation time increases drastically. To address the computation problem, a two-stage method is used. In the two-stage method, the original graph and the hyper-graph methods are applied sequentially. In the first stage, a shortest path is found in the original graph. The curve segments that are far way from the initial detection are removed, therefore the hyper-graph built in the second stage is more compact. By using this two-stage detection strategy, the method can enjoy both the fast computation in the original graph and the benefits of geometric constraints in the hyper-graph. Figure 1(c) and (d) show an example of the two-stage detection.

4 Evaluations

The presented interactive method can be applied to detect various curve structures, such as catheters, vessels, and guidewires in 2D fluoroscopy. For the quantitative evaluations, we select a set of 500 frames over 50 fluoroscopic sequences, which mainly contain guidewires. The image set includes challenging scenes for curve detection, such as various shapes and low signal-to-noise ratio. Some of testing images and results are shown in Figure 5(a). All the wires in the

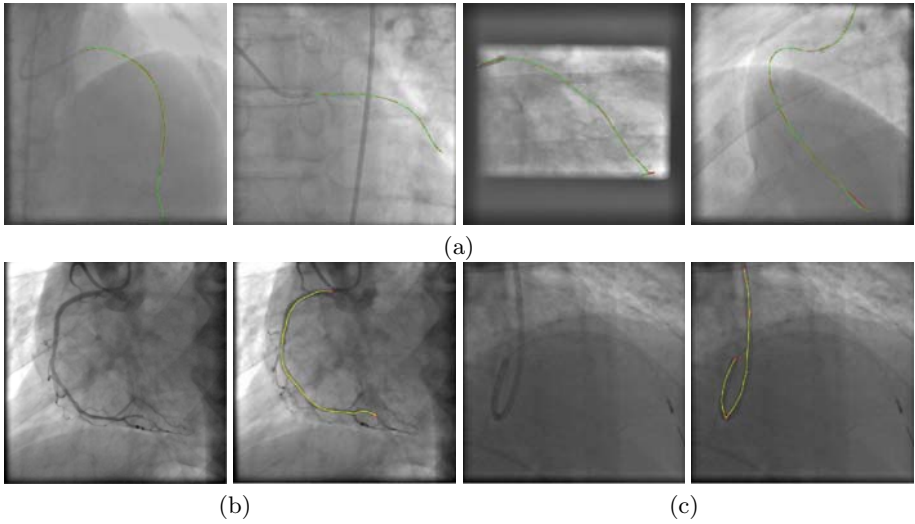


Fig. 5. Exemplar interactive detection results. (a) detected guidewires, where the green lines represent interactive detections with only two clicks, and red lines are annotations; (b) an image with a vessel branch in it, and the interactive detection result; (c) an image with a guiding catheter in it, and the interactive detection result.

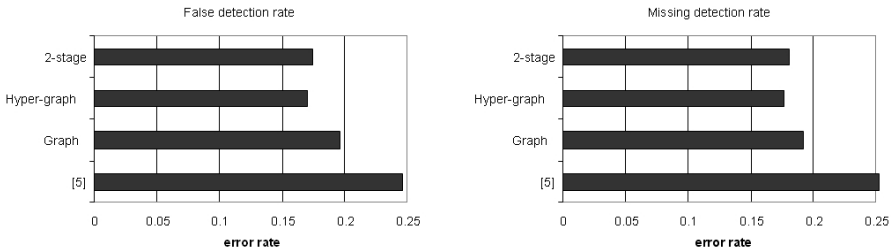


Fig. 6. Accuracy comparison of different methods with only 2 clicks

sequences are manually annotated as the ground truth. To measure the interactive detection accuracy, we compute two quantities: the false detection rate, and the missing detection rate. The false detection rate is the percentage of false detections that are at the distance of at least 3 pixels away from annotations. The missing detection rate is the percentage of annotated pixels that are at the distance of at least 3 pixels away from detection curves. We also measure the mean, standard deviation, and median of the distance between the detection results and annotation results. The experiments are conducted on a dual core 2.4 GHz Desktop PC with 3 GB of RAM.

In the first experiment setting, only two user clicks are provided, and the detection accuracy is measured under the minimum interaction. Figure 6 and Table 1 show the accuracy comparison between different methods with only two

Table 1. Quantitative evaluations on interactive detections with only two clicks

	Detection Error (in pixels)			Running time (in seconds)
	median	mean	std	(excluding segment detections)
5	10.7	18.7	6.7	0.19
Graph	8.3	15.1	6.7	0.02
Hyper-Graph	6.8	12.5	5.7	1.23
2-stage	7.2	13.2	6.1	0.78

Table 2. Quantitative evaluation on interactive detections with multiple clicks

Number of user clicks	2	3	4	5	6	7	8	9	10
False detection rate	17.8%	13.1%	8.4%	5.4%	3.8%	2.6%	2.0%	1.4%	1.0%
median (pixel error)	6.51	2.29	1.25	0.99	0.95	0.92	0.91	0.88	0.87
mean (pixel error)	7.58	3.72	2.06	1.36	1.19	1.07	1.06	0.96	0.94
std (pixel error)	13.60	8.76	5.16	2.19	1.83	1.14	1.55	0.66	0.74

user clicks. For a fair comparison, we select optimal parameters for each method. In our sensitivity test, the performance of all the methods are not sensitive to the parameter settings in a neighbor of optimal parameters. From Figure [6](#), it can be observed that the graph based methods perform better than the previous work [5](#), which gets an average error rate of 25% for the testing set containing diverse shapes. The method based on the hyper-graph, with an error rate of 17%, has further improved from the original graph based method which has an error rate of 19%. The computation time for the hyper-graph is slightly higher than that of the original graph. To get the advantage of accuracy from the hyper-graph and the advantage of efficiency from the graph, the two-stage algorithm gets the error rate of 17% with acceptable running time.

In the second experiment setting, multiple user clicks are provided. The users are asked to continuously provide clicks until the detection results are satisfactory. Table [2](#) summarizes the detection rates and pixel errors with multiple user clicks. The detection accuracy progresses with more user inputs. Usually after 4 clicks, the pixel error converges to less than 1 pixel. In the work [5](#), its performance does not improve after 7 clicks, which means that the performance of their algorithm may reach its limits due to the lack of geometric constraints. The method can also be applied to detect other curve structures such as vessel, and catheter. Two examples are shown in Fig. [5](#)(b) and (c).

5 Conclusion

The paper provides a novel and practical method that can effectively and efficiently detect curve structures in 2D images. A novel graph based framework has been presented to combine automatically detected curve segments, user interactions, and geometric constraints for a smooth interactive detection. The method

has been applied to the guidewire and vessel detections, and encouraging results have been obtained. The future work will be to apply the method to detect more types of curve structures, and to handle complex curve structures, such as loops.

Acknowledgement

We would like to thank Dr. Xun Xu for the discussions on the hyper-graph in Siemens Corporate Research.

References

1. Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: CVPR, Minneapolis, Minnesota, USA (2007)
2. Katz, N., Nelson, M., Goldbaum, M., Chaudhuri, S., Chatterjee, S.: Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Trans. Med. Imaging* 8(3), 263–269 (1989)
3. Mortensen, E.N., Barrett, W.A.: Intelligent scissors for image composition. In: SIGGRAPH 1995: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, pp. 191–198. ACM, New York (1995)
4. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: CVPR, Washington, DC, USA, pp. 755–762. IEEE Computer Society, Los Alamitos (2005)
5. Mazouer, P., Chen, T., Zhu, Y., Wang, P., Durlak, P., Thiran, J.-P., Comaniciu, D.: User-constrained guidewire localization in fluoroscopy. In: Proc. SPIE of Medical Imaging: Image Processing (2009)
6. Viola, P., Jones, M.: Robust real-time object detection. *International Journal of Computer Vision* 57(2), 137–154 (2004)
7. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: ICCV, pp. 1589–1596 (2005)
8. Corme, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: *Introduction to Algorithms*, 2nd edn. MIT Press / McGraw-Hill (2001)
9. Diestel, R.: *Graph Theory*, 3rd edn. Springer, Heidelberg (2005)
10. Guy, G., Medioni, G.: Inferring global perceptual contours from local features. *Int. J. Comput. Vision* 20(1-2), 113–133 (1996)

Automated Digital Dental Articulation

James J. Xia¹, Yu-Bing Chang^{1,2}, Jaime Gateno¹,
Zixiang Xiong², and Xiaobo Zhou¹

¹ The Methodist Hospital Research Institute, Houston, Texas, USA

{JXia, JGateno, XZhou}@tmhs.org

² Texas A&M University, College Station, Texas, USA

yubingchang@tamu.edu, zx@ece.tmau.edu

Abstract. Articulating digital dental models is often inaccurate and very time-consuming. This paper presents an automated approach to efficiently articulate digital dental models to maximum intercuspation (MI). There are two steps in our method. The first step is to position the models to an initial position based on dental curves and a point matching algorithm. The second step is to finally position the models to the MI position based on our novel approach of using iterative surface-based minimum distance mapping with collision constraints. Finally, our method was validated using 12 sets of digital dental models. The results showed that using our method the digital dental models can be accurately and effectively articulated to MI position.

Keywords: digital dental models, automated, digital dental articulation, collision avoidance.

1 Introduction

With the giant leap of computer technology, more and more dental offices are going to digital and replacing their traditional stone dental casts with digital dental models. The digital dental models are the exact replica of the teeth. They are usually generated by scanning dental impressions or stone dental models, or directly scanning the teeth intraorally. By incorporating the digital dental models into a 3D head model [1, 2], the orthodontic and orthognathic treatment can be entirely planned within a computer, and thus significantly improve the treatment outcome and decrease the planning time [3]. However, the utilization of digital dental models also creates a new problem in which the reestablishment of the dental occlusion to a maximum intercuspation (MI) position has become more difficult and time consuming than before. A main goal of the orthodontic and orthognathic treatment is to reestablish patient's occlusion. When doctors use plaster dental models to establish the occlusion, the physical action of aligning upper and lower dental models into MI position is quick and accurate, usually in a matter of seconds. The same is not true in the virtual world, where the dental arches are represented by two 3D images that lack collision constraints. The computer system does not stop the images from moving through each other once the models have made contact. In addition, the operator has no tactile feedback when articulating the digital models. Virtual articulation of an arch of 14 upper teeth against 14 lower teeth into

their best possible intercuspation is a complex task. Ideally, the 14 buccal cusps and 4 incisal edges of the mandibular teeth will make maximal contact against the corresponding fossae, marginal ridges and lingual surfaces of the maxillary teeth at MI position. At the same time, the palatal cusps of the maxillary teeth also need to make contact against the fossae and marginal ridges of the lower teeth. Moreover, the dental midlines should be coincidental, and the transverse relationship between the teeth should be appropriate. Finally, all of this needs to be accomplished without creating unwanted areas of overlap. Because of these difficulties, it usually takes close to an hour to achieve the “visually best possible” intercuspation in the computer. More importantly, it is almost impossible to be certain that what is seen in the computer represents the true best possible alignment. Therefore, there is an urgent need to develop a method that is capable of efficiently articulating the upper and lower digital dental models to MI position.

To this end, the purpose of this study is to develop an effective approach to automatically articulate the digital dental models. Our approach includes two steps. The first step is to position the models to an initial position based on dental curves and point matching algorithm. The second step is to finally position the models to the MI position based on iterative surface-based minimum distance mapping (ISMDM) algorithm with collision constraints. Finally, our method was validated using 12 sets of digital dental models.

2 Algorithm Development

2.1 Data Acquisition and Preparation

Three sets of stone dental cases were randomly selected from our dental model archive. An experienced doctor first hand articulated each set of the stone casts to MI position. They were then mounted on a specially designed mounting jig to keep their MI relationship. The mounted models were finally scanned using a 3D laser scanner (0.1mm of scanning accuracy) by a commercially available service (GeoDigm Corp, Chanhassen, MN). This resulted in a set of digital dental models at MI position and saved in stereolithography (.STL) format. Because of the triangulated surface of the model, the upper and lower digital models were slightly outwards expanded and penetrated to each other with a range of 0.08–0.20 mm at MI position, which did not exist in their physical form (stone casts). These scanned models served as a gold standard. Prior to the algorithm development, the lower models were disarticulated in 3D by randomly rotational and translational transformations.

2.2 Initial Alignment

The main purpose of initial alignment is to obtain approximate dental occlusion before two dental models are finally articulated to an accurate and collision-free position and orientation. When the two dental models are initially disarticulated and located at an arbitrary orientation and position in a *Cartesian* coordinate system, it is necessary to estimate a transformation to bring them relatively close to each other. Two pairs of corresponding curves are extracted and matched from the upper and lower dental models. The first pair is the buccal cusps of the lower arch (Fig.1b) corresponding to the

central groove of the upper arch (Fig.1a), while the second pair is the palatal cusps of the upper arch corresponding to the central groove of the lower arch. These curves can be viewed as 3D continuous curves (not necessarily fitting polynomial curves) along the dental arches. In this article we only use the first pair as an example.

Based on the above assumptions, we developed an automated approach to initially position the models. In the first step, we identify the feature points on cusps, incisal edges, central grooves, pits, and fossae to approximately represent the dental curves along the arches. In the second step, the dental curves of the upper and lower arches are matched using a point matching algorithm to complete the initial alignment.

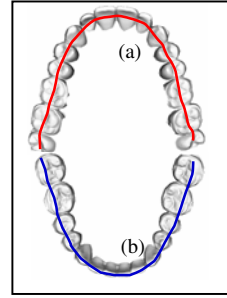


Fig. 1. Upper (a) and lower (b) dental curves

Step 1: Identification of Feature Points on Upper and Lower Occlusal Surfaces

A 2D range image (the heights of the digital model in the z - coordinate) is first calculated for lower arch. Based on the range image and the two-step curve fitting approach [4], we can then compute a 2D fourth-order polynomial dental fitting curve that fits buccal cusps and incisal edges of lower arch in least square. Finally, with the aid of 2D fitting curves, the 3D feature points of the lower occlusal surface can be extracted by detecting the peaks along the lower fitting curve. Similarly, the 3D feature points of upper occlusal surfaces can be identified by detecting the central groove along the upper fitting curve.

Step 2: Point Matching Algorithm

Let $\{p_i\}_{i=0}^{N-1}$ and $\{q_j\}_{j=0}^{K-1}$ be sets of 3D feature points of the upper and lower dental models obtained in Step 1. The two sets of points are matched by applying our point matching approach as if the dental curves fit together when dental models are in the MI. Our point matching algorithm (an improved iterative closest points algorithm [5]) is based on graduated assignment combining “softassign” method [6] and a weighted least squares optimization [7]. The initial alignment becomes to find a transformation (R, t) and a correspondence between two sets of feature points $\{p_i\}$ and $\{q_j\}$, and to minimize an energy function in the standard point matching algorithm [6].

2.3 Final Alignment

After the dental models are aligned to an approximate occlusion, they can be finally digitally articulated using our algorithm, Iterative Surface-based Minimum Distance Mapping (ISMDM). The criterion based on maximal contact of the teeth at MI is the key to develop the ISMDM. The dental model articulation can be modeled by consecutive executions of translations and rotations with continuous changes of rotational origin on the lower dental model. In order to automatically achieve maximal contact between upper and lower teeth and reach the MI position, we model this movement by iteratively minimizing distance of surfaces between lower and upper teeth. This method is based on the idea of the iterative closest point algorithm [8] that is generally

used in shape matching, registration, and alignment of two similar datasets from the same object. In addition, an important component in our ISMDM method is that we add constraints to prevent the two surfaces from overlap [9]. The detailed computational algorithms are described as follows.

The Modelling of Dental Occlusion

Let $\{\mathbf{u}_i\}_{i=0}^{M-1}$ and $\{\mathbf{v}_j\}_{j=0}^{J-1}$ be 2 sets of M and J vertices in the meshes of the digital upper and lower dental models, respectively. In the following, we assume the upper model is in a static position. The transformations and rotations are only performed on the lower dental model. The transformed \mathbf{v}_j is modeled as:

$$\mathbf{v}'_j(\tilde{\mathbf{R}}, \tilde{\mathbf{t}}) \equiv \tilde{\mathbf{R}}(\mathbf{v}_j - \tilde{\mathbf{o}}) + \tilde{\mathbf{o}} + \tilde{\mathbf{t}} \quad (1)$$

where $\tilde{\mathbf{o}}$ is a rotational origin (the pivot point) of the rotation matrix $\tilde{\mathbf{R}}$, and $\tilde{\mathbf{t}}$ is the translation vector. Maximizing contact area is equivalent to maximizing the number of contacting vertices in $\{\mathbf{v}_j\}$. However, not every vertex in $\{\mathbf{v}_j\}$ will make contact when the models are in the MI. Those contact areas are even more difficult to be predicted precisely. Therefore, we model the distance of surfaces between lower and upper teeth as:

$$d_S \equiv \sqrt{\frac{1}{J} \sum_{j=0}^{J-1} \|\mathbf{u}_{i_j} - \mathbf{v}_j\|^2}. \quad (2)$$

\mathbf{u}_{i_j} is a point closest to \mathbf{v}_j and given by:

$$\mathbf{u}_{i_j} = \arg \min_{\mathbf{u} \in \{\mathbf{u}_i\}} \|\mathbf{u} - \mathbf{v}_j\| \quad (3)$$

where $i_j \in \{0, 1, \dots, M-1\}$. Instead of directly maximizing contact area, we increase the chances of making contact by minimizing d_S .

Because the 2 digital models should not penetrate to each other, adding collision constraints is the most important step in digital dental occlusion. The avoidance of collision is formulated as constraints and will be incorporated into the optimization programming. Let \mathcal{P}_j be a plane with a unit normal vector \mathbf{n}_j and a point \mathbf{r}_j on it. When the transformed vertex $\mathbf{v}'_j(\tilde{\mathbf{R}}, \tilde{\mathbf{t}})$ is not allowed to be at the opposite side of the plane \mathcal{P}_j , the constraint can be expressed as

$$\left(\mathbf{v}'_j(\tilde{\mathbf{R}}, \tilde{\mathbf{t}}) - \mathbf{r}_j\right)^T \mathbf{n}_j > 0. \quad (4)$$

\mathbf{r}_j can be given by

$$\mathbf{r}_j = \mathbf{u}_{i_j} - \delta \mathbf{n}_j \quad (5)$$

where δ is allowable penetration depth (0.1mm of tolerance to compensate outwards expansion of triangulated surface) for the lower teeth. The unit normal vector \mathbf{n}_j can be chosen by calculating the average normal at the vertex \mathbf{u}_{i_j} . Since some areas

between upper and lower teeth will never make contact during the MI, a large number of constraints added to the algorithm may be redundant. In order to reduce the number of constraints, it is not necessary to add a constraint to a point pair v_j and u_i , if the distance between them is beyond a threshold ρ .

Minimization of the Distance of Occlusal Surfaces and the Algorithm

Given a rotational origin \tilde{o} , we calculate the rotation matrix \tilde{R} and the translation vector \tilde{t} which minimize

$$d_S^2(\tilde{R}, \tilde{t}) \equiv \sum_{j=0}^{J-1} \|u_{i_j} - v'_j(\tilde{R}, \tilde{t})\|^2, \quad \text{subject to } (v'_j(\tilde{R}, \tilde{t}) - r_j)^T n_j > 0 \tag{6}$$

The rotation matrix consists of non-linear terms which can be linearized by small-angle approximation [10, 11]. When the 2 dental models are getting occluded, the increment needed to seat the lower model will gradually become smaller. Therefore, errors caused by this approximation will become less significant. Approximate the rotational matrix \tilde{R} by linearizing it as

$$\tilde{R} \equiv \begin{pmatrix} 1 & -\theta_z & -\theta_y \\ \theta_z & 1 & -\theta_x \\ \theta_y & \theta_x & 1 \end{pmatrix} \tag{7}$$

where θ_x , θ_y , and θ_z are rotational angles with respect to x -, y -, and z - axes. Define $\theta = (\theta_x, \theta_y, \theta_z)^T$ and L_j as:

$$L_j \equiv \begin{pmatrix} 0 & -v_{j,z} + \tilde{o}_z & -v_{j,y} + \tilde{o}_y \\ -v_{j,z} + \tilde{o}_z & 0 & v_{j,x} - \tilde{o}_x \\ v_{j,y} - \tilde{o}_y & v_{j,x} - \tilde{o}_x & 0 \end{pmatrix}. \tag{8}$$

$(\tilde{R} - I)(v_j - \tilde{o})$ can be rewritten as: $(\tilde{R} - I)(v_j - \tilde{o}) = L_j \theta$. Let $b_j = v_j - u_{i_j}$, $x \equiv (\tilde{t}^T, \theta^T)^T = (\tilde{t}_x, \tilde{t}_y, \tilde{t}_z, \theta_x, \theta_y, \theta_z)^T$, and $\tilde{L}_j \equiv [L_j \ I]$. The objective function in Equation (6) becomes:

$$d_S^2(\tilde{R}, \tilde{t}) = x^T \left(\sum_{j=0}^{J-1} \tilde{L}_j^T \tilde{L}_j \right) x + 2 \left(\sum_{j=0}^{J-1} b_j^T \tilde{L}_j \right) x + \sum_{j=0}^{J-1} \|b_j\|^2. \tag{9}$$

With the linearization of rotation matrix, the objective function $d_S^2(\tilde{R}, \tilde{t})$ becomes a quadratic form, and (4) becomes a linear constraint. The minimization of (6) can be solved by quadratic programming.

3 Validations and Results

Twelve sets of the stone dental casts were used to validate our approach. They were randomly selected from our dental model archive. The models used for algorithm development were excluded. The selection criteria included no early contact and a

stable occlusion. All the models had relatively normal and full dentition except 4 pairs were partial edentulous. The models were scanned at MI position using the method described in 2.1. These digital dental models at MI position served as a control group.

The models were positioned in *Cartesian* coordinate system. The origin O was the centroid of the boundary box of the lower model (Fig 2). Three landmarks commonly used in clinic were digitized. They were mesiobuccal cusps of the first right and left molars (A and B), and central dental midline (C). The coordinates of each landmark were used later to compare with the same landmarks in the experimental group.

In order to establish an experimental group, lower models and their landmarks were duplicated. Because the models were scanned at MI position, it was necessary to disarticulate the duplicated lower model from its original position. The lower model was first randomly rotated around the X -, Y -, Z - axes angles between $[-\frac{1}{3}\pi, \frac{1}{3}\pi]$, respectively. It was then randomly translated in millimeters between $[-20, 20]$ along the X -, Y -, and Z - axes, respectively. During the rotational and translational transformations, the landmarks were “glued” on the lower model and transformed with it. We believed that these transformations were randomly enough to disarticulate the lower models from their MI position. These disarticulated lower models served as an experimental group. The centroid o , the x -, y - and z - coordinates, and the landmarks A , B and C in the control model became centroid O' , the x' -, y' - and z' - coordinates, and the landmarks A' , B' and C' in the experimental model.

In the experimental group, the upper and lower models were first articulated using our initial alignment algorithm, followed by our ISMDM algorithm with the following parameters: $S = 60$ (iterations of ISMDM), $\rho = 0.2$ mm and $\delta = 0.1$ mm. Fig.3 showed a plot of average distance of surface d_s versus iteration k for all 12 sets of the models using the ISMDM algorithm. During the articulation process, the upper model remained static while the lower model sought its MI position. In the initial alignment, the point match algorithm is applied only to the feature points of the models. This may cause collisions between the lower and upper teeth. After they are fed into the ISMDM, the collision constrain will force the lower model out of the upper model until the models are free of collision. This is why the average distance d_s between surfaces of lower and upper models is smaller at the beginning than it is at the final position after 60 iterations. The entire process was completed using a regular office PC computer with an Intel P4 2.2GHz CPU and 4GB memories.

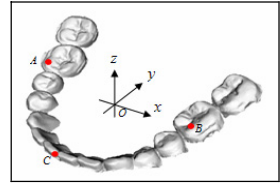


Fig. 2. Coordinate system

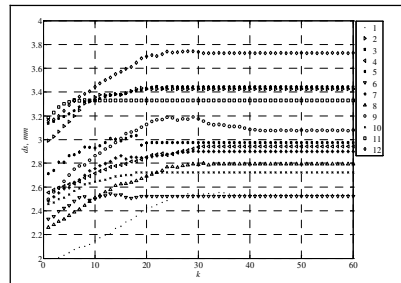


Fig. 3. Articulation results using ISMDM algorithm

Finally, the validation was completed by calculating the transitional and rotational deviations of the lower models between the experimental and control groups. Based on our clinical experience, there would be no clinical significance if the translational deviation of the lower dental models between the control the experimental groups is less than 0.5mm in each X , Y and Z direction and the angular deviation is less than 1° on sagittal, coronal, and axial planes, respectively.

Translational Deviations between the Experimental and Control Groups

The means and standard deviations (SD) of the translational differences between the experimental and control models were computed. The mesiobuccal cusp of the first right molar (A', A) and the first left molar (B', B), the central dental midline (C', C), and the centroid (O', O) were computed in x -, y -, and z -axis, respectively (Table 1). It indicated that the models were articulated successfully with a small degree of translational deviation with no clinical significance.

Table 1. Translational Deviations (mean \pm SD, calculated in mm)

	Initial alignment			Final alignment (ISMDM)		
	x	y	z	x	y	z
(A', A)	-0.0938 \pm 0.5873	0.4451 \pm 1.2834	1.6366 \pm 0.6841	-0.0661 \pm 0.3828	-0.1206 \pm 0.5621	0.1318 \pm 0.2185
(B', B)	-0.1471 \pm 0.5935	0.5708 \pm 1.0660	1.6332 \pm 0.5336	-0.0697 \pm 0.3794	-0.1793 \pm 0.4606	0.1707 \pm 0.1151
(C', C)	-0.0596 \pm 0.7824	0.5433 \pm 0.7382	0.5394 \pm 0.6998	-0.1088 \pm 0.3380	-0.1449 \pm 0.2059	0.1214 \pm 0.2830
(O', O)	-0.0917 \pm 0.4310	0.5912 \pm 0.6965	1.3714 \pm 0.4399	-0.0735 \pm 0.2926	-0.1481 \pm 0.1829	0.1464 \pm 0.1026

Angular Deviations between the Experimental and Control Groups

The angular differences between the experimental and control models were computed on sagittal ($Y-O-Z$), coronal ($X-O-Z$), and axial ($X-O-Y$) planes, respectively. In order to compute the angular deviation, the models in experimental group were moved translationally so that the centroid O' was matched to the centroid O in the control group. Afterwards, z' -axis was projected onto the $Y-O-Z$ plane. The sagittal angular deviation $\hat{\omega}_x$ was defined by the angle between the projected z' -axis and z -axis on the $Y-O-Z$ plane. By the same token, by projecting the z' -axis onto the $X-O-Z$ plane, the coronal angular deviation $\hat{\omega}_y$ was defined by the angle between the projected z' - and z -axis on the $X-O-Z$ plane. Furthermore, by projecting the y' -axis onto the $X-O-Y$ plane, the axial angular deviation $\hat{\omega}_z$ was defined by the angle between the projected y' -axis and y -axis on the $X-O-Y$ plane. Finally, the means and SDs of the angular deviations between the experimental and control

Table 2. Angular Deviations

	Initial alignment	Final alignment
$\hat{\omega}_x$	-2.3016 \pm 2.0818	-0.0850 \pm 0.7459
$\hat{\omega}_y$	0.0102 \pm 0.4752	-0.0624 \pm 0.2322
$\hat{\omega}_z$	-0.1507 \pm 2.2328	0.0717 \pm 1.1453

models were computed (Table 2). It indicated that the models are successfully articulated with only a small degree of rotational deviation that had no clinical significance.

4 Discussion

We have developed an automated approach to digitally articulate dental models. This approach consists of two major steps. The first step is the initial alignment, which using the point match algorithm to match the feature points of dental curves in order to bring the models relatively close to each other. The second step is the final alignment, which uses the ISMDM algorithm to minimize the average distance of surfaces of the models in order to articulate the upper and lower models to the MI without overlapping. This approach has been validated using 12 sets of the dental models. The results showed the models were successfully articulated with a small degree of deviation that did not have clinical significance.

Our approach is robust. First, the initial alignment algorithm itself can bring the models to a closed position to the final occlusion. This can significantly reduce the number of executions of nearest point searching of ISMDM. Second, each lower model can be effectively docked to a final occlusion after the average distance of surface d_s converges at $k < 30$ using our ISMDM algorithm. The deviations are small enough to meet clinical standard. Third, our ISMDM algorithm has successfully overcome the notorious uncontrollable overlapping problem between the upper and the lower models. This is done by applying linear constraints and allowable tolerance of penetration depth δ . In this experiment, we set the penetration depth was 0.1mm. Fourth, as indicated in validation, our ISMDM algorithm can also be used to articulate the partially edentulous models.

Finally, our approach is different from others. Hiew *et al.* [9] used the right and posterior surfaces of the model bases, rather than based on the occlusal criteria, to perform the dental model alignment. Zhang *et al.* [12] designed a two-stage occlusal analysis algorithm to manually alignment the models in the computer. Finally, DeLong *et al.* [13] utilized a “3-point alignment” method by first identifying 3 pairs of contacting points on both the upper and lower stone models, digitizing them onto the digital models, and finally matching the corresponding points using a fitting algorithm to bring the digital models together. The results were visually checked and process was repeated until the visual outcome was satisfactory. We found it is almost impossible to be certain that what is seen in the computer truly represents the best possible alignment. Comparing to the above methods, our method is practical and can be immediately used in the patient treatment planning process.

References

- [1] Gateno, J., Xia, J.J., Teichgraber, J.F., Christensen, A.M., Lemoine, J.J., Liebschner, M.A., Gliddon, M.J., Briggs, M.E.: Clinical feasibility of computer-aided surgical simulation (CASS) in the treatment of complex cranio-maxillofacial deformities. *J. Oral. Maxillofac Surg.* 65, 728–734 (2007)

- [2] Xia, J.J., Gateno, J., Teichgraber, J.F.: New Clinical Protocol to Evaluate Craniomaxillofacial Deformity and Plan Surgical Correction. *Journal of Oral and Maxillofacial Surgery* 67, 2093–2106 (2009)
- [3] Xia, J.J., Gateno, J., Teichgraber, J.F., Christensen, A.M., Lasky, R.E., Lemoine, J.J., Liebschner, M.A.: Accuracy of the computer-aided surgical simulation (CASS) system in the treatment of patients with complex craniomaxillofacial deformity: A pilot study. *J. Oral. Maxillofac Surg.* 65, 248–254 (2007)
- [4] Kondo, T., Ong, S.H., Foong, K.W.C.: Tooth segmentation of dental study models using range images. *IEEE Transactions on Medical Imaging* 23, 350–362 (2004)
- [5] Besl, P.J., McKay, H.D.: A method for registration of 3-D shapes. *IEEE Trans Pattern Analysis and Machine Intelligence* 14, 239–256 (1992)
- [6] Gold, S., Rangarajan, A., Lu, C.P., Pappu, S., Mjolsness, E.: New algorithms for 2D and 3D point matching pose estimation and correspondence. *Pattern Recognition* 31, 1019–1031 (1998)
- [7] Walker, M.W., Shao, L., Volz, R.A.: Estimating 3-D location parameters using dual number quaternions. *CVGIP: Image Understanding* 54, 358–367 (1991)
- [8] Hajnal, J.V., Hawkes, D.J., Hill, D.: *Medical image registration*. CRC, Boca Raton (2001)
- [9] Hiew, L.T., Ong, S.H., Foong, K.W.C.: Optimal Occlusion of Teeth. In: 9th International Conference on Control, Automation, Robotics and Vision, ICARCV 2006, pp. 1–5 (2006)
- [10] Milenkovic, V.J., Schmidl, H.: Optimization-based animation. In: *Proceedings of the 28th Annual Conference on Computer graphics and interactive techniques*, pp. 37–46 (2001)
- [11] Zhang, L., Kim, Y.J., Varadhan, G., Manocha, D.: Generalized penetration depth computation. *Computer-Aided Design* 39, 625–638 (2007)
- [12] Zhang, C., Chen, L., Zhang, F., Zhang, H., Feng, H., Dai, G.: A New Virtual Dynamic Dentomaxillofacial System for Analyzing Mandibular Movement, Occlusal Contact, and TMJ Condition. In: Duffy, V.G. (ed.) *HCI 2007 and DHM 2007*. LNCS, vol. 4561, pp. 747–756. Springer, Heidelberg (2007)
- [13] DeLong, R., Ko, C.C., Anderson, G.C., Hodges, J.S., Douglas, W.H.: Comparing maximum intercuspal contacts of virtual dental patients and mounted dental casts. *J. Prosth. Dent.* 88, 622–630 (2002)

Image-Based Respiratory Motion Compensation for Fluoroscopic Coronary Roadmapping

Ying Zhu*, Yanghai Tsin, Hari Sundar, and Frank Sauer

Siemens Corporate Research,
755 College Road East, Princeton, NJ 08540, USA
yingzhu@siemens.com

Abstract. We present a new image-based respiratory motion compensation method for coronary roadmapping in fluoroscopic images. A temporal analysis scheme is proposed to identify static structures in the image gradient domain. An extended Lucas-Kanade algorithm involving a weighted sum-of-squared-difference (WSSD) measure is proposed to estimate the soft tissue motion in the presence of static structures. A temporally compositional motion model is used to deal with large image motion incurred by deep breathing. Promising results have been shown in the experiments conducted on clinical data.

1 Introduction

Motion compensation is an important issue for image guided coronary angioplasty procedures. Besides live fluoroscopic images, imaging techniques such as coronary roadmapping and multimodal image fusion have been introduced to provide clinicians additional image guidance. In coronary roadmapping [1,2], a dynamic coronary roadmap is obtained from dye-injected images showing vascular structures under cardiac motion. When the contrast medium disappears, the 2D roadmap can be superimposed on live fluoroscopy to provide immediate feedback in properly directing guidewires or placing stents. Similarly, 3D vessel roadmap extracted from cardiac CTA or MRA can be overlaid on live fluoroscopy [3,4,7], to provide detailed vessel information such as calcification and tortuosity which are important success factors for percutaneous coronary interventions. In both cases, reliable motion compensation is required to dynamically move 2D or 3D coronary roadmaps to properly match live fluoroscopy, especially when the contrast medium has disappeared and the coronary structures are not visible from fluoroscopic images.

Cardiac and respiratory motion are the main sources of motion observed in coronary fluoroscopic images. The technique of ECG gating is commonly used to deal with cardiac motion, where roadmap images are overlaid on cardiac gated frames. Motion over the respiratory cycle is generally less reproducible and there are a few drawbacks associated with respiratory gating [5]. Different motion models including translation, rigid body and affine transformations

* Corresponding author.

as well as statistical models have been investigated by several studies [5,6,7], to characterize the effect of respiration on the heart and coronary arteries. In contrast to the previous work which rely on 3D volumetric data or biplane acquisition, we explore a 2D image-based approach and present a new method to estimate the image motion incurred by respiration from monoplane fluoroscopic images. The motion estimates allow dynamic overlay of coronary roadmaps to match live fluoroscopy.

Image-based motion estimation has been explored in the context of digital subtraction angiography [11,12] which dealt mainly with cardiac motion correction. Image-based cardiac and respiratory cycle synchronization techniques have been discussed in [8]. Other techniques rely on guidewire tracking [9,10] to locate a coronary artery in fluoroscopic images. An interesting observation about cardiac fluoroscopy is that regardless of whether the contrast medium, guidewires or other devices are present or not, image motion of the soft tissues of the heart is consistently visible and measurable, which suggests the motion of the coronary arteries especially when the contrast medium is not seen. This motivates us to explore the motion information embedded in the broader image region of the heart, rather than focusing on local anatomical landmarks or devices. In addition, mixed layers of static bone structures and moving soft tissues in cardiac fluoroscopic images pose a challenge in recovering soft tissue motion. Our proposed method of image-based respiratory motion compensation through recovering soft tissue motion is a new idea for coronary roadmapping using monoplane fluoroscopic images. To our knowledge, it is also the first attempt to recover soft tissue motion with a special handling of static structures.

2 Method

In X-ray imaging, image intensity is determined by the energy flux which undergoes exponential attenuation through layers of tissues. With logarithmic postprocessing, the intensity can be described as an additive superposition of multiple tissue layers undergoing different movements. Previous work focused on separating transparent layers from one or multiple images [13,15,16], based on the assumption that the layers remain static, or their motion is either known beforehand or irrelevant to layer separation. Jointly recovering layers and their motion from fluoroscopic images remains an open problem.

2.1 Image-Based Motion Estimation

For motion compensation, the goal is to extract the coronary motion incurred by respiration. We consider a simplified model with two main layers, a static structure layer including bone tissues and a moving soft tissue layer of the heart including coronary arteries. Denote $I_t(\mathbf{x})$, $I^s(\mathbf{x})$ and $I_t^d(\mathbf{x})$ as the intensity value of pixel \mathbf{x} in the fluoroscopic image, the static layer and the soft tissue layer at time t respectively. They are related through additive superposition.

$$I_t(\mathbf{x}) = I^s(\mathbf{x}) + I_t^d(\mathbf{x}) \quad (1)$$

In coronary roadmapping, a 2D or 3D coronary roadmap is initially overlaid on one or multiple dye-injected fluoroscopic images through vessel-based registration [4]. These fluoroscopic images are referred to as the reference images. Once the contrast medium disappears, we perform motion compensation by recovering soft tissue motion between the reference images and live fluoroscopic images. The recovered motion is used to appropriately place the roadmap in the dye-free live images. Denote $I_{t_0}(\mathbf{x})$ as a reference image. An incoming live image I_t is related to I_{t_0} through soft tissue motion

$$I_{t_0}(\mathbf{x}) = I^s(\mathbf{x}) + I_{t_0}^d(\mathbf{x}) \quad (2)$$

$$I_t(\mathbf{x}) = I^s(\mathbf{x}) + I_{t_0}^d(W(\mathbf{x}; P)) \quad (\mathbf{x} \in \Omega) \quad (3)$$

$W(\mathbf{x}; P)$ denotes the model characterizing soft tissue motion between t_0 and t with parameters P . Ω denotes the image area of the heart exclusive of the coronary arteries due to contrast disappearance. Our problem is to estimate the soft tissue motion $W(\mathbf{x}; P)$ from I_{t_0} and I_t .

Without the layer of static structures I^s , we have $I_{t_0} = I_{t_0}^d$, $I_t = I_t^d$ and $I_t(\mathbf{x}) = I_{t_0}(W(\mathbf{x}; P))$. Motion estimation can be solved by the well-known Lucas-Kanade (LK) algorithm [14], which uses a Gauss-Newton gradient descent method to minimize the sum of squared difference (SSD)

$$SSD(P) = \frac{1}{N_{\Omega, P}} \sum_{\mathbf{x} \in \Omega_P^{-1}} \|I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P))\|^2 \quad (4)$$

where $\Omega_P^{-1} = W^{-1}(\Omega; P)$ and $N_{\Omega, P}$ is the number of pixels in Ω_P^{-1} . The motion parameters are iteratively updated with small increments.

$$\begin{aligned} \delta P &= H^{-1} \sum_{\mathbf{x} \in \Omega_P^{-1}} \nabla I_{t_0}^T \left(\frac{\partial W}{\partial P} \right)^T (I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P))) \\ (H &= \sum_{\mathbf{x} \in \Omega_P^{-1}} \left(\frac{\partial W}{\partial P} \nabla I_{t_0} \right) \left(\frac{\partial W}{\partial P} \nabla I_{t_0} \right)^T \end{aligned} \quad (5)$$

This involves the calculation of image gradients ∇I_{t_0} and the Jacobian of the motion transformation function $\frac{\partial W}{\partial P}$.

2.2 Dealing with Static Structures

With a layer of static structures such as bone tissues, visible skin markers and devices, the image intensities I_t , I_{t_0} have an added component I^s which remains static over time. Without a proper treatment, the static structures would have an adverse effect on the estimation of soft tissue motion. Here we present a novel approach to deal with static structures in the estimation of soft tissue motion. Note that the motion parameters obtained by (5) are essentially determined by two gradient terms, the spatial image gradient ∇I_{t_0} and the temporal image gradient $I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P))$.

$$\begin{aligned} \nabla I_{t_0} &= \nabla I_{t_0}^d + \nabla I^s \\ I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P)) &= I_t^d(\mathbf{x}) - I_{t_0}^d(W(\mathbf{x}; P)) + I^s(\mathbf{x}) - I^s(W(\mathbf{x}; P)) \end{aligned} \quad (6)$$

Ideally the soft tissue motion should be estimated from the dynamic components $\nabla I_{t_0}^d$ and $I_t^d(\mathbf{x}) - I_{t_0}^d(W(\mathbf{x}; P))$, which would require knowing I^s exactly to remove ∇I^s and $I^s(\mathbf{x}) - I^s(W(\mathbf{x}; P))$ from the spatial and temporal gradient terms. Instead of requiring the full knowledge of I^s , we explore the gradient fields $\nabla I_t, \nabla I_{t_0}$ to identify local image structures with strong image gradients which are likely to come from static structures. This is motivated by the fact that salient image structures characterized by high gradient values are sparse and distributed locally in I^s . For the vast homogeneous areas in I^s , ∇I^s is negligible, i.e. $\nabla I^s \approx 0, I^s(W(\mathbf{x}; P)) \approx I^s(\mathbf{x})$ given that the motion is small, thus we have

$$\begin{aligned} \nabla I_{t_0}(\mathbf{x}) &\approx \nabla I_{t_0}^d(\mathbf{x}) \\ I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P)) &\approx I_t^d(\mathbf{x}) - I_{t_0}^d(W(\mathbf{x}; P)) \end{aligned} \tag{7}$$

This means that if we ignore the local image structures that have high gradient values and remain static across I_t and I_{t_0} , we can obtain an approximate solution for soft tissue motion using I_t and I_{t_0} .

To identify the local static structures with strong gradients, motivated by [16], we compute in the gradient field the local correlation coefficient between images acquired at different times.

$$\rho(\mathbf{x}) = \frac{Cov(|\nabla I_t(\mathbf{x}')|, |\nabla I_{t_0}(\mathbf{x}')|)}{\sqrt{Var(|\nabla I_t(\mathbf{x}')|)Var(|\nabla I_{t_0}(\mathbf{x}')|)}} \Big|_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} \quad (-1 \leq \rho(\mathbf{x}) \leq 1) \tag{8}$$

$Cov(|\nabla I_t|, |\nabla I_{t_0}|)$ and $Var(I_{t/t_0})$ are the covariance and variance terms computed in a local neighborhood $\mathcal{N}(\mathbf{x})$. The correlation coefficient $\rho(\mathbf{x})$ is bounded between -1 and 1 . The more consistent the local structures are across images, the higher the correlation coefficient is. $\rho(\mathbf{x}) = 1$ when the local structures are exactly the same in both images. In practice, when multiple incoming fluoroscopic images or reference images are available, the local correlation coefficient over multiple images is computed as the statistical mean of the local correlation coefficient between every image pair.

Using the local correlation coefficient, we propose to estimate soft tissue motion by minimizing a weighted SSD (WSSD) term defined as follows.

$$WSSD(P) = \frac{1}{\mathcal{K}} \sum_{x \in \Omega_P^{-1}} \kappa(\mathbf{x}) \kappa(W(\mathbf{x}; P)) \|I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P))\|^2 \tag{9}$$

$$\kappa(\mathbf{x}) = \frac{1}{2}(1 - \rho(\mathbf{x})) \quad (0 \leq \kappa(\mathbf{x}) \leq 1) \tag{10}$$

$$\mathcal{K} = \sum_{x \in \Omega_P^{-1}} \kappa(\mathbf{x}) \kappa(W(\mathbf{x}; P))$$

The weighting function $\kappa(\mathbf{x})$ determines the contribution from each pixel to the estimation of soft tissue motion. For static structures across multiple images, the local correlation coefficient $\rho(\mathbf{x})$ is close to 1 and $\kappa(\mathbf{x})$ is close to 0 , and in the gradient domain their contribution to the WSSD is negligible. Therefore the motion parameters obtained by the WSSD method are determined mainly by the gradient structures of the moving tissues. To solve the WSSD problem, we extend the LK algorithm and iteratively update the motion parameters by

$$\delta P = \tilde{H}^{-1} \sum_{x \in \Omega_P^{-1}} \nabla I_{t_0}^T \left(\frac{\partial W}{\partial P} \right)^T \kappa(\mathbf{x}) \kappa(W(\mathbf{x}; P)) (I_t(\mathbf{x}) - I_{t_0}(W(\mathbf{x}; P))) \tag{11}$$

$$\tilde{H} = \sum_{x \in \Omega_P^{-1}} \kappa(\mathbf{x}) \kappa(W(\mathbf{x}; P)) \left(\frac{\partial W}{\partial P} \nabla I_{t_0} \right) \left(\frac{\partial W}{\partial P} \nabla I_{t_0} \right)^T$$

Note the warped weighting function $\kappa(W(\mathbf{x}; P))$ suppresses the contribution from the static structures in the warped image $I_{t_0}(W(\mathbf{x}; P))$, similar to what $\kappa(\mathbf{x})$ does to $I_t(\mathbf{x})$. Affine motion model is used in our experiments to describe the image motion of the soft tissues incurred by respiration. To improve the motion capture range, we adopt a pyramid implementation to estimate and propagate image motion from coarse to fine resolutions.

Fig. (II) shows how the proposed WSSD-based method works with static structures. We started with a real coronary image as the reference and introduced rotation, scaling and translation to generate a test image. A layer of static grid structures was added to the reference image and the test image through (II). The weighting function computed from the local correlation coefficients well captures the static grid structures. A vesselness map generated by the approach presented in [12] was used to simulate the roadmap overlays. Compared to the standard SSD-based method, the WSSD-based method performs more accurate motion compensation when static structures are present and the roadmap overlay matches the actual coronary arteries more precisely.

2.3 Compositional Motion Model

To compensate large image motion incurred by breathing, we introduce a temporally compositional motion model. Assume that the roadmap image is initially overlaid on a reference image I_{t_0} acquired at t_0 and for incoming fluoroscopic images, respiratory motion compensation is carried out on cardiac gated frames. At time t_1 , motion compensation is performed between I_{t_1} and I_{t_0} to transform the roadmap image to match I_{t_1} . Denote the motion between I_{t_1} and I_{t_0} as $W(\mathbf{x}; P_{1,0})$. As the respiratory phase moves away from t_0 , for instance at t_2 , the soft tissue motion observed in the fluoroscopic image can be quite large, which makes direct motion estimation between I_{t_2} and I_{t_0} difficult. Instead, we first estimate the motion $W(\mathbf{x}; P_{2,1})$ between I_{t_2} and I_{t_1} assuming the motion between I_{t_2} and I_{t_1} is much smaller than the motion between I_{t_2} and I_{t_0} . We then use the compositional motion model $W(\mathbf{x}; P_{2,0}) = W(W(\mathbf{x}; P_{2,1}); P_{1,0})$ to initialize the motion parameters $P_{2,0}$, and continue with the extended LK algorithm to refine the motion model $W(\mathbf{x}; P_{2,0})$ between I_{t_2} and I_{t_0} . Essentially, I_{t_1} serves as an online reference image to relate the live fluoroscopic image I_{t_2} to the predetermined reference image I_{t_0} . In our implementation, multiple online reference images covering a normal respiratory cycle are used in the motion compensation of live fluoroscopic images.

3 Results

To test the accuracy of the proposed method for motion compensation, we used fluoroscopic images from 7 clinical cases of chronic total occlusion or stenosis treatment. The data was acquired on Angiographic C-arm systems (AXIOM Artis, Siemens Medical Solution) from different angles. Each image frame has 512×512 pixels and the pixel size is either $0.216mm$ or $0.308mm$. These cases were chosen because they all had guidewires present throughout the entire image

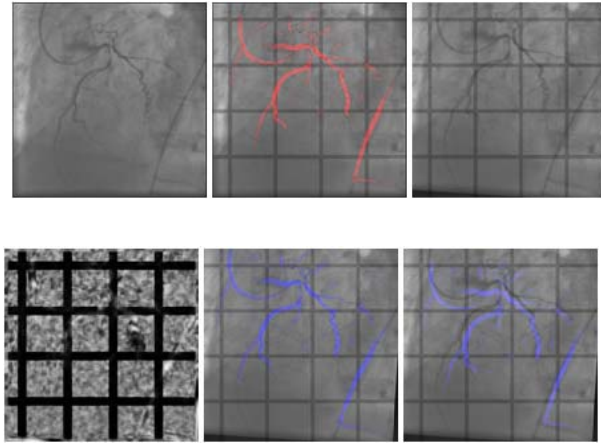


Fig. 1. WSSD-based motion estimation. Top row (left to right): original coronary image; reference image with a vesselness map shown in red simulating a coronary roadmap; test image. Bottom row (left to right): weighting function computed from local correlation coefficients; overlay of vesselness map after WSSD-based motion compensation; overlay of vesselness map after standard SSD-based motion compensation.

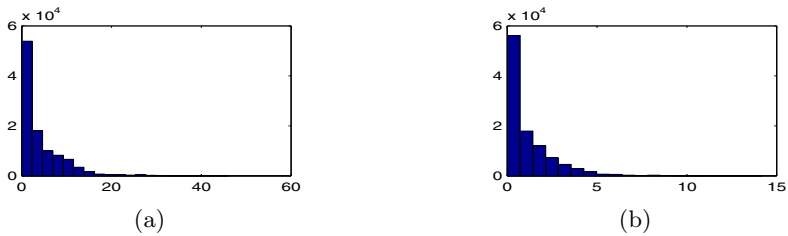


Fig. 2. Distribution of the misalignment error. (a) Error in pixels. (b) Error in millimeters.

sequences, which provided the ground truth of vessel centerlines for evaluation. Images with dye injection or visible guidewires were used as the reference images for the initial roadmap overlay. In each reference image, we manually labeled the centerline or the guidewire of one coronary artery to simulate the initial roadmap overlay. We performed motion compensation on a total of 106 frames, and used the estimated motion parameters to transform the initial roadmaps to match the test images. In each test image, we manually labeled the guidewire and used it as the ground truth for the coronary centerline. We then compared the ground truth of the coronary centerlines with the motion compensated roadmap overlays. As a misalignment measure, the distances between the motion compensated roadmap and the ground truth labels were calculated.

Since the fluoroscopic sequences used in testing were captured by long focal length perspective models, i.e., near orthographic projection of the 3D volume, pixel spacing in 2D images can be mapped to the distance between corresponding points in the 3D space through scaling by a factor. This is valid in the

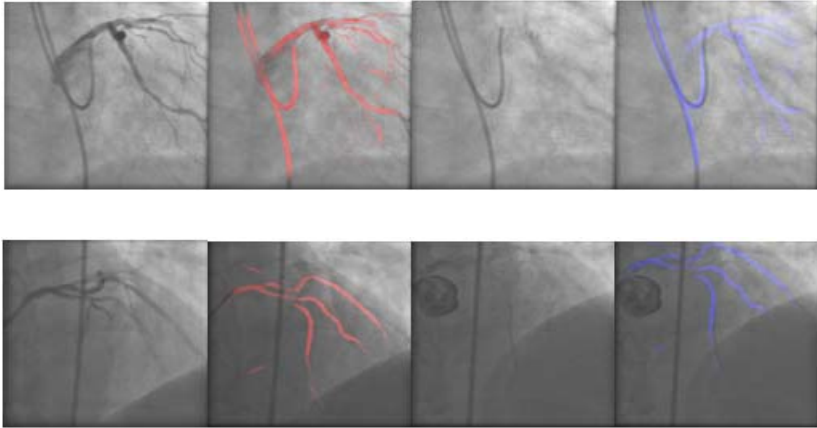


Fig. 3. Coronary roadmapping shown in pairs of an original image and the image with roadmap overlay. Roadmaps in reference images are shown in red. Motion compensated roadmaps in test images are shown in blue.

plane orthogonal to the viewing direction, regardless of the 3D geometry of the imaged subject or its depth. Fig. 2 shows the distribution of the misalignment error in pixels and in millimeters. The mean, median and standard deviation of the misalignment are 4.1064, 2.2361 and 4.9317 pixels, or 1.1443, 0.6469 and 1.3801mm respectively. The two error distributions are slightly different because of the different pixel sizes in the testing sequences. A few examples of motion compensated roadmap overlays are shown in Fig. 3.

4 Discussions

We have presented a novel image-based method of respiratory motion compensation for coronary roadmapping in fluoroscopic images. An extended LK algorithm based on weighted SSD is proposed to robustly estimate soft tissue motion when layers of static structures are present. For coronary roadmapping, the coronary locations are essentially inferred from the soft tissue motion when the contrast medium is not present. Even though a simplified two-layered compositional model and a global affine motion model are used to characterize the content and the dynamics of the fluoroscopic images of the heart over respiratory cycles, the initial results on real clinical data look very promising. Future directions include investigation of more elaborated models including local motion models to better characterize soft tissue motion incurred by respiration.

References

1. Tobis, J., Johnston, W., Montelli, S., Henderson, E., Roeck, W., Bauer, B., Nalcioğlu, O., Henry, W.: Digital coronary roadmapping as an aid for performing coronary angioplasty. *Am. J. Cardiol.* 56(4), 237–241 (1985)

2. Elion, J.L.: Dynamic coronary roadmapping. United States Patent 4878115, October 31 (1989)
3. A 3D CT Vessel Roadmap Over Live Fluoroscopy for Chronic Total Occlusion, Cath Lab Digest (February 2008), <http://www.cathlabdigest.com>
4. Ruijters, D., Romeny, B.M., ter, H., Suetens, P.: Vesselness-based 2D-3D registration of the coronary arteries. *Int. J. Computer Assisted Radiology and Surgery* 4(4), 391–397 (2009)
5. McLeish, K., Hill, D.L.G., Atkinson, D., Blackall, J.M., Razavi, R.: A study of the motion and deformation of the heart due to respiration. *IEEE Trans. Medical Imaging* 21(9), 1142–1150 (2002)
6. Shechter, G., Ozturk, C., Resar, J.R., McVeigh, E.R.: Respiratory Motion of the Heart From Free Breathing Coronary Angiograms. *IEEE Trans. Medical Imaging* 23(8), 1046–1056 (2004)
7. King, A.P., Boubertakh, R., Rhode, K.S., Ma, Y.L., Chinchapatnam, P., Gao, G., Tangcharoen, T., Ginks, M., Cooklin, M., Gill, J.S., Hawkes, D.J., Razavi, R.S., Schaeffter, T.: A subject-specific technique for respiratory motion correction in image-guided cardiac catheterisation procedures. *Med. Image Anal.* 13(3), 419–431 (2009)
8. 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, Part II. LNCS*, vol. 5762, pp. 381–388. Springer, Heidelberg (2009)
9. Baert, S.A.M., Viergever, M.A., Niessen, W.J.: Guide-wire tracking during endovascular intervention. *IEEE Trans. on Medical Imaging* 22(8), 965–972 (2003)
10. Wang, P., Zhu, Y., Zhang, W., Chen, T., Durlak, P., Bill, U., Comaniciu, D.: Hierarchical guidewire tracking in fluoroscopic sequences. In: *SPIE Medical Imaging*, vol. 7258, pp. 72591L–72591L-8 (2009)
11. Meijering, E.H.W., Niessen, W.J., Viergever, M.A.: Retrospective Motion Correction in Digital Subtraction Angiography: A Review. *IEEE Trans. Medical Imaging* 18(1), 2–21 (1999)
12. Zhu, Y., Prummer, S., Chen, T., Ostermeier, M., Comaniciu, D.: Coronary DSA: Enhancing Coronary Tree Visibility through Discriminative Learning and Robust Motion Estimation. In: *SPIE Medical Imaging*, vol. 7258, pp. 72591L–72591L-8 (2009)
13. Zhang, W., Ling, H., Prummer, S., Zhou, K.S., Ostermeier, M., Comaniciu, D.: Coronary tree extraction using motion layer separation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I. LNCS*, vol. 5761, pp. 116–123. Springer, Heidelberg (2009)
14. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework. *Int. J. Computer Vision* 56(3), 221–255 (2004)
15. Chen, Y., Chang, T.-C., Zhou, C., Fang, T.: Gradient domain layer separation under independent motion. In: *ICCV*, pp. 694–701 (2009)
16. Sarel, B., Irani, M.: Separating transparent layers through layer information exchange. In: Pajdla, T., Matas, J(G.) (eds.) *ECCV 2004. LNCS*, vol. 3024, pp. 328–341. Springer, Heidelberg (2004)

Surgical Task and Skill Classification from Eye Tracking and Tool Motion in Minimally Invasive Surgery

Narges Ahmidi¹, Gregory D. Hager², Lisa Ishii³, Gabor Fichtinger¹,
Gary L. Gallia⁴, and Masaru Ishii³

¹ Queen's University, Kingston, ON K7L3N6, Canada
{narges, gabor}@cs.queensu.ca

² Johns Hopkins University, Baltimore, MD 21211
hager@cs.jhu.edu

³ Johns Hopkins Medical Institutions, Baltimore, MD 21287
{learnes2, mishii3}@jhmi.edu

⁴ Department of Neurosurgery, Johns Hopkins University School of Medicine,
Baltimore, MD 21287
ggallia1@jhmi.edu

Abstract. In the context of minimally invasive surgery, clinical risks are highly associated with surgeons' skill in manipulating surgical tools and their knowledge of the closed anatomy. A quantitative surgical skill assessment can reduce faulty procedures and prevent some surgical risks. In this paper focusing on sinus surgery, we present two methods to identify both skill level and task type by recording motion data of surgical tools as well as the surgeon's eye gaze location on the screen. We generate a total of 14 discrete Hidden Markov Models for seven surgical tasks at both expert and novice levels using a repeated k -fold evaluation method. The dataset contains 95 expert and 139 novice trials of surgery over a cadaver. The results reveal two insights: eye-gaze data contains skill related structures; and adding this info to the surgical tool motion data improves skill assessment by 13.2% and 5.3% for expert and novice levels, respectively. The proposed system quantifies surgeon's skill level with an accuracy of 82.5% and surgical task type of 77.8%.

1 Introduction

The performance of a minimally invasive surgery highly depends on surgeons' dexterity in using surgical tools and their knowledge of the anatomy. This fact highlights the significance of Objective surgical skill evaluation.

A procedure of Functional Endoscopic Sinus Surgery (FESS) involves inserting an endoscope with a tiny camera on the end into the sinus cavity to provide the surgeon with a clear view of the surgical field and the ability to use instruments in treating the pathology. The surgeon's performance is limited due to indirect observation of the anatomy and inflexibility of the tools' movements inside the sinus cavity. FESS involves even higher risks due to the sinus' close proximity to the brain, major arteries and critical tissues such as optic nerves.

A skill evaluation system reveals the characteristics hidden in motion data to accurately associate given test sequences to their corresponding skill level. They have

been assessed in many studies by either tracking the surgeon's body motion in the operating room [1] or hand motion while performing a specific surgical task [2][3][4]. The Imperial College Surgical Assessment Device (ICSAD) system tracks the surgeon's hand motions during a surgery using electromagnetic (EM) markers [4]. In their system, they use a simple feature vector of the number of movements, hand speed, and procedure time to define an observed motion. Authors in [5] propose a system for scoring an image-guided percutaneous needle-based surgery by a feature vector of the successful trials, distance to target, and number of needle retractions. Here, we focus on the analysis of kinematic parameters of motion including translation and rotation of both the tool and the camera.

In laparoscopic surgeries, skill level is evaluated by measuring force and motion data [3]. The promising results with tele-operated robotic systems [2][6][7] show that Hidden Markov Models (HMM) enable us to recognize skill level and subtasks from motion data. Results in [8] show that rotated view of camera in laparoscopic surgeries increases the complexity of the task. In this paper, motion data is not recorded from a robotic system. We collect surgical tool motion data by attaching EM sensors to them.

An infrared-based eye tracking system can be used to measure the gaze position as an important factor in skill evaluation [9]. These trackers are used in other fields such as psychology [10]. Eye-gaze information plays a key role in eye-hand coordination, performance and adjustment of the surgical tools. To the best of our knowledge there is no prior published work in surgical skill evaluation featuring this system.

Our work differs from previous studies in that we generate 14 models (two skill levels in seven surgical tasks) and used them to recognize skill level and task type. Motions of the surgical tools and the surgeon's eye-gaze are recorded while performing different FESS tasks. We address the two following questions: First, is there any skill indicative structure in surgeon's eye-gaze motions? Second, how significant is the addition of eye-gaze data to surgical skill evaluation performance?

2 Experiment Setup

In this experiment, subjects are asked to find and touch a given anatomy inside the sinus cavity of a cadaver by using an endoscope and a nasal pointer. A group of seven anatomical targets are defined: Anterior Genu (AG), Eustachian Tube (ET), Fossa of Rosenmuller (FR), Opto Carotid Recess (OCR), Optic Nerve (ON), Pituitary Gland (PIT), and Superior Turbinate (ST). Figure 1 is the block diagram of the proposed experiment which illustrates the data collection setup and the proposed methodology.

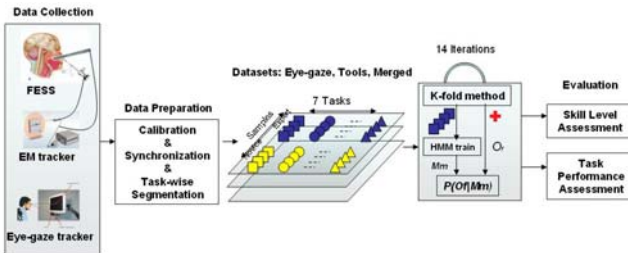


Fig. 1. System block diagram

Motion data from different trials of surgical tasks were collected from a group of 11 different subjects in two levels: 5 expert and 6 novices (Table 1). We defined an expert as a surgeon possessing knowledge of sinus anatomy structure and operation of the endoscope. Our novice subjects were those with no prior endoscopic experience.

Each subject performs two trials of surgeries. Each trial consists of 14 tasks: two sets of all seven tasks in random order. The duration of tasks vary between 5 to 46 seconds. We discard tasks involving irregular procedures (i.e. cleaning the endoscope tip, or leaving tools idle on the bed). The number of trials for each task in two possible skill levels is listed in the rows of Table 1. A total of 95 tasks are collected from expert surgeons and 139 tasks from novice surgeons.

The motion data of each trial is recorded using two trackers and a video stream: (1) The EM tracker collects motion data of both the endoscope and the nasal pointer at a frequency of 40 Hz, using two 5-DOF coil sensors attached to them. Sixteen variables are recorded per frame: a time stamp, a 3D translation and the four parameters of a rotation quaternion per sensor. (2) The eye-gaze tracker records 2D eye-gaze locations on the monitor (800x1200 pixel) at a frequency of 50 Hz. Three parameters are recorded per frame: a time stamp and a 2D eye-gaze location on the screen. (3) The video stream (352x240 pixel, 30 fps) is recorded from the endoscope tower.

Table 1. Number of performed tasks per skill level

Task \ Skill	AG	ET	FR	OCR	ON	PIT	ST
Expert	11	14	14	14	11	14	17
Novice	15	19	21	24	16	23	21

Later, we perform a pivot calibration for EM-tracker dataset and an eye-gaze calibration in order to register identified pupil position to the ground truth. In this experiment we initially study skill information of each of these trackers individually and then assess their aggregated performance.

3 Methodology: Hypotheses and Tests

In this paper we answer the aforementioned key questions: First, is there any skill indicative structure in surgeon's eye-gaze motions? Second, how significant is the addition of eye-gaze data to surgical skill evaluation performance?

We run two tests to find the answer for each question: a *Skill Level Assessment* (SLA) and a *Task Performance Assessment* (TPA). The former evaluates the skill level of the surgeon given a particular task, while the latter identifies the surgical task for a given expertise level. An additional *pre-observed test* is run on each dataset to verify that the generated models recognize their trained sequences properly. Referring to Figure 1, 14 configurations of SLA and 49 trials of TPA are carried out for a given skill level.

3.1 SLA: Skill Level Assessment

The intent of this test is to identify skill level of a given known task. To do so, we evaluate different trials of the task against HMMs of both expertise levels and measure True Positives, True Negatives, False Positives, and False Negatives ratios in a cross validation context. The Positives and Negatives are associated to expert and

novice skill level, respectively. Our first hypothesis is True Positives and True Negatives are higher than False Positives and False Negatives for each model. The more the disparity they have, the higher the skill level performance is.

3.2 TPA: Task Performance Assessment

We use this test to recognize the type of the performed task for a given expertise level. We find the most similar task to the test task by evaluating it against all HMMs in the same expertise level. A noise model REJ is generated from random parts of the datasets to measure the False Rejection Probability. Our second hypothesis is that the performance of TPA is high when we compare the test task against its matching model and is low when compared to other tasks models.

3.3 Data Preparation

To prepare motion data for training the task models, a *task-wise segmentation* is run on the synchronized datasets. We are able to split the videos based on the performed tasks. We need to mark the motion data to help synchronize with the corresponding video. During the clinical procedures, subjects are asked to look at a fixed point in the surgical field while touching it with the pointer tip for a few seconds in between the trials. To extract the fixed points in the motion data, we follow two assumptions: (1) the tool-tip is not in motion, and (2) the endoscope does not move significantly while touching the fixed points. The gradient of the motion is zero for a stationary object.

In equation 1, the function $f(x)$ is used to extract those stationary moments in the tool and endoscope motion datasets. The function $g(x)$ is a binary signal which represents the same moments in the video stream. Both f and g are variables of time and S is the convolution of fixed-point moments in the video and the tools. The global maxima of S is used to sync the tool motion data with the performed tasks in the video. We carry out the same procedure for the eye-gaze dataset to synchronize it with the performed task. This allows us to segment all of the datasets task-wise.

$$S = g(\text{video}) * \left[f(\nabla \text{endoscope}), f(\nabla \text{tool}) \right] \quad (1)$$

$$f(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases} \quad g(x) = \begin{cases} 1 & x = \text{fixpoint} \\ 0 & \text{otherwise} \end{cases}$$

In a clinical procedure, an expert surgeon tries to avoid critical tissues inside the sinus cavity by constantly monitoring them on the screen while holding the tools away from them inside the sinus. This leads to a group of imaginary points in the surgical field which constrain the surgeon's path toward the desired anatomical target. To discover the path, we apply *k-means* clustering algorithms to the motion datasets. We use a range of 4-9 for variable k in tool dataset, and 2-6 in eye-gaze dataset to determine the number of clusters with higher accuracy. Each frame of motion sequences is replaced by its corresponding cluster.

3.4 Evaluation Method

The HMMs are generated and tested using a k -fold method. The k -fold is run 14 times by changing the let-out task randomly. Each model is trained using the Baum-Welch algorithm with 100 iterations and an error tolerance of 0.01.

The HMMs are evaluated using Equation 2 to classify a test sequence. Probability function P is the log likelihood of an observation sequence (O_f) to a given model M_m , where the set O is generated from the trials of k -fold method. Comparing the resulting probabilities, the model C with the highest log likelihood is taken as the most probable source for that observed task. Then, we measure the percentage of similarity between the given test sequence O and a given HM model M_m by counting the number of identified models (Function V).

$$\forall O_f \in O \quad C(f) = \max_{m=1}^8 P(O_f | M_m) \quad 1 < f < |O|$$

$$V(O | M_m) = \sum_{f=1}^{|O|} [C(f) = m] / |O| \quad (2)$$

4 Results and Discussions

Question 1: Is there any skill indicative structure in the surgeon's eye-gaze motions?

As explained in the previous section, we execute both *SLA* and *TPA* tests for each skill level and performed task. Table 2 shows the results of skill level recognition for a given task. Additionally, Table 3 shows the accuracy of recognizing the type of the performed task for a given expertise level. To test the accuracy of the HMMs, the *pre-observed test* is run for each level of expertise and a result of 100% recognition is achieved for all the following datasets.

Our first hypothesis is confirmed by comparing the result of each column in Table 3. True Positives and True Negatives are significantly larger than False Positives and False Negatives, except for the expert level task PIT which is misclassified as the novice level. However Table 3 indicates that task PIT can be successfully classified at a given expertise level. The disparity between False and True classification in Table 2 shows that surgeon eye-gaze data includes structure for skill-level recognition.

Table 2. Skill Level Assessment of a given task, using eye-gaze dataset

Task \ Skill	AG	ET	FR	OCR	ON	PIT	ST
TPR	85%	95%	88%	87%	75%	42%	82%
FNR	15%	5%	12%	13%	25%	58%	18%
FPR	12%	25%	11%	4%	23%	7%	19%
TNR	88%	75%	89%	96%	77%	93%	81%

Table 3. Performed Task Assessment for a given skill level, using eye-gaze dataset

Skill level \ Test	Expert								Novice							
	AG	ET	FR	OCR	ON	PIT	ST	REJ	AG	ET	FR	OCR	ON	PIT	ST	REJ
AG	91.7	-	-	8.3	8.3	-	-	-	85.8	7.1	-	-	7.1	-	21.4	7.1
ET	-	100	-	-	8.3	8.3	-	8.3	-	85.8	-	-	7.1	-	-	7.1
FR	-	-	75	-	8.4	8.3	-	-	-	-	100	-	7.1	-	-	7.1
OCR	-	-	8.3	75	-	-	8.3	-	7.1	-	-	92.9	7.1	-	-	-
ON	-	-	-	-	50	-	-	-	-	-	-	-	64.5	-	-	-
PIT	8.3	-	8.3	-	-	75	-	-	-	-	-	-	7.1	100	14.3	-
ST	-	-	8.4	16.7	-	8.4	92.7	-	7.1	-	-	-	-	-	64.3	-
Reject (REJ)	-	-	-	-	25	-	-	91.7	-	7.1	-	7.1	-	-	-	78.7
%accuracy	91.7	100	75	75	50	75	92.7	91.7	85.8	85.8	100	92.9	64.5	100	64.3	78.7

The columns of Table 3 show that the performed task is always classified to the correct HMM (the diagonal of each square repeated in last row). The notable difference between the recognized models and the REJ model confirms the second hypothesis that eye-gaze data can be used for task evaluation as well.

Question 2: Can eye-gaze data improve surgical skill evaluation?

First, we execute SLA and TPA for the motion data of the tools to measure the accuracy of the system. Then, we make a new system of HMMs by combining both eye-gaze and tool datasets. To quantify the improvement achieved, we compare the performance of these two systems.

• *System1: skill evaluation using tool datasets*

Tables 4 and 5 show the results for SLA and TPA using EM tracked tool datasets. In Table 4, the columns show that the skill level is recognized correctly for each task, except for task AG. It reveals that both expert and novice participants perform that task with the same level of expertise. Carotid artery is one of the largest objects in the nose and one of the most prominent so we would expect it to be the easiest to find. It is also the most understood even by novices, since they are taught specifically from the start to always identify this structure. The anatomy was distorted by removing part of the skull base (planum) to make the task harder. Therefore, some subjects mis-identified structures.

Table 4. Skill Level Assessment of a given task, using tools dataset

Task \ Skill	AG	ET	FR	OCR	ON	PIT	ST
TPR	50%	79%	75%	80%	75%	82%	73%
FNR	50%	21%	25%	20%	25%	18%	27%
FPR	31%	18%	11%	25%	11%	11%	25%
TNR	69%	82%	89%	75%	89%	89%	75%

Table 5. Performed Task Assessment for a given skill-level, using tools dataset

Skill level \ Test	Expert								Novice							
	AG	ET	FR	OCR	ON	PIT	ST	REJ	AG	ET	FR	OCR	ON	PIT	ST	REJ
AG	100	43	50	28.6	42.9	35.7	14.4	50	64.4	7.1	-	7.1	-	7.1	-	14.3
ET	-	57	-	-	-	-	-	7.1	14.3	64.4	-	7.1	-	7.1	-	7.1
FR	-	-	50	-	-	-	7.1	-	-	14.3	78.6	-	7.1	7.1	7.1	-
OCR	-	-	-	64.3	-	-	7.1	-	7.1	-	-	71.6	-	-	7.1	-
ON	-	-	-	-	50	-	7.1	-	7.1	7.1	-	7.1	78.7	-	21.4	-
PIT	-	-	-	-	-	64.3	21.4	-	-	-	7.1	-	7.1	78.7	-	-
ST	-	-	-	7.1	7.1	-	42.9	7.1	7.1	-	-	7.1	-	-	50	-
Reject (REJ)	-	-	-	-	-	-	-	35.8	-	7.1	14.3	-	7.1	-	14.4	78.6
%accuracy	100	57	50	64.3	50	64.3	42.9	35.8	64.4	64.4	78.6	71.6	78.7	78.7	50	78.6

Overall accuracy of the models in skill level detection is 73.4% and 81.1% for expert and novice surgeons, respectively (Table 8). The columns of Table 5 show that tasks are always recognizable for novice datasets. The overall identification rate for task type recognition is 58.1% and 70.6% for expert and novice groups, respectively.

- *System2: skill evaluation using combined dataset*

Tables 6 and 7 show the results of executing SLA and TPA over the merged dataset. Table 8 shows that expertise level is identifiable in 82.9% of cases for expert surgeons and 82% for novice surgeons. The columns of Table 7 show that all of the tasks in the same level of expertise are recognized correctly.

Table 6. Skill Level Assessment of a given task, using merged dataset

Task \ Skill	AG	ET	FR	OCR	ON	PIT	ST
TPR	50%	89%	89%	85%	89%	85%	93%
FNR	50%	11%	11%	15%	11%	15%	7%
FPR	46%	13%	13%	13%	10%	13%	18%
TNR	54%	87%	87%	87%	90%	87%	82%

Table 7. Performed Task Assessment for a given skill-level, using merged dataset

Skill level \ Test	Expert								Novice							
	AG	ET	FR	OCR	ON	PIT	ST	REJ	AG	ET	FR	OCR	ON	PIT	ST	REJ
AG	100	27.3	27.3	27.3	18.2	27.3	36.4	36.3	71.4	-	-	-	-	-	-	-
ET	-	72.7	-	-	--	-	-	-	-	78.6	-	-	-	-	-	-
FR	-	-	72.7	-	-	-	-	-	-	-	78.6	-	-	-	-	-
OCR	-	-	-	72.7	-	-	-	-	-	-	-	71.4	-	-	-	-
ON	-	-	-	-	81.8	-	-	-	-	--	-	-	-	78.6	-	-
PIT	-	-	-	-	-	72.7	-	-	-	-	-	-	-	71.4	-	-
ST	-	-	-	-	-	-	63.6	-	-	-	-	-	-	-	85.7	-
Reject (REJ)	-	-	--	-	-	-	-	63.6	-	-	-	-	-	-	-	78.6
%accuracy	100	72.7	72.7	72.7	81.8	72.7	63.6	63.6	100	78.6	78.6	71.4	78.6	71.4	85.7	78.6

Table 8. Performance improvement in skill assessment adding eye-gaze dataset

Test	Skill	Surgical tools dataset	Merged with eye-gaze dataset	Performance Improvement
Avg TPA	Expert	58.1%	74.98%	16.9%
	Novice	70.7%	80.36%	9.7%
Avg SLA	Expert	73.4%	82.9%	9.5%
	Novice	81.1%	82%	0.9%

Each row in Table 8 is the average result of the corresponding method over the related models (an average of first and last row of SLA, and last row of TPA). Comparing the performance of the two systems reveals that adding eye-gaze information improves SLA performance by 9.5% for expert and 0.9% for novice surgeons. The average performance of TPA is 75% and 80.4% for expert and novice surgeons, respectively. This indicates that the new system improves the TPA performance by 16.9% for experts and by 9.7% for novice surgeons.

5 Conclusion

Precision of sinus surgery is critical due to its close proximity to the brain, major arteries and critical tissues. The work presented here is a promising statistical method to assess skill of a surgeon while performing a Functional Endoscopic Sinus Surgery. We make HMM models for seven different surgeries in two level of expertise using the eye-gaze locations and the surgical tools motions. Two metrics of SLA and TPA access skill level and task type of a given surgery. Results reveal that eye-gaze data contains skill related structures; and combining it with the surgical tool motion data improves the classifier performance. The proposed system improves SLA performance 9.5% for experts and 0.9% for novices, on average. Besides, TPA performance is improved 16.9%

for experts and 9.7% for novice surgeons. The proposed system quantifies surgeon's skill level with an accuracy of 82.5% and surgical task type of 77.8%.

Currently, we are exploring the methods to improve performance of the proposed skill assessment techniques. One is to unify coordination systems of the tools and eye-gaze motions by registering them to the CT image volume of the target anatomy.

Acknowledgments

The authors gratefully acknowledge support from NIH 5R01CA118371, NSF CDI-0941362 and NSF CPS 0931805 as well as Johns Hopkins internal funding. They would like to thank Daniel Abretske, Daniel Mirota, and Kelleher Guerin for their time and help with the data collection.

References

1. Padoy, N., Mateus, D., Weinland, D., Berger, M.O., Navab, N.: Workflow Monitoring based on 3D Motion Features. In: ICCV Workshop on Video-oriented Object and Event Classification, Kyoto, Japan (September 2009)
2. Lin, H.C., Shafran, I., Yuh, D., Hager, G.D.: Towards Automatic Skill Evaluation: Detection and Segmentation of Robot-Assisted Surgical Motions. *Computer Aided Surgery* 11(5), 220–230 (2006)
3. Rosen, J., Brown, J.D., Chang, L., Sinanan, M., Hannaford, B.: Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete Markov model. *IEEE Trans. Biomed. Eng.* 53(3), 399–413 (2006)
4. Munz, Y., Almoudaris, A., Moorthy, K., Dosis, A., Liddle, A., Darzi, A.: Curriculum-based solo virtual reality training for laparoscopic intracorporeal knot tying: objective assessment of the transfer of skill from virtual reality to reality. *The American Journal of Surgery* 193(6), 774–783 (2007)
5. Ahmidi, N., U-Thainual, P., Vikal, S., Mousavi, P., Iordachita, I., Fichtinger, G.: A System for Performance Analysis of Surgeon Dexterity in Percutaneous Needle-based Interventions. *Imaging Network Ontario*. University of Toronto, Canada (2008)
6. Reiley, C., Hager, G.D.: Task versus subtask surgical skill evaluation of robotic minimally invasive surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5761, pp. 435–442. Springer, Heidelberg (2009)
7. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G.D.: Data-Derived Models for Segmentation with Application to Surgical Assessment and Training. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I*. LNCS, vol. 5761, pp. 426–434. Springer, Heidelberg (2009)
8. Leong, J.J., Nicolaou, M., Atallah, L., Mylonas, G.P., Darzi, A.W., Yang, G.Z.: HMM assessment of quality of movement trajectory in laparoscopic surgery. *Comput. Aided Surg.* 12, 335–346 (2007)
9. Nicolaou, M., James, A., Darzi, A., Yang, G.Z.: A Study of Saccade Transition for Attention Segregation and Task Strategy in Laparoscopic Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3217, pp. 97–104. Springer, Heidelberg (2004)
10. Martín-Loeches, M., Schacht, A., Casado, P., Hohlfeld, A., Abdel Rahman, R., Sommer, W.: Rules and heuristics during sentence comprehension: evidences from a dual-task brain potential study. *Journal of Cognitive Neuroscience* 21(7), 1380–1395 (2009)

Micro-force Sensing in Robot Assisted Membrane Peeling for Vitreoretinal Surgery

Marcin Balicki^{1,*}, Ali Uneri¹, Iulian Iordachita¹,
James Handa², Peter Gehlbach², and Russell Taylor¹

¹ ERC for Computer Integrated Surgery, Johns Hopkins University

² Wilmer Eye Institute, Johns Hopkins Medical, Baltimore, MD, USA

{marcin, iordachita, rht}@jhu.edu

Abstract. Vitreoretinal surgeons use 0.5mm diameter instruments to manipulate delicate tissue inside the eye while applying imperceptible forces that can cause damage to the retina. We present a system which robotically regulates user-applied forces to the tissue, to minimize the risk of retinal hemorrhage or tear during membrane peeling, a common task in vitreoretinal surgery. Our research platform is based on a cooperatively controlled microsurgery robot. It integrates a custom micro-force sensing surgical pick, which provides conventional surgical function and real time force information. We report the development of a new phantom, which is used to assess robot control, force feedback methods, and our newly implemented auditory sensory substitution to specifically assist membrane peeling. Our findings show that auditory sensory substitution decreased peeling forces in all tests, and that robotic force scaling with audio feedback is the most promising aid in reducing peeling forces and task completion time.

Keywords: Robotic Microsurgery, Vitreoretinal Surgery, Cooperative Robot Control, Force Sensing, Force Scaling, Auditory Sensory Substitution.

1 Introduction

Microsurgical manipulation requires precise manual dexterity, fine visual-motor coordination, and application of forces that are well below human tactile sensation [1]. Imprecise movements during these operations are further attributed to physiological hand tremor, fatigue, poor visual and kinesthetic feedback, as well as patient movement. In vitreoretinal surgery, the common microsurgery risks and limitations are further extended by surgical inaccessibility, poor visibility, tissue fragility and the flexibility of delicate (20–25 Ga) surgical instruments.

Epiretinal membrane (ERM) peeling is a typical task where a thin membrane is carefully delaminated off the surface of the retina by grasping the membrane's edge with micro-forceps and pulling it free from the retina. Due to unstable manipulation and unknown forces applied to the tissue, the maneuver is associated with the risks of retinal hemorrhage and tearing, leading to potentially irreversible damage that results

* Supported in part by the NSF EEC9731748, NIH 1R01 EB007969-01 and ARCS Foundation.

in vision loss. Surgeons manipulate the peeling tissue at very slow instrument velocities, observed to be within 0.1–0.5 mm/s range, and simultaneously visually monitor local surface deformation that may indicate undesirable forces between the surgical instrument, the membrane and the retina. A capable surgeon reacts to such cues by retracting the instrument and regrasping the tissue for an alternate approach. This task is extremely difficult to master due to nearly imperceptible visual cues, and a requirement for very precise visuomotor reflexes. Factors such as involuntary patient motion, inconsistent tissue properties, high or variable tool velocities, and changing manipulation directions can dramatically increase undesirable forces applied to the delicate retinal tissue. Actively sensing and limiting these forces has the potential to significantly improve surgical precision and diminish surgical complications.

One approach is to reduce physiological hand tremor and actively improve the surgeon's fine motion control. An example is MICRON, a hand held device that uses piezo actuators to counteract undesirable hand tremor and scale input motions [2]. Another approach is a use of a tele-operated robot, such as [3], to reduce hand tremor through remote control and motion scaling. The cooperatively controlled JHU Steady-hand Eye Robot [4] relies on its stiff structure and non back-drivable actuators with high resolution encoders to provide stability and precision. To our knowledge, these systems do not provide any tool-tissue interaction force feedback to the surgeon. Logically, combining hand tremor reduction with precise motion control and end-point force sensing feedback is a more complete approach.

There have been some attempts to measure tool to tissue interaction forces with instruments that had sensing elements built into the handle [5] and to incorporate sensed tool-tissue forces into cooperative control (e.g. [7,8]). Handle mounted tool force sensors are not practical for vitreoretinal surgery, because the shaft of the instrument is inserted through the sclera. The friction between the tool and the trocar, and lateral forces from the transacting sclera with the tool can significantly attenuate or distort the propagation of the forces to the tissues inside of the eye. This limitation may be addressed by incorporating force sensing elements into the shaft section of the instrument that is typically located inside the eye [6].

In this paper we investigate and compare micro-force sensing feedback to specifically assist in membrane peeling, for which we developed a novel phantom that approximates peeling forces encountered in retinal surgery. We summarize our microsurgical research

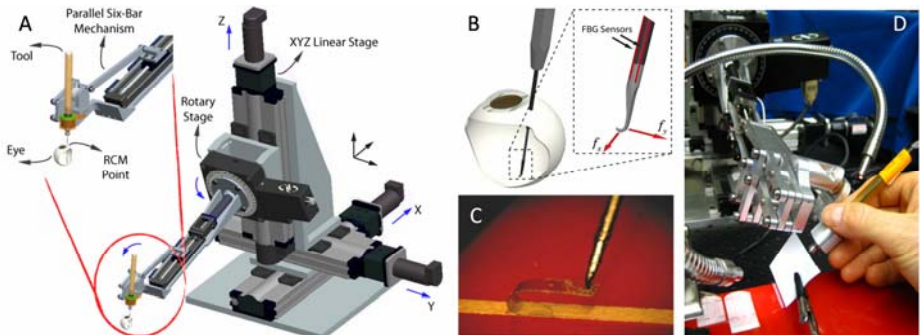


Fig. 1. A) Robot with RCM mechanism [11]; B) Force Sensor Instrument Concept [6]; C) Peeling sample and hooked force sensor instrument; D) Experimental setup

platform integrating a 2-DOF force-sensing vitreoretinal surgery instrument and the new cooperative control robot with a remote center-of-motion mechanism (RCM). In addition to the linear “force scaling” cooperative control in [7,8], we introduce a velocity limiting force-cooperative control mode. We also present our new task-specific auditory sensory substitution and assess its effects on task performance, both with freehand instruments and with our various robot cooperative control modes.

2 Experimental Platform

2.1 Robotic Assistant

Our cooperatively controlled “steady hand” robotic assistant is a 5-DOF system (Fig. 1A) similar to [4] and designed as a development platform for microsurgery research. It is an admittance-type system, collaboratively providing steady-hand motion by inherently filtering physiological hand tremor and low-frequency drift found in surgeons’ hand movements during microsurgery [9]. A 6-DOF force/torque sensor mounted at the tool holder senses forces exerted by the surgeon on the tool, for use as command inputs to the robot. This design improves on similar work described in [4] by increasing the range of motion, including a custom mechanical RCM and improving the stiffness and precision of the stages. The RCM mechanism improves the general stability of the system by reducing range of motion and velocities in the Cartesian stages when operating in virtual RCM mode [10], which constrains the tool axis to always intersect the sclerotomy opening on the eye. This is a necessary safety feature to minimize undesirable translations of the eye and also simplifies software control. The actuator motion controller servo loop frequency is 4 kHz, while the robot control loop operates at 400 Hz. A more detailed description is in [11].

2.2 Micro-force Sensing Instrument

Vitreoretinal microsurgical applications introduce certain limitations on the exact choice of force sensor by demanding sub-mN accuracy required to sense forces that are routinely less than 7.5 mN [1]. A miniature instrument size is necessary to be inserted through a 25 Ga sclerotomy opening and the force sensor must be able to obtain measurements at the instrument’s tip, below the sclera.

A tool with integrated fiber Bragg grating (FBG) sensors was manufactured for this purpose, following the design [6] of Iordachita *et al.* FBGs are robust optical sensors capable of detecting changes in strain, without interference from electrostatic, electromagnetic or radio frequency sources. Three optical fibers are placed along the tool shaft (Fig. 1B) and by measuring the bending of the tool they allow for calculation of the force in the transverse plane with a sensitivity of 0.25 mN. The tip of the tool is a simple hook, and the tool is mounted in the robot tool holder in a calibrated orientation relative to the robot. The sensor data was collected and processed at 2 kHz and transmitted over TCP/IP.

2.3 Membrane Peeling Phantom

To develop and assess the performance of control and auditory feedback methods, we required a consistent and easily fabricated phantom model, behaving within the

parameters of vitreoretinal surgery. The actual peeling procedure involves grasping or hooking a tissue layer and slowly delaminating it, often in a circular pattern. To reduce the factors needed to be controlled, we simplified the target maneuver to a straight line peel using a hooked instrument.

Video recordings of actual epiretinal membrane peeling procedures were analyzed and tool velocities during membrane peeling were found to be 0.1–0.5 mm/s. Gupta *et al.* found that retinal tissue manipulation forces are likely to be below 7.5 mN [1], while Jagtap *et al.* found them to be comparable but slightly higher [12]. With these values in mind, after extensive search and trial-and-error testing of many materials, we identified that sticky tabs from 19 mm Clear Bandages (RiteAid brand) to be a suitable and repeatable phantom for delaminating. The tab was sliced to produce 2 mm wide strips (Fig. 1C) that can be peeled multiple times from its backing, with predictable behavior showing increase of peeling force with increased peeling velocity. The plastic peeling layer is very flexible but strong enough to withstand breaking pressures at the hook attachment site. A 10 mm section of the strip is considered, requiring 20 mm of tool travel to complete a peel. Fig. 2 shows the forces observed at various velocities.

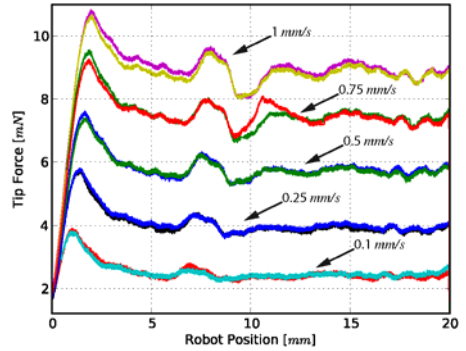


Fig. 2. Peeling sample repeatability tests

3 Methods

We have employed three cooperative control methods which modulate the behavior of the robot based on operator input and/or tool tip forces. The control method parameters considered handle input force range (0–5 N), and peeling task forces and velocities. Audio sensory substitution serves as a surrogate or complementary form of feedback and provides high resolution real-time tool tip force information.

Proportional Velocity Control (PV) paradigm is described in [7], where the velocity at the tool (V) is proportional to the user's input force at the handle (F_h). A gain of $\alpha = 1$ was used, which translates handle input force of 1 N to 1 mm/s tool velocity.

$$\dot{x} = \alpha F_h \quad (1)$$

Linear Force Scaling Control (FS) maps, or amplifies, the human-imperceptible forces sensed at the tool tip (F_t) to handle interaction forces by modulating robot velocity. Prior applications used $\gamma = 25$ and $\gamma = 62.5$ scale factors [7, 8] which are low for the range of operating parameters in vitreoretinal peeling. Scaling factor of $\gamma = 500$ was chosen to map the 0–10 mN manipulation forces at the tool tip to input forces of 0–5 N at the handle.

$$\dot{x} = \alpha(F_h + \gamma F_t), \quad \alpha = 1 \quad (2)$$

Proportional Velocity Control with Limits (VL) increases maneuverability when low tip forces are present. The method uses PV control but with an additional velocity

constraint that is inversely proportional to the tip force. With such scaling, the robot response becomes very sluggish with higher tool tip forces, effectively dampening manipulation velocities. The constraint parameters were chosen empirically to be $m = -180$ and $b = 0.9$. To avoid zero crossing instability, forces lower than $f_1 = 1 \text{ mN}$ in magnitude do not limit the velocity. Likewise, to provide some control to the operator when tip forces are above a high threshold ($f_2 = 7.5 \text{ mN}$), a velocity limit ($v_2 = 0.1$) is enforced.

$$\dot{x} = \begin{cases} V_{\text{lim}}(F_t), & -F_h < V_{\text{lim}}(F_t) \wedge F_t < 0 \\ V_{\text{lim}}(F_t), & -F_h > V_{\text{lim}}(F_t) \wedge F_t > 0 \\ \alpha F_h, & \text{otherwise} \end{cases} \quad (3)$$

Where $V_{\text{lim}}(F_t)$, is the function in Fig. 3A.

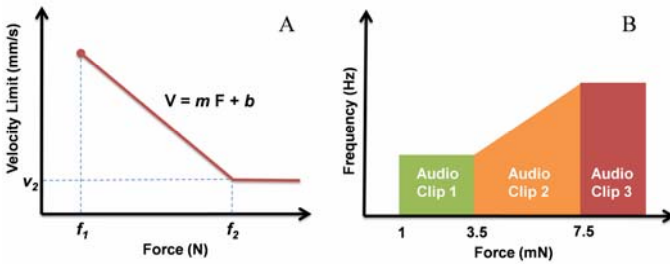


Fig. 3. A) Velocity limiting function (symmetric about $V = -F$); B) Audio feedback zones

Force-to-Auditory Sensory Substitution. In current practice, surgeons indirectly assess the relative stress applied to tissue via visual interpretation of changing light reflections from deforming tissue. This type of “visual sensory substitution” requires significant experience and concentration, common only to the expert surgeons. To provide more clear and objective feedback, we measure these forces directly and convey them to the surgeon in real time with auditory representation. Kitagawa *et al.* showed that auditory feedback representing force in a complex surgical task improves robot assisted performance and suggests that continuous and real-time feedback is more effective than discrete, single event information [13]. We designed auditory feedback that modulates the playback tempo of audio “beeps” in three force level zones which represent force operating ranges chosen to be relevant in typical vitreoretinal operations. The audio is silent until 1 mN or greater force is measured. A constant slow beeping is emitted until 3.5 mN, which is designated to be a “safe” force operating zone. A “cautious zone” was designated as 3.5–7 mN, and has a proportionally increasing tempo followed by a “danger zone” that generates a constant high tempo beeping.

Experiments. Our objective is to compare the effectiveness of the above methods in decreasing mean and maximum peeling forces while minimizing time taken to complete the task. A single subject was tested in this preliminary experiment, which was configured in the following ways. The phantom was adhered to a stable platform with double-stick tape and the robot was positioned so the hook is ~1.5 mm above the

peeling surface. The orientation of the handle was perpendicular to the peeling direction and comfortable to the operator. To eliminate force cues from tool bending, the visibility of the tool shaft was obstructed with the exception of the tool tip (Fig. 2D). The test subject was trained extensively (~3 hours) prior to the trials. Five minute breaks were allowed between trials. The operator was directed to peel the membrane steadily and as slow as possible without stopping. To simplifying the experiments we limited the robot motion to Cartesian translations only; experiments showed no noticeable difference between trials with and without rotational DOFs. No visual magnification was provided to the operator. For all trials we used the same sample and, for consistency, we have tested the behavior of the sample before and after the experiment. For comparison, we have included freehand peeling tests where the operator peeled the sample without robot assistance. Five trials of each method were performed with audio feedback, and five without for the following: *Freehand* (FH/FHA), *Proportional Velocity Control* (PV/PVA), *Linear Force Scaling Control* (FS/FSA), *Velocity Limiting Control* (VL/VLA).

4 Results

In every method tested, audio feedback decreased the maximum tip forces, as well as tip force variability (Fig. 4). It significantly increased the task completion time for freehand and proportional velocity control trials while the time decreased slightly for the others. The operator was naturally inclined to “hover” around the discrete audio transition point corresponding to 3.5 mN, which was observed in all cases except freehand. This was particularly prominent in force scaling, where the operator appears to rely on audio cues over haptic feedback (see Fig. 5C, time 60–80 s). In velocity limiting trials, audio reduced mean input handle forces by 50% without compromising performance. This indicates that the user consciously attempted to use audio feedback to reduce the forces applied to the sample.

Forces(mN)	FH	FHA	PV	PVA	FS	FSA	VL	VLA
Mean	4.11	3.80	4.20	3.64	3.34	3.22	3.58	3.45
StdDev	0.97	0.59	0.95	0.51	0.54	0.40	0.36	0.33
Max	7.85	6.21	6.93	4.74	4.10	3.59	4.03	3.83
Time(s)	93.03	125.25	62.30	85.98	103.80	96.80	88.67	80.58

Fig. 4. Table of tip force results from all trials

Freehand (Fig. 5A) trials showed considerable high force variation due to physiological hand tremor. The mean force applied was around 5 mN, with maximum near 8 mN. Audio feedback helped to reduce large forces but significantly increased task completion time.

Proportional Velocity (Fig. 5B) control performance benefited from the stability of robot assistance and resulted in a smoother force application, while the range of forces was comparable to freehand tests. Likewise, audio feedback caused a decrease in large forces but increased time to complete the task.

Force Scaling (Fig. 5C) control yielded the best overall performance in terms of mean forces with and without audio. Although, the average time to completion was the longest, except for freehand with audio.

Velocity Limiting (Fig. 5D) control resulted in a very smooth response except for the section that required higher absolute peeling forces at the limited velocity. This

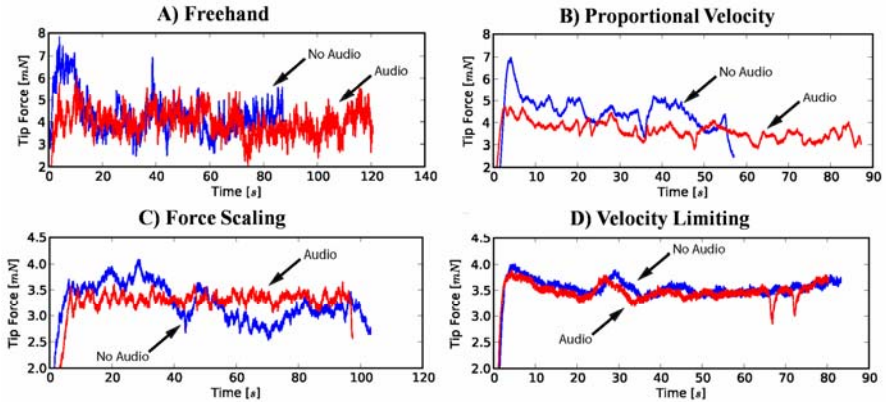


Fig. 5. Plots of representative trials of each mode showing tip forces, with/without audio feedback

had an effect of contouring “along” a virtual constraint. Due to matching thresholds, audio had very little effect on the performance.

5 Discussion and Conclusions

Robotic assistance combining a cooperatively controlled manipulation with real time tool-to-tissue force sensing has significant potential to improve surgical practice, especially when combined with audio sensory substitution. We have shown that our experimental system is capable of measuring and reacting to forces under 7.5 mN, a common range in microsurgery. We have found that force scaling with audio feedback provides the most intuitive response and force-reducing performance in a simulated membrane peeling task, where the goal is to apply low and steady forces to generate a controlled delamination.

We designed robot control and audio feedback parameters specifically for our membrane peeling phantom which, according to our expert surgeon co-authors, is a satisfactory surrogate for the target task. The feedback parameters can also be tuned in real-time to accommodate for operator preference and optimize the system for other microsurgical tasks. In our experiment, the force scaling yielded the best results, however the handle input forces were high (~ 2.5 N) enough to cause fatigue and decreased precision due to prolonged stress. Lowering the force scaling gain will lower the handle input forces, but this may compromise sensitivity since the human finger has a force sensing resolution of 0.5 N [14].

Preliminary user feedback has indicated that auditory sensory substitution can be very useful, especially when combined with cooperative robot control. Our surgeon coauthors believe that continuous audio feedback may be disruptive or overwhelming in already noisy operating room, but occasional and unique sounding feedback during critical moments can be very valuable. During operator training, we have observed significant improvement in task completion rates and decrease in force variation stemming from audio feedback alone. This concept has potential to be a very useful training tool in providing immediate and objective tissue manipulation force information which can be correlated with other cues.

The results from this single-user pilot study are encouraging and provide the basis for a planned multi-user study. Other next steps involve characterizing in-vivo membrane peeling forces and tool trajectories to verify and improve our artificial phantom, and to optimize our robot control parameters for in-vivo system assessment. We are also planning a human factors study to identify intuitive, effective and operating room compatible auditory sensory substitution methods. Finally, to cover a variety of peeling approaches and force application directions, we plan to build a microsurgical forceps that measures axial force in addition to the two axis transverse forces used in this work.

References

1. Gupta, P.K., Jensen, P.S., de Juan, E.: Surgical forces and tactile perception during retinal microsurgery. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 1218–1225. Springer, Heidelberg (1999)
2. Riviere, C., Ang, W.T., Khosla, P.: Toward active tremor canceling in handheld microsurgical instruments. *IEEE Tr. Rob. & Aut.* 19(5), 793–800 (2003)
3. Nakano, T., Sugita, N., Ueta, T., Tamaki, Y., Mitsuishi, M.: A parallel robot to assist vitreoretinal surgery. *JCARS* (2009)
4. Mitchell, B., Koo, J., Iordachita, M., Kazanzides, P., Kapoor, A., Handa, J., Hager, G., Taylor, R.: Development and application of a new steady-hand manipulator for retinal surgery. *IEEE ICRA*, 623–629 (2007)
5. Berkelman, P., Whitcomb, L., Taylor, R., Jensen, P.: A miniature microsurgical instrument tip force sensor for enhanced force feedback during robot-assisted manipulation. *IEEE TRA* 19(5), 917–921 (2003)
6. Iordachita, I., Sun, Z., Balicki, M., Kang, J., Phee, S., Handa, J., Gehlbach, P., Taylor, R.: A sub-millimetric, 0.25 mN resolution fully integrated fiber-optic force-sensing tool for retinal microsurgery. *JCARS* 4(4), 383–390 (2009)
7. Kumar, R., Berkelman, P., Gupta, P., Barnes, A., Jensen, P.S., Whitcomb, L.L., Taylor, R.H.: Preliminary Experiments in Cooperative Human/Robot Force Control for Robot Assisted Microsurgical Manipulation. *IEEE ICRA* 1, 610–617 (2000)
8. Berkelman, P., Rothbaum, D., Roy, J., Lang, S., Whitcomb, L., Hager, G., Jensen, P., de Juan, E., Taylor, R., Niparko, J.: Performance “Evaluation of a Cooperative Manipulation Microsurgical Assistant Robot Applied to Stapedotomy. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 1426–1429. Springer, Heidelberg (2001)
9. Riviere, C., Rader, R., Khosla, P.: Characteristics of hand motion of eye surgeons. *IEEE EMBS* 4, 1690–1693 (1997)
10. Kapoor, A., Li, M., Taylor, R.H.: Constrained control for surgical assistant robots. *IEEE ICRA*, 231–236 (2006)
11. Uneri, A., Balicki, M., Handa, J., Gehlbach, P., Taylor, R., Iordachita, I.: New Steady-Hand Eye Robot with Microforce Sensing for Vitreoretinal Surgery Research. In: *IEEE BioRob*. 2010 (2010)
12. Jagtap, A.S., Riviere, C.N.: Applied force during vitreoretinal microsurgery with handheld instruments. In: *Proc. IEEE Conf. on Eng. Med. Biol. Soc.*, vol. 4, pp. 2771–2773 (2004)
13. Kitagawa, M., Dokko, D., Okamura, A., Yuh, D.: Effect of sensory substitution on suture manipulation forces for robotic surgical systems. *JTCS* 129(1), 151–158 (2005)
14. Taylor, C.L.: The biomechanics of the normal and of the amputated upper extremities. In: Klopsteg, P.E., Wilson, P.D. (eds.) *Human limbs and their substitutes*, ch. 7, pp. 169–221. Hafner Pub. Co., New York (1968)

C-arm Pose Estimation in Prostate Brachytherapy by Registration to Ultrasound

Pascal Fallavollita¹, Clif Burdette², Danny Song³,
Purang Abolmaesumi⁴, and Gabor Fichtinger¹

¹ Queen's University, Canada
pascal@cs.queensu.ca

² Acoustic MedSystems Inc., Illinois, USA

³ Johns Hopkins Hospital, Baltimore, USA

⁴ University of British Columbia, Canada

Abstract. In prostate brachytherapy, transrectal ultrasound (TRUS) is used to visualize the anatomy, while implanted seeds can be seen in C-arm fluoroscopy. Intra-operative dosimetry optimization requires reconstruction of the implanted seeds from multiple C-arm fluoroscopy images, which in turn requires estimation of the C-arm poses. We estimate the pose of the C-arm by two-stage registration between the 2D fluoroscopy images to a 3D TRUS volume. As single-view 2D/3D registration tends to yield depth error, we first estimate the depth from multiple 2D fluoro images and input this to a single-view 2D/3D registration. A commercial phantom was implanted with seeds and imaged with TRUS and CT. Ground-truth registration was established between the two by radiographic fiducials. Synthetic ground-truth fluoro images were created from the CT volume and registered to the 3D TRUS. The average rotation and translation errors were 1.0° (STD= 2.3°) and 0.7mm (STD=1.9 mm), respectively. In data from a human patient, the average rotation and lateral translation errors were 0.6° (STD= 3.0°) and 1.5 mm (STD=2.8 mm), respectively, relative to the ground-truth established by a radiographic fiducial. Fully automated image-based C-arm pose estimation was demonstrated in prostate brachytherapy. Accuracy and robustness was excellent on phantom. Early result in human patient data appears clinically adequate.

1 Introduction

Prostate cancer is the second most common cancer in men, diagnosed in 192,280 new patients each year in North America [1]. Brachytherapy is a definitive treatment of early stage prostate cancer, chosen by over 50,000 men each year. The procedure entails permanent implantation of small radioactive isotope capsules (a.k.a. seeds) into the prostate to kill the cancer with radiation. Success hinges on precise placement of the implants to provide the needed dose distribution. Unfortunately, primarily due to tissue motion, organ deformation, and needle deflection, actual seed positions never turn out to be as planned. Intraoperative dose optimization during the procedure would allow for correcting deviations from the plan and thus tailor the dose to cancer without harming surrounding normal tissues. This requires localization of the prostate and implanted seeds; a much coveted function that is not available today [2]. Prostate

brachytherapy is performed with transrectal ultrasound (TRUS) guidance that provides adequate real-time visualization of the prostate but not of the implanted seeds. C-arm fluoroscopy is widely used for gross visual assessment of the implanted seeds (Figure 1) but it cannot show the prostate and other relevant structures. Fusion of these complementary modalities would enable dynamic dosimetry. A variety of implant reconstruction techniques have been investigated [3,4,5] which share one common requirement: the relative poses of the fluoroscopy images must be known prior to reconstruction. The fluoroscopy pose is usually determined in one of three ways. (a) Electronic joint encoder [3], rarely available as most facilities do not upgrade their vintage C-arms. (b) Optical tracker could localize the C-arm [4], which

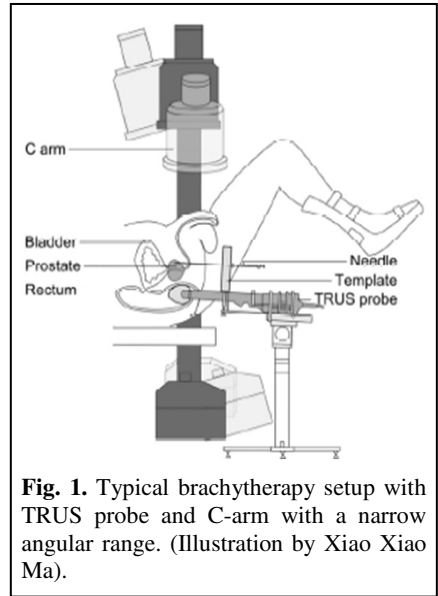


Fig. 1. Typical brachytherapy setup with TRUS probe and C-arm with a narrow angular range. (Illustration by Xiao Xiao Ma).

introduces prohibitive devices and logistical complexity in the otherwise streamlined clinical procedure. (c) Radiographic fiducials placed in the field of imaging, from which the pose of the fluoroscope can be discerned in relation to the fiducial structure [5]. Radiographic fiducials are independent of any C-arm brand or type. Unfortunately, fiducials must be segmented in fluoroscopy, a perennial issue for clinical practice. Also, the fiducials occupy priceless real estate in the image, forcing the prostate toward the edges where image distortion tends to be more severe, which in turn demands online distortion correction. Finally, mounting the fiducial to be visible in all C-arm poses is a major procedural challenge and requires a large and thus very expensive image detector.

We propose a method that is radically different from the prior art. We estimate the relative pose of C-arm images by the registration of the 2D fluoroscopy images to the 3D TRUS volume, and by doing so we estimate the poses of C-arm images in a coordinate system fixed to the prostate. As a byproduct, we receive an estimate of the registration between the C-arm and ultrasound spaces, which is required for dynamic dosimetry. The objective is to recover the C-arm poses with an accuracy that is sufficient for subsequent reconstruction of the implanted seeds like in [4] or [5].

Our contribution is the first report of C-arm pose estimation by registration of 2D C-arm images to 3D ultrasound. The apparent straightforwardness of our approach should not misrepresent the investment of creative effort needed to make it a workable clinical tool, despite the availability of underlying technical components. We devised an elegant and entirely novel solution for a longstanding clinical problem by adapting available techniques. Besides registering the C-arm space directly to the prostate without surrogate markers, the most salient feature is that we avoid adding any instrumentation to the standard clinical setup. Our solution blends seamlessly with the current clinical install base and so it could be rapidly introduced to community care with minimal cost and make a positive impact in the very near future.

Prior work in 2D/3D registration can be divided into two categories: feature-based and intensity-based methods. Feature-based methods [6] use distance between corresponding point pairs or surfaces as a measure to be optimized. Establishing point correspondences and minimizing the distance between them is alternated and repeated iteratively until convergence. Consequently, a segmentation of the data is required. Intensity-based methods compare the 2D image with a digitally reconstructed radiograph (DRR) created from the 3D volume. One can compare the imprints of anatomical structures obtained from either gradient information or voxel intensity [7-9]. Literature on registering a 3D ultrasound volume and 2D images has been scarce. Hummel *et al.* [10] used 2D ultrasound to 3D CT registration, where fiducial spheres served as markers for alignment. Leung *et al.* [11] reported rigid registration of 2D cardiac X-ray images with 3D echocardiography based on intensities, with cross-correlation and sum of squared distances metrics. They also used *a priori* knowledge of the full pose to initialize the registration. They report rather large lateral translation and rotation errors of about 8 mm and 8°, respectively.

2 Methodology

2.1 Central Intuition

While seeds in an implanted prostate show up well in fluoroscopy, TRUS images of the prostate are saturated with artifacts emanating from seeds. In both modalities but especially in TRUS, artifacts often masquerade as seeds, an effect called *false positive appearances*. Seeds may also obscure one another in both modalities, an effect called *hidden seeds*. Although there is no exact matching between true positive appearances of seeds in TRUS and fluoroscopy, seeds carry enough common information for an intensity-based 2D/3D registration to “hone in” on the correct pose between the two. Due to false positives, exact segmentation of the seeds in TRUS is unattainable and it cannot be used for registration. Figure 2-left shows that even phantom images contain many false positives and there are more of them in human images.

2.2 Single-View 2D/3D Registration

We apply 2D/3D registration considering the 3D TRUS as the moving volume and the 2D fluoro as the fixed image. As we only use one fluoro image at a time, we termed this single-view registration or shortly SVR.

Metric: We implemented the normalized cross correlation (NCC) metric that considers all pixel values in the images during registration. Fixed image pixels and their positions are mapped to the moving image. The correlation is normalized by the auto-correlations of both the fixed and moving images. **Transform:** After TRUS imaging, the probe is retracted from the rectum, so as not to block seeds during fluoroscopy. This causes the prostate to relax posteriorly, but usually without any apparent deformation. Good clinical practice requires minimal rectal pressure, to prevent deformation during TRUS imaging involving probe translation [2]. Prostate deformation is curtailed by minimizing rectal pressure. Since there is no delay between 3D TRUS and 2D fluoro acquisition, our assumption of 2D/3D rigid registration should suffice.

We implemented a transformation of six parameters, three for Euler angles and three for translation.

Initial Guess: In the operating room, we have an accurate and consistent initial guess for the registration. Standard patient positioning allows for aligning the main axes of the TRUS and the C-arm. We also place the prostate around the center of the C-arm that yields an estimate for the translation component of the transformation [5].

DRR Interpolator: A ray casting interpolator is used to project the moving TRUS volume onto the fixed 2D fluoro image. We cast rays from the X-ray source of the

C-arm through the moving TRUS volume to each pixel of the fixed fluoro image. As seeds in C-arm show up as point-like structures we introduced a threshold during ray casting to ensure that only seed-like appearances from TRUS contribute to the DRR.

Optimizer: The optimizer chosen to optimize the NCC similarity metric is the (1+1) Evolutionary Strategy [13]. In this strategy, both the number of parents and the population size are set to one: $\mu = \lambda = 1$. Mutation is accomplished by adding a vector of usually uncorrelated Gaussian random numbers, i.e. $\Sigma = \text{diag}(\sigma^2)$ is a diagonal matrix. Step size adaptation can be performed according to Rechenberg's 1/5-rule: if less than 20% of the generations are successful then decrease the step size for the next generation; if more than 20% are successful, then increase the step size in order to accelerate convergence. As TRUS and fluoro are both spotted with positive appearances in low-signal areas, the registration could be trapped in local minima. To counteract this problem, we restart the registration several times with slightly changing the initial pose and then take the median.

2.3 Multi-view 2D/3D Registration

As single-view 2D/3D registration is prone to depth error along the X-ray beam, we decided to bootstrap the registration in its weakest dimension, depth. In clinical practice, we can position the C-arm so that the center of the prostate (judged by the extent of seeds cloud in the images) is near the C-arm's isocenter. With non-isocentric C-arms, we can use coplanar rotation and set the axis of rotation in the prostate. With this setup, the prostate appears at a constant depth, especially at small (maximum $\pm 15^\circ$) rotations. We estimate this average depth from multiple fluoroscopy views. We pick pairs of fluoro images with wide angular separation and one-axis rotation between the two. We register them simultaneously to 3D TRUS in a much simplified manner,

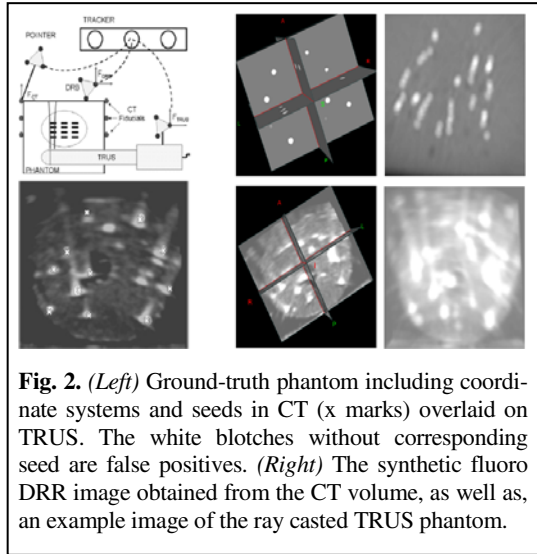


Fig. 2. (Left) Ground-truth phantom including coordinate systems and seeds in CT (x marks) overlaid on TRUS. The white blotches without corresponding seed are false positives. (Right) The synthetic fluoro DRR image obtained from the CT volume, as well as, an example image of the ray casted TRUS phantom.

where we optimize only the depth and one rotation. The cost function implemented computes the sum of NCC between the fixed image and the DRR. Then we feed the approximate depth to the single-view 2D/3D registration (Sec. 2.2.)

2.4 Ground-Truth Phantom

A commercial brachytherapy phantom (CIRS Inc., Virginia) was implanted with 48 non-radioactive Pd103 seeds according to a clinically realistic implant plan. Six CT fiducials were mounted on the opposite walls of the phantom, three on each side. This configuration of fiducials guarantees maximum target registration accuracy in the center where the prostate is located, Figure 2 (left). A dynamic reference body (DRB) optical marker (Traxtal, Versa Trax, model TT002-B) was affixed to the phantom, to define a coordinate frame that was used for constructing the 3D TRUS volume. The fiducials were localized with a calibrated pointer, relative to the DRB. The TRUS probe was tracked optically with Polaris (Northern Digital, Waterloo, Canada.), relative to the DRB. The CT fiducials were localized by segmentation and registered to their respective locations determined earlier by the tracked pointer. This yields ground-truth registration between the TRUS and CT.

2.5 Ground-Truth 2D X-ray and TRUS Images

In order to have perfectly accurate reliable ground-truth image poses, we decided to derive fluoroscopy images from high resolution CT volume image of the phantom. Further, the DRR-s of the phantom computed are similar to C-arm images, but with the immediate advantage that CT has higher geometrical accuracy with less experimental complexity. We acquired CT images of the phantom, with 0.3×0.3 mm in-plane resolution and 0.6 mm slice thickness. As seeds are very prominent in CT, we only clip a region of interest and use window-level scaling to create the 8-bit CT. Window/level scaling is a method for mapping a range of intensity to a different scale. Generally, CT data has intensity values between -2000 to 4095. In our case, true seeds (intensities above 1700) were mapped to 8-bit grayscale. Then a suitable threshold of 100 was applied to suppress all remaining artifacts, thus yielding ground-truth seed locations in CT that was registered to TRUS earlier. We created DRR-s from the 8-bit CT, such as in Figure 2-right. The angular range of the

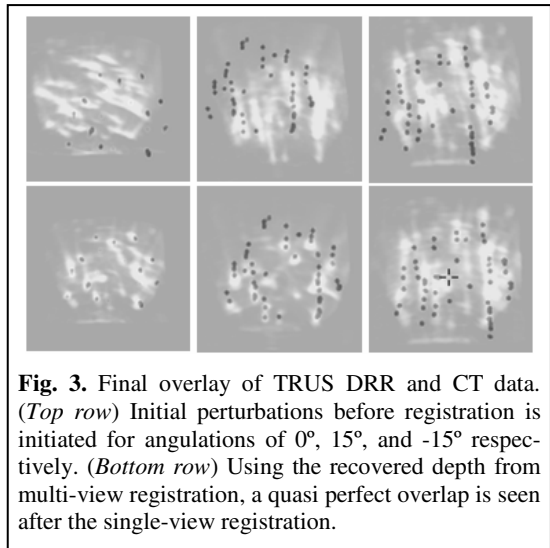


Fig. 3. Final overlay of TRUS DRR and CT data. (*Top row*) Initial perturbations before registration is initiated for angulations of 0° , 15° , and -15° respectively. (*Bottom row*) Using the recovered depth from multi-view registration, a quasi perfect overlap is seen after the single-view registration.

of the

C-arm is constrained by the patient, table, or brachytherapy mount. In clinical practice, the maximum in-plane and out-of-plane rotation is about $\pm 15^\circ$. We created DRR-s at 15° increments, using the geometry of a clinical C-arm from [5]. Again, the DRR-s played the role of the 2D fluoro images in subsequent phantom experiments. Using tracked freehand TRUS acquisition, we scanned the entire phantom volume systematically with translational motion, with continuous image capture. 2D pixel spacing was 0.14×0.13 mm. Interframe spacing was 0.5 mm. An 8-bit 3D TRUS volume was compounded from the 2D images. A distinctive feature of our approach is that false positive seed appearances are allowed to remain in the TRUS data. As seeds create strong sonic impression in TRUS, as seen in Figure 2-right, it is not necessary to filter soft tissue signal from the images.

3 Results and Discussion

The clinical degree of accuracy for pose estimation and implant 3D reconstruction is presented in the paper by Jain *et al.* [12]; thus we require curtailing the rotation error to $\pm 4^\circ$ and lateral translations to about ± 2 mm. In 2D/3D registration, the cost metric usually has difficulties with properly “driving” the depth component of the pose. In C-arm reconstruction, however, the exact same effect is working for our advantage, because the reconstruction metric is similarly insensitive to the depth component of the C-arm pose. Jain *et al.* [14] found that “reconstruction error is insensitive to mis-calibration in origin and focal length errors of up to 50 mm”, inferring that huge depth errors are permissible if image poses shift together. What follows is that if the prostate is kept near the isocenter, projection and reconstruction are both insensitive to depth; a fact that we exploited in designing the multi-view registration scheme.

3.1 Phantom Studies

We used 20 random perturbations of maximum ± 5 mm translation and $\pm 5^\circ$ rotation about the ground-truth poses, and then we repeated the same with ± 10 mm and $\pm 10^\circ$. These perturbation values are justified since standard patient positioning constrains the rotation of the C-arm and allows for quasi iso-centric positioning of the prostate. For each case we run a single-view (SVR) and multi-view (MVR) registration. The threshold in the DRR interpolator was set at 150, after trying several values. The results are summarized in Figure 3 and Table 1. All results are reported as absolute distances. With ± 5 mm and $\pm 5^\circ$ perturbations, SVR performed within clinical limits. All runs converged and are reported. Average rotation and lateral translation (Tx/Ty) errors were 0.8° (STD= 2.3°) and 0.5 mm (STD=1.6 mm), respectively. The depth error (Tz) was 3.5mm (STD=3.5 mm). Then MVR reduced the depth error to 1.2 mm (STD=2.8 mm). Note that lateral translation and rotation errors remained about the same, as MVR affects only the depth (Section 2.3.) With ± 10 mm and $\pm 10^\circ$ perturbation, pose recovery was less accurate. Still, all runs converged and are reported. SVR achieved average rotation and lateral translation errors of 3.8° (STD= 3.4°) and 2.3 mm (STD=2.1 mm), respectively. Then MVR reduced the depth error to 2.9 mm (STD=2.1 mm).

Table 1. Ground-truth phantom results. Average translation and rotation error with STD, for single-view and multi-view registration.

	± 5 mm and $\pm 5^\circ$ perturbation			± 10 mm and $\pm 10^\circ$ perturbation		
	Tx/Ty	Tz	Rotation	Tx/Ty	Tz	Rotation
SVR	0.5 \pm 1.6	3.5 \pm 3.5	0.8 \pm 2.3	2.3 \pm 2.1	5.7 \pm 4.7	3.8 \pm 3.4
MVR	0.4 \pm 1.5	1.2 \pm 2.8	1.0 \pm 2.3	2.2 \pm 2.4	2.9 \pm 2.1	3.9 \pm 3.4

We implemented the single-view and multi-view registration technique using the Insight toolkit (ITK). We used an Intel Core2, 2.4 GHz dual-core computer. The average speed of the SVR registration was 60 seconds which is feasible in clinic.

3.2 Clinical Results

Clinical patient data was collected under ethics board approval. Here we report results on the first patient dataset in the trial. Nine C-arm fluoroscopy images were acquired and their relative poses recovered with a precision-machined radiographic fiducial [5], serving as ground-truth. For the 2D/3D registration, the fluoro images were dewarped and a 256 \times 256 pixel ROI was cut around the prostate's center. The registration parameters were retuned: since the rotation has a stronger initial guess in the actual clinical setup we assigned higher optimization weights to rotation than to translation. We used 20 random perturbations of maximum ± 5 mm translation and $\pm 5^\circ$ rotation about the true poses, and then we repeated the same with ± 10 mm and $\pm 10^\circ$. We only ran the single-view registration (SVR). MVR could not be tested, because the fiducial prevented us from setting the prostate in the isocenter. At ± 5 mm and $\pm 5^\circ$ perturbations, human patient data, the average rotation and lateral translation errors were 0.6° (STD= 3.0°) and 1.5 mm (STD=2.8 mm), respectively, relative to the ground-truth established by a precise radiographic fiducial. The average depth recovered was 3.8 mm (STD=4.2 mm). After doubling the maximum perturbation, the average rotation and lateral translation errors were 1.7° (STD= 5.7°) and 1.4 mm (STD=4.0 mm), and average depth of 5.1 mm (STD=6.7 mm). Although average errors grew, they still remained below the clinically acceptable limits.

Visual observation is not sufficient for more precise evaluation, due to the concurrent effects of true seeds, false positives and hidden seeds. When these are all compounded in ray casting, the best matching is not perceivable to the human eye. Nonetheless, true seeds carry sufficient information for the intensity-based metric to lock on the pose. This phenomenon underscores why explicit segmentation of the seeds in transrectal ultrasound cannot be used for registration or for the evaluation thereof. Our experience with early clinical data clearly and forcefully underlines the inherent difficulty of reliable validation based on explicit segmentation of seeds in TRUS. For many seeds, the expert clinician could not tell apart true seeds from noise in TRUS. After two weeks, the clinician repeated the task of seed identification in the same patient data and nearly half of all seed locations were picked differently, suggesting unreliable consistency in visual seed localization. A possible workaround might be applying multiple segmenters, but that is likely to fail as well. Earlier, Orio *et al.* reported the same difficulty [15], as they were able to visually identify 20–25% of all the implanted seeds in TRUS. We established registration ground-truth as suggested by Jain *et al.* in [5], by pre-registration of a

radiographic fiducial and the TRUS coordinate space. This approach, however, is not generally robust and, as Jain *et al.* mentioned, it may require compensation for biases emanating from multiple sources.

In summary, we presented the first application of 2D/3D registration in prostate brachytherapy for estimating the C-arm pose. Overall accuracy and robustness were excellent on phantom data and adequate on human data. The ultimate test, still, is whether this pose estimation will prove to be sufficient in brachytherapy implant reconstruction such as [4] or [5] – an issue of great clinical interest.

References

1. Jemal, A., Siegel, R., Ward, E., et al.: Cancer Statistics. *Cancer J. Clin.* 59(4), 225–249 (2009)
2. Nag, S., Ciezki, J.P., Cormack, R., et al.: Intraoperative planning and evaluation of permanent prostate brachytherapy: report of the American Brachytherapy Society. *Int. J. Radiat. Oncol. Biol. Phys.* 51(5), 1422–1430 (2001)
3. Westendorp, H., Hoekstra, C.J., van't Riet, A., et al.: Intraoperative adaptive brachytherapy of iodine-125 prostate implants guided by C-arm cone-beam computed tomography-based dosimetry. *Brachytherapy* 6(4), 231–237 (2007)
4. Su, Y., Davis, B.J., Furutani, K.M., et al.: Seed localization and TRUS- fluoroscopy fusion for intraoperative prostate brachytherapy dosimetry. *Computer Aided Surgery* 12(1), 25–34 (2007)
5. Jain, A.K., Deguet, A., Iordachita, I., et al.: Intra-operative Guidance in Prostate Brachytherapy Using an Average C-arm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II. LNCS*, vol. 4792, pp. 9–16. Springer, Heidelberg (2007)
6. Yamazaki, T., Watanabe, T., Nakajima, Y., et al.: Improvement of depth position in 2d/3d registration of knee implants using single-plane fluoroscopy. *IEEE Transactions on Medical Imaging* 23(5), 602–612 (2004)
7. Livyatan, H., Yaniv, Z., Joskowicz, L.: Gradient-based 2-D/3-D rigid registration of fluoroscopic X-ray to CT. *IEEE Trans. Med. Imaging* 22, 1395–1406 (2003)
8. Mahfouz, M., Hoff, W., Komistek, R., Dennis, D.: A robust method for registration of three-dimensional knee implant models to two-dimensional fluoroscopy images. *IEEE Transactions on Medical Imaging* 22, 1561–1574 (2003)
9. Lau, K., Chung, A.: A global optimization strategy for 3d-2d registration of vascular images. In: *Proceedings of 17th British Machine Vision Conference*, pp. 489–498 (2006)
10. Hummel, J., Figl, M., Bax, M., Bergmann, H., Birkfellner, W.: 2D/3D registration of endoscopic ultrasound to CT volume data. *Physics in Medicine and Biology* 53(16), 4303–4316 (2008)
11. Leung, K.Y.: Registration of 2D cardiac images to real-time 3D ultrasound volumes for 3D stress echocardiography. In: *SPIE Medical Imaging*, vol. 6144, pp. 405–416 (2006)
12. Jain, A., Fichtinger, G.: C-arm Tracking and Reconstruction without an External Tracker. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006. LNCS*, vol. 4190, pp. 494–502. Springer, Heidelberg (2006)
13. Arnold, D.V., Brauer, D.: Local performance of the (1 + 1)-ES in a noisy environment. *IEEE Transactions on evolutionary computation* 6(1), 30–41 (2002)
14. Jain, A., Kon, R., Zhou, Y., Fichtinger, G.: C-arm calibration - is it really necessary? In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005. LNCS*, vol. 3749, pp. 639–646. Springer, Heidelberg (2005)
15. Orto, P.F., Tutar, I.B., Narayanan, S., et al.: Intraoperative ultrasound-fluoroscopy fusion can enhance prostate brachytherapy quality. *Int. J. Radiat. Oncol. Biol. Phys.* 69(1), 302–307 (2007)

Cognitive Burden Estimation for Visuomotor Learning with fNIRS

David R.C. James, Felipe Orihuela-Espina, Daniel R. Leff, George P. Mylonas,
Ka-Wai Kwok, Ara W. Darzi, Guang-Zhong Yang

Imperial College London, United Kingdom

{d.james, f.orihuela-espina, d.leff, george.mylonas}@imperial.ac.uk
{k.kwok07, a.darzi, g.z.yang}@imperial.ac.uk

Abstract. Novel robotic technologies utilised in surgery need assessment for their effects on the user as well as on technical performance. In this paper, the evolution in ‘*cognitive burden*’ across visuomotor learning is quantified using a combination of functional near infrared spectroscopy (fNIRS) and graph theory. The results demonstrate escalating costs within the activated cortical network during the intermediate phase of learning which is manifest as an increase in cognitive burden. This innovative application of graph theory and fNIRS enables the economic evaluation of brain behaviour underpinning task execution and how this may be impacted by novel technology and learning. Consequently, this may shed light on how robotic technologies improve human-machine interaction and augment minimally invasive surgical skills acquisition. This work has significant implications for the development and assessment of emergent robotic technologies at cortical level and in elucidating learning-related plasticity in terms of inter-regional cortical connectivity.

Keywords: Robotic surgery, cognitive burden, graph theory, near infrared spectroscopy, neuroergonomics.

1 Introduction

The use of robots for enhancing surgical performance, particularly for minimally invasive surgery (MIS) has opened a number of new opportunities. In addition to tremor filtration, motion scaling, and virtual fixtures (active constraints), it also provides a natural framework for integrating pre- and intra-operative imaging combined with augmented reality. Recently, the concept of gaze-contingent motor channelling (GCMC) has been proposed as a way of potentially improving hand-eye coordination for certain aspects of surgery where movement impacts on the surgical site, *e.g.* beating heart surgery or targeted therapy and focused energy delivery of robotically assisted MIS [1]. The method relies on eye gaze and its dynamic pursuit movement to form constraints that can effectively direct hand movements for accurate targeting in the presence of large tissue deformation. It is intuitive that in robotic tasks where surgeons are effectively ‘*perceptually docked*’ with their environment [2], subject-specific neurocognitive behaviour plays a critical role that influences the final quality

of the procedure. The paradigm of *neuroergonomics* utilizes neuroimaging to assess the impact of undertaking tasks at brain level and uses this information to aid the development of performance enhancing tools [3]. Methods of assessing task-induced mental workload are directed towards the amount of mental resources required versus those available *i.e.* ‘functional segregation’ [4]. However, this approach is unable to capture interactions between recruited cortical areas, *i.e.* ‘functional integration.’ The concept of ‘*cognitive burden*’ relates to the most efficient use of cortical interactions and deviations thereof. Neuroplasticity underpins the learning process [5], therefore it is necessary to study the evolution of learning related cortical interactions over time.

The purpose of this paper is to evaluate economic changes in a cortical network associated with the acquisition of a complex simulated beating heart procedure with and without the use of GCMC. At each phase of learning, the activated network is evaluated to assess the degree of task-induced *cognitive burden*.

Functional Near Infrared Spectroscopy (fNIRS) [6, 7] allows for the assessment of brain function in a more realistic human-robotic environment. The prefrontal cortex (PFC) and parietal cortex (PC) are interrogated since a frontoparietal (F-P) network underlies visuomotor tracking [8]. Task evoked brain activation is identified using a classical statistical approach, leading to the construction of an activity guided map. A cortical network defined by phase locked cooperation between channels is analysed in terms of its economy within the graph theory framework to unveil longitudinal fluctuations in task related *cognitive burden*. In this paper, a novel use of graph theory in the analysis of fNIRS data enables appreciation of the interaction between cortical regions subserving task execution. Ensuing information may be valuable in ensuring that novel technology is beneficial to the user and that performance improvements are not detrimental to other facets of task execution. This study may have potential implications for guiding the design and use of performance enhancing technology not only in surgery but in other safety critical tasks such as aviation. It is hypothesized that performance improvements evoked by GCMC will be secondary to enhanced interaction within a F-P network.

2 Methods

2.1 Subjects, Task Paradigm and fNIRS

A total of 21 healthy right-handed subjects were recruited to track a moving target on a simulated beating heart using a virtual tool controlled by a haptic manipulator (SensAble Tech. MA). Participants performed this without (control group, n=11) or with GCMC (experiment group, n=10). An eyetracking device (Tobii, Sweden) distinguishes the subject’s fixation point. This information is used to define the force required to constrain the subject’s hand allowing accurate target localization as illustrated in Fig. 1 [1]. A block design experiment was conducted comprising baseline rest (30s), followed by five blocks of the task (20s) and inter trial rest (30s). Performance was determined as the distance from the tool tip to the target (pixels). Subjects undertook task sessions on six separate days. Cortical haemodynamic data was acquired with a fNIRS system (ETG-4000, Hitachi Medical Corp, Japan). Twenty four channels in two 3 x 3 arrays were positioned over the left PC and PFC using the UI

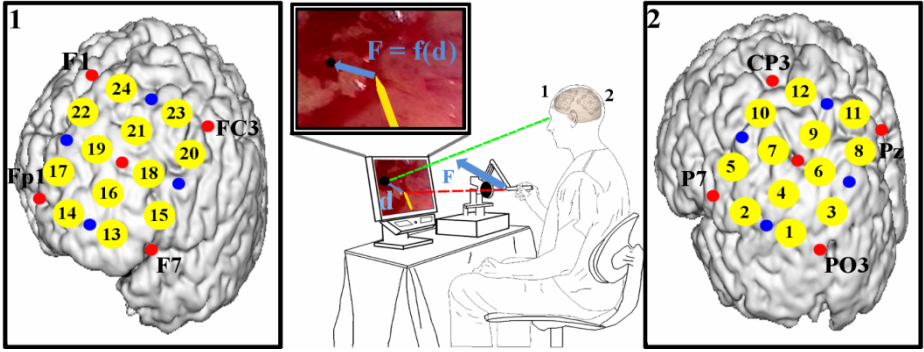


Fig. 1. Experimental set up (central panel) illustrating the moving target (black dot) and the virtual tool (yellow line) controlled by the haptic manipulator. A force (F) proportional to the distance (d) from the tool tip to the fixation point is fed back to the subject (via the haptic manipulator) to diminish this separation [1]. The movement space of the haptic phantom was restricted to 2D as the monitor screen is inherently a 2D space. Optode arrays can be appreciated on the subject's head and the approximate channel location overlain onto a reference atlas for PFC and PC (subplots 1 and 2 respectively). Channels (yellow circles) and optode emitters and detectors (red and blue circles respectively) are displayed in relation to the UI 10/10 coordinate systems [9].

10-10 system [9] as illustrated in Fig. 1. Raw optical data was converted into relative changes in oxygenated haemoglobin (ΔHbO_2) and deoxygenated haemoglobin (ΔHb) using the modified Beer-Lambert Law, then decimated, linearly detrended, and subjected to data integrity checks. Representative haemodynamic data are demonstrated in Fig. 2. Task induced ΔHbO_2 and ΔHb (t-test rank sign, $\alpha < 5\%$) were determined for each session and channel for each group.

2.2 Cross Correlation and Activity Guided Network Formation

Let ΔHb_i and ΔHb_j be the bidimensional $\langle \Delta HbO_2, \Delta Hb \rangle$ haemodynamic response at channels i and j respectively and T_i and T_j their corresponding signal length in samples. Let $R_{i,j}(\tau_t, \tau_{Hb})$ be the normalised 2D cross-correlation between the responses at channels i and j , where τ_t is the temporal lag and $0 \leq \tau_t < T_i + T_j - 1$ and τ_{Hb} is the haemodynamic dimensional lag and $0 \leq \tau_{Hb} < 3$. An $m \times m$ phase locked zero lagged correlation matrix $C = \{c_{ij} = R_{i,j}(\tau_t = (T_i + T_j - 1)/2, \tau_{Hb} = 1)\}$ can be formed between each pair of channels. A pseudo-metric of signal similarity can be constructed as $d_{ij} = 1 - R_{i,j}$, and the accompanying distance matrix $d = \{d_{i,j}\}$. The fNIRS channels can be regarded as nodes in a graph. A fully connected weighted undirected graph $G(C)$ can be constructed using c_{ij} as the weight of the connecting edge between channels i and j , and d_{ij} representing the ease of information flow between the channels. The lower the d_{ij} , the easier to transmit information between nodes i and j . $G(C)$ is then pruned retaining only activated channels as defined below. The haemodynamic activity signature as recorded by fNIRS is described as a task-evoked increase in HbO_2

and a concomitant decrease in HHb . Each haemoglobin species can vary in either direction with respect to the pre-stimulus baseline, and any change may or may not reach the statistical threshold. The following four patterns describe cortical activity by means of a t-test ($\alpha < 5\%$): (A) ΔHbO_2 increment and ΔHHb decrement both reaching statistical significance, (B and C) ΔHbO_2 increment and ΔHHb decrement with only one reaching significance (D) ΔHbO_2 increment and ΔHHb decrement with neither species reaching significance. A threshold was then applied to the connected graph, $G(C)$ permitting only those links between any two channels representing the activity patterns (A-D), thereby capturing the evolving relationships within an activated F-P network. It is customary to modulate the strength of the activity guided edges according to the statistical significance of the pattern at each of the edge end-nodes. Here we use scales 1, 0.75, 0.75 and 0.5 for patterns A to D respectively, thus:

$$\widehat{C} = \{c_{ij} \cdot scaleFactor_i \cdot scaleFactor_j\} \tag{1}$$

$$\widehat{D} = \{d_{ij} \cdot scaleFactor_i \cdot scaleFactor_j\} \tag{2}$$

The activity guided brain network is then $G(\widehat{C})$. This serves as the substrate for calculating the measures of network cost and efficiency.

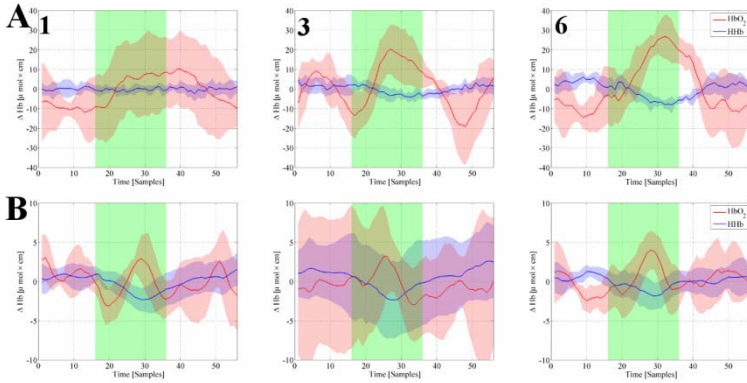


Fig. 2. Session averaged haemodynamic data from representative control (A) and GCMC (B) subjects for sessions 1, 3 and 6. Mean (bold line) and standard deviation (shaded region) are indicated for both HbO_2 (red) and HHb (blue). The 20s task period (green shaded region) is bounded by rest and recovery periods (white).

2.3 Network Efficiency and Cognitive Burden

Excluding self connections (intra-channel) for which the flow of information between neurons is assumed infinite, the cost of a weighted graph or network is defined as per Eq. 3, and the network global efficiency has been defined as in Eq. 4 [10]:

$$K(G) = \sum_{i \in G} \sum_{j \in G} d_{ij} |i \neq j| \tag{3}$$

$$E_{glob}(G) = \frac{1}{N(N-1)} \sum_{i \in G} \sum_{j \in G, j \neq i} \frac{1}{d_{ij}} \quad (4)$$

Both quantities, cost and global efficiency are normalised in proportion to the maximum efficiency and cost of a comparable network comprising all possible connections [11]. The economy of a cortical network has been defined as:

$$Economy(G) = E_{glob}^{norm}(G) - K^{norm}(G) \quad (5)$$

From its definition, it is intuitive that the cognitive burden is inversely related to the network economy. Furthermore, if the cost dominates the equation, the cognitive burden will be positive and high. Conversely, if the efficiency dominates the equation, the cognitive burden will be negative and low.

3 Results

One subject withdrew from the study after the second session and the remaining 20 subjects completed all six sessions conducted over a median of eight days. Fig. 3 illustrates the evolution in F-P excitation across the experiment for gaze-assisted and unassisted learners.

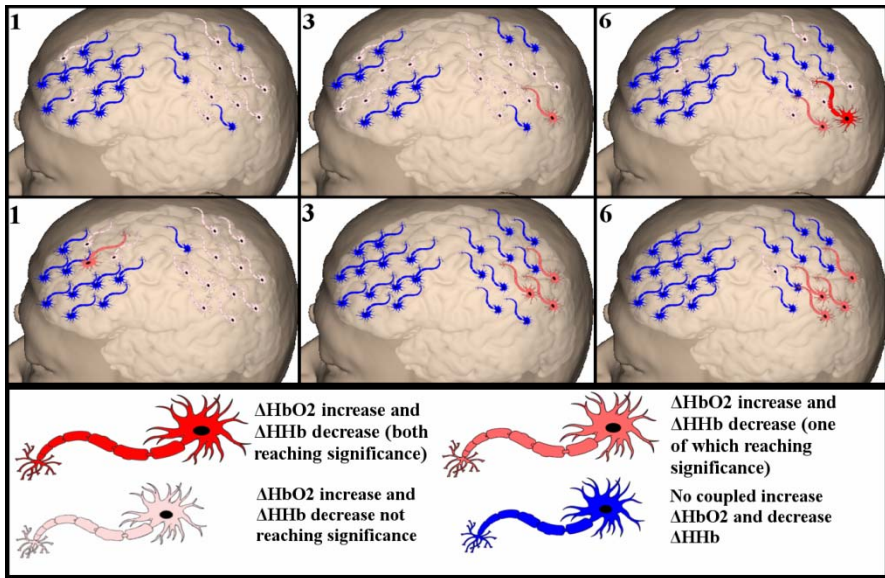


Fig. 3. Groupwise statistical analysis of cortical haemodynamic responses in unassisted (top) and GCMC-assisted (bottom) learners across practice (Sessions 1, 3 and 6). Channels are represented by coloured neurons according to the represented haemodynamic pattern (figure key). The spatial attenuation in PFC and PC responses and magnification of activity in the PPC can be appreciated.

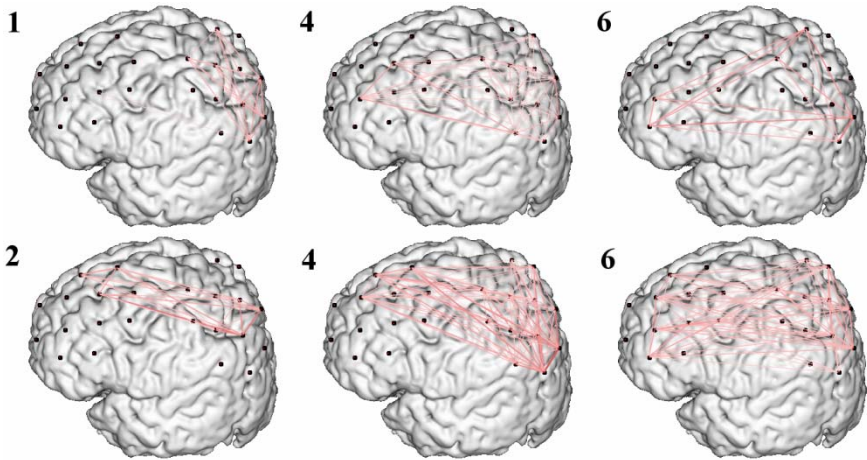


Fig. 4. F-P activated network graphs overlaid onto an MRI image of the cortical surface. Graph nodes (black circles) indicate approximate channel location and connections are weighted according to the strength of the association between nodes. A representative control (top) and GCMC (bottom) subject is illustrated at practice session time points (labelled black). The majority of connections are formulated around the intermediate phase of practice.

Longitudinal changes in brain behaviour may be appreciated as a spatial attenuation in the F-P response and a magnification of excitation within the posterior parietal cortex (PPC). These longitudinal changes appear to occur more rapidly in GCMC enabled learners. Fig. 4 displays the evolution in regional connectivity within the activated F-P network for GCMC-assisted and unassisted learners. During the intermediate phase of learning, a greater number of intra- and inter-regional functional connections are evident, which may reflect formation of new cortical networks. In general, at each phase, a greater number of cortical connections were observed in assisted learners, reflecting the impact of GCMC assistance on F-P connectivity. Fig. 5 depicts the progression of the cost, efficiency and therefore *cognitive burden* evoked by the task at each time point across learning. Both groups progressed from a region of low cost and low efficiency (session 1) to a region of low cost but increased efficiency (session 6). This is achieved with different network development strategies. Evidently, an escalation in activated network cost accompanies the functional increase in cortical connections during intermediate learning. Furthermore, the magnitude of this cost increase is greater in GCMC-assisted learners, and occurs just after the sessions associated with maximal improvement in technical skill in this group (1st session accuracy (pixels) median (SD) = 46.3 (22.4); 2nd session = 30.2 (15.8) ($p < 0.000$), and significantly better accuracy compared to the control group until practice termination (session 3-6, $p = 0.019, 0.023, 0.001, 0.000$). In contrast to the changes observed in early practice, a reduction in the cognitive burden and an increase in F-P network efficiency seem to define ongoing performance enhancement in late practice (sessions 3-6).

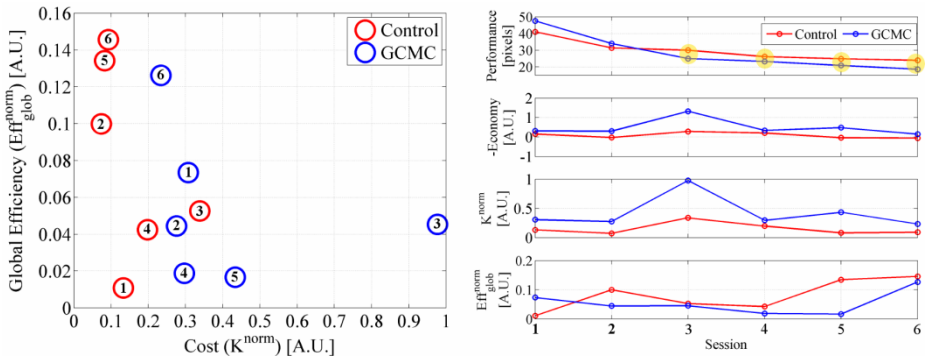


Fig. 5. Evolution in the *cognitive burden* across practice for control (red) and GCMC (blue). The cost efficiency plane (left panel) demonstrates the progression of these parameters across practice sessions (1-6) and that the foundation for the increase in cognitive burden demonstrated in session 3 (right panel) is due to an increase in cost. By session 6 both groups have migrated to a region indicative of a high efficiency and low cost. GCMC users significantly outperformed the control group in session 3-6 (highlighted yellow).

4 Discussion and Conclusion

In this paper, we have proposed a novel framework for analysing longitudinal fNIRS data in order to estimate the *cognitive burden* associated with learning visuomotor tracking. Using this strategy, it has been possible to unveil transformations within an activated F-P network as a result of deliberate practice. Enhanced tracking accuracy is indexed by a functional increase in *cognitive burden* and attenuation in cortical haemodynamics across a F-P network. More importantly, fundamental differences in the pattern of cortical excitation change and variation in cognitive burden have been exposed depending upon the mode of skills learning. Specifically, GCMC learners demonstrate the greatest increase in network connections and cognitive burden, and more rapid F-P attenuation. Therefore, the technical improvements induced by GCMC over and above unaided learning may be secondary to improved F-P communication. This may have relevance to training on tasks that require precise visuomotor skill and also in the assessment of emergent surgical technologies.

Acknowledgements. The authors would like to thank Daniel Ramirez-Cano for useful discussions.

References

1. Mylonas, G.P., Kwok, K.W., Darzi, A., Yang, G.Z.: Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery. In: Med. Image Comput. Comput. Assist. Interv. Int. Conf., vol. 11, pp. 676–683 (2008)
2. Yang, G.Z., Mylonas, G.P., Kwok, K.W., Chung, A.: Perceptual docking for robotic control. In: Dohi, T., Sakuma, I., Liao, H. (eds.) MIAR 2008. LNCS, vol. 5128, pp. 21–30. Springer, Heidelberg (2008)

3. Parasuraman, R.: Neuroergonomics: research and practice. *Theoretical Issues in Ergonomics Science* 4, 5–20 (2003)
4. Just, M.A., Carpenter, P.A., Miyake, A.: Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science* 4, 56–88 (2003)
5. Kelly, A.M., Garavan, H.: Human functional neuroimaging of brain changes associated with practice. *Cereb Cortex* 15, 1089–1102 (2005)
6. Leff, D.R., Elwell, C.E., Orihuela-Espina, F., Atallah, L., Delpy, D.T., Darzi, A.W., Yang, G.Z.: Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual coordination task: an fNIRS study. *Neuroimage* 39, 805–813 (2008)
7. Leff, D.R., Orihuela-Espina, F., Atallah, L., Darzi, A., Yang, G.Z.: Functional near infrared spectroscopy in novice and expert surgeons—a manifold embedding approach. In: *Med. Image Comput. Comput. Assist. Interv. Int. Conf.*, vol. 10, pp. 270–277 (2007)
8. Hill, H., Raab, M.: Analyzing a complex visuomotor tracking task with brain-electrical event related potentials. *Human Movement Science* 24, 1–30 (2005)
9. Jurcak, V., Tsuzuki, D., Dan, I.: 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage* 34, 1600–1611 (2007)
10. Latora, V., Marchiori, M.: Efficient behaviour of small-world networks. *Physical Review Letters* 87 (2001)
11. Achard, S., Bullmore, E.: Efficiency and cost of economical brain functional networks. *PLoS Computational Biology* 3, 174–183 (2007)

Prediction Framework for Statistical Respiratory Motion Modeling

Tobias Klinder¹, Cristian Lorenz², and Jörn Ostermann³

¹ Philips Research North America - Briarcliff Manor, NY, USA
tobias.klinder@philips.com

² Philips Research Europe - Hamburg, Germany

³ Institut für Informationsverarbeitung, Leibniz University of Hannover, Germany

Abstract. Breathing motion complicates many image-guided interventions working on the thorax or upper abdomen. However, prior knowledge provided by a statistical breathing model, can reduce the uncertainties of organ location. In this paper, a prediction framework for statistical motion modeling is presented and different representations of the dynamic data for motion model building of the lungs are investigated. Evaluation carried out on 4D-CT data sets of 10 patients showed that a displacement vector-based representation can reduce most of the respiratory motion with a prediction error of about 2 mm, when assuming the diaphragm motion to be known.

1 Introduction

Respiratory motion is a significant source of error for image guided interventions working on the thorax and upper abdomen. To reduce the uncertainties caused by breathing, different approaches for motion compensation have been presented, e.g., breathing coaching, patient-specific motion field extraction on the basis of pre-operatively acquired four-dimensional images, image-guided tracking throughout treatment or a combination of methods [1]. However, current approaches typically result in additional workload increasing treatment time.

One possibility to reduce breathing-induced problems without a significant increase in effort is by means of a statistical motion model, which is learned in advance. The prior knowledge about likely deformation can provide valuable information for image-guided interventions, image acquisition, but also to support image processing algorithms. Furthermore, the statistics covered in the model and thus the variability in breathing is interesting in itself.

Different motion models have been proposed in the past for cardiac motion [2,3] as well as for respiratory motion [4,5,6]. Besides the target organ, the motion models differ in the representation of the extracted data. However, different representations lead to different results and it is mostly not known in advance, which is best suited for the problem.

In this paper, a prediction framework for statistical lung motion modeling is proposed. Based on this general framework, different representations are presented and their characteristics and applicability for respiratory motion modeling

is investigated. The different models are compared by evaluating the model-based prediction of an unknown motion field.

Focus in this paper is set on lung motion modeling on the basis of 4D-CT. The created models are thereby patient-specific, i.e., the images for model building all belong to the same patient. As the purpose of this study was to analyse different model representations, patient-specific models are preferred over general motion models [4], due to the fact that the establishing of correspondence is much better defined. With a comparably high spatial resolution focus is set on 4D-CT.

2 Methods

2.1 Image Data

Free-breathing 4D-CT thorax data sets of 10 patients were acquired for radiotherapy treatment. In each treatment week, a ten-phase 4D-CT was acquired representing a complete breathing cycle. During acquisition of the 4D-CT data, patient respiratory traces were stored while the patient is breathing freely. The projection images were then retrospectively sorted into ten respiratory phase-based bins of 3D-CT image data based on the respiratory signal (i.e., from 0 % to 90 % phase at 10 % intervals with 0% being typically maximum inhale and 50 % maximum exhale). Depending on the course of treatment, 4D-CT data of up to eight weeks plus two weeks of preliminary examination was available per patient. For each patient at least nine 4D-CT scans were acquired. All images had an in-plane resolution of 0.85-0.97 mm and a slice thickness of 2.5 mm.

2.2 Motion Field Estimation

Lung motion fields are derived from the four-dimensional images using a surface-based tracking technique [7]. By propagating a topologically identical patient-specific lung surface mesh through all phases, anatomical point correspondences are assumed to be preserved. The trajectories of corresponding points of the adapted meshes thus provide a sparse motion field, which is finally interpolated using thin-plate-splines. The estimated continuous motion fields are denoted as T_{ij} defined between image i and j . We extensively evaluated the proposed method on lung 4D-CT images on the basis of publicly available data sets [8] achieving an accuracy of 1.0-1.7 mm in average.

For motion estimation, the lungs are at first segmented in one selected exhale image per patient. Afterwards patient-specific lung surface meshes are obtained by triangulating the thresholded image (cf. Fig. 1 (a)). The meshes cover the outer border of the lungs and also the surfaces of the pulmonary vessel tree as well as potential tumor surfaces [7]. By adapting the patient-specific reference lung mesh to the exhale images of all weeks of treatment, vertex correspondences are preserved for all images of the same patient. From the exhale phase, all meshes are tracked through the 4D-CT data set (cf. Fig. 1 (b) and (c)).

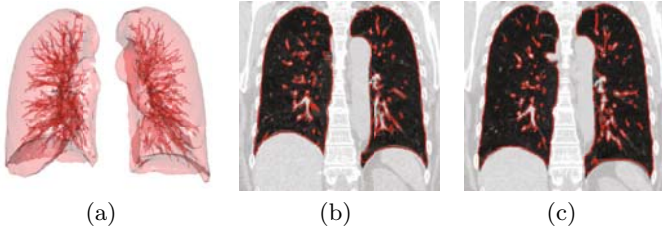


Fig. 1. Surface-based motion field estimation. Patient-specific lung mesh (a) tracked through all phases, exemplarily shown for end-exhale (b) and end-inhale (c). Note that inside structures are covered in the mesh.

2.3 Intra-week Alignment

The 4D-CT data are acquired at different treatment days. For statistical modeling, the data has to be aligned. To remove differences in patient position and image orientation but to remain differences in breathing level, all 4D-CT data sets are at first aligned by means of rigid transformations based on the segmented spinal column.

2.4 Statistical Motion Model

Statistical modeling is based on expressing the extracted data as a random vector \mathbf{z} with different possible representations proposed in the next section. A set of N corresponding vectors \mathbf{z}_n from a learning set, e.g., obtained from N different 4D-CT data sets, is combined in a matrix $\mathbf{Z} = [\mathbf{z}_1 | \mathbf{z}_2 | \dots | \mathbf{z}_N]$ and statistics can be applied on \mathbf{Z} . Note that the rows of the individual vectors \mathbf{z}_n contain corresponding information.

From the learning set, the mean is given as $\bar{\mathbf{z}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{z}_n$ and higher order statistics are calculated using the covariance matrix of the data

$$\Sigma_{\mathbf{z}\mathbf{z}} = \tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^T, \tag{1}$$

which gives the first order *statistical motion model*. The matrix $\tilde{\mathbf{Z}}$ contains the learning set after subtracting the mean.

2.5 Data Representation

The extracted information extracted from the four-dimensional images covers: i) adapted surface meshes in each phase, where each mesh is given as a set of K vertices building a vector $\mathbf{s} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_K, y_K, z_K)^T$, ii) deformations \mathcal{T}_{ij} between the images of phase i and j , iii) volume of the lungs V obtained from a lung segmentation.

From this extracted information different representations can be defined that are based on the absolute vertex locations of the segmentations or on the displacement vectors. Furthermore, time can be modeled implicitly based on the

temporal binning of the data (e.g., for 4D-CT the binning into ten phases), or explicitly. Four representations are given in the following.

Shape-Based. The shape-based representation takes the absolute vertex positions provided in all images of a dynamic sequence. Given a ten-phase 4D-CT, the vector \mathbf{z}_{SM} can be defined as

$$\mathbf{z}_{SM} = [\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_8, \mathbf{s}_9]^T. \quad (2)$$

where \mathbf{s}_i is the segmentation i -th phase.

Volume-Binned Shape-Based. In the case of the \mathbf{z}_{SM} , temporal corresponding phases are correlated in the statistical model that nevertheless can show high differences in shape due to differences in breathing. The idea of the volume-based representation is to correlate shapes that are more similar by replacing the temporal binning of the data by a volume binning (cf. Fig. 2 (a)). At first, the patient-specific volume range defined by a maximal and minimal volume is separated into a predefined number of bins. Then, the adapted meshes are sorted by there corresponding breathing volume as

$$\mathbf{z}_{VOL_i} = [\mathbf{s}], V_{li} \leq V < V_{ui}, \quad (3)$$

where V_{li} and V_{ui} define the lower and upper volume range of the i -th volume bin. The statistical model is then built for each bin from all samples over the weeks of treatment that belong to the corresponding volume range. As this representation cannot express the effect of hysteresis, separate statistical ensembles have to be built for inhalation and exhalation. Assuming K bins, this results in K models for inhalation and exhalation, respectively.

Displacement Vector-Based. Both proposed representations are so far based on absolute landmark locations. Alternatively, the representation can also be based on the displacement field. Motion field extraction results in a set of continuous transformation \mathcal{T}_{ij} that explain the deformation between the images. In the case of a ten-phase 4D-CT, \mathcal{T}_{ij} is given with respect to a reference phase (commonly the end-exhale phase which is denoted as phase 5). Thus, the \mathcal{T}_{ij} , become \mathcal{T}_{5j} with $j = 0, 1, \dots, 4, 6, \dots, 9$. The displacement vector-based representation \mathbf{z}_{MF} contains the deformations defined at a set of K vertices. In contrast to the continuous transformation \mathcal{T} , a motion field defined at a given set of locations is denoted as $\mathbf{m} = [\delta x_1, \delta y_1, \delta z_1, \dots, \delta x_K, \delta y_K, \delta z_K]^T$ where δx_k , δy_k , δz_k give the displacement in \mathbf{x} -, \mathbf{y} -, and \mathbf{z} -direction of the k -th vertex location. The vector \mathbf{z}_{MF} is then the concatenation, e.g., if ten phases should be taken into account

$$\mathbf{z}_{MF} = [\mathbf{m}_{50}, \mathbf{m}_{51}, \dots, \mathbf{m}_{54}, \mathbf{m}_{56}, \dots, \mathbf{m}_{59}]^T. \quad (4)$$

The proposed vector is based on differences and can be considered as a deformation model (cf. Fig. 2 (b)). It represents the deformation of a discrete set of landmarks. Thus, for motion prediction, reference locations have to be provided. Compared to \mathbf{z}_{SM} , the vector \mathbf{z}_{MF} contains $3 \times K$ less entries.

Parametric Displacement Vector-Based. The vector \mathbf{z}_{MF} models the time implicitly. The temporal binning of the model is defined by the temporal binning of the data. Another way to represent the trajectory is by means of a parametrisation of the trajectory. By describing the displacement over time along one coordinate as a discrete signal r_k , the discrete fourier transform $R_u = \frac{1}{M} \sum_{k=0}^{M-1} r_k e^{-j2\pi uk/M}$ can be calculated, where R_u are called fourier coefficients. The trajectory can be described with less parameters when using only the first $L \leq M$ complex coefficients. As the input signal is real, the first fourier coefficient is real. Thus, taking exemplarily only the first two fourier coefficients, the vector \mathbf{z}_{DFT} is defined as

$$\mathbf{z}_{DFT} = [Re\{R_{0,x1}\}, Re\{R_{1,x1}\}, Im\{R_{1,x1}\} \dots, Re\{R_{0,zK}\}, Re\{R_{1,zK}\}, Im\{R_{1,zK}\}]^T, \tag{5}$$

where $Re\{R_{1,xk}\}$ and $Im\{R_{1,xk}\}$ are the real and imaginary part of the second complex fourier coefficient in \mathbf{x} -direction at the k -th vertex position. Although losing some accuracy compared to the true trajectory, this representation describes the trajectory continuously and reduces the number of parameters per location (cf. Fig. 2 (c)). Due to the fact that the trajectory over the breathing cycle is rather elliptical shaped, it is mostly sufficient to choose $L = 2$ or $L = 3$.

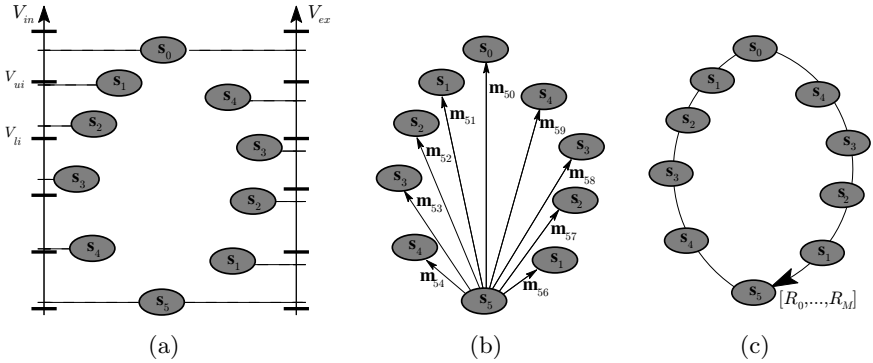


Fig. 2. Different model representations. The data is sorted in (a) into bins based on the breathing volume which is done separately for inhalation and exhalation. In (b), motion fields from a reference phase are calculated to the individual phases. The trajectory over the breathing cycle is described in (c) by fourier coefficients $R_0, \dots, R(M)$ as a continuous curve.

2.6 Prediction

Independent of the exact representation, the basic idea of the prediction framework is to use the statistics covered in \mathbf{Z} to predict a patient-specific motion, e.g., at a new treatment day. For that purpose, the vector \mathbf{z} will be partitioned into two parts $\mathbf{z} = [\frac{\mathbf{y}}{\mathbf{x}}]$, where \mathbf{y} gives the part to be predicted and the \mathbf{x} give the

information available during application to drive the model. Motion prediction is formulated as a linear multivariate regression problem

$$\mathbf{y}_{\text{pred}} = \mathbf{B} \cdot \mathbf{x}, \quad (6)$$

where \mathbf{B} can be found from the learning set using ordinary least squares

$$\text{argmin}(\mathbf{Y} - \mathbf{B}\mathbf{X})^2 \Rightarrow \mathbf{B} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} = \Sigma_{YX}\Sigma_{XX}^{-1} \quad (7)$$

where Σ_{XX} is the covariance matrix of \mathbf{X} and Σ_{YX} the cross-covariance matrix of \mathbf{Y} and \mathbf{X} . However, while the solution is optimal in a least squares sense, it is highly sensitive to overfitting. When the dimension of \mathbf{X} is higher than the number of observations, $\mathbf{X}\mathbf{X}^T$ is likely to be singular and the regression approach is no longer feasible, i.e., because of multicollinearity. We therefore perform Principal Component Regression (PCR) to reduce the noise contained in the matrix $\Sigma_{XX} = \mathbf{X}\mathbf{X}^T$ by expressing Σ_{XX} as

$$\Sigma_{XX} = \sum \phi_p \lambda_p \phi_p^T, \quad (8)$$

where ϕ_p is the p -th eigenvector with eigenvalue λ_p . By taking only the first P principal components that belong to the P largest eigenvalues, only the directions of largest variance are considered preventing the singularity of Σ_{XX}^{-1} [9].

3 Results

The different representations are evaluated by using the respective models for motion prediction. As a motion stimulator that drives the model, motion information at a small part close to the diaphragm was assumed to be known, as shown in Fig. 3. From our point of view, using a statistical motion model for motion prediction to support image guided interventions is one of the most relevant scenarios. The diaphragm is chosen as it is the main breathing motor and it can be easily tracked in many imaging modalities.

Evaluation is carried out in a leave-one out manner, i.e., the model is built from all weeks but the one under consideration. For the shape-based representations, the prediction result is given in the form of absolute landmark locations and the L_2 -norm between predicted and propagated vertices is calculated. The displacement vector-based prediction is evaluated by calculating the L_2 -norm between predicted and estimated motion vectors. For PCR regression, the first $N/2$ of N possible principal components were chosen which resulted in best results.

The results for the different representations are shown in Fig. 4 (a) and (b). As for the displacement vector-based prediction, the end-exhale was chosen as



Fig. 3. Labelled lung meshes. Motion information at small parts of the diaphragm (blue) assumed to be known.

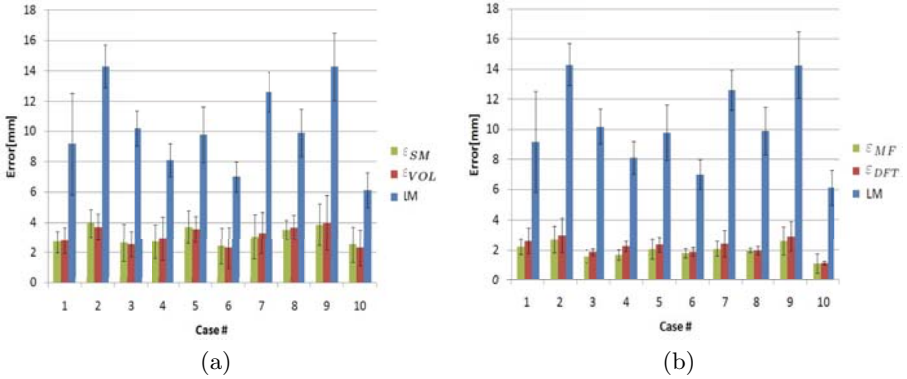


Fig. 4. Results of motion prediction in end-inhale phase over weeks of treatment for (a) shape-based representation and (b) displacement-vector based representation. In addition, mean lung amplitude (LM) is given.

the reference phase. Evaluation was focused on the error in the end-inhale phase being the phase with largest distance. In both Fig. 4 (a) and (b), the mean motion amplitude from end-exhale to end-inhale is given for comparison.

The results of the two shape-based representations, \mathbf{z}_{SM} and \mathbf{z}_{VOL_i} with $K = 5$ are given in Fig. 4 (a). Both representations can compensate most of the motion amplitude. However, setting the number K of volume bins, turned out to be complicated. If the bin size is too small, only few samples might be available for a certain volume and if it is too large, shapes with very different volumes will be correlated. As \mathbf{z}_{VOL_i} is based on the volume, it can nevertheless be easily adapted to data with a different temporal binning. The results of the two displacement vector-based representations, \mathbf{z}_{MF} and \mathbf{z}_{DFT} using three complex fourier coefficients, is shown in Fig. 4 (b). As both representations are based on the displacement from exhalation to inhalation, reference locations of the corresponding exhalation state have to be provided. Due to this additional information, the error is much lower compared to a shape-based representation. Although \mathbf{z}_{DFT} provides a more compact representation of the trajectory, the results are slightly worse compared to \mathbf{z}_{MF} . This was caused by the fact that the approximation using three fourier coefficients introduced an additional error of 0.40 mm in average. However, the advantage of \mathbf{z}_{DFT} lies in the continuous representation. Both \mathbf{z}_{MF} and \mathbf{z}_{DFT} achieved an error of about 2 mm.

4 Conclusion

Prior knowledge provided by a statistical motion model can help to overcome the problems caused by breathing. The main contributions of this paper are i) the design of a general framework for motion prediction and ii) the proposition and comparison of different representations. In addition to representations that are based on the temporal binning of the data, more flexible representations are

proposed that describe the trajectory continuously or that are oriented on the breathing volume. In principle, adaptation to data, that does not show a fixed temporal binning is thus possible. For motion prediction, the diaphragm motion was assumed to be known. The respective models were evaluated by comparing predicted and estimated motion fields. Although breathing motion is known to be highly variable, it was in each case possible to compensate for most of the breathing amplitude. Different representations showed different characteristics and it thus depends on the clinical application and the available data which representation fits best. In general, deformation models that are based on the displacement vectors perform better than those that are based on the absolute landmark locations. On the basis of 4D-CT data sets of 10 patients, a prediction error of about 2 mm could be obtained when using a displacement vector-based representation. Although the framework was evaluated on the basis of 4D-CT, it can be easily adapted to other modalities as, e.g., 4D-MR.

References

1. Keall, P., Mageras, G., Balter, J., et al.: The management of resp. motion in radiation oncology report of AAPM Task group. *Med. Phys.* 33(10), 3874–3900 (2006)
2. Bosch, J., Mitchell, S., Lelieveldt, B., et al.: Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Transactions on Medical Imaging* 21 (11), 1374–1383 (2002)
3. Casero, R., Noble, J.: A novel explicit 2D+t cyclic shape model applied to echocardiography. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 527–534. Springer, Heidelberg (2008)
4. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., et al.: Generation of a mean motion model of the lung using 4D-CT image data. In: *Eurographics Workshop on Visual Computing for Biomedicine* (2008)
5. McClelland, J., Blackall, J., Tarte, S., et al.: A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. *Med. Phys.* 33 (9), 3348–3358 (2006)
6. Zhang, Q., Pevsner, A., Hertanto, A., et al.: A patient-specific respiratory model of anatomical motion for radiation treatment planning. *Medical Physics* 34(12), 4772–4782 (2007)
7. von Berg, J., Barschdorf, H., Blaffert, T., et al.: Surface based cardiac and respiratory motion extraction motion extraction for pulmonary structures from multi-phase CT. In: *Proc. of SPIE*, vol. 6511, pp. 65110Y1–11 (2007)
8. Castillo, R., Castillo, E., Guerra, R., et al.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology* 54, 1849–1870 (2009)
9. Rosipal, R., Krämer, N.: Overview and recent advances in partial least squares. In: Saunders, C., Gribelnik, M., Gunn, S., Shawe-Taylor, J. (eds.) *SLSFS 2005*. LNCS, vol. 3940, pp. 34–51. Springer, Heidelberg (2006)

Image Estimation from Marker Locations for Dose Calculation in Prostate Radiation Therapy

Huai-Ping Lee¹, Mark Foskey¹, Josh Levy³, Rohit Saboo¹, and Ed Chaney^{2,3}

¹ Dept. of Computer Science,

² Dept. of Radiation Oncology, University of North Carolina at Chapel Hill, USA

³ Morphormics Inc, Chapel Hill, USA

Abstract. Tracking implanted markers in the prostate during each radiation treatment delivery provides an accurate approximation of prostate location, which enables the use of higher daily doses with tighter margins of the treatment beams and thus improves the efficiency of the radiotherapy. However, the lack of 3D image data with such a technique prevents calculation of delivered dose as required for adaptive planning. We propose to use a reference statistical shape model generated from the planning image and a deformed version of the reference model fitted to the implanted marker locations during treatment to estimate a regionally dense deformation from the planning space to the treatment space. Our method provides a means of estimating the treatment image by mapping planning image data to treatment space via the deformation field and therefore enables the calculation of dose distributions with marker tracking techniques during each treatment delivery.

1 Introduction

In order to deliver a high dose of radiation to cancerous tissue while sparing nearby normal tissues, modern radiotherapy protocols create steep dose gradients near the boundary of the target volume (we consider the prostate in this paper). These techniques, however, are very sensitive to treatment uncertainties such as day-to-day changes in the geometry of the internal organs, because they have only a narrow margin of high dose around the target volume. To cope with this difficulty, image-guided radiotherapy (IGRT) [1] uses a CT image taken at planning time (the *planning image*) and a CT image taken immediately before a dose fraction (a *treatment image*) to detect potential positioning errors and changes in anatomic geometry relative to the planning image. Patients are realigned and beam apertures are reshaped to correct for these errors prior to treatment. Treatment images can also be used to calculate and assess the delivered dose in the context of adaptive radiation therapy, a form of treatment that compensates for differences between planned and delivered doses. The delivered dose is calculated in the treatment space given the known treatment parameters. A non-rigid transformation from the treatment space to the planning space is established by mapping the treatment image to the planning image using an image registration method [2,3]. The transformation for each treatment day is

then used to deform the radiation dose calculated for each treatment day to the planning space, where it can be added to dose delivered on other days. The result can then be used if necessary to modify the treatment plan to compensate for any discrepancy between the planned and actual cumulative delivered doses [4]. A problem with this approach is that the patient geometry is not sampled during actual treatment delivery [5,6].

An alternative method to monitor the prostate motion is by tracking several (usually three) markers that are implanted in the prostate. For example, in the Calypso system, the markers take the form of electromagnetic transponders that can be tracked to sub-millimeter accuracy at a sample rate of 10 Hz, using a GPS-like system [7,8]. The locations of the markers are used to position the patient accurately in the treatment machine, thus eliminating the need for image guidance or external marks on the patient's skin or fixation device. Although the tracking information is accurate and takes intra-treatment motion into account, the image data needed for dose calculation is missing, eliminating the promise of calculating and accumulating delivered dose. In this paper, we demonstrate that the non-rigid mapping from the planning space to the treatment space can be inferred in a neighborhood of the prostate from the three marker locations and used to estimate the missing treatment image data from the planning image.

Our problem is different from the image registration problem in that the image data on the treatment day is missing, therefore the methods based on voxel-scale intensity matching [3] or based on surface matching [9] cannot be applied. Landmark-based registration methods [10] require much more than three marker locations to interpolate the deformation field and are not directly applicable to our problem. Our approach is to use a statistical shape model for the prostate so that the most likely prostate shape can be estimated given the measured marker locations. As our shape model we choose to use the *m-rep* (medial representation) [11] because it provides a coordinate system that can represent the interior and the nearby exterior of the object, so that the markers in the prostate can be also represented relative to an *m-rep* model of the prostate.

We create a reference *m-rep* from the planning image and one for each treatment day, based on the marker locations. The deformation between the reference *m-rep* and estimated treatment *m-rep* provides a transformation between the planning space and the treatment space that can be used to estimate the treatment image to calculate delivered dose. We compare the estimated images with the real treatment images to demonstrate the ability to reconstruct treatment images, and we compare the dose histograms computed using the estimated treatment images to the histograms generated by an IGRT procedure (using actual CT images) to establish the feasibility of dose calculation using estimated treatment images.

2 Method

Given the planning image, a reference *m-rep* of the prostate is created based on the gray-scale image and several user-specified landmarks on the organ

boundary [12]. Calypso markers are ~ 8.5 mm in length and ~ 1.5 mm in diameter that appear as bright spots in the image data. The location of each marker is computed as the center of mass of the bright spot, which closely corresponds to the origin of signals emitted by the miniature electromagnetic coil housed in the marker. For each treatment fraction, the three marker locations are given by the Calypso system, and the reference m-rep is then fitted to those locations by an optimization procedure. This procedure is similar to organ segmentation except that the seed locations, rather than image intensities, are what drive deformation within the trained shape space describing day-to-day shape variations of the prostate [12]. The diffeomorphism implied by the two m-reps is then computed with a shape interpolation method called *rotational flows* [13]. The deformation is used to transform the planning image to generate an estimated treatment image for dose calculation.

2.1 M-reps

The discrete m-rep [11] can be thought of as a discrete generalization of the medial axis [14]. A simple m-rep model consists of a discretely sampled medial surface. Sample points are referred to as *hubs*, and associated to each hub are two vectors (known as *spokes*) extending, on opposite sides of the medial sheet, from the hub to the boundary of the modeled object. The model is trained from a population of shapes by adjusting the parameters (the locations of the hubs and the directions and lengths of the spokes) to make the model fit each shape. From this data, a mean shape and its modes of variation can be determined [15]. This statistical representation can be used to reduce the effective dimension of the shape space, so that the limited information available from the three marker locations can be applied most effectively. For instance, changes orthogonal to the plane of the three points may be implied by their relative motions, given the statistical behavior of the shape space, even though that orthogonal direction is not explicitly sampled. At planning time, a patient-specific m-rep is created from the planning image for each patient. For each treatment fraction, the patient-specific m-rep is deformed using the modes of variation to fit the treatment image (in an IGRT) or the marker locations measured for each treatment fraction (in the case of using markers instead of treatment images).

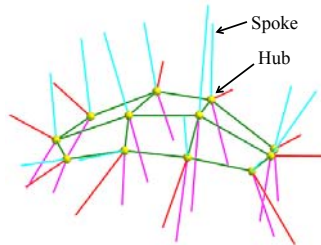


Fig. 1. A simple 3×4 m-rep. A single shape is represented by the locations of the hubs and the lengths and directions of the spokes. The boundary of the object (not shown) passes through the ends of the spokes.

An object-relative coordinate representation (*figural coordinates*) of the interior and nearby exterior of the organ is used to interpolate the locations of unsampled spokes. Two coordinates, u and v , reflect the arrangement of the hubs on the medial surface. Each hub, along with its associated spokes determines integer values for u and v , and points between them have non-integer values. An additional coordinate, ϕ , is equal to 1 along the entire top surface (bounded by the upper spokes extending from the outermost hubs) and -1 along the bottom surface, varying smoothly between those values over the intermediate portion of the boundary. A fourth coordinate, τ , gives the fraction of the distance from the medial sheet to the boundary of the organ. This fraction may be greater than 1 for points outside the organ.

2.2 Rotational Flows and Deformation Field Generation

Given the reference m-rep m_0 and the target m-rep m_1 created by fitting m_0 to the marker locations, the deformation from m_0 into m_1 is computed by interpolating between the sets of boundary points b_0 and b_1 , which are sampled using the same set of figural coordinates $\{(u_i, v_i, \phi_i) | i = 1, \dots, n\}$, on the boundary of m_0 and m_1 , respectively. Thus for each point \mathbf{x}_0 in the sets b_0 , there is a point \mathbf{x}_1 in b_1 that has the correspondence with \mathbf{x}_0 given by the m-reps and the figural coordinates. The goal of the rotational flows method is to interpolate between each pair of corresponding oriented vertices in \mathbb{R}^3 , $(\mathbf{x}_0, \mathbf{E}_0)$ and $(\mathbf{x}_1, \mathbf{E}_1)$, where $\mathbf{E}_i = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is the set of orthonormal bases representing the orientation for the point \mathbf{x}_i . In 3D cases, each oriented vertex moves along a helical path along the axis of rotation. It can be shown that the interpolation is shape-maintaining if the two objects are similar and is size-maintaining if the two models are congruent [13].

Once the curved path for each pair of corresponding points is computed, the deformation for each voxel is computed by numerically integrating the subdivided displacement along each curved path. At each integration step, a radial basis function (RBF) interpolation is used to compute the deformation field in a rectangular box containing the prostate with a small margin, using the collection of landmarks. Some static landmarks are added on the boundary of the box to make the transition across the box boundary smoother. The deformation outside of the rectangular box is assumed to be zero. Due to the physics of dose deposition at high energies, this assumption has negligible effect on the accuracy of dose calculation.

2.3 Dose Calculation

Calculation of delivered dose is a well-established part of radiation therapy planning. CT intensities are based on the absorptivity of tissue to radiation, so it is possible to calculate the amount of energy a given beam deposits in each portion of tissue it passes through or near, taking into account the attenuation of the beam and the scattering of radiation into neighboring tissue. We use the treatment planning system known as PlanUNC [16] to calculate dose.

Dose calculations for treatment planning are expected to be $\pm 3\%$ compared to measurements.

These calculations are most commonly used in treatment planning. But when images are available for treatment fractions, dose can be calculated for those days too. Then, if a mapping can be constructed between the image of each treatment fraction and a common reference frame, the dose can be accumulated to determine the total amount of radiation received by each portion of the tissue. Our approach is a reversal of the most common methods of dose accumulation. Taking the planning image as given, we derive a deformation first and derive the treatment image from it, rather than acquiring a treatment image first and then deriving the deformation.

3 Experiments

We evaluated our method on four patient data sets, which contain one, eight, six, and five treatment fractions, respectively. The data for each patient consist of a planning image, treatment images, and Calypso marker locations during treatment. Note that for each treatment fraction, the patient is positioned in the treatment machine according to the Calypso origin, which also serves as the isocenter of the treatment plan (i.e., the point in the space where the central beam of radiation passes), and treatment images are taken only for comparison with estimated images. Since the treatment images are taken before the patient is positioned for treatment, they are visually aligned to the Calypso marker locations before the comparison with estimated images and the non-rigid image registration for dose calculation.

In order to assess the similarity between the estimated image and real treatment image, we create an m-rep of the prostate for each treatment image based on the gray-scale image and user-specified landmarks. The estimated m-rep (m_{marker} , reference m-rep fitted to the marker locations) is then compared to the treatment m-rep (m_{treat} , m-rep created with the treatment image) by computing the distance between sample points on the surfaces. For each sample point p on the surface of m_{treat} , its distance to the m-rep m_{marker} is approximated as $D(p, B_{\text{marker}}) = \min \{d(p, q) | q \in B_{\text{marker}}\}$, where B_{marker} is the set of sample points on the surface of m_{marker} , and $d(p, q)$ is the Euclidean distance between the two points p and q . The average distances from the estimated m-rep to the treatment m-rep are 0.15 cm, 0.12 cm, 0.13 cm, and 0.10 cm for the four patient data sets, respectively. Note that these average distances are within the image resolution of the four data sets; two of them are $0.12 \times 0.12 \times 0.3$ cm, and the other two are $0.12 \times 0.12 \times 0.3$ cm. Fig. 2 shows an example visual comparison of the estimated image and corresponding treatment image.

The dose calculation using the estimated image is compared against the dose calculated with real treatment images to evaluate the accuracy of our method. The non-rigid transformation with the treatment images is computed in the same manner as with the estimated image, as described in Section 2.2. For the first patient data set, we consider the dose in the prostate. For the other data

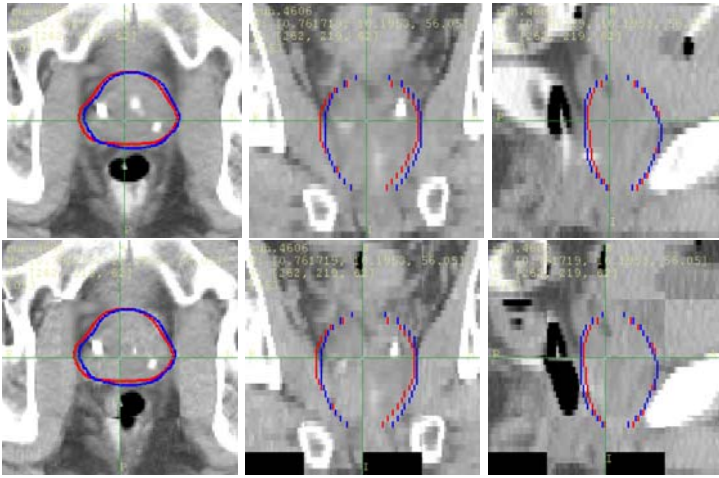


Fig. 2. Axial (left), coronal (center), and sagittal (right) views of an example m-rep comparison; blue contour: m-rep fitted to Calypso marker locations during the treatment fraction; red contour: m-rep created with real treatment image; the green crosshairs show the Calypso origin, which also serves as the isocenter of the treatment plan; the CT image shown are the estimated image (top row) and the 4×4 checkerboard image comparing the estimated image to the real treatment image (bottom row).

sets, we consider the dose in the prostate and in the anterior rectal wall (the part of the rectum right next to the prostate). Fig. 3 shows the differential and cumulative dose-volume histograms (DVH) of one of the data sets using the estimated images and real treatment images. Besides visual similarity, we also numerically compare the DVHs by considering the equivalent uniform dose (EUD) [17] of the differential DVHs, as shown in Table 1, along with the errors of the values given by our method relative to those given by the real images. There are no generally accepted standards for errors in calculating delivered dose, but $\pm 5\%$ is a reasonable goal. We observed relative errors of less than 3% in EUD for the four patient data sets we experimented on.

Table 1. Comparison of equivalent uniform dose (normalized to percentage of reference dose) of the dose-volume histograms for the four patient data sets, and the errors relative to the values given by the IGRT

Patient	Prostate EUD			Anterior Rectal Wall EUD		
	Tx Image	Est. Image	Error	Tx Image	Est. Image	Error
1	96.93	98.17	1.28%	N/A	N/A	N/A
2	102.54	102.51	-0.02%	102.04	101.55	-0.49%
3	98.68	100.71	2.06%	92.42	92.20	-0.24%
4	102.62	102.59	-0.03%	101.72	101.21	-0.50%

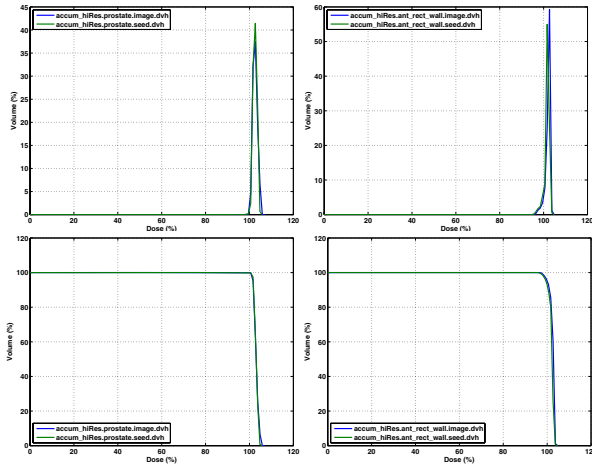


Fig. 3. Differential (top row) and cumulative (bottom row) dose-volume histograms for the prostate (left) and anterior rectal wall (right); blue lines represent results with real treatment images, and green lines represents results with estimated images

4 Conclusion

We presented a method to estimate the treatment image using the planning image and locations of implanted markers during the treatment. We demonstrated the feasibility of image estimation by comparing the estimated images to real treatment images, and we also showed that the calculated dose histograms using the estimated images are close to those using real treatment images. Our method complements the Calypso system, where the prostate motion can be tracked accurately at a high frequency during the treatment but the image data is missing, so that the delivered dose distribution can be calculated and a safe delivery can be insured.

In the future, we will experiment on more patient data and assess different deformation methods such as FEM-based methods. We will also investigate the feasibility of adjusting the treatment machine to compensate for the intra-treatment motion observed.

Acknowledgements. This work was supported by funds from grant 1R43CA141941 to Morphormics, Inc. from the US National Cancer Institute, and by a research agreement with Calypso Medical.

References

1. Wong, J.R., Grimm, L., Uematsu, M., Oren, R., Cheng, C.W., Merrick, S., Schiff, P.: Image-guided radiotherapy for prostate cancer by CT-linear accelerator combination: Prostate movements and dosimetric considerations. *International Journal of Radiation Oncology*Biography*Physics* 61(2), 561–569 (2005)

2. Maintz, J., Viergever, M.A.: A survey of medical image registration. *Medical Image Analysis* 2(1), 1–36 (1998)
3. Foskey, M., Davis, B., Goyal, L., Chang, S., Chaney, E., Strehl, N., Tomei, S., Rosenman, J., Joshi, S.: Large deformation Three-Dimensional image registration in Image-Guided radiation therapy. *Physics in Medicine and Biology* 50, 5869–5892 (2005)
4. Yan, D., Lockman, D., Brabbins, D., Tyburski, L., Martinez, A.: An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer. *International Journal of Radiation Oncology*Biography*Physics* 48(1), 289–302 (2000)
5. Litzenberg, D.W., Hadley, S.W., Tyagi, N., Balter, J.M., Haken, R.K.T., Chetty, I.J.: Synchronized dynamic dose reconstruction. *Medical Physics* 34(1), 91–102 (2007)
6. Noel, C., Parikh, P.J., Roy, M., Kupelian, P., Mahadevan, A., Weinstein, G., Enke, C., Flores, N., Beyer, D., Levine, L.: Prediction of intrafraction prostate motion: Accuracy of pre- and Post-Treatment imaging and intermittent imaging. *International Journal of Radiation Oncology Biology Physics* 73(3), 692–698 (2009)
7. Langen, K.M., Willoughby, T.R., Meeks, S.L., Santhanam, A., Cunningham, A., Levine, L., Kupelian, P.A.: Observations on real-time prostate gland motion using electromagnetic tracking. *International Journal of Radiation Oncology, Biology, Physics* 71(4), 1084–1090 (2008); PMID: 18280057
8. Balter, J.M., Wright, J.N., Newell, L.J., Friemel, B., Dimmer, S., Cheng, Y., Wong, J., Vertatschitsch, E., Mate, T.P.: Accuracy of a wireless localization system for radiotherapy. *International Journal of Radiation Oncology, Biology, Physics* 61(3), 933–937 (2005); PMID: 15708277
9. Kaus, M.R., Brock, K.K., Pekar, V., Dawson, L.A., Nichol, A.M., Jaffray, D.A.: Assessment of a Model-Based deformable image registration approach for radiation therapy planning. *Int. J. Radiation Oncology Bio. Phys.* 68(2), 572–580 (2007)
10. Rohr, K., Stiehl, H.S., Sprengel, R., Buzug, T.M., Weese, J., Kuhn, M.H.: Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging* 20(6), 526–534 (2001)
11. Pizer, S.M., Fletcher, P.T., Joshi, S., Thall, A., Chen, J.Z., Fridman, Y., Fritsch, D.S., Gash, A.G., Glotzer, J.M., Jiroutek, M.R., Lu, C., Muller, K.E., Tracton, G., Yushkevich, P., Chaney, E.L.: Deformable M-Reps for 3D medical image segmentation. *Int. J. Comput. Vision* 55(2-3), 85–106 (2003)
12. Pizer, S.M., Fletcher, P.T., Joshi, S., Gash, A.G., Stough, J., Thall, A., Tracton, G., Chaney, E.L.: A method and software for segmentation of anatomic object ensembles by deformable m-reps. *Medical Physics* 32(5), 1335–1345 (2005)
13. Levy, J.H., Foskey, M., Pizer, S.M.: Rotational flows for interpolation between sampled surfaces. In: *Computer Vision and Pattern Recognition Workshop*, pp. 1–8. IEEE Computer Society, Los Alamitos (2008)
14. Blum, H., Wathen-Dunn, W.: A transformation for extracting new descriptors of shape. *Models for the Perception of Speech and Visual Form*, 362–380 (1967)
15. Merck, D., Tracton, G., Saboo, R., Levy, J., Chaney, E., Pizer, S., Joshi, S.: Training models of anatomic shape variability. *Medical Physics* 35(8), 3584–3596 (2008)
16. Schreiber, E., Xu, Z., Lorenzen, A., Foskey, M., Cullip, T., Tracton, G., Chaney, E.: SU-FF-T-362: PPlanUNC as an Open-Source radiotherapy planning system for research and education. *Medical Physics* 33(6), 2129 (2006)
17. Wu, Q., Mohan, R., Niemierko, A., Schmidt-Ullrich, R.: Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. *International Journal of Radiation Oncology, Biology, Physics* 52(1), 224–235 (2002)

A Machine Learning Approach for Deformable Guide-Wire Tracking in Fluoroscopic Sequences

Olivier Pauly, Hauke Heibel, and Nassir Navab

Computed Assisted Medical Procedures,
Technische Universität München, Germany

Abstract. Deformable guide-wire tracking in fluoroscopic sequences is a challenging task due to the low signal to noise ratio of the images and the apparent complex motion of the object of interest. Common tracking methods are based on data terms that do not differentiate well between medical tools and anatomic background such as ribs and vertebrae. A data term learned directly from fluoroscopic sequences would be more adapted to the image characteristics and could help to improve tracking. In this work, our contribution is to learn the relationship between features extracted from the original image and the tracking error. By randomly deforming a guide-wire model around its ground truth position in one *single* reference frame, we explore the space spanned by these features. Therefore, a guide-wire motion distribution model is learned to reduce the intrinsic dimensionality of this feature space. Random deformations and the corresponding features can be then automatically generated. In a regression approach, the function mapping this space to the tracking error is learned. The resulting data term is integrated into a tracking framework based on a second-order MAP-MRF formulation which is optimized by QPBO moves yielding high-quality tracking results. Experiments conducted on two fluoroscopic sequences show that our approach is a promising alternative for deformable tracking of guide-wires.

1 Introduction

During the last decade, the success of angiographic interventions relied on the ability of physicians to navigate in the patient's anatomy based only on their mental three-dimensional representation of the human body as well as on the haptic feedback from the instruments. Recent advances in computer aided planning and navigation techniques offer great potential of minimizing the risk of complications and improving the precision. In the case of angiographic applications, the most common imaging modality is X-ray fluoroscopy. Currently, in order to monitor guidance procedures, a roadmap, e.g. a digital subtracted angiography (DSA) showing vessel anatomy, is computed during the intervention. Unfortunately, such roadmaps cannot directly be fused with the intra-operative fluoroscopic sequence due to misalignment caused by respiratory motion. A fundamental step toward a successful integration of any navigation application into clinical routine is the estimation and compensation of such respiratory motion.

Determining this spatio-temporal information is a challenging task due to the fact that fluoroscopic X-ray images have a low signal to noise ratio, are subject to big changes in contrast and suffer from background clutter in the abdominal area. Moreover, the apparent motion of the guide-wire is a combination of multiple components. The major motion in the chest is caused by patient breathing. A second, deformable

component results from forces applied to the guide-wire by the physician and by surrounding organs which are subject to non-uniform motions during the breathing cycle. Furthermore, the guide-wire may sometimes partially vanish.

A recent approach dealing with the problem of guide-wire tracking in fluoroscopy is [5]. In this work, Heibel et al. proposed a scheme for deformable tracking based on a MAP-MRF formulation. However, their data term does not differentiate well between medical tools and anatomic background such as ribs and vertebrae. A learned data term being more robust and adapted to the image characteristics of fluoroscopic sequences could help to further improve the tracking. Since MRF formulations are derivative free optimization procedures, they ease the integration of such learning based energies for which analytical derivatives are hard to derive if possible at all. Learning permits to model complex relationships between the information contained in the images and the quality of alignment.

In the context of guide-wire tracking, we can distinguish two kinds of learning approaches: First, methods for the detection of the guide-wire in each frame and second, methods used for learning a data driven energy. A learning-based tracking approach by detection based on marginal space learning was presented by Barbu et al. in [1]. Later, Wang et al. proposed in [11] the combination of learning-based detectors and online appearance models. In the case of energy learning, Nguyen et al. [7] addressed the problem of modeling the error surface of parametric appearance models in order to minimize the number of local minima for image alignment and recently Pauly et al. suggested in [8] to learn the statistical relationship between two different imaging modalities to model a data term for multi-modal rigid registration.

In this work, our contribution is a learning approach for deformable tracking: we propose to learn a data term based on the relationship between features extracted from the original image and the tracking error. As features, we introduce local mean orthogonal intensity profiles representing information contained in the original image. Since deformable transformations have a high number of degrees of freedom, the intrinsic dimensionality of the space spanned by these features is high. However, typical guide-wire deformations are lying on a subspace we propose to learn to reduce the complexity of our problem. A set of random deformations is then generated automatically and applied to the ground truth position of the guide-wire on a single reference image. A training set of data points from the corresponding local mean orthogonal profiles and their associated tracking error values is thereby created. Learning is then performed on this dataset with a support vector regression. The resulting data term is integrated into a tracking framework based on a MAP-MRF formulation which is solved with higher-order clique reduction techniques. Due to the higher-order nature of our problem and since we are dealing with non-submodular energy functions we chose a combination of the recently proposed reduction scheme of Ishikawa [6] and the QPBO [4] optimizer supporting improvements in order to deal with unlabeled nodes [9].

The remainder of the paper is organized as follows: Section 2 presents our regression approach to define an optimal data term for guide-wire tracking. Section 3 reports experiments performed on two fluoroscopic sequences. Results show that our approach presents a promising alternative for guide-wire tracking in fluoroscopic sequences. Section 4 concludes the paper and gives an outlook on future work.

2 Methods

2.1 Problem Statement

The goal of tracking is to identify the relative motion of an object in a series of consecutive frames. In most tracking algorithms, we can distinguish two phases: first, the detection of the object of interest in the initial frame followed by the actual tracking in each new frame given previous positions. In this paper, we focus on the problem of tracking a guide-wire through a fluoroscopic sequence knowing its initial position. Let us denote \mathcal{C} our guide-wire model and $\{I_t\}_{t \in \{0, \dots, T\}}$ the set of consecutive images in which we want to track the guide-wire. In fluoroscopic images, guide-wires appear as curvilinear structures which can be represented as B-spline curves. The advantage of such a representation is its low-dimensionality, its implicit smoothness and its local support of control points. Our guide-wire model \mathcal{C} is defined as the following linear combination of control points:

$$\mathcal{C}(s) = \sum_{i=1}^M N_i(s) P_i \quad \text{where } s \in [0, 1] \quad (1)$$

where N_i denote the basis functions and P_i the positions of M control points. By using this model, we want to estimate the optimal curve parameters, i.e. the best configuration of the control points, to match the visible structures in an image, and this, knowing its previous position. The tracking problem can be then formulated as a maximum a posteriori estimation:

$$\mathcal{C}_t^* = \operatorname{argmax}_{\mathcal{C}_t} P(I_t | \mathcal{C}_t) P(\mathcal{C}_t) \quad (2)$$

where \mathcal{C}_t^* is the best curve estimate at instant t . $P(I_t | \mathcal{C}_t)$ is the likelihood of observing the data knowing the model and $P(\mathcal{C}_t)$ the prior or probability of the current curve configuration. Let us assume the likelihood to follow a Gaussian distribution and the prior a Gibbs' distribution, we can then reformulate Eq. 2 as an energy minimization:

$$\mathcal{C}_t^* = \operatorname{argmin}_{\mathcal{C}_t} (E_{data}(I_t | \mathcal{C}_t) + E_{reg}(\mathcal{C}_t)) \quad (3)$$

$E_{reg}(\mathcal{C}_t)$ is a regularization term which constraints the space of possible model configurations. Assuming constant length of guide-wire segments in fluoroscopic sequences, we define the regularization term in order to penalize changes in length:

$$E_{reg}(\mathcal{C}_t) = \int_0^1 \left(1 - \frac{\|C'_t(s)\|}{\|C'_0(s)\|} \right)^2 ds \quad (4)$$

where C'_t and C'_0 are the first derivatives at instant t and 0 respectively. Thank to the inherent smoothness of a B-spline representation, higher-order terms can be discarded. $E_{data}(I_t | \mathcal{C}_t)$ can be seen as a data term which drives the model according to the current image:

$$E_{data}(I_t | \mathcal{C}_t) = \int_0^1 \Phi(I_t(\mathcal{C}_t(s))) ds \quad (5)$$

A common choice for Φ is a function which enhances tubular structures similar to the ridgeness measure proposed by Frangi et al. [3]. Such measures can be tuned to emphasize only structures of the scale of the guide-wire and to remove outliers such as ribs

or vertebrae. However, since the data term is only evaluated along the current position of the curve, the main drawback is a very low capture range and a lack of robustness in terms of outliers or partial occlusions.

Instead of relying on the feature image intensities along the curve profile, we propose to extract features from the unprocessed image orthogonally to the curve, namely *local mean orthogonal intensity profiles*. We can then model a data term by learning a function Ψ relating the space \mathcal{M} spanned by these features and the tracking error. By using a single fluoroscopic image and a set of local displacements around the ground truth position of our guide-wire, we can sample the space \mathcal{M} by extracting the local mean orthogonal intensity profiles associated to each displaced curve. Each of these “points” of \mathcal{M} is then associated to a tracking error derived from the corresponding curve parameters, hereby generating a set of data points. Finally Ψ is modeled by performing a regression on these points. The following section presents how to extract the mentioned features.

2.2 Local Mean Orthogonal Profiles

In a fluoroscopic image, a human being may recognize the guide-wire because of its curvilinear aspect and its darker intensities compared to its environment. For this reason, a common method would be to enhance this structure and to keep track of it along the sequence by using a data term based on the intensity profile along the curve. Unfortunately, in the case of larger displacements between two consecutive frames, it is hard to relocate the guide-wire in a heterogeneous region containing outliers without any information about the search direction. Indeed, such data terms suffer from an extremely narrow valley around the global extremum. To overcome this problem and benefit from an increased capture range, we propose features which describe the intensity profiles *orthogonally* to the curve. First, we subdivide our curve \mathcal{C}_t into n segments $\{S_t^k\}_{k \in \{1, \dots, n\}}$. Each segment S_t^k is a spline we characterize by the following descriptor \mathcal{J}_t^k :

$$\mathcal{J}_t^k = \frac{1}{q} \sum_{j=1}^q \Lambda_t^{k,j}, \quad (6)$$

with q being the number of sample points along this segment. $\Lambda_t^{k,j}$ is an orthogonal intensity profile whose r^{th} element is defined as:

$$\Lambda_t^{k,j}(r) = I_t(S_t^k(u) + r \cdot \mathbf{n}(u)) \quad (7)$$

where $\mathbf{n}(u)$ is the normal vector at point $u = (j - 1)/(q - 1)$ and $r \in \{-R, \dots, R\}$. The dimensionality of this vector is $2R + 1$ which corresponds to the length of the profile centered on the segment. Note that since only the profile’s shape is of interest, each profile $\Lambda_t^{k,j}$ is normalized between 0 and 1. Taking the mean over the segment provides a feature vector which is more robust to noise and outliers. Each curve \mathcal{C}_t is then described by the following set $\{\mathcal{J}_t^k\}_{k \in \{1, \dots, n\}}$.

2.3 Data Points Generation by Motion Learning

The goal of our approach is to learn a function Ψ relating the local mean orthogonal profiles and the tracking error:

$$\Psi : \mathcal{M} \rightarrow \mathbb{R}, \quad (8)$$

with good characteristics for tracking purposes, namely convexity and smoothness. Therefore, the space \mathcal{M} spanned by these features needs to be sampled thoroughly as a function of the relative displacement. Since the guide-wire is a deformable structure, the intrinsic dimensionality of our features according to free deformations would be high and thus, hard to sample. However, in a real fluoroscopic sequence, a guide-wire is not subject to free deformations. Indeed, main displacements are due to breathing motions and additional small deformations. This means that in reality, our features do not describe the full space \mathcal{M} but lie on a lower dimensional subspace. To reduce the complexity of our problem, we propose to learn the deformation probability distribution from a real sequence. Thus, random displacements can be automatically generated to build our training dataset.

Learning guide-wire Motions: During a sequence, each segment S_t^k of our curve \mathcal{C}_t is subject to a series of consecutive displacements we denote $\{D_t^k\}_{t \in \{0, \dots, T-1\}}$. Each D_t^k is modeled by a vector containing the displacements of sample points of the segment between 2 consecutive frames. Its j^{th} element is defined as:

$$D_t^k(j) = S_{t+1}^k(u) - S_t^k(u) \tag{9}$$

These vectors are collected for all segments along the whole sequence and grouped in a training set $\mathcal{D} = \{D_t^k\}_{t \in \{0, \dots, T-1\}}^{k \in \{1, \dots, n\}}$. To learn the underlying probability distribution of these displacements, we propose to model it with a gaussian mixture model \mathcal{G} . The parameters of \mathcal{G} can be estimated by using Expectation-Maximization. Once we have learned our gaussian mixture model, we can generate random segment displacements $\{D_i\}_{i \in \{1, \dots, Q\}}$ from this probability distribution.

Data points generation: As shown on Fig 1 by using a reference fluoroscopic image, e.g. the first frame of the sequence, we can generate local mean orthogonal profiles $\{\mathcal{J}_i\}_{i \in \{1, \dots, Q\}}$ by perturbing the segments of the ground truth curve with the

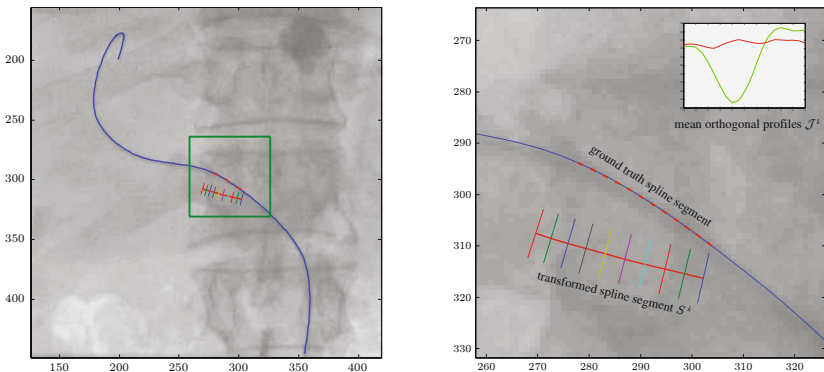


Fig. 1. Data term learning: by perturbing the ground-truth curve from a single frame with random displacements, we can build a training set of local mean orthogonal profiles with their associated tracking errors

randomly generated displacements $\{D_i\}_{i \in \{1, \dots, Q\}}$. The corresponding tracking error \mathcal{E}_i associated to each \mathcal{J}_i is computed as follows:

$$\mathcal{E}_i = \|D_i\|^2 \tag{10}$$

This procedure permits us to generate the set of pairs $\{(\mathcal{J}_i, \mathcal{E}_i)\}_{1, \dots, Q}$, on which the regression will be performed to learn our function Ψ .

2.4 Learning Data Term through Support Vector Regression

From previously generated data points, the function Ψ can be learned through non-parametric support vector regression. Let us consider the problem of fitting a function on the set of Q data points $\{(\mathcal{J}_i, \mathcal{E}_i)\}_{i \in \{1, \dots, Q\}}$. Ψ is modeled as the following function:

$$\Psi(\mathcal{J}) = \langle w, \mathcal{J} \rangle + b, \tag{11}$$

where w is a weighting vector of dimensionality $\dim(\mathcal{M})$ and b a bias. This can be written as a convex optimization problem [10]:

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|w\|^2 \\ &\text{subject to: } \begin{cases} \mathcal{E}_i - \langle w, \mathcal{J}_i \rangle - b \leq \epsilon \\ \langle w, \mathcal{J}_i \rangle + b - \mathcal{E}_i \leq \epsilon \end{cases} \end{aligned} \tag{12}$$

This aims at minimizing the norm of w to penalize the model complexity and the regression errors on the data points with a regression tolerance denoted by ϵ . Equation (12) corresponds to minimizing the following functional:

$$\begin{aligned} &\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^Q (\xi_i^+ + \xi_i^-) \\ &\text{subject to: } \begin{cases} \mathcal{E}_i - \langle w, \mathcal{J}_i \rangle - b \leq \epsilon + \xi_i^+ \\ \langle w, \mathcal{J}_i \rangle + b - \mathcal{E}_i \leq \epsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0 \end{cases} \end{aligned} \tag{13}$$

where C weights the impact of the errors and thus the flexibility of the model. According to the Representer theorem, a solution w_{opt} of this minimization is always a linear combination of the training vectors in \mathcal{M} with weights $\{\alpha_i\}_{i \in \{1, \dots, Q\}}$:

$$w_{opt} = \sum_{i=1}^Q \alpha_i \mathcal{J}_i \tag{14}$$

which leads to the following model:

$$\Psi(\mathcal{J}) = \sum_{i=1}^Q \alpha_i \langle \mathcal{J}_i, \mathcal{J} \rangle + b \tag{15}$$

Finally, the global data term computed on all segments can be written as:

$$E_{data}^{learn}(C_t) = \frac{1}{n} \sum_{k=1}^n \Psi(\mathcal{J}_t^k) \tag{16}$$

3 Experiments and Results

In the following experiments, we show the successful application of our machine learning approach for the tracking of guide-wires in fluoroscopic images. The two sequences we used for our experiments were acquired during liver chemoembolizations. In this procedure, a guide-wire is inserted into the femoral artery and threaded into the aorta. The catheter is then advanced into the hepatic artery. Once the branches that feed the liver cancer are reached, the chemotherapy is infused. In both sequences, the catheter is already inserted in the artery and we aim at recovering from breathing motions.

Motion learning: A set of inter-frame segment displacements is computed from a reference sequence where the guide-wire positions were manually annotated. A gaussian mixture model is then fitted to this dataset by using EM algorithm. The analysis of Bayes' Information Criterion leads to the choice of two gaussian components.

Data term learning: A quadratic B-spline is fit to each hand-labeled point set by minimizing discontinuities in the second derivative [2]. Given the previously learned gaussian mixture model, a set of $Q = 3000$ random segment displacements is automatically generated. By perturbing the ground-truth curve from a single frame with these random displacements, we can build a training set of 3000 local mean orthogonal profiles with their associated tracking errors. Note that the choice of Q is a compromise between complexity and accurate modeling of the data term. During the experiments profiles with different radii are evaluated. Finally, the data term is learned by performing a support vector regression.

Tracking experiments: Experiments are conducted on two clinical sequences of 142 and 228 frames with a resolution of 512×512 pixels and respective pixel spacings of 0.432×0.432 mm and 0.308×0.308 mm. In order to evaluate the tracking results, guide-wires are manually annotated in each frame. The following distance measure has been used throughout all experiments to assess the quantitative tracking quality:

$$\chi = \frac{1}{2} \left(\frac{1}{|\mathcal{C}_t|} \sum_{x_i \in \mathcal{C}_t} \min_{y \in \mathcal{C}_{GT}} d(x_i, y)^2 + \frac{1}{|\mathcal{C}_{GT}|} \sum_{y_j \in \mathcal{C}_{GT}} \min_{x \in \mathcal{C}_t} d(x, y_j)^2 \right). \quad (17)$$

Here \mathcal{C}_{GT} is the manually annotated curve and \mathcal{C}_t the tracking result of an individual frame.

Results: Tab I shows mean errors on whole sequences where the data term is trained on the first frame of one sequence, and tested in tracking in both sequences. Submillimeter yet subpixel tracking accuracy can be achieved with our learned data-term and this, for a frame rate of 1.5 frame/s on a 3 Ghz duo core. Moreover, cross-validation illustrates the robustness of our approach even if it has been trained on another sequence showing different contrasts, motions and background. Note that since the Seq.1 presents motions of higher amplitude, its mean error is slightly bigger than for the other sequence. The great advantage is the ability to model the convexity and smoothness of this term. Indeed, its convexity properties can be designed by replacing the tracking error function I0. The choice of hyper-parameter C from equation I3 influences the flexibility of the regression and thus the smoothness of the resulting function.

Table 1. Tracking errors in real fluoroscopic sequences

Tracking Results								
Trained on	Seq.1				Seq. 2			
Tested on	Seq.1		Seq.2		Seq.1		Seq.2	
Profile Radius	5 pixels	10 pixels	5 pixels	10 pixels	5 pixels	10 pixels	5 pixels	10 pixels
χ mean (mm ²)	0.7115	0.5249	0.1636	0.1622	0.6632	0.5815	0.1796	0.1700
χ std dev (mm ²)	0.4289	0.2715	0.1633	0.1185	0.6184	0.3366	0.1771	0.1645

4 Discussion and Conclusion

In this work, our contribution was to learn the relationship between features extracted from an unprocessed image and the tracking error in order to model a data term. Experiments conducted on two fluoroscopic sequences show that our approach is a promising alternative for deformable guide-wire tracking. Indeed, our method is robust to changes in contrast, background clutter and partial occlusions of the guide-wire during the sequence, and this, even if training was performed on another dataset. Since the feature space under free deformations is high-dimensional, we proposed to model the distribution of the reduced space of typical guide-wire motions with a gaussian mixture model. In turn, this permitted us to automatically generate random guide-wire deformations from this distribution for the sake of regression. Going further, the space of relative motions between consecutive frames could be constrained during tracking to expected guide-wire motions. In future work, we will explore the possibility of deriving an adapted regularization term from this motion distribution model.

References

1. Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: CVPR (2007)
2. Dierckx, P.: Curve and surface fitting with splines. Oxford University Press, Inc., New York (1993)
3. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, p. 130. Springer, Heidelberg (1998)
4. Hammer, P.L., Hansen, P., Simeone, B.: Roof duality, complementation and persistency in quadratic 01 optimization. *Mathematical Programming* 28, 121–155 (1984)
5. Heibel, T.H., Glocker, B., Groher, M., Paragios, N., Komodakis, N., Navab, N.: Discrete tracking of parametrized curves. In: CVPR (2009)
6. Ishikawa, H.: Higher-order clique reduction in binary graph cut (2009)
7. Nguyen, M.H., de la Torre, F.: Metric learning for image alignment. In: IJCV (2009)
8. Pauly, O., Padoy, N., Poppert, H., Esposito, L., Navab, N.: Towards application-specific multimodal similarity measures: a regression approach. In: MICCAI Workshop, PMMIA (2009)
9. Rother, C., Kolmogorov, V., Lempitsky, V., Szummer, M.: Optimizing binary mrfs via extended roof duality (2007)
10. Smola, A.J., Schölkopf, B.: A tutorial on support vector regression. *Statistics and Computing* (2004)
11. Wang, P., Chen, T., Zhu, Y., Zhang, W., Zhou, S.K., Comaniciu, D.: Robust guidewire tracking in fluoroscopy. In: CVPR (2009)

Collaborative Tracking for MRI-Guided Robotic Intervention on the Beating Heart

Y. Zhou¹, E. Yeniaras¹, P. Tsiamyrtzis², N. Tsekos¹, and I. Pavlidis¹

¹ Department of Computer Science, University of Houston, Houston, TX 77024, USA

² Department of Statistics, Athens University of Economics, Athens 10434, Greece
{yzhou9,eyeniaras2,nvtsekos,ipavlidis}@uh.edu, pt@aueb.gr

Abstract. Magnetic Resonance Imaging (MRI)-guided robotic interventions for aortic valve repair promise to dramatically reduce time and cost of operations when compared to endoscopically guided (EG) procedures. A challenging issue is real-time and robust tracking of anatomical landmark points. The interventional tool should be constantly adjusted via a closed feedback control loop to avoid harming these points while valve repair is taking place in the beating heart. A Bayesian network of particle filter trackers proves capable to produce real-time, yet robust behavior. The algorithm is extremely flexible and general - more sophisticated behaviors can be produced by simply increasing the cardinality of the tracking network. Experimental results on 16 MRI cine sequences highlight the promise of the method.

1 Introduction

Image guided and robot-assisted (IGRA) surgeries are evolving and may selectively replace endoscopically guided (EG) surgeries in the future. Research is motivated by several IGRA advantages, such as wide field-of-view with planar or volumetric appreciation of the area of operation, minimally invasive processes, and reduction of operating time [1]. A grand challenge in IGRA surgeries is the compensation of tissue motion. This is particularly true in heart operations.

Among the most promising types of IGRA surgeries are the interventional magnetic resonance imaging (MRI) surgeries; distinct advantages include lack of ionizing radiation, a wide range of soft tissue contrast mechanisms, 3D data acquisition, and operator-independent image quality [2]. MRI-guided robotic interventions have a wide range of potential applications [3], including cardiac procedures, such as aortic valve repair [4]. Tracking cardiac motion is a very active field within the cardiovascular MRI community [5]. Nevertheless, a method to estimate the motion of specific anatomical landmarks needed in surgical procedures has not yet been proposed.

The research described in this paper is motivated by the need to develop alternatives to the highly invasive and long (several hours) surgical procedures related to heart valve repairs. An MRI-guided robotic intervention will obviate the need to open the thorax and stop the beating heart, thus potentially completing the surgery within minutes. For this, the motion of specific anatomical

landmarks should be tracked in real-time to close the feedback control loop. As a result, the robotic interventional tool will be held in place during the procedure without harming healthy tissue and vital structures.

To locate a specific anatomical landmark, one approach is to segment the underlying structures, i.e., accurately determine the boundary of the endocardium and the left ventricle on short and long axis views of the heart. Such approach may entail defining an appropriate optimization scheme to iteratively minimize a cost function, and the computational cost can be significant. For example, in [6], while the proposed boundary segmentation method achieved very high accuracy, the computational time was 10 seconds per slice. In contrast, Yuen et al. [7] achieved real-time performance in Ultrasound images with an extended Kalman filter tracker.

We adopted a particle filtering approach to estimate landmark motion. Particle filtering is a general tracking mechanism [8], free of strong modeling, which can accommodate very efficiently the predict-update loop. A loner particle filter tracker, however, may become unstable in sudden motion or large appearance changes. This is risky in cardiac surgeries. To maintain robustness we propose a collaborative tracking framework, which coordinates multiple particle filter trackers. Some of these trackers may fail when confronting challenging conditions, such as sudden motion or significant appearance changes, but others that are less affected will survive and yield good state estimates. The latter can be used to produce a reliable overall estimate and eventually recover the failed trackers. For the first time, we use a Bayesian network method to decide which trackers fail and which ones survive at each time step.

The methodology is quite elegant, as it provides a unified framework to tackle a wide variety of tracking problems, from the most mundane to the most difficult, by simply adjusting the number of trackers n in the probabilistic network. For easy problems, we typically choose $n = 1$, while for more challenging problems $n > 1$. The algorithm achieves robust and real-time performance, as the experimental results indicate. Specifically, the algorithm's ability to monitor the motion of the apex, the center of the left ventricle, and the aortic annulus has been tested with success on cine long and short axis MRI imagery.

2 Methodology

2.1 Single Particle Filter Tracker ($n = 1$)

The particle filter tracker that we use features 100 particles and performs a single iteration per frame. We denote the motion state of an individual tracker T_i at time t by $\theta_{i,t}$ and its observations by $z_{i,t}$. The state transition model is:

$$\theta_{i,t} = \theta_{i,t-1} + N, \quad (1)$$

where N is the noise subscribing to a Normal distribution. Please note that the motion state itself is characterized by three variables: x, y for translation and ϕ for rotation.

The algorithm approximates the posterior distribution $p(\theta_{i,t}|z_{i,1:t})$ via a set of weighted particles $S_{i,t} = \{\theta_{i,t}^r, \omega_{i,t}^r\}_{r=1}^J$, where $\sum_{r=1}^J \omega_{i,t}^r = 1$; $S_{i,t}$ is properly weighted with respect to $p(\theta_{i,t}|z_{i,1:t})$. Then, we use the maximum a posteriori (MAP) estimate to determine the state of the tracker:

$$\hat{\theta}_{i,t} = \arg \max_{\theta_{i,t}} p(\theta_{i,t}|z_{1:t}) \approx \arg \max_{\theta_{i,t}} \omega_{i,t}^r. \quad (2)$$

The weight values of the particles are proportional to the posterior probability:

$$\omega_{i,t} \propto p(\theta_{i,t}|z_{i,t}). \quad (3)$$

In our implementation, the particle weight is computed as the correlation coefficient of the sampled region of interest (ROI) with an appearance template. The appearance template is composed of intensity values inside the ROI. The choice of pixel-based template ensures generality, rendering the method applicable beyond the MRI modality. Also, intensity blocks are computationally efficient. We adopt the spatio-temporal matte (STM) template described in [9]. The strong point of STM is that updating is based both on pixel dependence (spatial smoothness) and temporal dependence (temporal smoothness).

2.2 Collaborative Tracker Network ($n > 1$)

Particle filter trackers trade sophistication for generality and efficiency. This approach works well in simple motion scenarios, but may reduce robustness in challenging ones. By forming a collaborative network of particle filter trackers we aim to increase sophistication without sacrificing generality and efficiency. Figure 1(a) shows as an example a 3×3 tracker network ($n = 9$), where each tracker is assumed to interact with its neighbors. After having all tracker states computed via the corresponding particle filters, a survivor group is formed consisting of all the well performing trackers. For each individual tracker T_i , a decision is made whether to include it or not into the survivor group. The adjacent trackers $\{T_j, \dots, T_m\}$ provide evidence to make this decision.

The effects of the adjacent trackers on tracker T_i are modeled via a Bayesian network (Figure 1(b)). $\hat{\Theta}_{i,t} = \{\hat{\theta}_{i,t}, \hat{\theta}_{j,t}, \dots, \hat{\theta}_{m,t}\}$ and $Z_{i,t} = \{z_{i,t}, z_{j,t}, \dots, z_{m,t}\}$ are the estimated states and observations of T_i and its adjacent trackers. $W_{i,t}$ represents the event that tracker T_i is in the survivor group at time t . $G_{i,t-1}$ is the Bayesian network at time $t-1$, whose probability $p(G_{i,t-1})$ is known at time t . The arrows in Figure 1(b) indicate dependency relationships. Two underlying assumptions for this Bayesian network are:

- An individual tracker is likely to have motion similar to its adjacent trackers.
- A tracker included in the survivor group at the previous time step is likely to be in the current survivor group.

The joint probability of the Bayesian network $G_{i,t}$ is:

$$p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t}) \propto p(W_{i,t}|G_{i,t-1}, \hat{\Theta}_{i,t}) \prod_k p(\hat{\theta}_{k,t}|z_{k,t}), \quad (4)$$

where k identifies each of the trackers in the Bayesian Network and $p(W_{i,t}|G_{i,t-1}, \hat{\Theta}_{i,t})$ is a probability function of $W_{i,t}$.

Given $G_{i,t-1}$ and $\hat{\Theta}_{i,t}$ as parameters, the function p is defined as:

$$p(W_{i,t}|G_{i,t-1}, \hat{\Theta}_{i,t}) \propto p(G_{i,t-1}) \prod_k N(\hat{\theta}_{i,t}|\hat{\theta}_{k,t}, \sigma^2), \tag{5}$$

where $p(G_{i,t-1})$ is computed at time $t - 1$ and is known at time t . $N(\hat{\theta}_{i,t}|\hat{\theta}_{k,t}, \sigma^2)$ is the probability density of $\hat{\theta}_{i,t}$ on the Normal distribution centered at $\hat{\theta}_{k,t}$ with variance σ^2 . According to Equation (3) each tracker $p(\hat{\theta}_{k,t}|z_{k,t})$ in Equation (4) is proportional to the particle weight of the estimated tracker state. This equals to the normalized highest matching score among all the particles. Thus, the conditional probability $p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t})$ on the left hand side of Equation (4) can be easily computed.

$p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t})$ serves as evidence in deciding whether to include or not tracker T_i in the survivor group. If the evidence exceeds a minimum threshold, then it is included in the survivor group, otherwise it is excluded.

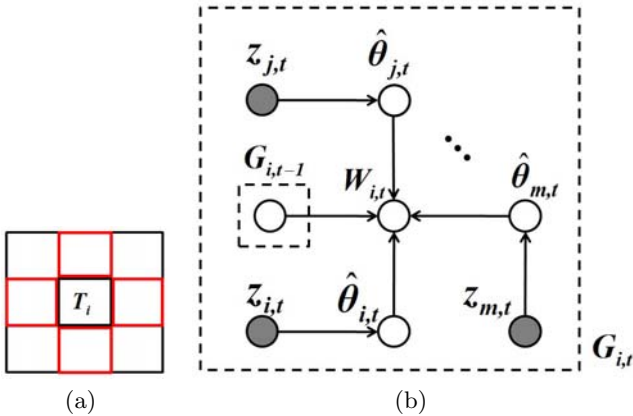


Fig. 1. (a) Layout of tracker network ($n = 9$). (b) Bayesian network for tracker T_i .

The overall motion state is determined by the trackers in the survivor group and computed as follows:

$$\hat{\theta}_{overall} = \frac{1}{\sum_{i=1}^{|W|} \beta_i} \sum_{i=1}^{|W|} \hat{\theta}_i \beta_i, \tag{6}$$

where $|W|$ is the cardinality of the survivor group and β_i is the impact factor of each linked tracker. The latter is determined from:

$$\beta_i = \frac{p(W_{i,t}, \hat{\Theta}_{i,t}|G_{i,t-1}, Z_{i,t})}{D_{i,t}}, \tag{7}$$

where the enumerator is the conditional probability computed in Equation (4) and $D_{i,t}$ is the Euclidean distance from the point of interest (landmark) to the center of the tracker. Trackers closer to the landmark point weigh more on motion estimation with respect to those that are further away. The overall state is used to determine the new position of the landmark point as well as relocate the failed trackers.

3 Experimentation

3.1 Experimental Design

Data were acquired with a 1.5T Espree Siemens MRI scanner. The collected cine sets include short and long axis views on normal volunteers ($N = 2$) with a true fast imaging with steady-state precession (TrueFISP) sequence (TR/TE/a = 60.3ms/1.4ms/80°; slice thickness = 6mm; acquisition matrix = 256 × 256). We are interested in tracking anatomical landmarks on the heart, as a way to close the feedback control loop in MRI-guided robotic surgery on the valve. The interventional tool should be constantly adjusted with respect to the anatomical points to avoid harming critical structures of the beating heart. Specifically, there are four landmarks of interest (see Figure 2):

- Apex (A): The apical point of the left ventricle selected in long axis view.
- Medium (M): The center of the left ventricle at a basal level, i.e., just below the valves, in long axis view.
- Valve (V): The center of the entrance of the aortic valve annulus in long axis view.
- Centroid (C): The center of the left ventricle in short axis view.

Tracking experiments were performed on a set of 16 MRI cine sequences. Each sequence had 25 heart phases (frames) and total duration of 1 sec (approximately one heart-beat). Each of the 16 sequences had one landmark that belonged to one of the above categories. The goal for the experiments was to compare the speed, accuracy, and robustness of different tracking network configurations.

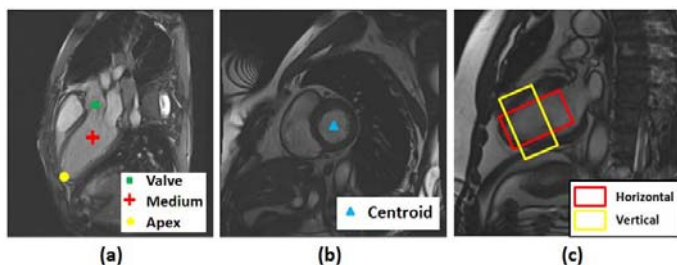


Fig. 2. (a)-(b)Anatomical landmarks of interest. (c) Tracker orientations: Horizontal (along the elongation axis of the moving part) and Vertical (across the elongation axis of the moving part).

These configurations featured different number, size, and orientation of trackers. Specifically, the following cardinalities and sizes were considered:

- Small single tracker (S1): One small tracker that tracks only the moving region around/near the landmark.
- Big single tracker (B1): One big tracker that in addition to the moving region includes some static surrounding structures.
- Collaborative trackers (B4): A 2×2 collaborative tracker network ($n = 4$) that has the same initialization as B1.

Each of these trackers can be applied in either horizontal (H) or vertical (V) orientation (Figure 2 (c)), leading to the following 6 configurations: $HS1$, $HB1$, $HB4$, $VS1$, $VB1$, and $VB4$. For the Centroid point, we make use of a square (instead of rectangular) tracker because the shape of the left ventricle that encloses, is approximately circular (instead of elliptical) in the short axis view. Thus, for this landmark type only, we have 3 instead of 6 configurations: S1, B1, and B4.

Ground-truth was manually labeled by an expert. The labeled points were recorded as time series of the coordinates of the relevant anatomical landmarks $\{x_g, y_g\}_t$, where t is the index of the heart phase ($1 \leq t \leq 25$). Then, all tracking configurations (6 for Apex, Medium, Valve and 3 for Centroid) were applied on the cine sequences to obtain time series of tracked estimates $\{x, y\}_t$ for the relevant landmarks. The closer the tracked time series $\{x, y\}_t$ to the corresponding ground-truth time series $\{x_g, y_g\}_t$, the better.

3.2 Experimental Results

In terms of speed, all proposed tracker configurations are computationally light and achieve real-time performance (25-40 fps) on a standard PC. In terms of accuracy, we use the Euclidean (L2) distance metric to measure the distance between the tracking results and the ground-truth values for all 16 cine sequences. We examine the following two questions: (a) For each landmark (Apex, Medium, Valve, or Centroid), which tracking configuration performs the best? (b) Of the three tracking configurations (S1, B1, or B4), which is the best overall?

In the table of Figure 3, we provide the L2 distances that each tracking configuration (columns) achieves on every cine sequence (rows). Cine sequences are named as $c\#$: [Landmark Designator], where $c\#$ stands for the cine sequence code number (c1-c10), while the Landmark Designator is A for Apex, M for Medium, V for Valve, and C for Centroid. For left and right views in Apex/Valve the designator becomes L-A or L-V and R-A or R-V correspondingly.

The cells with red numbers in the table indicate the configurations that achieve the minimum L2 distance for each cine sequence. We observe that the network of collaborative trackers ($n = 4$) clearly outperforms the single trackers ($n = 1$). This is true for all types of landmarks. Only in the case of Apex, the easiest of the landmarks, the single trackers appear to be somewhat competitive. Indeed, the Apex landmark point features the least amount of motion and typically sits on a well contrasted tissue area. The big single trackers, have better contrasting support but some complex tissue movements from outlying areas

introduce ambiguity at times. The collaborative trackers distribute the risks and advantages. Therefore, failures are locally isolated and overall success is achieved through optimal probabilistic reasoning (Bayesian network). Another interesting observation is that the horizontal configurations fare a lot better than the vertical ones. This is probably to be expected, as the horizontal configuration is along the major motion axis, which is what the tracker strives to capture in the first place. The poor performance of the vertical configurations (which are limit cases), indicates that is not a good idea to orient rectangular trackers away from the major motion axis of the tissue.

		L2 Distance					
		HS1	HB1	HB4	VS1	VB1	VB4
Apex	c1:A	4.98	6.18	4.16	4.53	4.31	4.99
	c2:A	1.68	7.01	2.41	2.33	2.20	2.66
	c3:A	5.84	8.47	8.76	7.20	8.18	7.48
	c4:L-A	7.18	9.37	7.03	—	—	—
	c5:R-A	7.00	5.62	5.23	—	—	—
Medium	c2:M	30.74	5.25	3.91	6.14	4.20	4.08
Valve	c1:V	3.18	5.53	2.60	5.40	3.78	3.99
	c2:V	22.39	3.92	3.83	4.58	4.35	3.42
	c4:L-V	2.86	2.59	2.46	4.39	1.93	1.82
	c4:R-V	3.53	4.58	3.57	3.95	4.26	3.73
Centroid		S1	B1	B4			
	c5:C	16.11	2.39	1.65			
	c6:C	13.54	2.56	2.60			
	c7:C	31.87	3.07	3.05			
	c8:C	5.23	2.15	2.08			
	c9:C	3.59	3.09	1.58			
c10:C	2.96	1.94	1.38				

The right panel of Figure 3 shows annotated results from the tracking of the Centroid point for all three tracking configurations. As it is evident in the third column of images, the collaborative tracker configuration outperforms the single trackers; the reported landmark completely coincides with ground-truth. All the annotated experimental results can be found at <http://ourpapers.info/miccai10-mri>.

Fig. 3. Left: L2-based tracking performance. Right: Centroid tracking. (a) Initialization; (b) Ventricular contraction; (c) Ground-truth(blue) versus tracking results(red).

The right panel of Figure 3 shows annotated results from the tracking of the Centroid point for all three tracking configurations. As it is evident in the third column of images, the collaborative tracker configuration outperforms the single trackers; the reported landmark completely coincides with ground-truth. All the annotated experimental results can be found at <http://ourpapers.info/miccai10-mri>.

4 Conclusion

We have presented a collaborative tracking algorithm that can handle robustly heart motion as appearance changes (due to blood) in MRI cine sequences. The algorithm provides real-time information about landmark points through which robotic interventional tools can be compliantly guided in future valve repair operations. The individual trackers in the algorithm are unimpressive particle

filter trackers. When combined, however, under a Bayesian network framework, they produce sophisticated behaviors without losing efficiency. The framework is flexible enough and general enough to be applied beyond MRI intervention studies.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under grants No. #IIS-0812526 and #CNS-0932272. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

References

1. Woo, Y.J.: Robotic cardiac surgery. *International Journal of Medical Robotics and Computer Aided Surgery* 2(3), 225–232 (2006)
2. Saikus, C.E., Lederman, C.J.: Interventional cardiovascular magnetic resonance imaging: A new opportunity for image-guided interventions. *JACC Cardiovascular Imaging* 2, 1321–1331 (2009)
3. Tsekos, N.V., Khanicheh, A., Christoforou, E., Mavroidis, C.: Magnetic resonance-compatible robotic and mechatronics systems for image-guided interventions and rehabilitation: A review study. *Annual Review of Biomedical Engineering* 9, 351–387 (2007)
4. McVeigh, E.R., Guttman, M.A., Lederman, R.J., Li, M., Kocaturk, O., Hunt, T., Kozlov, S., Horvath, K.A.: Real-time interactive MRI-guided cardiac surgery: Aortic valve replacement using a direct apical approach. *Magnetic Resonance Medicine* 56(5), 958–964 (2006)
5. Dowsey, A.W., Keegan, J., Lerotic, M., Thom, S., Firmin, D., Yang, G.: Motion-compensated MR valve imaging with COMB tag tracking and super-resolution enhancement. *Medical Image Analysis* 11(5), 478–491 (2007)
6. Fradkin, M., Ciofalo, C., Mory, B., Hautvast, G., Breeuwer, M.: Comprehensive Segmentation of Cine Cardiac MR Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 178–185. Springer, Heidelberg (2008)
7. Yuen, S.G., Kesner, S.B., Vasilyev, N.V., Nido, P.J., Howe, R.D.: 3D Ultrasound-Guided Motion Compensation System for Beating Heart Mitral Valve Repair. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 711–719. Springer, Heidelberg (2008)
8. Isard, M., Blake, A.: Condensation - Conditional density propagation for visual tracking. *International Journal of Computer Vision* 29(1), 5–28 (1998)
9. Zhou, Y., Tsiamyrtzis, P., Pavlidis, I.: Tissue Tracking in Thermo-physiological Imagery through Spatio-temporal Smoothing. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 1092–1099. Springer, Heidelberg (2009)

Calibration and Use of Intraoperative Cone-Beam Computed Tomography: An In-Vitro Study for Wrist Fracture

Erin Janine Smith¹, Anton Oentoro², Hisham Al-Sanawi³, Braden Gammon³, Paul St. John⁴, David R. Pichora^{3,4}, and Randy E. Ellis^{1,2,3,4}

¹ Mechanical Engineering, Queens University at Kingston, Canada K7L 3N6

² School of Computing, Queen's University, Kingston Ontario, Canada K7L 3N6

³ Department of Surgery, Kingston General Hospital, Kingston, Canada K7L 2V7

⁴ Human Mobility Research Centre, Queen's University, Kingston, Canada K7L 2V7
ellis@cs.queensu.ca

Abstract. The standard workflow in many image-guided procedures, preoperative imaging followed by intraoperative registration, can be a challenging process and is not readily adaptable to certain anatomical regions such as the wrist. In this study we present an alternative, consisting of a preoperative registration calibration and intraoperative navigation using 3D cone-beam CT. A custom calibration tool was developed to preoperatively register an optical tracking system to the imaging space of a digital angiographic C-arm. This preoperative registration was then applied to perform direct navigation using intraoperatively acquired images for the purposes of an in-vitro wrist fixation procedure. A validation study was performed to assess the stability of the registration and found that the mean registration error was approximately 0.3 mm. When compared to two conventional techniques, our navigated wrist repair achieved equal or better screw placement, with fewer drilling attempts and no additional radiation exposure to the patient. These studies suggest that preoperative registration coupled with direct navigation using procedure-specific graphical rendering, is potentially a highly accurate and effective means of performing image-guided interventions.

1 Introduction

Minimally invasive computer-assisted interventions in orthopedics have, to date, largely relied on preoperative computed tomography (CT) scans. In the usual paradigm, the images are segmented, a plan is made, intraoperative registration of the image to the patient is performed, and only then can the surgeon navigate the procedure. There are many drawbacks to this paradigm, including: logistical challenges associated with preoperative CT; time-consuming segmentation, with skilled technicians often requiring an hour or more; attachment of a tracking device to the anatomy can be difficult or impossible; and achieving the registration is invasive, time-consuming, and fraught with errors.

Recent developments in flat-panel fluoroscopy have made 3D cone-beam CT (CBCT) feasible for intraoperative use in an operating room. Such 3D

volumetric data can be used, with direct navigation, in a system that requires neither image segmentation nor registration during the surgical procedure. Instead, a direct-navigation system can use a preoperative calibration stage to register the coordinate frame of the image space to the coordinate frame of a fixed tracking system. The resulting workflow needs only a minimally invasive tracking device attached to a patient, with no segmentation or surface-based registration – nearly ideal for computer-assisted minimally invasive procedures.

Direct navigation is especially attractive for wrist fractures, particularly for the scaphoid, one of the eight carpal bones that is frequently fractured following a fall on an outstretched palm [1]. Proximal scaphoid fractures often fail to unite, due to the unique blood supply to this region. As such, primary internal fixation is often a preferred method of treatment, and typically involves the placement of a Kirschner wire (K-wire) or surgical screw along the central axis of the bone to unite the bony fragments.

Clinical studies have indicated the importance of an accurate placement along the central axis for a successful outcome [2], with minimally invasive percutaneous technique preferred to reduce the risk of infection and compromising nearby tissue structures [3]. However, percutaneous pinning is challenging because of the small size of the scaphoid, and requires many intra-operative X-ray images to ensure an accurate placement of the fixation device. As fluoroscopic images are 2D, it is also difficult to correctly interpret the three-dimensional anatomy in order to locate the central axis of the scaphoid.

It is difficult to adapt the traditional computer-assisted surgical workflow to navigate scaphoid pinning because a surface-based or fiducial-based registration cannot be used successfully in the region of the wrist [4]. Attempts at image-based registration using fluoroscopic images to match extracted contours [5] or using the data directly with mutual information [6] have so far been insufficiently accurate for clinical use. Ultrasound techniques [4] have not been shown to improve the application accuracy over conventional fluoroscopy.

Here, we present a novel technique for preoperative calibration and results of a preliminary validation assessment. We then apply our preoperative calibration to perform intraoperative image guidance for volar percutaneous scaphoid pinning, and compare this technique to conventional fluoroscopic guidance using a standard C-arm and guidance using a digital angiographic C-arm.

2 Preoperative Calibration: Materials and Methods

Our direct navigation solved the image-to-patient registration problem by splitting the chain of coordinate transformations into two pieces, preoperative and intraoperative. Preoperatively, the pose of the 3D CBCT was found in tracking coordinates by using a custom calibration device; later, the pose of the patient was captured simultaneously with the CBCT image acquisition. The key technical question was the accuracy and reliability of the preoperative calibration.

All experiments were performed in a recently-constructed operating room at Kingston General Hospital. This room was integrated with an Optotrak Certus

(Northern Digital Inc., Waterloo, Canada) tracking system and a flat-panel 3D fluoroscope (Innova 4100, GE Healthcare, Buc, France) that acquired CBCT images. A custom surgical navigation system was used to provide image guidance.

For full details of the preoperative calibration process, the reader is referred to a recent Master’s thesis [7]. A custom Multi-Modal Calibrator (MMC) was developed to conduct a preoperative calibration in order to relate the coordinate frames of the Optotrak {C} and that of the Innova 3D spin image {S}. The device was shaped like a hollow cube (shown in Fig. 1). To be sensed by both the imaging device and the position sensor of the Optotrak, the MMC contained respectively metallic radio-opaque markers and infrared light-emitting diodes.

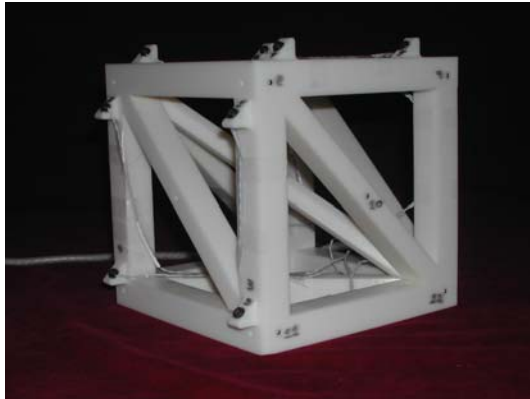


Fig. 1. A photograph of the multi-modal calibration device that was used for this study

Prior to a navigated surgical session, the registration was performed by positioning the MMC on the operating table. The CBCT image was acquired and, simultaneously, the pose ${}^C_M T$ was acquired to find the MMC in Optotrak Certus coordinates {C}. The MMC beads were found in the CBCT image using a validated interpolation algorithm [8]; because the algorithm also detected wires and IREDS, a variant of the robust RANSAC algorithm [9] was used to register the detected beads, yielding the transformation ${}^S_M T$. This gives the critical calibration equation

$${}^C T = {}^C_M T * {}^S_M T \tag{1}$$

that coupled any subsequent CBCT image to Optotrak coordinates.

The precalibrated registration was studied in an extensive experiment over 3 weeks. During this time, the MMC was left undisturbed on the operating table and 14 CBCT scans were acquired periodically. The analysis of the transformations used the differential transform ΔT between transforms T_A and T_B

$$\Delta T = T_B * T_A^{-1} \tag{2}$$

Two sets of comparisons were made: pairwise ${}^C_M T$ using $\Delta^C T_i$, and pairwise ${}^S_M T_i$ using $\Delta^S T_i$. The rotational component of each ΔT was converted to an angle/axis notation, and for the translational component the norm was calculated.

3 Preoperative Calibration: Results

The 14 analyses of the transforms ${}^C_M T_i$ and ${}^M_S T_i$ were examined statistically. The mean rotational and translational deviations in the dimensions of the operating room are presented in Table 1. The mean rotational and translational deviations in the movement of the Innova system are presented in Table 2.

Table 1. Average rotational and translational errors calculated from $\Delta^C T$ for the Innova registration experiment. μ is the mean and σ is the standard deviation.

Type of Error	μ	σ	Range
Rotational Error(degrees)	-	-	-
Translational Error for X (mm)	0.19	0.059	0.0035–0.44
Translational Error for Y (mm)	0.059	0.0062	0.00010–0.18
Translational Error for Z (mm)	0.19	0.029	0.00070–0.67
Norm of Translational Error (mm)	0.30	0.176	0.028–0.74

Table 2. Average rotational and translational errors calculated from $\Delta^S T$ for the Innova registration experiment; μ is the mean and σ is the standard deviation

Component	μ	σ	Range
Rotational Error(degrees)	0.27	0.027	0.0065–0.61
Translational Error for X (mm)	0.036	0.0040	0.00030–0.12
Translational Error for Y (mm)	0.053	0.0080	0.0016–0.14
Translational Error for Z (mm)	0.058	0.0083	0.00050–0.15
Norm of Translational Error (mm)	0.097	0.0076	0.020–0.17

The mean registration error was approximately 0.3mm, or 300 microns. This compares favorably with the stated accuracy of the Optotrak (150 microns) and is unlikely to be improved upon using our equipment. We concluded that preoperative calibration is a fast and highly accurate method for direct navigation.

4 In-Vitro Application: Materials and Methods

Direct navigation was applied to a volar percutaneous pinning procedure. This technique was compared to two conventional imaging techniques: (i) “C-arm”, in which a standard fluoroscopic C-arm was used, and (ii) “Innova”, in which the Innova imager was programmed with orthogonal 2D views.

A 3D model of the human wrist was constructed to provide greater consistency between the trials than permissible using cadaveric anatomy (Fig. 2). The model wrist featured an interchangeable scaphoid which allowed the wrist to be reused for each trial by replacing only the scaphoid. A molded recess was integrated into the model which served to firmly seat the scaphoid, and to which the scaphoid was secured using a plastic fastener.

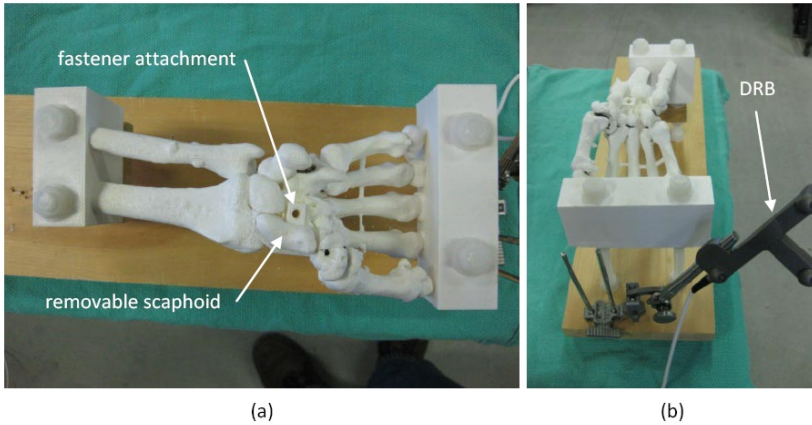


Fig. 2. 3D Wrist Model. (a) A removable scaphoid was seated in a molded recess and secured via a fastener hole. (b) The navigated trial setup with a patient coordinate frame \mathbf{P} established by an optical tracking array (DRB).

Both the model wrist and removable scaphoids were fabricated using a 3D rapid prototyping printer (SST 1200es, Dimension/Stratasys, Eden Prairie, USA). To facilitate drilling, the scaphoid models were printed with a medium density polymer then coated with barium-infused paint. The wrist was printed with lower density polymer. During each trial, the entire apparatus was covered to prevent direct visualization of the bony anatomy but to permit realistic palpation.

Randomized trials were conducted in which three surgeons each performed 24 trials – 8 using each imaging technique. The surgical goal was to insert a 1.6mm K-wire along the central axis of the model scaphoid. In each technique a volar approach was employed in which the wire was drilled from the distal end until it breached the proximal end. This exit hole was created to allow for post-trial data analysis. For each trial, the number of drill passes was recorded. To assess radiation exposure, an exposure meter was positioned adjacent to the wrist model to estimate a patient's radiation dose.

An optical tracking array was attached at the base of the plastic model to establish a patient coordinate frame $\{\mathbf{P}\}$. An intraoperative image of the wrist was captured using the Innova to perform a 3D acquisition and, simultaneously, the navigation system captured the pose of $\{\mathbf{P}\}$ in the Optotrak coordinate frame $\{\mathbf{C}\}$ as the transformation ${}^{\mathbf{C}}T$. The combined registration transformation was the composition of the preoperative calibration transform ${}^{\mathbf{C}}T$ (Eq. [1](#)) with the transformation ${}^{\mathbf{P}}T$ to find the overall image-to-patient registration as:

$${}^{\mathbf{P}}T = {}^{\mathbf{C}}T * {}^{\mathbf{S}}T \quad (3)$$

The navigation system provided digitally reconstructed radiographs to the surgeon (Fig. [3](#)). In the planning phase, the surgeon positioned a desired drill path on the rendered views. During the drilling phase, the real-time orientation of a tracked drill guide relative to this planned path was shown on the overhead OR

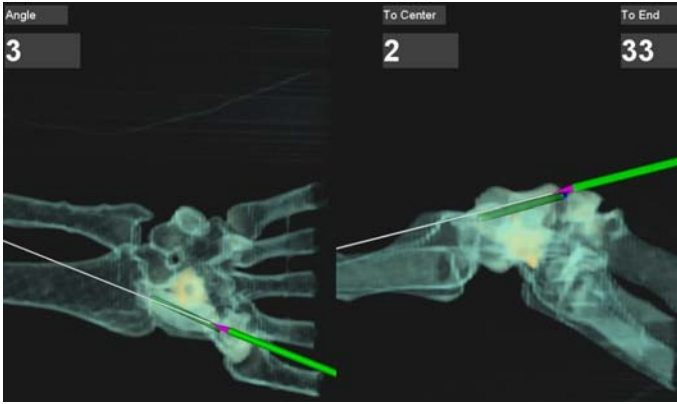


Fig. 3. Screen shot of the software during a navigated procedure showing orthogonal views. The dark green cylinder depicts the target plan; the purple and green stylus represents the real-time tracked drill guide.

monitors. In conventional techniques, surgeons were given the option of using an identical untracked drill guide.

Following the trials, CT scans of the drilled scaphoids were obtained and segmented (Mimics v13, Materialise, Leuven, Belgium) in order to generate 3D surface models for data analysis. Algorithms were developed to determine the line of best fit through the drill hole and compute the shortest distance from this line to the exterior surface of the scaphoid. This measure was extrapolated to determine the location of the shortest distance and to assess whether a placed screw would have breached the scaphoid. Two-sample F-tests were used to assess equal variances and select appropriate t-tests for wire placement centricity, drill passes, and radiation exposure.

5 In-Vitro Application: Results

The 24 scaphoid drilling trials were analyzed to assess screw placement, number of drill passes required, and x-ray exposure to the patient (Table 3). The minimum distance from the drill path to the scaphoid surface was deemed to be credible a measure of centricity as well as a safety factor for screw breach. Although no potential screw breaches were noted in any of the trials, the minimum distance from the drill path to the scaphoid surface was significantly higher using navigation when compared to the Innova ($p=0.02$, $\alpha = 0.05$).

There was no significant difference in the number of drill passes between the navigated and Innova techniques, however the C-arm technique required more attempts ($p=0.01$, $\alpha = 0.05$). Imager output was significantly higher in the navigated cases (mean = $11.05 \text{ cGy}/m^2$, SD = $1.67 \text{ cGy}/m^2$) than with the Innova (mean = $2.78 \text{ cGy}/m^2$, SD = $2.13 \text{ cGy}/m^2$) or C-arm (mean = $0.03 \text{ cGy}/m^2$, SD = $0.03 \text{ cGy}/m^2$). Conversely, there was no significant difference

Table 3. Summary of post-trial analysis of the three drilling methods: Navigated (N), Innova (I) and C-arm (C)

	N	I	C
Safety [mm]	2.8 ± 0.5	2.6 ± 0.5	2.8 ± 0.5
X-ray Exposure[mrad]	1.1 ± 1.4	0.5 ± 0.9	2.3 ± 3.3
Drill Passes	1.5 ± 0.7	1.7 ± 1.0	2.2 ± 0.8

in the mean imaging time. More importantly, no difference in patient x-ray exposure (as recorded by the exposure meter) was observed.

6 Discussion

An important consideration in our workflow was the overall speed of the pre-operative registration step. The entire registration process, including scanning, transfer, and computations, took less than 15 minutes to perform. In an ideal scenario, where the relative position and orientation of the imaging modalities and the position sensor remain perfectly fixed, the registration process would need to be performed only once. This is unlikely in a realistic clinical setting, so the registration process is expected to be performed more often – perhaps daily – and thus needs to be reasonable efficient.

The results from our preliminary trial demonstrate that we can achieve equal or better safety of screw placement, with fewer drill passes and, no additional radiation exposure to the patient. Of a potentially greater concern is the radiation exposure to the surgeon, which we are currently in the process of evaluating in a secondary round of trials, through the use of radiation badges. It is worth noting that we expect the radiation exposure to the surgeon to be minimal, as the initial 3D acquisition spin is acquired while the surgeon is outside of the radiation field, and the only notable exposure would be the potential use of supplementary fluoroscopic images to verify screw placement.

One of the main limitations of our current navigation system is that it only provides the orientation of the drill guide and does not provide any depth cues. Feedback from the surgeons indicated that this was indeed an impediment, and forced them to rely on using more intraoperative 2D fluoroscopic images to verify wire position. Amending the navigation to include depth cues is now a focus for upcoming trials. Another area of improvement for the next version of the navigation system concerns the techniques used to render the 3D image. Again, surgeons have expressed difficulty with viewing the margins of the scaphoid on the rendered views, and thus planning the procedure. We are currently experimenting with new rendering and visualization techniques for upcoming trials.

7 Conclusion

A novel workflow for image-guided surgery was presented that does not require conventional intraoperative image registration. The system has a much simpler

and efficient workflow than conventional CT image acquisition and segmentation without compromising accuracy. We have also demonstrated that this technique can be adapted to anatomical regions, such as the wrist, that cannot be readily registered and spatially tracked intraoperatively. Our future challenges will be to make this system practical for routine clinical use.

Acknowledgements. This work was supported in part by the Canada Foundation for Innovation, the Canadian Institutes for Health Research, Kingston General Hospital, the Ontario Innovation Trust, and the Natural Sciences and Engineering Research Council of Canada.

References

1. Leslie, I., Dickson, R.: The fractured carpal scaphoid. *J. Bone Joint Surg. Br.* 63-B, 225–230 (1981)
2. McCallister, W., Knight, J., Kaliappan, R., Trumble, T.: Central placement of the screw in simulated fractures of the scaphoid waist. *J. Bone Joint Surg. Am.* 85-A, 72–77 (2003)
3. Filan, S., Herbert, T.: Herbert screw fixation of scaphoid fractures. *J. Bone Joint Surg. Br.* 78-B, 519–529 (1996)
4. Beek, M., Abolmaesumi, P., Luenam, S., Ellis, R., Sellens, R., Pichora, D.: Validation of a new surgical procedure for percutaneous scaphoid fixation using intraoperative ultrasound. *Medical Image Analysis* 12, 152–162 (2008)
5. Feldmar, J., Ayache, N., Betting, F.: 3D-2D projective registration of free-form curves and surfaces. *Comput. Vis. Image Underst.* 65(3), 403–424 (1997)
6. Dalvi, R., Abugharbieh, R., Pickering, M., Scarvell, J., Smith, P.: Registration of 2D to 3D joint images using phase-based mutual information. *Proc. Soc. Photo Opt. Instrum. Eng.* 651209, 1–9 (2007)
7. Oeontoro, A.: A system for computer-assisted surgery with intraoperative ct imaging. Master's thesis, Queen's University, Kingston, Canada (2009)
8. Ellis, R.E., Toksvig-Larsen, S., Marcacci, M., Caramella, D., Fadda, M.: Use of a biocompatible fiducial marker in evaluating the accuracy of CT image registration. *Investigative Radiology* 31(10), 658–667 (1996)
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(3), 381–395 (1981)

A Strain Energy Filter for 3D Vessel Enhancement

Changyan Xiao^{1,2}, Marius Staring¹, Denis Shamonin¹, Johan H.C. Reiber¹,
Jan Stolk¹, and Berend C. Stoel¹

¹ Division of Image Processing, Department of Radiology, Leiden University Medical Center, P.O. Box 9600, 2300 RC, Leiden, The Netherlands

² College of Electrical and Information Engineering, Hunan University, China

Abstract. The traditional Hessian-related vessel filters often suffer from the problem of handling non-cylindrical objects. To remedy the shortcoming, we present a shape-tuned strain energy density function to measure vessel likelihood in 3D images. Based on the tensor invariants and stress-strain principle in mechanics, a new shape discriminating and vessel strength measure function is formulated. The synthetical and clinical data experiments verify the performance of our method in enhancing complex vascular structures including branches, bifurcations, and feature details.

1 Introduction

Pulmonary vessel detection plays an important role in computer analysis of lung CT images. Evaluating vessels is of considerable value to diagnosing for example pulmonary emboli and hypertension. Additionally, there is an interest in identifying the vascular trees as landmarks for matching lungs across variety [1]. However, accurate and robust detection of pulmonary vessels still remains a problem because of the geometrical complexity and fine characteristics of details. Especially, with non-contrast-enhanced images being widely used in CT densitometry, the vascular detection becomes even more challenging.

In this paper, we present a 3D vessel enhancing filter with the main purpose to break the cylinder limits of traditional Hessian filters [2,3,4,5], improve their ability in preserving more general vascular structures like bifurcations. The method is initially motivated by a recent achievement of stress and strain theory in solid mechanics [6], which was introduced previously to fMRI imaging by Ennis et al. [7]. We establish the idea basically on a link between image structures and locally compressed states of material, due to the mathematical equivalence between the image Hessian matrix and the mechanical stress or strain tensor.

2 Theoretical Background

The Hessian matrix H is equivalent to the stress tensor in solid mechanics in that it is also a second-order symmetrical tensor. Then, H can be decomposed into

an eigensystem of three real eigenvalues λ_i and three orthogonal eigenvectors \mathbf{e}_i ($i = 1, 2, 3$).

A common decomposition of stress tensor is $H = \bar{H} + \tilde{H}$, where \bar{H} corresponds to the so-called “isotropic” or spherical component. \tilde{H} indicates the “deviatoric” or anisotropic component, and contains the directional inequality information. They are calculated as

$$\bar{H} = \frac{1}{3}tr(H)\mathbf{1} = \frac{1}{3}(H : \mathbf{1})\mathbf{1}, \quad \tilde{H} = H - \frac{1}{3}tr(H)\mathbf{1} = H - \frac{1}{3}(H : \mathbf{1})\mathbf{1}, \quad (1)$$

where $\mathbf{1}$ is the identity matrix and the colon ($:$) denotes the tensor contraction operator. According to the stress principle, only the isotropic component causes volume changes, while the deviatoric one exclusively accounts for the distortion or shear without volume changes. Since $\bar{H} : \tilde{H} = 0$, the two components are mutually orthogonal. From this, Criscione et al. [6] introduced a set of tensor invariants K_i :

$$K_1 = tr(H), \quad K_2 = \|\tilde{H}\|_F, \quad \text{and} \quad K_3 = 3\sqrt{6} \det(\tilde{H}/\|\tilde{H}\|_F), \quad (2)$$

where $\det()$ is the determinant operator. Due to orthogonal decomposition, the three invariants reflect mutually independent properties of local deformation. The intuitive physical meaning behind them is that K_1 represents the amount of dilatation (with a negative value corresponding to compression), K_2 the magnitude of distortion, and K_3 the mode or type of distortion.

3 Method

The development of our vessel filter mainly originates from the above stress tensor invariants, and is divided into five parts:

3.1 Measure of Brightness Contrast

We first develop a contrast term from the invariance K_1 . Since K_1 is a measure of local volume change, it indirectly reflects the density variation inside an infinitesimal element of material by the mass conservation law. If the density is assumed to be the image intensity, K_1 will be an indicator of relative intensity change, i.e. brightness contrast.

Generally, we only need to confine the sign of K_1 , with negative (positive) values corresponding to bright (dark) objects. For pulmonary CT images, we add a relative threshold in proportion to the maximum magnitude of eigenvalues (λ_m) to ensure noise immunity, i.e.

$$|K_1| = |\lambda_1 + \lambda_2 + \lambda_3| > 3\alpha\lambda_m. \quad (3)$$

Here, the parameter $\alpha > 0$ is used to adjust sensitivity.

3.2 Measure of Structure Strength

In mechanics, the strain energy density is a term to quantify the local energy stored in solids due to stress effects after mechanical loading. Since the procedure simultaneously results in an uneven mass distribution, the strain energy also indirectly reflects the material density inhomogeneity. Based on the similarity of density variation to intensity concentration around image structures, we introduce an strain energy density term to measure structure strength.

Consider an isotropic and linear elastic material, the strain energy density is defined as $U = \frac{1}{2} \sum_{i=1}^3 \sum_{j=1}^3 \sigma_{ij} \eta_{ij}$, with σ_{ij} and η_{ij} the elements of stress and strain tensors. We rewrite it in the form of orthogonal components of the stress tensor: $U(H) = \frac{1-2\nu}{2\epsilon} (\bar{H} : \bar{H}) + \frac{1+\nu}{2\epsilon} (\tilde{H} : \tilde{H})$. Here, ϵ and ν are Young’s modulus and Poisson’s ratio, respectively. Thus, the energy density function is divided into two independent parts: volume changing (isotropic) energy and distortion deforming (anisotropic) energy. We further omit the fixed ϵ , and adopt the square root to keep the same power order of the original intensity, i.e.

$$\rho(H) = \sqrt{(1 - 2\nu)(\bar{H} : \bar{H}) + (1 + \nu)(\tilde{H} : \tilde{H})}. \tag{4}$$

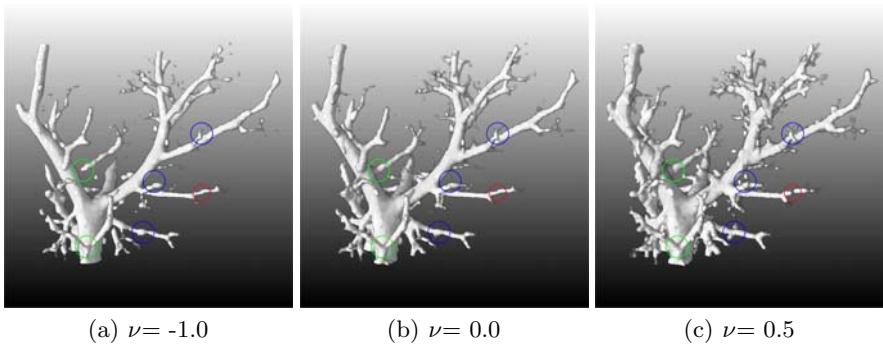


Fig. 1. Vascular structure strength measure, see eq. (7), with varying parameter ν

Notice the Poisson’s ratio $\nu \in [-1, 0.5]$ essentially describes the mutual influence between deformations in different directions. For image processing, if $\nu < 0$, the intensity concentrating in one direction will cause the relative intensity to increase in the perpendicular directions, which is known to generate isotropic “blob” structures. Inversely, $\nu > 0$ will adversely affect the intensity concentration in orthogonal directions, and then encourages anisotropy. If $\nu = 0$, we have $\rho(H) = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$, which corresponds to the “second order structureness” used by Frangi et al. [4]. We verified the effect of ν through a real image experiment in Fig.1. It can be found that a low ν tends to develop smooth surfaces, the anisotropic structures (labeled with “green” circles) and feature details (“blue” circle) are gradually enhanced with increasing ν . However, a very large ν is at the risk of exaggerating blob-like deformations (“red” circle).

3.3 Intensity Continuity Constraint

To understand the Hessian eigenvalue distribution of general vascular shape, we introduce an intensity continuity constraint. As observed, most vessel structures including the branch and junction, share one common characteristic of intensity continuity, i.e. there exists at least one direction in which the intensity variation is very small. The intensity continuity at pixel x_0 in direction \mathbf{r} is equivalent to $I_\sigma(x_0 + h\mathbf{r}) - I_\sigma(x_0) \approx 0$, with σ the observing scale, \mathbf{r} a unit vector and h the magnitude. Then, the corresponding Taylor expansion is $\nabla I_\sigma(x_0) \cdot \mathbf{r} + \frac{h}{2} \mathbf{r}^t H_\sigma(x_0) \mathbf{r} \approx 0$. If the first-order derivative is negligible, the Hessian term will mainly account for the intensity variation, i.e. $\mathbf{r}^t H_\sigma(x_0) \mathbf{r} \approx 0$.

To ensure the dominance of the Hessian term, we additionally define a relative Hessian strength function

$$G_{rel}(x) = \exp\left(-\beta \frac{\|\nabla I\|}{\lambda_m}\right). \quad (5)$$

Here, $\beta > 0$ is used to adjust the sensitivity of response. $G_{rel}(x)$ will tend to 1 (or 0), while the gradient (or Hessian) term is negligible. This relative strength function is useful in suppressing step edges, which also responds strongly to eigenvalue detection like the second-order structures, but takes a strong gradient.

Based on the continuity constraint, the Hessian eigenvalue relation can be formulated as $c_1^2 \lambda_1 + c_2^2 \lambda_2 + c_3^2 \lambda_3 = 0$, where c_i indicate the coefficients of \mathbf{r} under the eigen-system. The potential meaning is not only to explain the eigenvalue distribution of traditional line and sheet shapes, but also understand more general vascular structures like bifurcations and stenoses. The latter are then verified to have non-zero and differently signed eigenvalues.

3.4 Vessel Shape Discrimination

A vascular shape discriminating function can be formulated by combining the above tensor invariants. The *mode* of distortion (invariant K_3) provides a natural measure of shape anisotropy. But its discriminating ability will gradually decrease when the anisotropic extent becomes weaker. To remedy this drawback, we introduce the fractional anisotropy $FA = \sqrt{3(\tilde{H} : \tilde{H}) / (H : H)}$, which is actually a ratio of invariant K_1 and K_2 , as an additional indicator. Here, a constant 3 is added to ensure that $FA = 1$ for an ideal vessel branch ($|\lambda_1| \approx 0$ and $\lambda_2 \approx \lambda_3 \gg 0$ or $\ll 0$).

Generally, FA and mode both measure the shape variety, but have different sensitivity as a function of anisotropy. Based on their response curves, we merge them in the final vessel shape discriminating function:

$$V(x) = \begin{cases} \sqrt{FA}, & FA < 1 \\ \frac{1}{2} [mode(x) + 1], & \text{otherwise} \end{cases}. \quad (6)$$

It can be verified that $V(x)$ approaches 0 for blob ($\lambda_1 \approx \lambda_2 \approx \lambda_3 \gg 0$ or $\ll 0$) and sheet shapes ($\lambda_1 \approx \lambda_2 \approx 0$ and $\lambda_3 \gg 0$ or $\ll 0$), whereas it takes 1 for line shapes

corresponding to vascular branches. The most attractive property is that $V(x)$ also responds with high values to locally deformed structures like junctions and stenoses, which for example may have one positive and two negative eigenvalues.

3.5 Multi-scale Vesselness Measure

To generate a shape-tuned strain energy function, we combine the previous vascular structure related terms as

$$\varphi(\sigma, x) = \begin{cases} 0, & \text{if } \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3) > -\alpha\lambda_m \\ \exp\left(-\beta\frac{\|\nabla I\|}{\lambda_m}\right) V^\kappa(x)\rho(H, \nu), & \text{else.} \end{cases} \quad (7)$$

The power coefficient $\kappa \in [0, 1]$ of $V(x)$ is added to adjust the sharpness of shape selectivity. The details of multi-scale integration can be referred to [3,4]. We also adopt the γ -normalization and rescaling operation to compensate intensity changing across scales.

4 Experimental Results

The proposed method is validated by using synthetical and clinical datasets. The filtering results were compared with three traditional Hessian-based filters [3,4,8]. The quantitative evaluation is conducted by calculating the *precision – recall* (PR) curves after binarizing the filtered images with different global thresholds. Two versions of *recall* (sensitivity) are defined, namely in terms of volume and skeleton of true vessels. The latter is specifically designed to emphasize the presence of vessels and neglecting its radial size, which is uncertain for a vessel with a Gaussian profile. The free parameters of all methods are optimized, based on the area under curve (AUC) of the PR curves.

4.1 Synthetic Data

As shown in Fig. 2(a), a digital phantom was generated with six objects to simulate different kinds of vessel structures, which resemble stenoses, varying diameters, bifurcations, and curved or touching branches.

We compared the four methods on the synthetic image with 16% variance noise in Fig. 2(b). The scale range used in this experiment is 1 – 6 pixels, and is further divided logarithmically into 10 steps. The parameters of our proposed method were optimized to $\alpha = 0.2$, $\beta = 0.02$, $\kappa = 0.4$, ν was set to 0 without priority to specific structures. The filtering results are given in Fig. 2(c)–(f). As expected, the traditional Hessian filters were not good in preserving the junctions and local deformations, where disconnections in varying degrees can be observed. The proposed method enjoyed more merits in enhancing both vascular branches and bifurcations. Additionally, detailed features like thin vessels and local diameter variety are well preserved. For quantitative evaluation, we refer to Fig. 3. The performance of our method belongs to the best two with Frangi’s

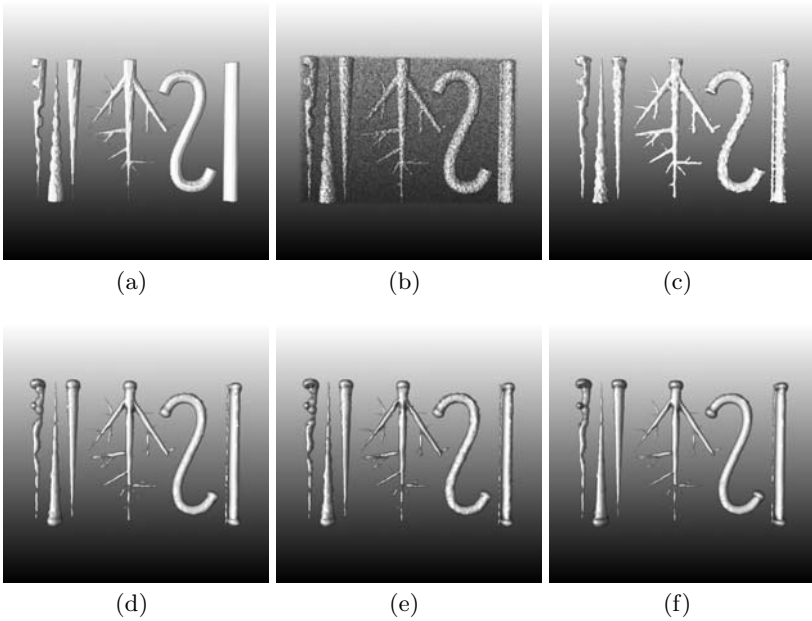


Fig. 2. Synthetic dataset experiment. (a) Original image, (b) synthetic image, enhanced images from (c) the proposed method, (d) Frangi [4], (e) Li [9], and (f) Sato [3] filters

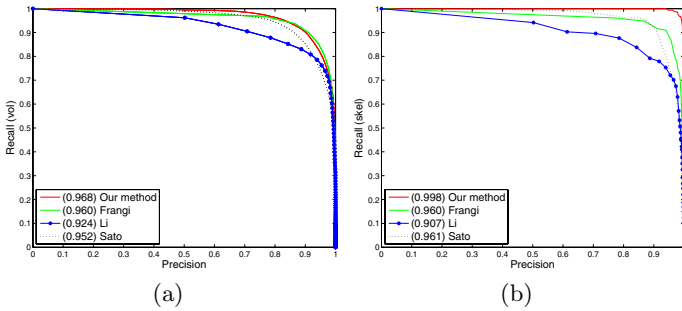


Fig. 3. PR curves of filtered synthetic images. (a) and (b) respectively correspond to the volume- and skeleton-based *recall*, where the numbers in legends are AUCs.

filter on the volume-based PR measure, while our advantage is dominant on the skeleton-based one. It can be understood from the distribution and AUCs of PR curves that our result takes higher completeness (*recall*) under the majority of volume fidelity (*precision*) levels compared with the other three methods. This just reflected the merit of our method in preserving thin vessels and connectivity of vascular networks.

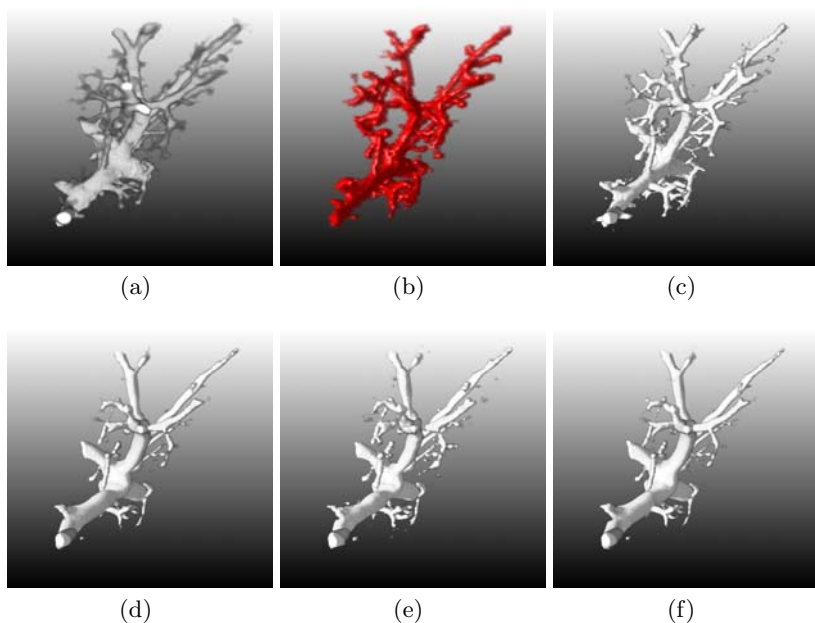


Fig. 4. Non-contrast enhanced pulmonary CT dataset experiment. (a) original image; (b) manually segmented “ground truth”; (c)-(f) give in order the filtered images from the proposed, Frangi^[4], Li^[8] and Sato^[3] methods

4.2 Clinical Data

Our method was further validated by using a cropped clinical pulmonary CT dataset. The images were acquired without contrast media injection, and the resolution was $0.7 \times 0.7 \times 0.5$ mm. The parameters of our filter were optimized to $\alpha = 0.1$, $\beta = 0.06$ and $\kappa = 0.5$, and a positive $\nu = 0.10$ was used to enhance junctions. The selected vascular scales were 0.5-3.0 mm, and 10 steps were used.

For better observation of details, a region of interest was extracted and shown in Fig. 4(a), together with a manually segmented “ground truth” by experts in Fig. 4(b) for reference. The extraction was drawn first by a radiologist, then a pulmonologist and a surgeon were asked to verify it. Fig. 4(c)-(f) are the corresponding results of the four filters. It is clear that the traditional filters improve the visualization of main vascular branches at the cost of weakening junctions and details. Our method keeps most vascular structures while suppressing the unwanted noise. Both branches and bifurcations are enhanced without distortion. In particular, thin vessels and details are clearly preserved with the continuity to main vessels. The merits can be further verified by the quantitative evaluation in Fig. 5.

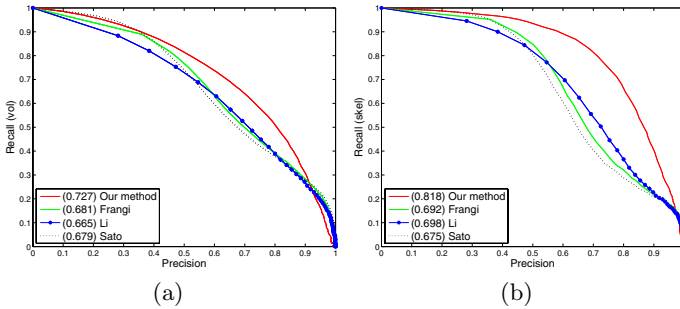


Fig. 5. PR curves of filtered clinical images. (a) and (b) respectively correspond to the volume- and skeleton-based *recall* with AUCs shown in the legends

5 Conclusions

In this paper, we have presented a 3D vessel enhancing filter based on the tensor invariants and strain energy density theory. The main feature is that we directly generalized the Hessian-based vesselness filters to non-tubular shapes and realized the enhancement of anisotropic vascular structures like junctions. The preliminary results verified the performance of our model in preserving locally deformed vessels and detailed features.

Acknowledgements. This work was funded by STW (grant LPG.07998) of the Netherlands and NSFC (No. 60835004, 60871096, 60872130) of China.

References

1. Sluimer, I., et al.: Computer analysis of computed tomography scans of the lung. *TMI* 25(4), 385–405 (2006)
2. Lorenz, C., et al.: Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images. In: *CVRMed/MRCAS 1997*, pp. 233–242 (1997)
3. Sato, Y., et al.: 3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In: *CVRMed/MRCAS 1997*, pp. 213–222 (1997)
4. Frangi, A.F., et al.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
5. Krissian, K., et al.: Model-based detection of tubular structures in 3d images. *CVIU* 80(2), 130–171 (2000)
6. Criscione, J.C., et al.: An invariant basis for natural strain which yields orthogonal stress response terms in isotropic hyperelasticity. *J. Mech. Phys. Solids* 48(12), 2445–2465 (2000)
7. Ennis, D.B., Kindlmann, G.: Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magn. Reson. Med.* 55(1), 136–146 (2006)
8. Li, Q., et al.: Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional ct scans. *Med. Phys.* 30(8), 2040–2051 (2003)

Virtual Stent Grafting in Personalized Surgical Planning for Treatment of Aortic Aneurysms Using Image-Based Computational Fluid Dynamics

Guanglei Xiong¹ and Charles A. Taylor²

¹ Biomedical Informatics Program, Stanford University, CA, USA

² Departments of Bioengineering and Surgery, Stanford University, CA, USA
{glxiong, taylorca}@stanford.edu

Abstract. Image-based computational fluid dynamics provides great promise for evaluation of vascular devices and assessment of surgical procedures. However, many previous studies employ idealized arterial and device models or patient-specific models with a limited number of cases, since the model construction process is tedious and time-consuming. Moreover, in contrast to retrospective studies from existing image data, there is a pressing need of prospective analysis with the goal of surgical planning. Therefore, it is necessary to construct models with implanted devices in a fast, virtual and interactive fashion. The goal of this paper is to develop new geometric methods to deploy stent grafts virtually to patient-specific models constructed from direct 3D segmentation of medical images. A triangular surface representing vessel lumen boundary is extracted from the segmentation. The diseased portion is then clipped and replaced by the surface of a virtual stent graft following the centerline obtained from the clipped portion. A Y-shape stent graft is employed in case of bifurcated arteries. A method to map a 2D strut pattern on the stent graft is also presented. We demonstrate the application of our methods to quantify wall shear stresses and forces acting on stent grafts in personalized surgical planning for endovascular treatment of thoracic and abdominal aortic aneurysms. Our approach enables prospective model construction and may help to increase its throughput required by routine clinical uses in the future.

Keywords: model construction, geometric processing, surgical planning, stent graft, aortic aneurysm, computational fluid dynamics.

1 Introduction

Image-based computational fluid dynamics, once emerged, is widely used as a valuable tool for investigating the role of local hemodynamics and the development of vascular diseases, such as atherosclerosis and aneurysms [1, 2]. Nowadays, increasing attention has been focused on applications of this tool to evaluate the safety and efficacy of vascular devices [3] and assess the outcomes of surgical procedures [4]. Considerable insights were gained by studying flow through stents in occlusive [5] and aneurysmal diseases [6] and computing fluid forces on stent grafts [7]. However,

many early studies employ idealized arterial and device models, which may be inadequate for replicating realistic hemodynamic conditions. Some recent studies have shifted to construct and analyze patient-specific models [8, 9], which provide great promises for inferring realistic conditions on an individual basis. Nevertheless, a majority of previous studies employ a limited number of cases stemming from the fact that the model construction process from 3D medical images is still tedious. Furthermore, in contrast to retrospective studies of post-operative states from existing image data, there is an obvious requirement of prospective analysis to predict the benefits or harms of a particular procedure in surgical planning.

To date, there are two major approaches to construct vascular models from medical images. One is the NURBS (Non-uniform Rational Basis Spline)-based modeling [10]. Usually, one parametric surface is reconstructed from sweeping a series of contours by segmenting in 2D over cross sectional images along the length of a vessel branch. Surfaces for multiple branches are then lofted together to represent the entire vascular tree of interest. The other is the triangle-based (sometimes polygon-based) modeling, where 3D segmentation is directly applied and triangulated into a single surface [11]. The former approach preceded the latter and remains quite useful for small vessels in images of poor contrast. However, it has been considered to be less favorable to the latter, which is more efficient, less subjective, and of more fidelity, especially for large arteries in decent-quality images. On the other hand, The NURBS-based approach yet shows great advantages in the case of surgical planning that not only arteries but also implanted devices are simultaneously modeled and ultimately combined [12]. This benefit is due to the fact that NURBS, a common representation for medical devices, offers a wealth of inherent features, such as interactive editing and Boolean operations, of which the triangle-based approach lacks. In order to fill the gap between triangle-based modeling and surgical planning, there is a pressing need for developments of new geometry processing techniques that are robust, efficient, and ideally usable in clinical settings.

In this paper, we present novel geometric methods to deploy stent grafts virtually to triangle-based patient-specific models constructed from direct 3D segmentation of medical images. Both unbranched and bifurcated stent grafts are produced for repairing thoracic and abdominal aortic aneurysms, respectively. Our approach allows the user to modify the preoperative model interactively to incorporate the effects of a virtual stent graft. We also describe a method to map a strut pattern defined on a 2D image onto the stent graft. Note we are here interested in the geometric effects of a deployed state rather than the deployment process itself, which takes account of mechanical interaction between stent graft and vessel wall. To demonstrate, we explore, using computational fluid dynamics, the wall shear stresses of pre-operative and virtual post-operative models as well as forces acting on stent grafts. They are informative to assess outcomes for endovascular treatment of aortic aneurysms.

2 Methods

The images used in this paper are from de-identified pre-existing imaging studies at Stanford University Medical Center. The image of a thoracic aorta was acquired using CT (resolution: $0.58 \times 0.58 \times 0.7$ mm, size: $512 \times 512 \times 375$). The image of an abdominal aorta was acquired using MRI (resolution: $0.78 \times 0.78 \times 1.5$ mm, size: $512 \times 512 \times 124$).

2.1 Geometric Model Construction and Centerline Extraction

After the image was preprocessed, the vessel lumen was segmented using region-based and edge-based level set methods [13]. A triangular surface mesh was extracted from the segmentation using the method proposed in [14], which produces provably good sampling and meshing. Geometric processing techniques, such as smoothing, trimming, pruning, circularization, and elongation, were performed to yield a suitable model. A method to extract the multi-branch centerline is then used. Two distance transforms of the model: distance to inlet/outlet and distance to vessel wall, were computed using the fast marching method [15]. Using the former distance, the model was divided. Centerline nodes with the maximum of the latter distance in every region were selected and connected to yield a raw centerline using adjacency relations among regions. The centerline was then refined and resampled to reach a higher quality. The detailed algorithms were reported elsewhere in [16]. Figure 1 shows the image of the abdominal aorta with an aneurysm, its geometric model and centerline.

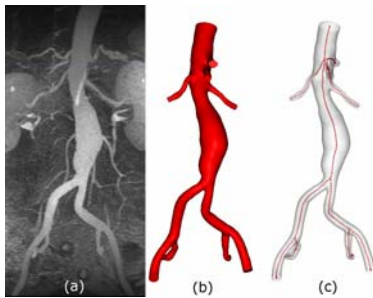


Fig. 1. An image of abdominal aorta (a), the geometric model (b), and its centerline (c).

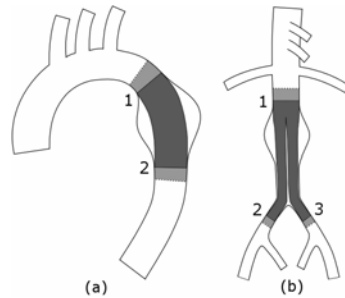


Fig. 2. Illustration of stent grafts for thoracic aortic aneurysm (a) and abdominal aortic aneurysm (b).

2.2 Virtual Stent Grafting for Thoracic Aortic Aneurysm

An illustration of the virtual stent graft for a model of thoracic aortic aneurysm is shown in Fig. 2(a). The aneurysm (the bulged region) is replaced by a tubular structure (in gray) connecting the proximal and distal portions of the aorta. The lighter gray regions depict fixation zones, where the graft is attached to the vessel wall. The requirements of the modified model are: (a) is a “water-tight” triangular mesh; (b) has the graft inside the original model and following a centerline; (c) keeps the portions other than the aneurysm intact; (d) has smooth transitions in fixation zones. We propose a novel algorithm to solve this problem:

(1) Trim the model at proximal and distal ends of the aneurysm (boundaries of light and dark gray regions at locations 1 and 2). It leads to two open contours, which are then circularized.

(2) Morph the surface at light gray regions by weighting the original surface with a perfectly tubular shape. The surface is weighted more to the original as further from the dark gray region.

(3) Extract the centerline of the portion between locations 1 and 2. Our algorithm in Section 2.1 ensures the starting and ending nodes are centers of contours in step (1).

(4) Construct a tubular structure as the stent graft by creating a series of circular triangle strips between neighboring pairs of contours, guided by the centerline from step (3). We here make the radius and number of vertices linearly varied from location 1 to 2. To establish the correspondences between vertices, a starting vertex needs to be identified on either neighboring contour. In Fig. 3, s and c denote a starting vertex and a centerline node at the current contour, respectively. $\vec{n}_{k-1} = (c_k - c_{k-1}) / \|c_k - c_{k-1}\|$ is the normal of the previous contour. \vec{x}, \vec{y} are local axes with $\vec{n} \perp \vec{x}, \vec{y}$ and $\vec{x} = (s - c) / \|s - c\|$. The starting vertex of the next contour s_k can be determined by $s_k = c_k + r_k \vec{n}_{k-1} \times \vec{x}_{k-1} \times \vec{n}_k$, where r_k is the radius. The starting vertex of the initial contour at location 1 is arbitrarily selected. The starting vertex of the final contour at location 2 is the vertex closest to that of the previous contour.

The results of our algorithm on the model of thoracic aorta are illustrated in Fig. 4.

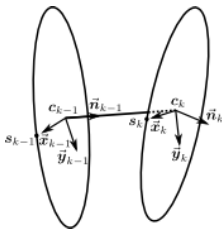


Fig. 3. Geometry of two neighboring contours to create a triangle strip.

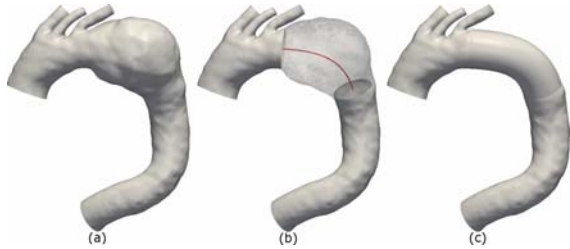


Fig. 4. Virtual stent grafting for thoracic aortic aneurysm. (a) The pre-operative model; (b) Aneurysm removal and the centerline; (c) The post-operative model.

2.3 Virtual Stent Grafting for Abdominal Aortic Aneurysm

In case of abdominal aortic aneurysm, the stent graft may be in Y shape as shown in Fig. 2(b). A special treatment of mesh shape close to the bifurcation is needed. We extend our algorithm as follows:

(1) Trim the model at the proximal end of the aneurysm (location 1) and at distal iliac arteries (locations 2 and 3). Morph the surfaces as before.

(2) Extract the centerline of the portion between locations 1, 2, and 3. Split the centerline inside the aneurysm corresponding to two iliac branches. It leads to three centerline segments

(3) Extend the surfaces along the centerline segments using tubular structures towards the bifurcation similar to before.

(4) Construct a transition surface at the bifurcation among three ending contours from last step. The transition essentially consists of a series of contours that varies from one circle in the parent vessel to two circles in the child vessels. Figure 5(a)-(d) lists four typical contours in the order from parent vessel to the child vessels. Finally, triangle strips between neighboring pairs of contours are generated to complete the transition.

The results on the model of abdominal aorta are illustrated in Fig. 6(a)-(d). Figure 6(e), (f) demonstrate the mesh at the bifurcation before and after surface transition.

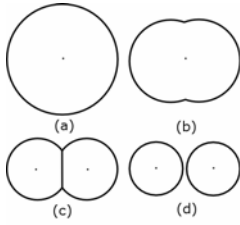


Fig. 5. Four typical contours for the surface transition at the bifurcation.

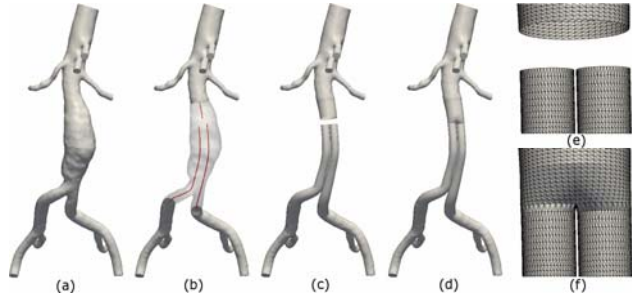


Fig. 6. Virtual stent grafting for abdominal aortic aneurysm. (a) The pre-operative model; (b) Aneurysm removal and the center-lines; (c) Surface extension; (d) The post-operative model; Mesh at the bifurcation before (e) and after (f) surface transition.

2.4 Strut Pattern Mapping

To make the stent graft more realistic, a particular strut pattern can be placed on the modified model to replicate the effects of a deployed stent graft. The designed stent pattern is usually described as a 2D image. Therefore, there is a need to map the pattern in 2D rectangular coordinate to 3D cylindrical coordinate. We assume that the struts can follow any bends in the graft, which is a first-order approximation. In addition, a typical strut width (e.g. 0.5mm) is generally smaller than the triangle size (e.g. 1.0mm) of the surface mesh, mesh refinement is necessary to represent the shape of the strut. Instead of refining the stent graft portion globally, a method to refine the mesh locally along the strut path is favorable to avoid excessively increasing the number of mesh elements and thus the computational cost. Furthermore, vertices along the strut path are displaced according to the intensity in the 2D image which characterizes specified strut thickness. Notice that strut is modeled as concavity on the surface since the stent is normally located inside the graft and the geometric model represents the channel of blood flow.

We propose a two-way algorithm that refines mesh first and then displaces vertices:

(1) Locally refine the triangles on the virtual stent graft along the strut path using a $\sqrt{3}$ -subdivision scheme [17]. To ensure a full coverage of the stent graft, we define the mapping inversely as a 3D cylindrical coordinate (r, ϕ, z) to a 2D rectangular coordinate (x, y) in image domain: $x = \frac{\phi}{2\pi}W$, $y = \frac{z}{L}H$, where W, H, L are the image width, height, and the total centerline length, respectively. A triangle is chosen to be refined if and only if its map in the image domain overlaps with the strut pattern. Since the strut pattern is sparse in the image, we first check whether there is any strut pixel in the bounding box of the triangle. If so, we further check whether the pixel region intersects the triangle.

(2) Displace vertices on the strut path radially towards the centerline by a distance depending on the intensity of its map in the 2D image. The linear interpolation of pixel intensity is used since maps are usually at subpixel locations.

(3) Smooth the mesh along the strut path using a diffusion-based method [18].

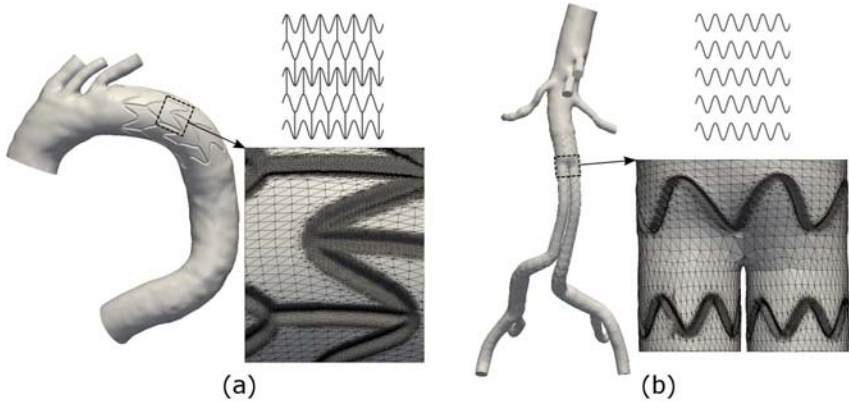


Fig. 7. Strut pattern mapping for the models of thoracic aorta (a) and abdominal aorta (b)

The results on strut pattern mapping are illustrated in Fig. 7 with given patterns and mesh details along the strut paths.

3 Simulation Results

Three models (pre-operative, post-operative without and with strut) were discretized into tetrahedral meshes for thoracic aorta (0.92M, 0.66M, 2.05M) and abdominal aorta (1.83M, 1.57M, 2.80M). Blood flow was solved using a custom stabilized finite element solver developed in our group [19]. Five cardiac cycles were simulated with assumptions of rigid wall and Newtonian fluid behavior. All results presented were from the last and converged cycle.

In Fig. 8, we show the mean wall shear stress (MWSS) of the three models, which is computed as the time-averaged amplitude of shear stress on the wall. For both cases, low MWSS are observed in the aneurysm regions. After virtual stent grafts are deployed, MWSS is higher in the same regions due to reduced calibers and increased blood velocity. This suggests stent grafts may help to reduce recirculation and their design should include the consideration of the raised levels of shear stress. In addition, strut patterns apparently influence local shear stress in vicinity. Therefore, it is necessary to include the modeling of strut to accurately quantify shear stress.

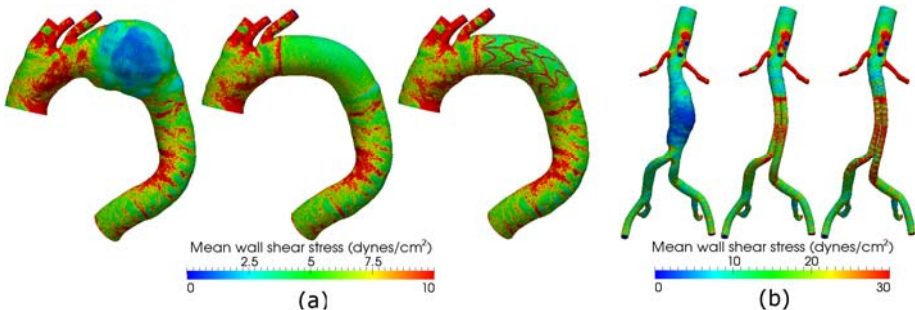


Fig. 8. Mean wall shear stress for the models of thoracic aorta (a) and abdominal aorta (b)

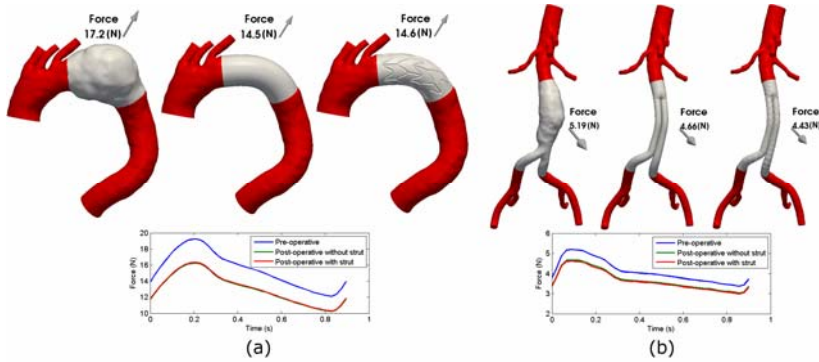


Fig. 9. Forces on aneurysm and stent graft for the models of thoracic aorta (a) and abdominal aorta (b).

The forces acting on the aneurysmal walls and stent grafts are believed to have causal effects of aneurysm rupture and stent graft migration. In Fig. 9, we show these forces on the aneurysm and stent graft (gray regions), which are calculated as the surface integral of tensile and shear forces. The arrows depict the forces at peak systole. In both cases, the forces are directed radially outwards with respect to curved paths due to the area difference between interior and exterior led by the curvature. The magnitude of the force is smaller in the stent graft than that in the aneurysm due to reduced areas. This holds true for the whole cardiac cycle as shown in the plots. The strut patterns do not have a significant impact on the magnitudes of the forces as their contributions to the surface integration are minimal.

4 Discussion and Conclusion

In this paper, we described new geometric methods to construct virtual post-operative models for simulation-based personalized planning of stent grafting of aortic aneurysms. The methods fill the gap between triangle-based modeling and prospective analysis. Our experiences show it normally takes less than ten minutes to construct the post-operative model from a pre-operative model. Our approach thus shows promises for its future use in clinical settings. We demonstrated simulations results of wall shear stress and force acting on stent grafts.

Our approach is generalizable to handle other models of vascular diseases, such as aortic coarctation, coronary stenosis, and cerebral aneurysms. Future works will focus on testing robustness and efficiency of our methods on a large number of cases and performing validations using experimental studies and some evaluations on dealing with more challenging cases. We aim at providing a user-friendly simulation tool for vascular surgeons to plan interventions.

References

1. Steinman, D.A., Taylor, C.A.: Flow Imaging and Computing: Large Artery Hemodynamics. *Annals of Biomedical Engineering* 33, 1704–1709 (2005)
2. Taylor, C.A., Steinman, D.A.: Image-Based Modeling of Blood Flow and Vessel Wall Dynamics: Applications, Methods and Future Directions. *Annals of Biomedical Engineering* (2010)

3. Zarins, C.K., Taylor, C.A.: Endovascular Device Design in the Future: Transformation from Trial and Error to Computational Design. *Journal of Endovascular Therapy* 16(suppl. 1), 112–21 (2009)
4. Taylor, C.A., Draney, M.T., Ku, J.P., Parker, D., Steele, B.N., Wang, K., Zarins, C.K.: Predictive Medicine: Computational Techniques in Therapeutic Decision-Making. *Computer Aided Surgery* 4, 231–247 (1999)
5. LaDisa, J.F., Guler, I., Olson, L.E., Hettrick, D.A., Kersten, J.R., Warltier, D.C., Pagel, P.S.: Three-Dimensional Computational Fluid Dynamics Modeling of Alterations in Coronary Wall Shear Stress Produced by Stent Implantation. *Annals of Biomedical Engineering* 31, 972–980 (2003)
6. Stuhne, G.R., Steinman, D.A.: Finite-Element Modeling of the Hemodynamics of Stented Aneurysms. *Journal of Biomechanical Engineering-Transactions of the Asme* 126, 382–387 (2004)
7. Li, Z., Kleinstreuer, C., Farber, M.: Computational Analysis of Biomechanical Contributors to Possible Endovascular Graft Failure. *Biomechanics and Modeling in Mechanobiology* 4, 221–234 (2005)
8. Fung, G.S., Lam, S.K., Cheng, S.W., Chow, K.W.: On Stent-Graft Models in Thoracic Aortic Endovascular Repair: A Computational Investigation of the Hemodynamic Factors. *Computers in Biology and Medicine* 38, 484–489 (2008)
9. Figueroa, C.A., Taylor, C.A., Chiou, A.J., Yeh, V., Zarins, C.K.: Magnitude and Direction of Pulsatile Displacement Forces Acting on Thoracic Aortic Endografts. *Journal of Endovascular Therapy* 16, 350–358 (2009)
10. Wang, K.C., Dutton, R.W., Taylor, C.A.: Improving Geometric Model Construction for Blood Flow Modeling. *IEEE Engineering in Medicine and Biology Magazine* 18, 33–39 (1999)
11. Antiga, L., Ene-Iordache, B., Remuzzi, A.: Computational Geometry for Patient-Specific Reconstruction and Meshing of Blood Vessels from MR and CT Angiography. *IEEE Transactions on Medical Imaging* 22, 674–684 (2003)
12. Wilson, N.M., Arko, F.R., Taylor, C.A.: Predicting Changes in Blood Flow in Patient-Specific Operative Plans for Treating Aortoiliac Occlusive Disease. *Computer Aided Surgery* 10, 257–277 (2005)
13. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic Active Contours. *International Journal of Computer Vision* 22, 61–79 (1997)
14. Boissonnat, J.D., Oudot, S.: Provably Good Sampling and Meshing of Surfaces. *Graphical Models* 67, 405–451 (2005)
15. Sethian, J.A.: *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge (1999)
16. Xiong, G., Figueroa, C.A., Xiao, N., Taylor, C.A.: Simulation of Blood Flow in Deformable Arteries Using Subject-Specific Geometry and Variable Vessel Wall Properties. In: *Summer Bioengineering Conference*, Lake Tahoe, CA (2009)
17. Kobbelt, L.: Sqrt 3-Subdivision. In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 103–112 (2000)
18. Desbrun, M., Meyer, M., Schröder, P., Barr, A.H.: Implicit Fairing of Irregular Meshes Using Diffusion and Curvature Flow. In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 317–324 (1999)
19. Stanford Cardiovascular Biomechanics Lab, <http://taylorlab.stanford.edu>

MRI-Guided Robotic Prostate Biopsy: A Clinical Accuracy Validation

Helen Xu¹, Andras Lasso¹, Siddharth Vikal¹, Peter Guion², Axel Krieger³, Aradhana Kaushal², Louis L. Whitcomb⁴, and Gabor Fichtinger^{1,4}

¹ Queen's University, Kingston, Canada

² National Institutes of Health, Bethesda, USA

³ Sentinelle Medical Inc., Toronto, Canada

⁴ Johns Hopkins University, Baltimore, USA

helen@cs.queensu.ca

Abstract. Prostate cancer is a major health threat for men. For over five years, the U.S. National Cancer Institute has performed prostate biopsies with a magnetic resonance imaging (MRI)-guided robotic system. *Purpose:* A retrospective evaluation methodology and analysis of the clinical accuracy of this system is reported. *Methods:* Using the pre and post-needle insertion image volumes, a registration algorithm that contains a two-step rigid registration followed by a deformable refinement was developed to capture prostate dislocation during the procedure. The method was validated by using three-dimensional contour overlays of the segmented prostates and the registrations were accurate up to 2 mm. *Results:* It was found that tissue deformation was less of a factor than organ displacement. Out of the 82 biopsies from 21 patients, the mean target displacement, needle placement error, and clinical biopsy error was 5.9 mm, 2.3 mm, and 4 mm, respectively. *Conclusion:* The results suggest that motion compensation for organ displacement should be used to improve targeting accuracy.

1 Introduction

In the United States, prostate cancer accounts for 25% of cancer incidents in men, making it the second most common cancer among American men. There was an estimated 192,280 new cases and 27,360 deaths in 2009 [1]. The two most common screening methods for prostate cancer are the prostate-specific antigen (PSA) test and the digital rectal exam (DRE). When either test shows abnormal results, needle biopsy is often recommended to determine if a tumor exists and whether it is malignant based on histological analysis.

Each year approximately 1.5 million prostate biopsies are performed and a positive case is found in every 6-8 biopsies. Transrectal ultrasound (TRUS) is currently the standard imaging modality for guiding biopsy due to its low cost and ease-of-use [2]. However, due to the poor image quality of ultrasound, TRUS only has a detection rate of 20-30% [3]. Studies have shown that this method misses the cancer in at least 20% of the cases [4].

Magnetic resonance imaging (MRI) provides an alternative approach to the detection and diagnosis of prostate cancer. It has high spatial resolution, excellent

soft tissue contrast, and volumetric imaging capabilities [5]. MRI provides clear visualization of the prostate and its substructures including the peripheral zone (PZ), which is the most common location of cancer [6]. It allows suspicious lesions to be identified and guides biopsies at these targeted sites. MRI has not been widely adopted for prostate interventions due to its strong magnetic fields, confined physical space, and high cost.

Krieger *et al.* developed an MRI-guided transrectal robotic prostate biopsy system [7] that has been used in over 200 biopsies to date at the U.S. National Cancer Institute. This paper reports a quantitative longitudinal evaluation of the clinical accuracy of this robotic biopsy system under MRI-guidance. In addition to reporting the difference between the planned and actual biopsy location, this study takes into account organ motion during the procedure, thus quantifying targeting accuracy with respect to the tissue target itself.

A much limited preliminary study was reported previously [8]. In this paper, we present major improvements to the validation framework, which include a three-stage deformable registration of an ensemble of organs and a longitudinal accuracy validation study of a much larger National Cancer Institute data set.

2 Method

2.1 Data Acquisition

During the prostate biopsy procedure, the patient was placed inside the MRI scanner in prone position to acquire a series of 2D high resolution T2 weighted axial volumetric slices of the prostate. From the pre-needle insertion volume, the clinicians selected the biopsy location(s) in RAS (Right, Anterior, Superior) coordinates, where the origin was approximately the center of the prostate. The robot was then used to guide a biopsy needle through the rectum into the target sites within the prostate to collect tissue samples. After the needle was in place, another set of 2D axial volumetric slices was taken to confirm needle placement. We used these pre and post-needle insertion image sets to validate the clinical accuracy of the robotic biopsy system.

2.2 Three-Stage Deformable Target Registration

Developing a target registration algorithm that works well for all patient data sets was a challenging task. The prostate motion upon needle insertion can be extremely complex since it is able to dislocate differently from the surrounding structures. The extent of the movement also varies from patient to patient. In addition, the data was gathered over a period of five years in different clinical trials, under different imaging protocols, by different clinicians. Variations in image resolution, field strength, etc. further increased the difficulty of the task. Our main goal was to find a method that would capture most of the prostate movement for the majority of patients during the biopsy procedure.

The pre and post-needle insertions images were examined and it was found that the main transformation between the two volumes was rigid. Although tissue

deformation may be present, it is expected to provide only minor adjustment to the rigid transformation. Karnik *et al.* have also concluded that the results from rigid and non-rigid registrations were not statistically significantly different ($p > 0.05$) in their transrectal prostate biopsies [9]. This observation is consistent with the conclusions of Misra *et al.* that boundary conditions surrounding the organ dominate the deformation more than the constitutive behavior of the tissue itself [10]. The major body structures around the prostate that are relevant to this study are the rectum and pubic bone.

A two-step 3D to 3D rigid registration was developed to capture prostate motion (Figure 1). Using the Insight Toolkit (ITK), the registration is performed between the pre and post-needle insertion volumes using mutual information. In the first step of our implementation, the pre and post-insertion volumes are used as the fixed and moving images, respectively. The region of interest consists of the rectum, prostate and pubic bone. This step compensates for prostate motion in coherence with the device and patient. To correct for residual decoupled prostate motion, the resulting image is registered again with the original fixed image using only the prostate as the region of interest. Movement in the superior and inferior direction is penalized because the first step should already have corrected for it. Finally, a B-spline deformable registration using grid size $5 \times 5 \times 5$ on the prostate is performed to serve as fine tuning. This compensates for any possible tissue deformation during needle placement.

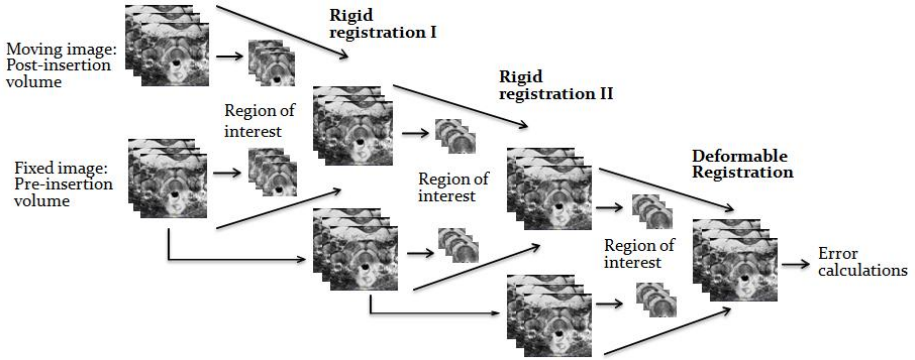


Fig. 1. Three-stage target registration between pre and post-needle insertion images

2.3 Validation of the Three-Stage Registration Scheme

The prostate usually does not show any apparent anatomical feature in the MRI images and it can move practically independently of bony structures. Therefore typical validation methods, such as using landmarks to evaluate the accuracy of the registration are not applicable.

To validate the three-stage registration, we chose to manually segment the prostate, rectum and pubic bone from both the fixed and moving image volumes using ITK-SNAP. Each segmented model was then registered manually

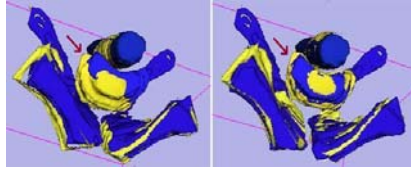


Fig. 2. Contour overlays of the segmented rectum, prostate and pubic bone from moving and fixed images before (left) and after (right) automatic three-stage registration

in 3D-Slicer by aligning the surfaces of the segmented objects in 3D. The results of contour-based prostate registration were compared with the three-stage automatic registration. The bone and rectum indicate the amount of patient movement during the procedure. Figure 2 shows the overlay of a segmented model before and after the automatic three-stage registration.

2.4 Biopsy Accuracy Calculations

Target displacement: The distance between the original and transformed target is calculated as the target displacement (Figure 3). To obtain the transformed target, transformations from the registrations are applied to the original target. To determine whether this movement is related to the needle insertion direction, the displacement is decomposed into two components: one parallel and one orthogonal to the needle vector. A Wilcoxon Signed Rank Test is conducted to see whether prostate movement in the needle direction was significantly larger than the orthogonal one.

Needle placement error: The distance from the original target to the biopsy needle trajectory line is used to represent the needle placement error (Figure 3). This is how much the robot had missed the intended target assuming no prostate motion. The needle trajectory line is obtained by using two needle tip coordinates from the post-insertion volume.

Biopsy error: The distance from the transformed target to the needle trajectory line is defined as the biopsy error (Figure 3). It represents the difference between planned and actual biopsy locations. This measurement is relevant for assessing biopsy accuracy. Since the tissue biopsy core is about 15 mm long, insertion depth is a less important factor.

3 Results

3.1 Registration Accuracy

We selected patients done at 3T that had usable biopsy needle confirmation images and original biopsy target coordinates. A total of 82 biopsies from 21 patients were evaluated. At least one biopsy for each patient was validated using manual contour-based registration method. In addition, all registrations that contained a translation of more than 10 mm were also validated. The prostate contour segmentation error was about 2 mm. When the automatic three-stage

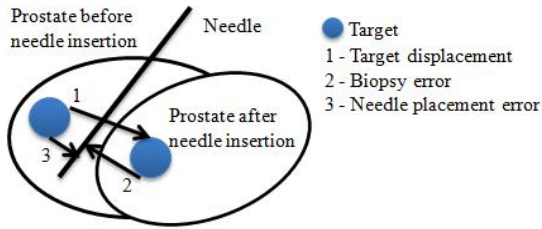


Fig. 3. Diagram illustrating prostate dislocation during needle insertion and biopsy error calculations

registration was off by 3 mm or more, results from the manual registration were used. The registration inaccuracy was mainly due to poor image quality and patient motion. A total of 11 biopsies contained patient movement that was greater than 5 mm. Following contour-based adjustments, all registrations were accurate up to 2 mm. To check the impact of deformable registration on the outcome, it was performed on 20 biopsies from various patients. A Wilcoxon Signed Rank Test was conducted, which showed that the results from deformable registration is not significantly different from the rigid one ($p = 0.54$). This is fully consistent with recent findings of Karnik *et al.* [9].

3.2 Biopsy Accuracy

Table 1 summarizes the mean, range and standard deviation for target displacement, needle placement error, and biopsy error (Figure 3) in all biopsies. To study the impact of patient movement on biopsy accuracy, the results from 11 biopsies which had more than 5 mm patient motion were grouped separately. Figure 4a-c shows the histograms of these three measurements. Lilliefors tests were conducted and it was found that only needle placement errors follow a normal distribution ($p = 0.06$). What follows is that any future needle placement error will have a 95% probability of falling between two standard deviations above or below the mean.

Table 1. The data statistics for biopsy accuracy

	Target Displacement (mm)		Needle Placement Error(mm)	Biopsy Error (mm)	
Mean	5.9	7.2*	2.3	4	4.8*
Range	1-13.4	3.7-11.2*	0.1-6.5	0.5-14.1	1.4-8.8*
Standard Deviation	3.5	2.9*	1.3	2.1	2.3
* Biopsies for patient motion > 5mm only					

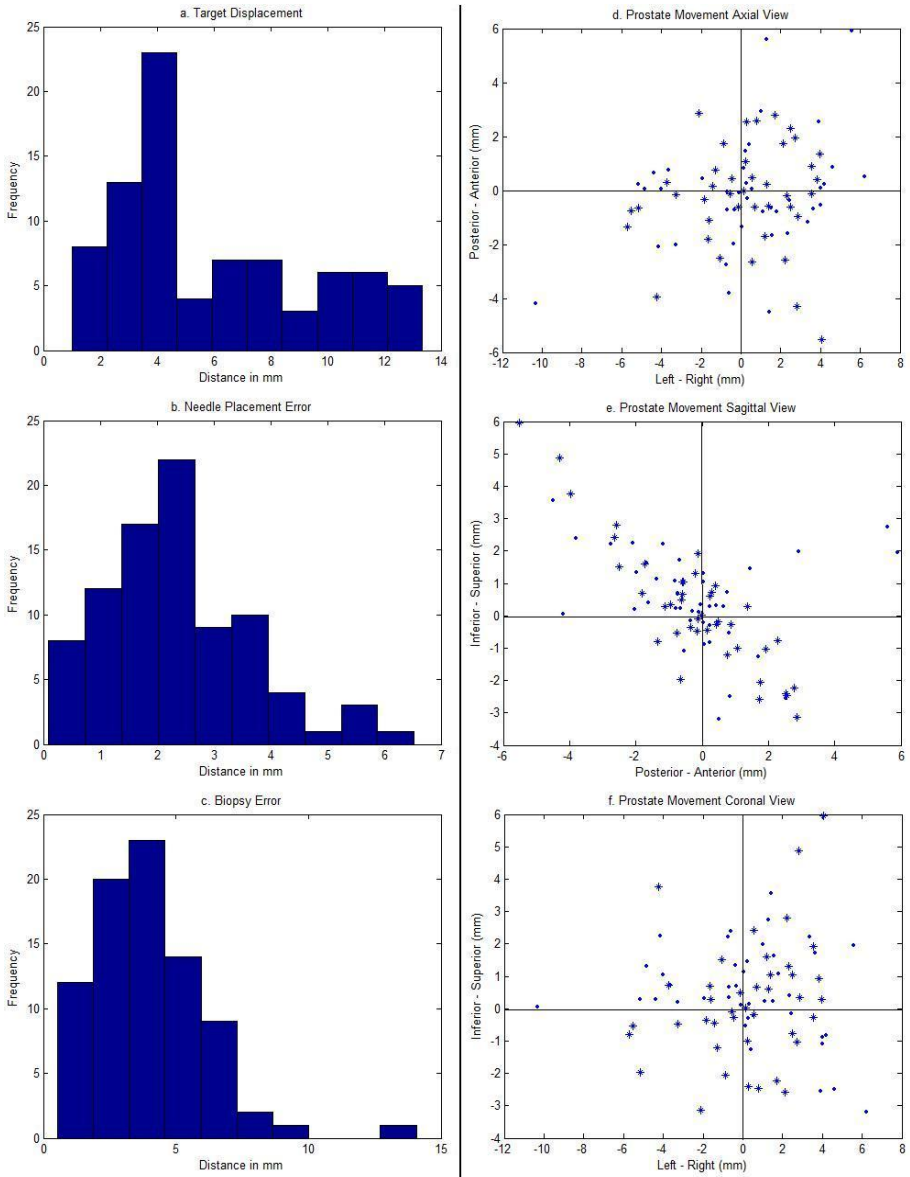


Fig. 4. Left: Histograms of target displacements (top), needle placement errors (middle), and biopsy errors (bottom) of the 82 biopsies. Right: Axial (top), sagittal (middle), and coronal (bottom) view of prostate movement orthogonal to the needle direction. ‘*’ and ‘.’ indicate biopsies taken on the left and right side of the prostate, respectively.

3.3 Target Displacement

Target displacement parallel and orthogonal to the needle direction was also computed. For the parallel component, 46% of the biopsies moved towards the

needle insertion direction (mean: 5.7 mm) and 54% went in the opposite direction (mean: 2.9 mm). The overall average was 4.2 mm in the parallel and 3.4 mm in the orthogonal direction. Results from the signed rank test showed that movement in the parallel direction was not significantly greater than the orthogonal one ($p = 0.36$). For the patient movement larger than 5 mm group, average parallel and orthogonal movement was 3.9 mm and 5.3 mm, respectively.

To analyze the orthogonal displacement component, it was further resolved into movement in RAS coordinates. 73% of the biopsies showed a target movement either towards the superior-posterior (SP) or inferior-anterior (IA) direction (Figure 4e). However, the correlation coefficient between SI and AP was only 0.56. The biopsies were also divided into two categories: left and right side biopsy, with 41 biopsies in each category. 59% of the left biopsies had a positive movement towards right, and 39% of the right biopsies had a positive movement towards left (Figure 4d).

As part of the registration validation process, the segmented rectum and pubic bone were also registered separately. We found that their movements were different from those of the prostate. However, prostate movement was more similar to the bone movement than the rectum movement.

4 Discussion

The mean needle placement error is considered as clinically acceptable since it is less than a clinically significant tumor (2.3 vs. 5 mm). The low error confirmed that the robot is accurate enough in positioning the biopsy needle to hit the intended target. However, this measurement assumes no prostate or patient motion during the procedure. In reality, these two factors usually result in some dislocation of the prostate, causing the target to move. This is evident by the 5.9 mm mean target displacement from the 82 biopsies we studied. The dislocation caused a mean biopsy error of 4 mm, which is on the verge of clinical acceptance.

The 11 biopsies which contained patient movement larger than 5 mm were studied again separately to observe the impact of patient movement on biopsy error. The slight increase in both target displacement and biopsy error suggests that fixating the patient during the procedure may help to decrease biopsy error by only about 1 mm.

The biopsy needle was inserted through the rectum towards the prostate in a mainly superior-anterior direction. It is intuitive to assume that the target should move in a direction along the needle path. Since the statistical test showed that there was no significant difference between target displacement parallel and orthogonal to the needle direction, it means that half of the displacements were in the needle direction. The other half could be due to patient movement during the procedure in addition to the impact of needle insertion.

The separate registrations of the rectum and bone showed that the prostate can move quite independently of these two structures. Since the robotic device was placed inside the patient's rectum, it can limit the rectum's ability to move,

which also explains the observation that prostate movement is more similar to movement of the bone than that of the rectum.

In conclusion, based on validation with segmented prostate contours, our registration algorithm captures the prostate motion during a biopsy with an accuracy of 2 mm. The non-significant impact of deformable registration on the final refinement stage indicates that prostate deformation is less of a factor than organ displacement during the needle placement process. We also found that the pre-planned biopsy target dislocated during the procedure and the prostate motion does differ from both the patient and robot motions. The exact amounts of these motions cannot be known without prostate fiducials or finer volume images. However, even taking into account the imperfection of the segmentation-based validation approach, the results still suggest that further research in organ motion and prostate tracking will be useful to reduce MRI-guided biopsy targeting error. From a sufficient number of biopsy error observations, a statistical model might be built to predict prostate movement during needle placement. Such a model could be used as a reference for clinicians to compensate the insertion plan for predicted movement prior to needle insertion on a prospective patient.

Acknowledgments. We thank C. Menard, A. Singh, J.A. Coleman, R.L. Grubb, J.B. Latouf, and P. Pinto for clinical data collection at the U.S. National Institutes of Health. This work is supported by NIH 5R01CA111288-04 and 5R01EB002963-05.

References

1. Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Thun, M.: Cancer statistics, 2009. *CA Cancer J. Clin.* 59(4), 225–249 (2009)
2. Presti Jr., J.: Prostate cancer: Assessment of risk using digital rectal examination, tumor grade, prostate-specific antigen, and systematic biopsy. *Radiol. Clin. North Amer.* 38(1), 49–58 (2000)
3. Terris, M., Wallen, E., Stamey, T.: Comparison of mid-lobe versus lateral systematic sextant biopsies in detection of prostate cancer. *Urol. Int.* 59, 239–242 (1997)
4. Wefer, A., Hricak, H., Vigneron, D., Coakley, F., Lu, Y., Wefer, J., Mueller-Lisse, U., Carroll, P., Kurhanewicz, J.: Sextant localization of prostate cancer: comparison of sextant biopsy, magnetic resonance imaging and magnetic resonance spectroscopic imaging with step section histology. *J. Urol.* 163(2), 400–404 (2000)
5. Susil, R., Ménard, C., Krieger, A., Coleman, J., Camphausen, K., Choyke, P., Fichtinger, G., Whitcomb, L., Coleman, C., Atalar, E.: Transrectal prostate biopsy and fiducial marker placement in a standard 1.5T magnetic resonance imaging scanner. *J. Urol.* 175(1), 113–120 (2006)
6. Adusumilli, S., Pretourius, E.: Magnetic resonance imaging of prostate cancer. *Semin. Urol. Oncol.* 20, 192–210 (2002)
7. Krieger, A., Susil, R., Menard, C., Coleman, J., Fichtinger, G., Atalar, E., Whitcomb, L.: Design of novel MRI compatible manipulator for image guided prostate interventions. *IEEE Trans. Biomed. Eng.* 52(2), 295–304 (2008)

8. Xu, H., Lasso, A., Vikal, S., Guion, P., Krieger, A., Kaushal, A., Whitcomb, L., Fichtinger, G.: Accuracy validation for MRI-guided robotic prostate biopsy. In: SPIE Medical Imaging: Visualization, Image-Guided Procedure, and Modeling, vol. 7625, pp. 762517–762517–8 (2010)
9. Karnik, V., Fenster, A., Bax, J., Cool, D., Gardi, L., Gyacskov, I., Romagnoli, C., Ward, A.: Assessment of registration accuracy in three-dimensional transrectal ultrasound images of prostates. In: SPIE Medical Imaging: Visualization, Image-guided Procedures and Modeling, vol. 7625, pp. 762516–762516–8 (2010)
10. Misra, S., Macura, K., Ramesh, K., Okamura, A.: The importance of organ geometry and boundary constraints for planning of medical interventions. *Medical Engineering and Physics* 31(2), 195–206 (2009)

Online 4-D CT Estimation for Patient-Specific Respiratory Motion Based on Real-Time Breathing Signals

Tiancheng He^{1,2}, Zhong Xue^{1,*}, Weixin Xie², and Stephen T.C. Wong¹

¹The Center for Bioengineering and Informatics, The Methodist Hospital Research Institute and Department of Radiology, The Methodist Hospital, Weil Cornell Medical College, Houston, TX

²Intelligent Information Institute, Shenzhen University, Shenzhen, China

zxue@tmhs.org

Abstract. In image-guided lung intervention, the electromagnetic (EM) tracked needle can be visualized in a pre-procedural CT by registering the EM tracking and the CT coordinate systems. However, there exist discrepancies between the static pre-procedural CT and the patient due to respiratory motion. This paper proposes an online 4-D CT estimation approach to patient-specific respiratory motion compensation. First, the motion patterns between 4-D CT data and respiratory signals such as fiducials from a number of patients are trained in a template space after image registration. These motion patterns can be used to estimate the patient-specific serial CTs from a static 3-D CT and the real-time respiratory signals of that patient, who do not generally take 4-D CTs. Specifically, the respiratory lung field motion vectors are projected onto the Kernel Principal Component Analysis (K-PCA) space, and a motion estimation model is constructed to estimate the lung field motion from the fiducial motion using the ridge regression method based on the least squares support vector machine (LS-SVM). The algorithm can be performed onsite prior to the intervention to generate the serial CT images according to the respiratory signals in advance, and the estimated CTs can be visualized in real-time during the intervention. In experiments, we evaluated the algorithm using leave-one-out strategy on 30 4-D CT data, and the results showed that the average errors of the lung field surfaces are 1.63mm.

Keywords: Image-guided intervention, respiratory motion, 4-D CT, K-PCA, least squares support vector machine.

1 Introduction

Image-guided intervention has been widely used in different procedures such as lung, liver, and kidney intervention, bronchoscopy, and endovascular interventions. In percutaneous lung intervention, traditional systems use a static pre-procedural CT for guidance and might generate larger errors due to the poor reproducibility of breath-holding and the dynamic deformation of lung parenchyma during respiratory cycles. The problem is further multiplexed by the simultaneous deformation of airways, bronchi and vascular structures. Therefore, a precise prediction of the movement of

the lung or acquiring 4-D CT would be highly desirable for more accurate localization. For percutaneous lung intervention, 4-D CT acquisition is normally not performed in major hospitals. Therefore, it is necessary to simulate the serial CT images from a 3-D CT of that patient based on the real-time tracked signals such as the respiratory belt or the fiducial signals on the chest.

Although complex biomechanical models [1] can be constructed to model the lung motion, image processing is still necessary to adapt these models to a patient. In addition, most biomechanical models need patient-specific 4-D data [2], which means that in a clinical scenario, motion information can only be incorporated if the 4-D images of the patient are additionally acquired. Little work has been presented for building motion models based on the extracted motion fields from 4-D data without considering biomechanical tissue properties. Sundaram et al. [3] first classified the serial images from free-breathing lung MR scans according to the normalized lung capacity and then registered the images with similar capacities to create a dynamic model of average lung deformation. However, this method only constructs the respiratory template dynamics and does not discuss motion modeling and serial image estimation. Motion model adaptation of an inter-subject model using sparse motion information had been recently presented for liver deformation to predict the drift of the exhalation position of corresponding points inside the liver [4]. Daniel A. Low et al. [5, 6] did the work for using 4D-CT lung registration to estimate the 4-D CT lung field motion.

In this paper, we propose a framework for simulating patient-specific serial images based on the 3-D data and the real-time breathing signals of the patient. First, the 4-D CT data from a number of subjects are captured and aligned onto a template space according to their baselines (the first timepoint image), and K-PCA [7] is then applied on the lung field motion vectors derived from the extracted lung field surfaces to construct the nonlinear statistics of lung motion. To establish the relationship between the fiducials' movement and lung surfaces' motion in the template space, the ridge regression method using the least squares support vector machine (LS-SVM) [8] is employed. Then, during the intervention, a 3-D CT and the real-time tracking signal of fiducials and respiratory belts of the patient will be available. Thus, we can use the trained motion estimation model to estimate the lung field motion from the real-time fiducial signals. This motion estimation can be performed in the template space. Finally, the estimated lung field motion can be used to generate serial deformations and serial images for that patient. The estimated serial images can be pre-calculated right after obtaining the intra-procedure CT and before the intervention in order to visualize them in real-time during the intervention.

In experiments, we used thirty 4-D CT data to construct the motion estimation model and applied the leave-one-out method to validate the algorithm. The results showed that the average differences between the simulated data and the real data (lung field surfaces) are 1.63mm. Our future work is to simulate the 4-D CT under breath holding and to embed the 4-D CT estimation package in our image-guided intervention system under development [9].

2 Method

2.1 The Framework for 4-D CT Estimation

Fig. 1 illustrates the framework of the proposed 4-D CT estimation algorithm. The pre-processing consists of lung field segmentation, serial image registration for lung motion estimation, and registration of the first timepoint images of different subjects onto a template image. In the training stage, the normalized lung field surface motion vectors and the corresponding fiducial motion vectors of each subject are extracted. Kernel-PCA is then performed on the surface motion vectors to construct the lung motion statistical model and reduce the dimensionality of surface points. Then, a lung motion estimation model is trained using the LS-SVM algorithm to model the relationship between the fiducial signals and the lung motion feature vectors projected on the K-PCA space. In the estimation stage, an intra-procedural 3-D CT and the real-time tracked fiducial signals of a patient are available. The respiratory signals of that patient can be transferred onto the template space in order to use the motion estimation model to estimate the lung motion feature vectors and to reconstruct the lung motion vectors (surface motion vectors) of the patient. Serial deformations can be generated by using the surface motion vectors as constraints in the serial deformation simulator [10]. These serial deformations are finally transformed onto the subject space to generate the serial CTs for online visualization during intervention.

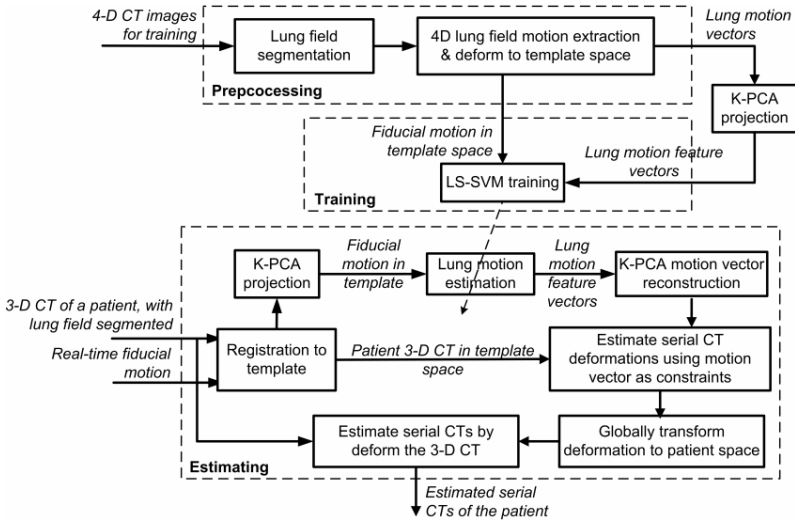


Fig. 1. The framework of the proposed 4-D CT estimation algorithm

2.2 Motion Field Extraction

Lung field segmentation is a critical pre-processing step in motion field extraction to limit the motion modeling step within the thorax field and for better visualization. For each subject, we applied a joint segmentation and registration algorithm [11] and extracted the lung field by first removing the background and the cavity areas using

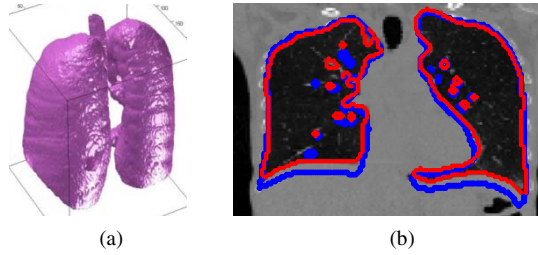


Fig. 2. An example of the lung field surface and the lung field shape changes

region growing and then performing 3-D morphological operations to clean up the segmented lung field. By registering the baseline image of each subject onto the template space, the surfaces of the lung field from different subjects and their motion vectors can be deformed onto the template space. In our study, serial CT images from thirty patients, each with twelve timepoints were used. For each subject the baseline image is the exhale image and the 7th image is the inhale image, and the 12th image is the exhale image again. Fig. 2(a) shows an example of a lung field extracted, and Fig. 2(b) shows the surfaces extracted from the 7th image before (blue) and after (red) registered to the 1st image (the 1st image is shown as the background). It can be seen that the deformable registration tracked respiratory motion well.

2.3 Motion Modeling Using Kernel-PCA

K-PCA is a nonlinear statistical modeling method and can capture the variations of shapes more accurately than PCA. The basic idea of K-PCA is that PCA computed in a high-dimensional implicit mapping function $\phi(\mathbf{v})$, or the feature space, of the surface motion vector \mathbf{v} can be replaced by a PCA of the kernel matrix. Let K denotes the kernel matrix of N sample surface motion vectors, i.e., $k_{i,j} = k(\mathbf{v}_i, \mathbf{v}_j)$, K-PCA can be computed in a closed form by finding the first M eigenvalues \mathbf{v}_i and eigenvectors \mathbf{a}_i of K , i.e., $KA = AV$. The corresponding eigenvectors in the feature space can be computed by multiplying the mapping function values of the samples with A , and they preserve the variance of data in the feature space. Therefore, given a surface motion vector \mathbf{v} , it can be projected onto the K-PCA space as,

$$\lambda = A^T (\mathbf{k} - \bar{\mathbf{k}}), \quad (1)$$

where $\bar{\mathbf{k}}$ is the mean of the kernel vectors, and \mathbf{k} is the kernel vector of \mathbf{v} , i.e., $k_i = k(\mathbf{v}, \mathbf{v}_i), i = 1, \dots, N$. Because in K-PCA the feature space is induced implicitly, reconstruction of a new vector \mathbf{v} given a feature λ is not trivial. Many methods were proposed for the K-PCA reconstruction, and different cost functions could lead to different optimization problems. In this work, we used Kwok & Tsang's algorithm for reconstruction [12].

2.4 Motion Prediction Modeling Using LS-SVM

The goal of motion estimation is to establish the relationship between the lung field surface motion \mathbf{v} (represented by λ in the K-PCA space) with the fiducials' motion

$\mathbf{v}^{(d)}$. Given N training sample-pairs $\{(\mathbf{v}_i^{(d)}, \lambda_i)\}, i=1, \dots, N$, the relationship between fiducial $\mathbf{v}_i^{(d)}$ and lung field motion feature vector λ_i needs to be established. In this work, we employ the ridge regression method with the LS-SVM model. Given the time series of the motion vectors $\lambda_{i,t}, i=1, \dots, N; t=1, \dots, T$ and those of the fiducial motion vectors $\mathbf{v}_{i,t}^{(d)}$, the goal is to estimate the motion estimation function, i.e., $\lambda(t) = \theta(\mathbf{v}(t)) + \mathbf{e}(t)$, where \mathbf{e} is a random process with zero mean and $\text{std } \sigma_e^2$. Because the elements of $\lambda(t)$ are independent each other in the K-PCA space, we can use the LS-SVM model to estimate each element of $\lambda(t)$. Denoting λ as one element of λ at time t , we can estimate it using:

$$\lambda = \mathbf{w}^T \phi(\mathbf{v}^{(d)}) + b, \tag{2}$$

where $\phi()$ denotes a potential mapping function. \mathbf{w} is the weighting vector, and b is the shifting vector. The regularized cost function of the LS-SVM is given by [8],

$$\min_{\mathbf{w}, \mathbf{b}, \mathbf{e}} \xi(\mathbf{w}, \mathbf{e}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^N \|e_i\|^2 \tag{3}$$

s.t. $\lambda_i = \mathbf{w}^T \phi(\mathbf{v}_i^{(d)}) + b + e_i, i=1, \dots, N.$

γ is referred to as the regularization constant. This optimization actually corresponds to a ridge regression in feature space. The Lagrangian method is utilized to solve the constrained optimization problem, and hence the new cost function becomes:

$$\zeta(\mathbf{w}, b, \mathbf{e}; \boldsymbol{\alpha}) = \xi(\mathbf{w}, \mathbf{e}) - \sum_{i=1}^N \alpha_i (\mathbf{w}^T \phi(\mathbf{v}_i^{(d)}) + b + e_i - \lambda_i), \tag{4}$$

with α_i as the Lagrange multipliers. According to [8], the conditions for optimality are equivalent to the following linear equation:

$$\begin{bmatrix} 0 & \mathbf{1}_N^T \\ \mathbf{1}_N & \Omega + \gamma^{-1} I_N \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \Lambda \end{bmatrix}, \tag{5}$$

where $\Lambda = [\lambda_1, \dots, \lambda_N]^T$ is the vector formed by the N samples of an element of vector λ , $\mathbf{1}_N = [1, \dots, 1]^T \in \mathbf{R}^N$, $\Omega_{i,j} = \Pi(\mathbf{v}_{i,t}^{(d)}, \mathbf{v}_{j,t}^{(d)}) = \phi(\mathbf{v}_{i,t}^{(d)})^T \phi(\mathbf{v}_{j,t}^{(d)}) \forall i, j=1, \dots, N$ with Π as the positive definite kernel function. Notice that because of the kernel trick, the feature mapping $\phi()$ is never defined explicitly, and we only need to define a kernel function $\Pi(\cdot, \cdot)$ of the fiducial vectors. The typical radial basis function (RBF) kernel $\Pi(\mathbf{v}_d^i, \mathbf{v}_d^j) = \exp(-\|\mathbf{v}_d^i - \mathbf{v}_d^j\|^2 / \sigma^2)$ is used in our study, where σ denotes the bandwidth of the kernel. After solving Eq.(5), we get $\boldsymbol{\alpha}$ and b , and the element of lung motion feature vector λ can be calculated for given fiducial motion vector $\mathbf{v}^{(d)}$:

$$\lambda = \sum_{i=1}^N \alpha_i \Pi(\mathbf{v}^{(d)}, \mathbf{v}_i^{(d)}) + b. \tag{6}$$

Notice that because different elements of the lung motion feature vector λ are independent, all of the elements of λ at different timepoints are calculated by this model separately, similar to model the motion according to different lung capacity.

3 Results

We used thirty 4-D CT datasets from 30 different patients in the experiments. Twelve 3-D images were acquired for each patient, and the images were aligned so that the first and the last images were exhale data and the 7th data was the inhale data. All the images have an in-plane resolution of 0.98x0.98mm and a slice thickness of 1.5mm. To ensure that we can get a consistent lung field surface representation, one image was randomly selected as the template, and its lung field surface was constructed first. Then, we applied the image segmentation and registration program to deform this lung field surface onto all the other images. In this way, we can obtain the lung field surface correspondences across different subjects and different timepoints to ensure the surfaces have the same trajectory. Using the same strategy, nine artificial fiducials were automatically put on the surface of the chest/belly of each CT image. In our system the EM-guidance is used to track the interventional needle, thus it is valuable to show the feasibility of the motion estimation using EM-tracked fiducials. We used the leave-one-out method to validate the proposed algorithm. Each time the baseline images from twenty eight subjects were registered onto the template image for training the lung motion estimation model. Then, the baseline image of the left-out subject and the fiducial movement signals of this subject were used to estimate the serial CT images.

Without loss of generality, suppose the baseline image of the patient to be tested is $I_1^{(p)}$, we can first register it onto the template image T using deformable registration $\Phi_{T-p} : T \rightarrow I_1^{(p)}$, and $\Phi_{T-p} = \{G, \mathbf{f}\}$ consists of both global G and deformable \mathbf{f} components of the registration. The corresponding lung field surface of the patient $v_1^{(p)}$ can also be aligned onto the template as v_1 . Similarly, the fiducial movement $v_i^{(d,p)}$ can be aligned onto the template space, denoted as $v_i^{(d)}$. Notice global transformation G needs to be applied to the fiducial motion because we are dealing with different spaces. We can then use Eq. (6) to estimate the serial lung motion feature vectors and reconstruct the lung field motion from K-PCA space to the template image space, denoted as $v_t, t = 2, \dots, T$. A lung field motion vector-constrained deformation simulation method is applied to generate the serial deformation fields [13]. Finally, the deformations are transformed onto the subject space using Φ_{T-p} . The values of these deformation vectors are subject to the global transformation G also.

The errors between the estimated lung field surface and the actual surface at each time point as well as the volumes of the lung fields were used to evaluate the accuracy of the estimation. The procedure was iterated 28 times with one subject left out each time after selecting one image as the template. The following equation was used to calculate the prediction errors for lung field surfaces:

$$\Delta_i = \frac{1}{28} \sum_{28 \text{ tests}} \text{dist}(\mathbf{v}, \hat{\mathbf{v}}), \text{ subject } i \text{ is left out.} \quad (7)$$

The distance between two surfaces is defined as the average of distances from all the points in one surface to the other surface [14]. Another quantitative measure is the volume of the lung field. Because the lung fields from both estimated CT and the original CT images are available, we simply calculated the lung volumes and compare whether they are close. The algorithm was implemented on an HP workstation with 4 Inter 2.5 GHz dual core Quad Q9300 processors and 4 GB RAM. The runtime for

training the 4D model using 30 samples was around 1 hour, and the runtime for testing each case was less than 1 minute.

Fig. 3 illustrates some examples of the results. From the second row we can see that the estimated lung field and the actual lung field match very well, and this can also be seen from the plots of the lung field volumes in the third row. In Table 1, we list the average lung field estimation errors for 8 experiments, and each result is evaluated on the left-out subject image. For the 8 tests, the maximum amount of lung motion (on the lung field surface) is 10.20cm, the minimum one is 0.048cm, and the average lung motion is 4.78cm. It can be seen that the average errors over the serial images are between 1.22mm and 2.18mm with an average of 1.63mm. If we look at the errors at different timepoints in detail, the largest errors happened at timepoint 7, which is the inhale stage due to relatively larger movement. Overall, an acceptable range of errors were obtained for predicting the lung motion. Our future work will focus on determining and tracking the location of the lung tumor.

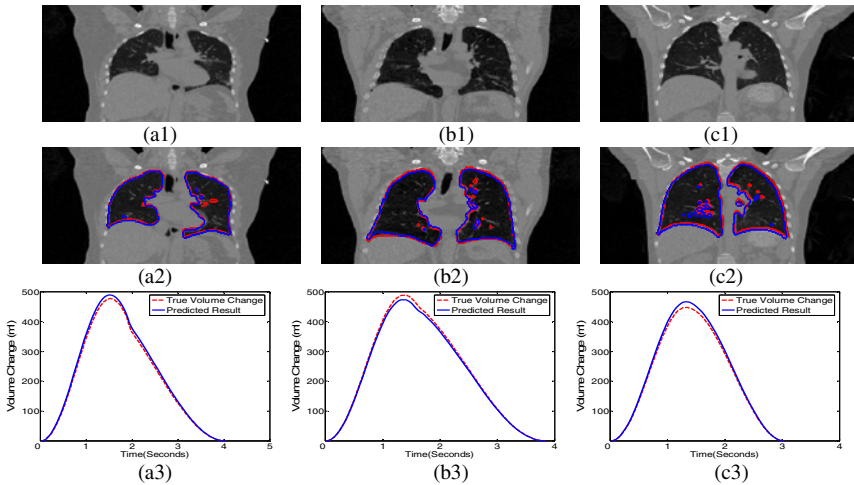


Fig. 3. Examples of the motion estimation results. Row 1: the exhale images; Row 2: the corresponding inhale images, blue contours denote the predicted lung field, and red contours are the actual lung field position; Row 3: predicted and actual changes of lung volumes.

Table 1. Average errors for lung motion estimation using leave-one-out validation (unit in mm)

Time	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	Mean
Pat. 1	1.20	1.39	1.61	1.64	2.69	3.49	2.75	2.75	2.88	1.94	1.63	2.18
Pat. 2	0.98	1.14	1.31	1.34	2.20	2.85	2.25	2.25	2.36	1.59	1.33	1.78
Pat. 3	1.08	1.25	1.44	1.47	2.42	3.14	2.47	2.48	2.59	1.75	1.46	1.96
Pat. 4	0.83	0.96	1.11	1.13	1.86	2.42	1.90	1.91	1.99	1.34	1.13	1.51
Pat. 5	0.90	1.04	1.20	1.23	2.01	2.62	2.06	2.06	2.16	1.46	1.22	1.63
Pat. 6	0.77	0.89	1.03	1.05	1.73	2.24	1.77	1.77	1.85	1.25	1.05	1.40
Pat. 7	0.67	0.78	0.90	0.92	1.51	1.96	1.54	1.55	1.62	1.09	0.92	1.22
Pat. 8	0.72	0.83	0.96	0.98	1.61	2.09	1.65	1.65	1.73	1.16	0.98	1.31

4 Conclusion

We proposed an online 4-D CT image estimation approach to patient-specific respiratory motion compensation. The idea is that a motion estimation model is first trained

using a number of 4-D CTs from different subjects. Then, this model can be used to simulate serial CTs if a 3-D image and the real-time tracked fiducial signals of a patient are given. Leave-one-out validation results from 30 4-D CT data showed the accuracy of the proposed algorithm. Our future work includes simulating 4-D CT under breath holding and implementing the algorithm in image-guided intervention.

Acknowledgement

The authors would like to thank Dr. Daniel Low and Dr. Wei Lu from Washington University in ST. Louis for providing the 4-D datasets.

References

1. McClelland, J.R., Blackall, J.M., Tarte, S., et al.: A Continuous 4D Motion Model from Multiple Respiratory Cycles for Use in Lung Radiotherapy. *Medical Physics* 33, 3348–3358 (2006)
2. Zhang, Q., Pevsner, A., Hertanto, A., Hu, Y.C., Rosenzweig, K.E., Ling, C.C., Mageras, G.S.: A Patient-Specific Respiratory Model of Anatomical Motion for Radiation Treatment Planning. *Medical Physics* 32, 4772–4782 (2007)
3. Sundaram, T.A., Avants, B.B., Gee, J.C.: A Dynamic Model of Average Lung Deformation Using Capacity-Based Reparameterization and Shape Averaging of Lung MR Images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3217, pp. 1000–1007. Springer, Heidelberg (2004)
4. Von Siebenthal, M., Szkely, G., Lomax, A., Cattin, P.: Inter-Subject Modeling of Liver Deformation During Radiation Therapy. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 659–666. Springer, Heidelberg (2007)
5. Lu, W., Song, J.H., Christensen, G.E., Parikh, P.J., Zhao, T., Hubenschmidt, J.P., Bradley, J.D., Low, D.A.: Evaluating Lung Motion Variations in Repeated 4D CT Studies Using Inverse Consistent Image Registration. *International Journal of Radiation Oncology Biology Physics* 66, S606–S607 (2006)
6. Yang, D.S., Lu, W., Low, D.A., Deasy, J.O., Hope, A.J., El Naqa, I.: 4D-CT Motion Estimation Using Deformable Image Registration and 5D Respiratory Motion Modeling. *Medical Physics* 35, 4577–4590 (2008)
7. Kim, K.I., Franz, M.O., Scholkopf, B.: Iterative Kernel Principal Component Analysis for Image Modeling. *IEEE Trans. on Patt. Anal. and Mach. Intell.* 27, 1351–1366 (2005)
8. An, S.J., Liu, W.Q., Venkatesh, S.: Fast Cross-Validation Algorithms for Least Squares Support Vector Machine and Kernel Ridge Regression. *Pattern Recognition* 40, 2154–2162 (2007)
9. He, T., Xue, Z., Wong, K., Valdivia y Alvarado, M., Zhang, Y., Xie, W., Wong, S.T.C.: Minimally Invasive Multimodality Image-Guided (MIMIG) Molecular Imaging System for Peripheral Lung Cancer Intervention and Diagnosis. In: Navab, N., Jannin, P. (eds.) *IPCAI 2010*. LNCS, vol. 6135, Springer, Heidelberg (2010)
10. Xue, Z., Shen, D.G., Davatzikos, C.: CLASSIC: Consistent Longitudinal Alignment and Segmentation for Serial Image Computing. *Neuroimage* 30, 388–399 (2006)
11. Xue, Z., Wong, K., Wong, S.T.C.: Joint Registration and Segmentation of Serial Lung CT Images for Image-Guided Lung Cancer Diagnosis and Therapy. *Computerized Medical Imaging and Graphics* 34, 55–60 (2010)
12. Kwok, J.T.Y., Tsang, I.W.H.: The Pre-Image Problem in Kernel Methods. *IEEE Transactions on Neural Networks* 15, 1517–1525 (2004)
13. Xue, Z., Shen, D., Karacali, B., Stern, J., Rottenberg, D., Davatzikos, C.: Simulating Deformations of MR Brain Images for Validation of Atlas-Based Segmentation and Registration Algorithms. *Neuroimage* 33, 855–866 (2006)
14. Gerig, G., Jomier, M., Chakos, M.: Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation. In: Niessen, W.J., Viergever, M.A. (eds.) *MICCAI 2001*. LNCS, vol. 2208, pp. 516–528. Springer, Heidelberg (2001)

Modeling and Segmentation of Surgical Workflow from Laparoscopic Video

Tobias Blum¹, Hubertus Feußner², and Nassir Navab¹

¹ Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany

² Department of Surgery, Klinikum Rechts der Isar, Technische Universität München, Germany

Abstract. Modeling and analyzing surgeries based on signals that are obtained automatically from the operating room (OR) is a field of recent interest. It can be valuable for analyzing and understanding surgical workflow, for skills evaluation and developing context-aware ORs. In minimally invasive surgery, laparoscopic video is easy to record but it is challenging to extract meaningful information from it. We propose a method that uses additional information about tool usage to perform a dimensionality reduction on image features. Using Canonical Correlation Analysis (CCA) a projection of a high-dimensional image feature space to a low dimensional space is obtained such that semantic information is extracted from the video. To model a surgery based on the signals in the reduced feature space two different statistical models are compared. The capability of segmenting a new surgery into phases only based on the video is evaluated. Dynamic Time Warping which strongly depends on the temporal order in combination with CCA shows the best results.

1 Introduction

Automatic analysis of surgical workflow is an important topic for assessment of surgical skills, analysis of surgical workflow and intelligent systems that need to be aware of the current state of an ongoing surgery. Work in this area usually involves signals that can be obtained in an automatic way. This can be video images, information about tools that are currently used, signals from robotic systems or additional sensors like force sensors, that are installed on surgical tools. These signals are used as input data for machine learning techniques or statistical modeling. For surgical skills assessment often simulators are used, where it is possible to attach sensors to tools or phantoms [1] or use tracking systems to record motion [2]. Other work uses signals from surgical robots [3] where sensors are often built-in and the data is easily accessible. However, in non-robotic surgery, the acquisition of signals is a more challenging problem. In laparoscopic surgery, video images are one important source of information. In [4] instrument segmentation and tracking, tissue deformation and changes in specular highlights are detected from laparoscopic video. This data has been used to classify four different states. In [5] five different laparoscopic tools have been recognized

based on color and shape using a stereo endoscope in a simulated setup. Both methods have not been used on whole surgeries, where a lot of different instruments are used that often only have subtle differences. Recognition of surgical phases for a whole surgery has been shown in [6], where video was used to detect the presence of surgical clips and whether the endoscopic camera is inserted or not. However, additional information about the use of instruments has been used, which has been obtained manually. In [7] four operating room states have been detected from a video camera mounted on the ceiling of an OR. In a simulate setup a more fine grained detection of surgical workflow has been shown by [8] using nine external cameras. In this work we present an approach to detect phases of a full real minimally invasive surgery (MIS) only from laparoscopic video. Instead of training classifiers for specific instruments, we use a supervised dimensionality reduction on simple image features. By using additional data for the dimensionality reduction, we extract features from the video that contain semantic information. First we will describe the signals that are used, then we will discuss statistical modeling based on these signals.

2 Method

2.1 Signals

The method we describe is applicable to every kind of laparoscopic surgery. For the experiments we used data from a laparoscopic cholecystectomy. This is a very common surgery that is performed minimally invasive in most of the cases. The surgery has a fixed workflow that we have split up into 14 phases that occurred in every instance of the surgery. Especially in MIS the workflow is strongly correlated with the instruments that are used. For every phase the ending point has been defined based on the use of a certain instrument or a combination of instruments. The starting point of a phase corresponds to the ending point of the previous phase. We have recorded the laparoscopic video and additional video from external cameras for ten surgeries. We present a method that tries to detect the 14 phases only based on the laparoscopic video. The external videos and the laparoscopic video have been used to manually annotate which instrument is used at which time. The information about the instrument use is taken for the dimensionality reduction that is described below but not for the detection itself. There are 17 different signals that have been obtained for every surgery. Most of them represent the use of surgical instruments. But also high-frequency coagulation and cutting, which is performed by applying current to an instrument, and the information which trocar is placed, are used. Every surgery i is represented by a multidimensional time series $O_i \in \mathbb{R}^{17 \times l_i}$, where l_i is the length of operation i in seconds. While the values of O_i will only be 0 and 1 in our case, we use \mathbb{R} in the formula, as the described methods are also applicable to real valued data without modifications. Such a representation of the surgical workflow by instrument vectors has been used before for segmenting a surgery into phases by [9]. An example of these signals can be seen in figure 1.

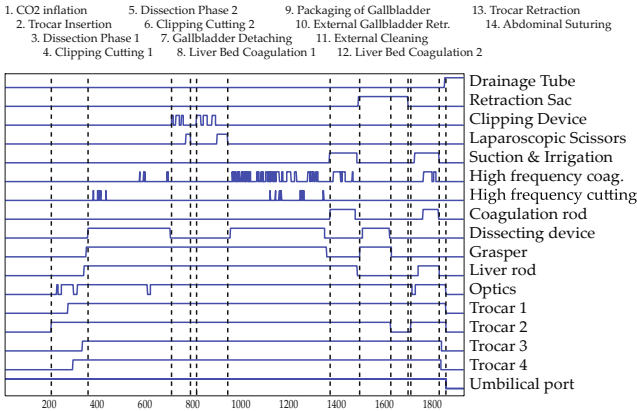


Fig. 1. The instrument use over time during one exemplary surgery. The time is given in seconds and the dotted lines indicate the phases.

From the video images a range of simple image features are computed for every image. The features are horizontal and vertical gradient magnitudes, histograms and the pixel values of a 16x16 version of the image. All of these features have been computed for all three RGB and all three HSV channels, resulting in a 1932-dimensional feature vector for each image. Sampling the features at 1 Hz we obtain the time series $V_i \in \mathbb{R}^{1932 \times l_i}$ for every surgery i .

Most machine learning methods do not perform well with high dimensional feature spaces. There are several ways to deal with this problem. One way is to design classifiers that detect certain instruments or aspects of a surgery as for example done by [45]. When developing such a classifier, the feature space is usually reduced by manually choosing features that work well for a certain instrument. While these methods work, it is tedious to design them and often they are only applicable to one certain kind of procedure. Other approaches that have been used in the domain of workflow analysis are unsupervised dimensionality reduction methods like PCA [3] which performs dimensionality reduction in a way to maintain the maximum variability in the data or feature weighting methods like Boosting [10] which select features based on their capability to discriminate between two classes. We use another approach that makes use of the additional information about the use of instruments. In contrast to the features that are extracted from the video, the instruments have an obvious strong semantic meaning. By using Canonical Correlation Analysis (CCA) the visual features are weighted based on their correlation with the manually annotated signals. By using CCA we perform a dimensionality reduction such that the resulting signals are correlated with semantic meaningful signals and thus also have an expressive power.

CCA takes two time series $\mathbb{O} \in \mathbb{R}^{o \times l}$ and $\mathbb{V} \in \mathbb{R}^{r \times l}$ and computes two projection matrices A and B where A_i , respectively B_i denote the i th row of the matrix. The two matrices project both time series to a new space with dimensionality $d = \min(o, r)$. This is done such that the correlation between every

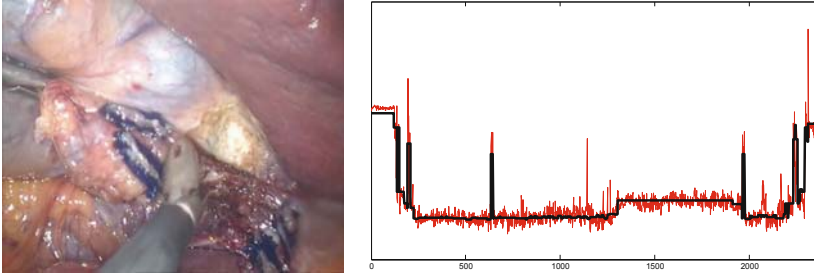


Fig. 2. Left: Example of a laparoscopic image. Right: Instruments (bold) and image features (fine) projected to a common space using CCA. The first dimension which has the highest correlation is shown here.

pair, $\text{corr}(A_i \odot, B_i \vee)$, is maximized while every linear combination A_i is orthogonal to all linear combinations A_j , $j < i$. The same condition holds for B . CCA can be seen as a method that takes two views of the same semantic object in order to extract a representation of the semantics [11]. It has been used e.g. for alignment of human behavior [12] and text based image retrieval [11].

By applying CCA to our data, we reduce the dimensionality of the image features to 17 and obtain a new 17-dimensional representation of the instrument use. The first dimension of the image features and the instrument signals projected to a common space is shown in figure 2. For the statistical modeling and the detection of phases that is described below, we completely discard $A \odot$ and only use $B \vee$. The correlation in the new space is decreasing with every dimension. Therefore we have chosen only to use dimensions, where the correlation $\text{corr}(A_i \odot, B_i \vee)$ is > 0.50 . We compare this method to a standard dimensionality reduction using PCA.

2.2 Modeling

For the segmentation we are using a 14-state left-to-right Hidden Markov Model (HMM), where each state represents one phase. To segment a surgery O_i into phases, we compute the Viterbi-path, that assigns one of the 14 states to each time step of O_i . As each HMM state corresponds to one phase, we can directly use this to segment the surgery. The HMM transition probabilities are simply estimated from the length of each phase in the training data. Defining the observation symbol probabilities i.e. the probability that one feature vector has been generated in one phase, is an important choice for an HMM. As we have real valued data, standard methods like counting the observation symbol frequency are not applicable. To be able to compare different advanced methods, we have chosen to use WEKA [13], a library that implements a wide range of standard machine learning approaches many of which can output probabilities that can be used as observation symbol probabilities. A first test was done using nine surgeries for training and one for testing. The segmentation results for a 14-state HMM using different classifiers have been computed. The best results have been

achieved using Support Vector Machines and the meta-classifiers RotationForest, Bagging and LogitBoost. These have been included in a full cross-validation that is described later.

While a 14-state left-to-right HMM takes into account the temporal order of the phases, it does not capture the whole underlying semantics of the workflow. Especially for the signals that are obtained using CCA the model should be able to represent as much semantic information as possible. One option would be constructing a HMM that has many states, modeling each surgical step. However constructing such a HMM is difficult. Instead we have chosen to use Dynamic Time Warping (DTW) to build a model of an average surgery that captures the underlying semantics. DTW is a method that warps one time series onto another one. This is done by generating a warping path that maps every time step i of one surgery to a time step j in the other surgery while minimizing the sum of distances between corresponding points. Similar as done by [10] we construct a model of an average surgery by warping all surgeries to a common timeline and averaging the signals for each time step. As we know the phase for every time step in the training surgeries, we can label the phase for every time step of the average surgery. To segment a new surgery we warp it to the average model using DTW and carry over the phase labels. Building the DTW average and warping a surgery to this average is done using the features obtained after applying CCA, respectively using PCA.

3 Results

For comparing the methods we have used data from ten surgeries. For three of the surgeries, parts of the surgery have not been recorded due to technical problems. These surgeries have been used only for training. We have performed a leave-one-out cross-validation always using one of the seven complete surgeries for testing and all other nine for training. Four different methods have been compared. DTW using the features obtained using CCA, DTW on the features after PCA, and HMM on the data from CCA and PCA. For the HMM observation probability distribution we have used RotationForest, Bagging, LogitBoost and SVM. For the three meta-classifiers which build a classifier based on simpler classifiers we have performed a full cross-validation using several choices of simple classifiers. We only provide the results obtained with the best classifier, which was in both cases Bagging using C4.5 decision trees. For the methods that use PCA, we tried different numbers of principal components. The results are presented in table 1. It can be seen that the standard deviation is very high. We think that this is a result of the small training set and is an indicator that results can be improved when working with more data. The confusion matrix of DTW + CCA can be seen in figure 3. Most errors are along the diagonal. This is because a strong temporal model is used.

Recording laparoscopic and additional external video images during a surgery and labeling the instrument use is difficult and tedious. Therefore we could only acquire a data set of limited size. One must be careful to draw conclusions from

Table 1. Means and standard deviations of the number of time-steps where the phase was classified correctly

	PCA	CCA
HMM + Bagging	47.12%(±12.67)	53.46%(±14.51)
DTW	62.90%(±18.30)	76.81% ± (12.42)

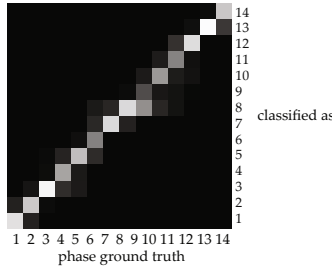


Fig. 3. This images shows the confusion matrix from the whole cross-validation visualized by the image brightness. It can be seen that in most cases of misclassification a phase is classified as neighboring phase.

such a data set. To be able to interpret the outcome, we compared the results from the different methods using the Wilcoxon signed-rank test, a statistical hypothesis test that can handle small data sets. The differences between the compared methods and the corresponding p-values are provided in table 2. The PCA data does not contain a large amount of semantic information. Therefore we did not expect improvement from using DTW compared to a 14-state HMM (1). While there is a difference of 15.8%, this results can not be considered as significant ($p > 0.03$) as the difference mainly results from only two surgeries. Also from using HMMs with PCA or CCA (2) we did not expect a big difference, as the HMM does not fully take advantage of the characteristic of the CCA data. While the significance here is high, the difference is only low. For using CCA + DTW we expected a significantly better result as PCA + DTW (3) or CCA + DTW (4), as only the combination CCA and DTW makes full use of the semantic information that is added by the dimensionality reduction using CCA. This assumption is supported by the results of our comparison.

Table 2. Comparison of the methods that have been used. Difference in percentage points (pp) and p-value is given

compared methods	difference in pp	p-value
(1) PCA + HMM \ PCA + DTW	15.8%	0.188
(2) PCA + HMM \ CCA + HMM	6.3%	0.016
(3) CCA + DTW \ PCA + DTW	29.7%	0.023
(4) CCA + DTW \ CCA + HMM	23.4%	0.008

One way to improve the performance would be to add additional information. As discussed before, other work on laparoscopic video used classifiers for special instruments or aspects of the surgery. In [6] we have presented classifiers to detect surgical clips or whether the camera is inserted into a trocar or not. We have added these two signals to the ones obtained using CCA. Using DTW we obtained a classification result of 79.12%. Other information that could be added with limited technical efforts are signals that are obtained from devices or machines used in the OR. As example we added two signals representing the use of high frequency coagulation and cutting. Using these signals we could further improve the classification to 81.36%.

4 Discussion

In this work we have presented a method that allows segmenting a laparoscopic surgery into phases, using only information from laparoscopic video. We have used a supervised dimensionality reduction method that makes use of additional semantic meaningful information to extract a new representation of the image features that also includes semantic information. In combination with a statistical model that can represent the semantics of time series, we have shown that this method performs better than standard machine-learning and dimensionality reduction methods. It has been shown that especially the combination of the supervised dimensionality reduction and an appropriate statistical model leads to better results. One shortcoming of this work is that the segmentation can only be performed after the whole video has been recorded as DTW requires the whole time-series. HMMs are capable of estimating the current state while the time series is not complete yet. However the results of this work have shown that a simple HMM topology can not achieve good results. One way to handle this would be to use more complex HMM topologies that take into account more of the semantics of the data. This could be achieved using HMMs that derive their topology from data as done by [14]. One advantage of DTW is that a warping path is obtained that assigns every time step of a surgery to a time step of the average model. By taking the warping paths of two surgeries their video can be synchronized e.g. to compare different surgeries or to show a set of synchronized surgeries for training. It can also be used to automatically search for a certain phase in the video.

We believe that methods like CCA will play an important role for workflow analysis. The amount of data that can be obtained from the OR is increasing. There is a growing number of cameras, data can be gathered from anesthesia devices and signals from robots, instrument tracking or people localization systems become available. To be able to combine data from several sources and to do sophisticated modeling and analysis, methods like CCA are well suited. An important advantage of the method that was presented here is that we only need the video to detect the current phase. By taking the approach of performing a supervised dimensionality reduction we add the additional information about instrument use while being able to segment a new surgery without needing this additional information. A

future goal is to extend the method to online use. This would allow monitoring, prediction the remaining duration of a surgery or offering context-sensitive user interfaces only by using the data from the laparoscopic video.

References

1. Mackel, T., Rosen, J., Pugh, C.: Markov Model Assessment of Subjects' Clinical Skill Using the E-Pelvis Physical Simulator. *IEEE Transactions on Biomedical Engineering* 54(12) (2007)
2. Leong, J., Nicolaou, M., Atallah, L., Mylonas, G., Darzi, A., Yang, G.: HMM assessment of quality of movement trajectory in laparoscopic surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 752–759. Springer, Heidelberg (2006)
3. Lin, H., Shafran, I., Murphy, T., Okamura, A., Yuh, D., Hager, G.: Automatic detection and segmentation of robot-assisted surgical motions. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3749, pp. 802–810. Springer, Heidelberg (2005)
4. Lo, B., Darzi, A., Yang, G.: Episode classification for the analysis of tissue/instrument interaction with multiple visual cues. In: Ellis, R.E., Peters, T.M. (eds.) *MICCAI 2003*. LNCS, vol. 2878, pp. 230–237. Springer, Heidelberg (2003)
5. Speidel, S., Benzko, J., Krappe, S., Sudra, G., Azad, P., Müller-Stich, B., Gutt, C., Dillmann, R.: Automatic classification of minimally invasive instruments based on endoscopic image sequences. In: *SPIE Medical Imaging*, vol. 7261 (2009)
6. Padoy, N., Blum, T., Feußner, H., Berger, M.O., Navab, N.: On-line recognition of surgical activity for monitoring in the operating room. In: *Innovative Applications of Artificial Intelligence* (2008)
7. Bhatia, B., Oates, T., Xiao, Y., Hu, P.: Real-Time Identification of Operating Room State from Video. In: *Innovative Applications of Artificial Intelligence*, pp. 1761–1766 (2007)
8. Padoy, N., Mateus, D., Weinland, D., Berger, M.O., Navab, N.: Workflow monitoring based on 3d motion features. In: *ICCV Workshop on Video-oriented Object and Event Classification* (2009)
9. Bouarfa, L., Jonker, P., Dankelman, J.: Surgical context discovery by monitoring low-level activities in the OR. In: *MICCAI Workshop on Modeling and Monitoring of Computer Assisted Interventions, M2CAI* (2009)
10. Padoy, N., Blum, T., Essa, I., Feußner, H., Berger, M.O., Navab, N.: A boosted segmentation method for surgical workflow analysis. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 102–109. Springer, Heidelberg (2007)
11. Hardoon, D., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: an overview with application to learning methods. *Neural Computation* 16(12), 2639–2664 (2004)
12. Zhou, F., de la Torre, F.: Canonical time warping for alignment of human behavior. In: *Neural Information Processing Systems Conference, NIPS* (2009)
13. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11(1) (2009)
14. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G.: Data-Derived Models for Segmentation with Application to Surgical Assessment and Training. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5761, pp. 426–434. Springer, Heidelberg (2009)

Fused Video and Ultrasound Images for Minimally Invasive Partial Nephrectomy: A Phantom Study

Carling L. Cheung^{1,2}, Chris Wedlake¹, John Moore¹,
Stephen E. Pautler³, and Terry M. Peters^{1,2}

¹ Imaging Research Laboratories, Robarts Research Institute

² Biomedical Engineering Graduate Program, The University of Western Ontario

³ Division of Urology, Department of Surgery and Division of Surgical Oncology,
Department of Oncology, The University of Western Ontario,
London, Ontario, Canada

Abstract. The shift to minimally invasive abdominal surgery has increased reliance on image guidance during surgical procedures. However, these images are most often presented independently, increasing the cognitive workload for the surgeon and potentially increasing procedure time. When warm ischemia of an organ is involved, time is an important factor to consider. To address these limitations, we present a more intuitive visualization that combines images in a common augmented reality environment. In this paper, we assess surgeon performance under the guidance of the conventional visualization system and our fusion system using a phantom study that mimics the tumour resection of partial nephrectomy. The RMS error between the fused images was 2.43mm, which is sufficient for our purposes. A faster planning time for the resection was achieved using our fusion visualization system. This result is a positive step towards decreasing risks associated with long procedure times in minimally invasive abdominal interventions.

1 Introduction

Many abdominal surgery procedures are now performed minimally invasively. In this paper we discuss laparoscopic partial nephrectomy, which is the resection of a renal tumour along with a margin of healthy tissue to ensure cancer control. During this procedure, the surgeon uses a laparoscopic camera to view the organ surface and a laparoscopic ultrasound (US) probe to visualize the tumour and other underlying structures to plan and perform the resection. In the conventional operating room (OR), the video and US images are presented separately and are typically in 2D (Fig. 1a-b). Therefore, the surgeon has to look back and forth between the images and mentally map the US image onto the video image to determine where the tumour is relative to the organ surface. Furthermore, the 2D nature of the images results in decreased depth perception.

A more intuitive visualization system would be advantageous for guidance during partial nephrectomy, especially in decreasing procedure time and in

achieving a more accurate margin. During partial nephrectomy, the kidney is often subject to warm ischemia [1]. Periods of warm ischemia greater than 30 minutes put the organ at risk of irreparable damage, therefore a short warm ischemic time is crucial. A 1cm margin of normal renal parenchyma is accepted as the standard to prevent the recurrence of cancer. More recently, it has been suggested that a smaller margin of 5mm or even less is acceptable for cancer control [2]. A smaller margin allows for more healthy tissue to be retained, increasing the nephron sparing and the likelihood of a positive postoperative prognosis. To attain a smaller margin without violating the tumour boundaries, it is crucial to know exactly where the tumour is located and how far it extends into the organ.

We developed and validated a visualization platform (Fig. 1c) that fuses video and US in a common 3D environment. This work is presented as a system that is unique from other work [3,4] through the use of magnetic tracking and a flexible-tip US probe, combined with stereoscopic video. The goal of this paper is to evaluate the efficacy of fused video and US by studying surgeon performance under the guidance of the conventional setup (video and US separate) and the fusion setup, presented both monoscopically and stereoscopically. This goal is achieved with a novel phantom study. Targeting has been shown to be faster with a stereoscopic visualization [5], and more accurate using multimodality fusion [6]. These studies provide the basis for exploring the use of a stereoscopic fusion environment for more complex tasks such as tumour resection.

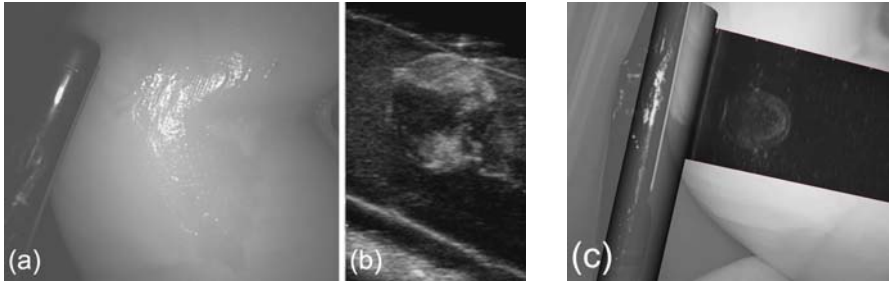


Fig. 1. (a)-(b) Separate video and US images, with the US probe visible in the video image. (c) Fused video and US images showing a tumour within a phantom.

2 Materials and Methods

2.1 Fusion System

Our image fusion platform used the Aurora magnetic tracking system (MTS) (Northern Digital Inc., Waterloo, ON) to determine, in real time, the location of images from a stereoscopic camera and an US probe. A 6 degree-of-freedom MTS sensor was affixed to each tool. The camera was calibrated using Zhang’s method [7] to determine its intrinsic and extrinsic properties. We adapted this method to integrate the camera with the MTS by determining the transform between the camera’s optical origin and its sensor. The US was calibrated [8] to

determine the transform between the center of the probe element array and its sensor. The tracking of the tools combined with calibration was the key to the fusion of the images [9]. Using the real-time positions and orientations of the images, they were rendered in a common coordinate system to create the fusion.

The visualization was run on an x86-based desktop personal computer with dual quad core Xeon processors at 2.66GHz and 4GB RAM, running Windows XP with an NVIDIA Quadro FX 4600 graphics card. We used an Intuitive Surgical stereoscopic laparoscope from Olympus plugged into a Snell & Wilcox vision cart with a Kudos Plus TBS100 synchronizer. An Aloka SD-1700 US system was used with a 7.5MHz laparoscopic probe.

Validation. The accuracy of tool calibration and image fusion was demonstrated with a target localization experiment using a tracked point source target. The target was placed at 20 different locations arranged in a grid pattern throughout tracking space. Each imaging modality was used separately to locate the target. The camera was used between 40 and 60mm from the target, a range within which it would be used clinically. We compared the target location as determined by each modality to the true location given by tracking the target, and to the location determined by the other modality. To calculate the 3D location of the target according to US, its 2D US image coordinates were multiplied by the calibrated US tool transform [10]. Triangulation [11] was used to compute the location of the target according to the stereoscopic camera images. This task demonstrated how well the camera and the US were calibrated relative to the tracking space coordinate system, as well as to each other.

2.2 Phantom Study

A study using kidney mimicking phantoms was performed to compare the fusion system mentioned in the previous section to the conventional visualization system that is currently implemented in the OR.

Phantom Development. Polyvinyl alcohol (PVA) was used in this study to create tissue mimicking phantoms. PVA, a synthetic polymer with high tensile strength and flexibility, is suitable for creating phantoms due to its nontoxic properties, ease of preparation and excellent echogenicity. The concentration of PVA relative to water and the number of freeze/thaw cycles (FTC) determines the rigidity and density of the final product. We used a 10% by mass solution of PVA and water. The PVA undergoes FTCs (12 hours freeze and 12 hours thaw per cycle) to create a cryogel (PVA-C). The process of creating PVA-C and the effect of the number of FTCs on its properties have been reported previously [12].

Our phantoms consisted of two components: tumour and normal parenchyma. Each was subject to a different number of FTCs so that they differed in density, enabling them to be distinguished under US. A custom designed mold was used to create 19.05mm diameter spheres. Gadolinium contrast agent was added to the PVA of the tumours for enhanced magnetic resonance imaging (MRI) visualization. The tumour-replicating spheres were frozen and thawed twice, then removed and placed into a larger kidney-sized mold. After the tumour phantoms

were inserted, the larger mold was filled with PVA, and the entire system was subject to one FTC. In total, the tumour-mimicking PVA underwent three FTCs while the normal renal parenchyma-mimicking PVA underwent one cycle. The difference in density between tumour and normal tissue resulted in realistic contrast under US (Fig. 1b). In addition to realistic imaging qualities, the phantoms were realistic in texture, as determined by our clinical collaborators.

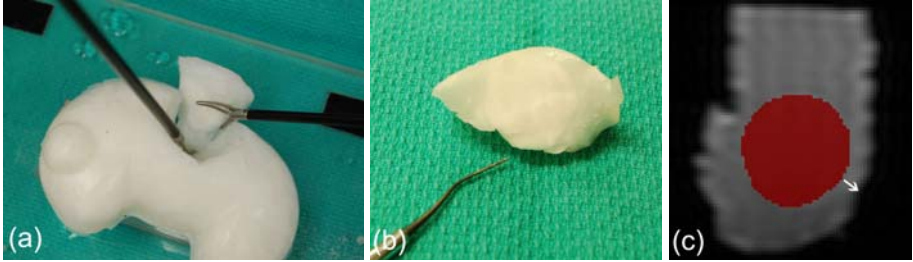


Fig. 2. (a) Tumour excision from phantom. (b) Excised segment. (c) MRI scan of excised section with segmented tumour (in red); white arrow denotes minimum margin.

User Study Procedure. We designed a task using the PVA-C phantoms to determine how well a surgeon is able to resect tumours under the guidance of fused video and US. The task replicated renal tumour resection using minimally invasive surgery (MIS) tools. This task was performed under the guidance of the conventional OR system with video and US displayed separately (conventional), the fused images presented in 2D (2D fusion), and the fused images presented in 3D (3D fusion). All of the visualizations were viewed on a novel autostereoscopic monitor (Dimension Technologies Inc., Rochester, NY) that can switch between 2D and 3D mode, eliminating any effect that may arise from variations in resolution and brightness caused by using a different monitor for each mode.

The resections were performed by a single experienced surgeon, who was instructed to resect the tumours as he would in the OR (Fig. 2a-b) and to aim for a 5mm margin. For this experiment, we focused on endophytic tumours (completely beneath the organ surface) because they are the most difficult to operate on. Multiple resections were performed, each time under the guidance of one of the three visualization systems. The order in which the visualization systems were presented was randomized to minimize learning effects. A laparoscopy box trainer (Fig. 3a-b) was used to replicate the minimally invasive nature of the procedure. The phantom was placed inside the box trainer and laparoscopic tools were inserted through ports on the surface. The tools provided to the surgeon included laparoscopic scissors, graspers, marker (used in place of cautery for marking the phantom surface during planning), US probe, and camera, arranged in a similar manner to the setup in the OR (Fig. 3c). The US monitor was occluded when the fusion environment was being used.

Performance Measures. Performance was assessed by measuring the procedure time and the accuracy of margin achievement. A single factor analysis of

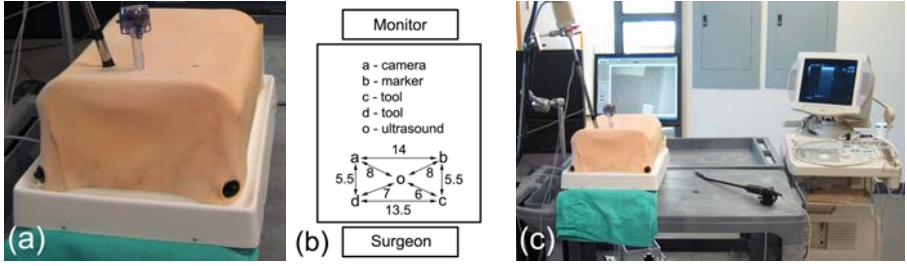


Fig. 3. (a) Laparoscopy box trainer used to replicate MIS environment. (b) Experiment port placement, numbers indicate distance in centimetres. (c) Experiment setup.

variance (ANOVA) was performed (GraphPad Prism 5.0) to determine significant differences in procedure duration between guidance modalities.

Time: The total procedure time was broken down into the planning stage and the excision stage. The planning stage included using the US to localize the tumour and using a marker to mark the surface of the phantom where the incision was to be made. The excision stage constituted the remainder of the procedure.

Accuracy: To evaluate margins, the excised sections were scanned with MRI. A T1 weighted sequence (T1 FLAIR) on a 3T GE MR750 Discovery scanner was used with the following parameters: TR=2300ms; TE=6.528ms; TI=850ms; frequency FOV=12.8; NEX=1; echo train length=2; BW=31.25; slice thickness=1.0mm; and spacing=0.2mm. The gadolinium contrast agent in the tumour PVA made it distinguishable from the normal parenchyma. This improved contrast allowed us to segment the tumours for post-operative evaluation. The volume of each segmented tumour was determined to validate the segmentation.

Clinically, margins are reported as positive (tumour extending to the edge) or negative (no cancer cells between the tumour and the edge). Some pathologists report the minimum margin from cut edges, although this is not consistent between all cases. We examined the MRI volumes of the excised sections and measured the shortest distance between the tumour and a cut edge to replicate the clinical scenario (Fig. 2c). We also reported whether the tumour boundary was exposed, indicating the potential for a positive margin.

3 Results

3.1 Fusion System Validation

The target localization error (Mean±Standard Error of the Mean(SEM)) was 1.20 ± 0.08 mm for the camera, 1.85 ± 0.14 mm for the US, and 2.38 ± 0.11 mm between the camera and the US. Furthermore, a positive correlation ($r=0.86$) was found between the error and the distance from the camera to the target, with the error being minimized in the camera focal range.

Table 1. Resection times (Mean \pm SEM), in seconds

	Planning	Excision	Total
Conventional	113.5 \pm 21.2	331.6 \pm 49.8	445.1 \pm 68.0
2D Fusion	80.62 \pm 6.1	393.8 \pm 62.9	474.4 \pm 64.5
3D Fusion	104.9 \pm 11.7	541.9 \pm 86.5	646.0 \pm 88.1

Table 2. Minimum margin thickness, in millimetres

	Mean	SEM	Max
Conventional	2.10	0.89	6.05
2D Fusion	1.11	0.39	2.22
3D Fusion	1.76	0.79	5.10

3.2 Phantom Study

The surgeon performed the task 6 times under each guidance modality, for a total of 18 trials. We report procedure times in Table 1 and margin accuracy in Table 2. These results show an improvement in the planning stage using 2D and 3D fusion. The tumour boundary was exposed in one case with the conventional system, in two cases with 2D fusion and in two cases with 3D fusion.

4 Discussion

The accuracy of the fusion is clinically sufficient as it allows surgeons to achieve acceptable margins without violating the tumour boundary. Furthermore, the accuracy of the MTS is on the order of 1-2mm within our working volume, therefore our accuracy is also acceptable from an engineering angle.

The user experienced a 29% decrease in planning time with 2D fusion and an 8% decrease with 3D fusion over the conventional system ($p=0.28$). Excision times using the fusion systems were greater than with the conventional system, with the user being 16% slower with 2D fusion and 39% slower with 3D fusion ($p=0.11$). Overall, this led to a total time that was 6% slower with 2D fusion and 31% slower with 3D fusion compared to the conventional system ($p=0.15$). Across all visualizations and procedure stages, time decreased with an increasing number of trials. This suggested a learning effect, which we expect will translate to improved procedure times with additional training.

The longer excision times with the fusion systems were due to latency in the visualization, especially with the second video image added for 3D fusion. The video was synchronized with the MTS, therefore its refresh rate was limited by tracking frequency. Our implementation would benefit from upgrades in the tracking system to provide a sufficiently high sampling rate when a greater number of tools is plugged in. With adequate hardware and software optimization, the excision times should be consistent over all visualizations, and the decreased planning time using the fusion systems should result in a decrease in overall procedure time. With such optimization, we suspect an even greater improvement

in the planning time when using fusion. Clinically, the tumour resection is under 10 minutes. In all cases except for one conventional, one 2D fusion and two 3D fusion examples, our total time was less than 10 minutes.

Psychophysical considerations must still be addressed when overlaying images from multiple modalities, particularly when using stereoscopic images, a factor that contributed to the longer times with 3D fusion. Further work is required in optimizing the display of the images to make their combination convincing. Image processing [13] needs to be performed to remove ambiguity as to where the US beam lies with respect to the organ surface shown in the video, and to provide a more realistic environment. In addition, the surgeon reported occasionally losing the 3D view when he moved, forcing him to become reacquainted with the stereoscopic monitor, therefore an improved display method is desired.

Margin thickness varies, both in real procedures and in our study, which makes systematic comparison between cases difficult. Results show that the surgeon maintained a relatively fixed accuracy and decreased procedure planning time. We found that tumour boundary exposure occurred with all visualization systems. However, positive margins are not necessarily an indication for disease recurrence, as many of the tumours removed with partial nephrectomy are well encapsulated (reflected in our phantom model) and it is safe to cut along the tumour edge to retain more tissue, as long as the tumour is not violated.

In addition to the reduction in procedure planning time that we achieved, we aim to reduce overall procedure time and provide a means for the surgeon to gain consistent margins of a specified thickness. These goals will be achieved via improvements in the visualization through system optimization and image enhancements. The necessary improvements were determined through identification of constraints, from both observation and surgeon feedback, encountered in clinical practice that can affect the procedure. Future studies will include evaluating the performance of surgeons with varying levels of training, and additional trials for stronger statistical power.

5 Conclusions

We evaluated the use of fused US and video images for tumour resection during minimally invasive partial nephrectomy. This preliminary study demonstrated system usability and found an improvement in resection planning times using fusion. Optimization of the system components motivated by limitations encountered during our experimental procedures, and enhancement of the images to address psychophysical considerations, is expected to inspire faster overall procedure times and a more effective method of establishing margins.

Acknowledgements

The authors thank Kevin Wang and Cyndi Harper-Little for help with MRI scanning, and Dr. Usaf Aladl, Cristian Linte and Dr. Elvis Chen for discussion. We acknowledge funding from Cancer Imaging Network of Ontario.

References

1. Gill, I.S., Desai, M.M., Kaouk, J.H., Meraney, A.M., Murphy, D.P., Sung, G.T., Novick, A.C.: Laparoscopic partial nephrectomy for renal tumor: duplicating open surgical techniques. *The Journal of Urology* 167(2), 469–476 (2002)
2. Sutherland, S.E., Resnick, M.I., MacLennan, G.T., Goldman, H.B.: Does the size of the surgical margin in partial nephrectomy for renal cell cancer really matter? *The Journal of Urology* 167(1), 61–64 (2002)
3. Leven, J., Burschka, D., Kumar, R., Zhang, G., Blumenkranz, S., Dai, X., Awad, M., Hager, G.D., Marohn, M., Choti, M., Hasser, C., Taylor, R.H.: 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)
4. Feuerstein, M., Reichl, T., Vogel, J., Traub, J., Navab, N.: New approaches to on-line estimation of electromagnetic tracking errors for laparoscopic ultrasonography. *Computer Aided Surgery* 13(5), 311–323 (2008)
5. Boritz, J., Booth, K.S.: A study of interactive 3D point location in a computer simulated virtual environment. In: *Proceedings of the ACM symposium on Virtual reality software and technology*, Lausanne, Switzerland, pp. 181–187. ACM, New York (1997)
6. Linte, C.A., Wiles, A., Moore, J., Wedlake, C., Peters, T.M.: Virtual reality-enhanced ultrasound guidance for atrial ablation: in vitro epicardial study. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II*. LNCS, vol. 5242, pp. 644–651. Springer, Heidelberg (2008)
7. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: *Proceedings of the 7th International Conference on Computer Vision*, Kerkyra, Greece, pp. 666–673. IEEE Computer Society Press, Los Alamitos (1999)
8. Muratore, D.M., Galloway, R.L.J.: Beam calibration without a phantom for creating a 3D freehand ultrasound system. *Ultrasound in Medicine & Biology* 27(11), 1557–1566 (2001)
9. Cheung, C.L., Wedlake, C., Moore, J.T., Pautler, S.E., Ahmad, A., Peters, T.M.: Fusion of stereoscopic video and laparoscopic ultrasound for minimally invasive partial nephrectomy. In: Miga, M., Wong, K. (eds.) *Medical Imaging 2009: Visualization, Image-Guided Procedures and Modeling*, Lake Buena Vista, FL, USA. SPIE 2009, vol. 7261, pp. 7261–7209 (2009)
10. Wiles, A.D., Moore, J., Linte, C.A., Wedlake, C., Ahmad, A., Peters, T.M.: Object identification accuracy under ultrasound enhanced virtual reality for minimally invasive cardiac surgery. In: Miga, M., Cleary, K.R. (eds.) *Medical Imaging: Visualization, Image-Guided Procedures, and Modelling*, San Diego, CA, USA. SPIE 2008, vol. 6918, p. 39180E (2008)
11. Hartley, R., Zisserman, A.: *Multiple view geometry in computer vision*, 2nd edn. Cambridge University Press, Cambridge
12. 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)
13. Lerotic, M., Chung, A.J., Mylonas, G., Yang, G.: 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)

Probabilistic 4D Blood Flow Mapping

Ola Friman¹, Anja Hennemuth¹, Andreas Harloff², Jelena Bock²,
Michael Markl², and Heinz-Otto Peitgen¹

¹ Fraunhofer MEVIS, Institute for Medical Image Computing, Germany

² University Hospital Freiburg, Germany

Abstract. Blood flow and tissue velocity can be measured using phase-contrast MRI. In this work, the statistical properties of 4D phase-contrast images are derived, and a novel probabilistic blood flow mapping method based on sequential Monte Carlo sampling is presented. The resulting flow maps visualize and quantify the uncertainty in conventional flow visualization techniques such as streamlines and particle traces.

1 Introduction

Phase-contrast (PC) MRI utilizes residual phase shifts of spins to quantify tissue motion and blood flow [1,2]. In cardiovascular PC MRI applications, visualization techniques, such as vector glyphs, streamlines, pathlines and particle traces, are employed for visualizing blood flow [3,4]. A 3D streamline illustrates the trajectory a zero-mass particle takes through a static vector field $\mathbb{V}(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, e.g., an instantaneous flow field, see Fig. 4a. Pathlines and particle traces show the particle trajectory in a vector field $\mathbb{V}(\mathbf{x}, t)$ that changes over time, e.g., in a 4D pulsative blood flow. Clinical applications that benefit from such flow pattern information include the assessment of stenoses, aneurysms, and heart valve function, the development of vessel plaque, and surgical planning and follow-up in congenital heart disease. While a visualization of the flow pattern using streamlines or particle traces indeed provide useful information, noise and uncertainty in the PC MRI measurements is not accounted for and the visualized traces may even give a false sense of precision. In this work, the uncertainty associated with a flow streamline or particle trace is addressed and visualized using the *distribution* of possible flow trajectories. To this end, the statistical properties of PC MRI images are first derived. The distribution of a streamline or a particle trace is then sampled and characterized using a sequential Monte Carlo approach.

2 Statistical Properties of Flow MRI Images

The statistical properties of PC MRI images in the presence of Gaussian measurement noise, particularly the statistical distribution of the estimated flow vectors in the estimated 4D flow vector field $\mathbb{V}(\mathbf{x}, t)$, are derived in this section. The derivation includes the use phased array MRI coils, which consist of a number of coils K (typically 4-16) arranged in an array. Such arrays permit fast parallel imaging and/or a better Signal-to-Noise ratio (SNR) over large spatial extents [5]. The derivation below focusses on non-parallel imaging, which provides the highest image SNR.

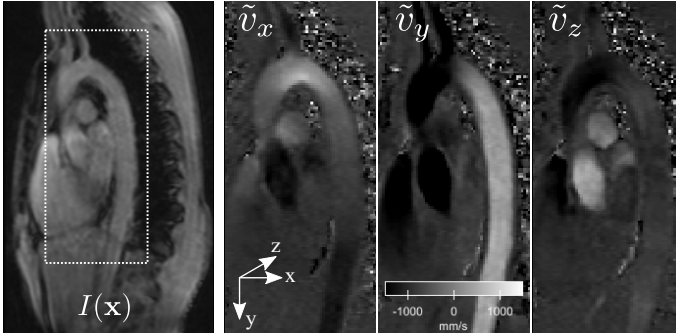


Fig. 1. Magnitude image $I(\mathbf{x})$ (left) and estimated velocity components \tilde{v}_x , \tilde{v}_y , and \tilde{v}_z in the zoomed area around the aorta

2.1 MRI Image Magnitude and Phase Modeling

In general, a complex MRI image in the spatial domain acquired with coil k in the phased array may be modeled as

$$S_k(\mathbf{x}) = A_k(\mathbf{x}) e^{i\theta_k(\mathbf{x})} + n_k^r(\mathbf{x}) + i n_k^i(\mathbf{x}), \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^3$ is a spatial voxel location, $A_k(\mathbf{x})$ represents the image magnitude weighted with the sensitivity profile of coil k , $\theta_k(\mathbf{x})$ is a coil-specific spatially varying phase, and $n_k^r(\mathbf{x})$ and $n_k^i(\mathbf{x})$ represent independent Gaussian noise $N(0, \sigma^2)$ in two quadrature channels [6]. From a statistical perspective, the magnitude image $|S_k(\mathbf{x})|$ follows a Rician probability distribution, which at high SNR ($A_k/\sigma > 5$) can be approximated by a Gaussian distribution [7] $|S_k(\mathbf{x})| \in N(A_k(\mathbf{x}), \sigma^2)$ for all practical purposes. This high-SNR assumption is generally true for voxels containing tissue or blood, which is shown in the Result section. A near-optimal-SNR magnitude image $I(\mathbf{x})$ can be reconstructed from all coils using the sum-of-squares method [5] (Fig. 1):

$$I(\mathbf{x}) = \sqrt{\sum_{k=1}^K |S_k(\mathbf{x})|^2}. \quad (2)$$

The statistical properties of expressions such as the one in Eq. 2 have been studied in [8], and again, for high SNR voxels, a Gaussian approximation applies:

$$I(\mathbf{x}) \in N(A(\mathbf{x}), \sigma^2), \text{ where } A(\mathbf{x}) = \sqrt{\sum_{k=1}^K A_k(\mathbf{x})^2}. \quad (3)$$

Note that $A(\mathbf{x})$ and σ^2 are unknown, but with the knowledge of the statistical distribution in Eq. 3, they can be estimated from the known $I(\mathbf{x})$. This estimation is described in Section 2.3.

The phase $\theta_k(\mathbf{x})$ in Eq. 1 is estimated with the argument operator $\arg(S_k(\mathbf{x}))$, and the exact probability distribution of this estimator is given in 7. For high SNR, however, one may again use a simpler Gaussian approximation

$$\arg(S_k(\mathbf{x})) \in N_w \left(\theta_k(\mathbf{x}), \frac{\sigma^2}{A_k(\mathbf{x})^2} \right), \tag{4}$$

where N_w denotes the wrapped Normal distribution, i.e., θ_k is calculated modulo 2π so that $\theta_k \in (-\pi, \pi]$.

2.2 Flow Vector Modeling

Four separate 3D PC MRI images must be acquired to reconstruct a 3D flow vector field $\mathbf{V}(\mathbf{x})$. A 4D flow field $\mathbf{V}(\mathbf{x}, t)$ is obtained by repeated acquisitions during the cardiac cycle synchronized with the RR-interval of the patient. The goal in this section is to derive the statistical properties of an estimated flow vector $\tilde{\mathbf{v}} \in \mathbf{V}(\mathbf{x}, t)$. All expressions in this section are applied voxelwise, and the spatial coordinate \mathbf{x} is omitted for clarity. Typically, an encoding scheme is used which is based on a baseline image S_k^0 and three images $S_k^x, S_k^y,$ and S_k^z that encode the flow velocity along the $x, y,$ and z directions in the image phase, cf. Eq. 1

$$\begin{aligned} S_k^0 &= A_k e^{i\theta_k} + \text{noise}, & S_k^y &= A_k e^{i(\theta_k - \frac{\pi}{v_{enc}} v_y)} + \text{noise}, \\ S_k^x &= A_k e^{i(\theta_k - \frac{\pi}{v_{enc}} v_x)} + \text{noise}, & S_k^z &= A_k e^{i(\theta_k - \frac{\pi}{v_{enc}} v_z)} + \text{noise}. \end{aligned} \tag{5}$$

In Eq. 5, $v_x, v_y,$ and v_z are the true velocities along the orthogonal coordinate axes, and the velocity encoding parameter v_{enc} is a sequence parameter that controls the upper limit of the velocity that can be measured without artifacts. A typical value is $v_{enc} = 1.5$ m/s. θ_k denotes an unknown and spatially varying phase for coil $k = 1 \dots K$, which is assumed constant over time and acquisitions. The flow velocity in the x -direction is found as the phase difference between the S_k^x image and the baseline image S_k^0 , i.e., following the sum-of-squares reconstruction in Eq. 2, a velocity estimate based on all K coils is calculated as

$$\tilde{v}_x = \frac{v_{enc}}{\pi} \arg \left(\sum_{k=1}^K S_k^0 S_k^{x*} \right), \tag{6}$$

where S_k^{x*} denotes the complex conjugate of S_k^x . The estimates of \tilde{v}_y and \tilde{v}_z are calculated analogously. Example velocity images for one slice are shown in Fig. 1. Using Eq. 4, one can show that \tilde{v}_x is distributed according to the following Gaussian distribution:

$$\tilde{v}_x \in N \left(v_x, \frac{v_{enc}^2}{\pi^2} \frac{2\sigma^2}{A^2} \right), \tag{7}$$

where A is given in Eq. 3. It is assumed that the velocity is smaller than v_{enc} so that no phase wrap occurs. In practise, a standard image preprocessing step corrects such wraps.

As each velocity component of the 3D flow vector $\tilde{\mathbf{v}} = [\tilde{v}_x, \tilde{v}_y, \tilde{v}_z]^T$ is Gaussian distributed according to Eq. 7, the joint distribution is multivariate Gaussian $\tilde{\mathbf{v}} \in N(\mathbf{v}, \mathbf{C})$. Furthermore, as the baseline images S_k^0 , $k = 1 \dots K$ are involved in the computation of all three velocity components \tilde{v}_x , \tilde{v}_y , and \tilde{v}_z , the covariance matrix \mathbf{C} will not be diagonal. For example, with a derivation similar to the one above, the covariance between \tilde{v}_x and \tilde{v}_y is

$$\text{Cov}(\tilde{v}_x, \tilde{v}_y) = \frac{v_{enc}^2 \sigma^2}{\pi^2 A^2}, \text{ giving the full covariance } \mathbf{C} = \frac{v_{enc}^2 \sigma^2}{\pi^2 A^2} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}. \quad (8)$$

To summarize, the measured flow velocity vector in each voxel may be seen as drawn from a multivariate Gaussian distribution with mean \mathbf{v} , i.e., the true velocity, and covariance matrix as given by Eq. 8. To fully specify the covariance matrix, estimates of the unknown parameters A and σ^2 are required. This is discussed in the next section.

2.3 Parameter Estimation

Methods for estimating the noise variance σ^2 and signal strength A_k from magnitude MRI images $|S_k|$ have previously been proposed in the literature [9, 7]. Most of the proposed methods require a homogenous image region. Typically an air region is selected in which the SNR is low and the Gaussian approximation of the Rician distribution is no longer valid. In contrast, here we propose using all 4 image volumes in the PC MRI acquisition (cf. Eq. 5) to perform a voxelwise parameter estimation: Let Ω denote a mask of high-SNR voxels in the four images $I^0(\mathbf{x})$, $I^x(\mathbf{x})$, $I^y(\mathbf{x})$, and $I^z(\mathbf{x})$, reconstructed from Eq. 5 with the sum-of-squares method in Eq. 2. The high-SNR voxels are found using a straightforward image thresholding, as these voxels correspond to voxels with high intensity, i.e., all non-air voxels. For the voxels in Ω , the Gaussian approximation in Eq. 3 can be applied, leading to an estimation procedure based on the well-known formulae for the mean and variance of Gaussian variables. First, an estimator of the signal strength $A(\mathbf{x})$ is obtained as

$$\tilde{A}(\mathbf{x}) = \frac{1}{4} [I^0(\mathbf{x}) + I^x(\mathbf{x}) + I^y(\mathbf{x}) + I^z(\mathbf{x})]. \quad (9)$$

According to Eq. 3, the noise variance σ^2 is equal for all voxels in Ω , so that a variance estimate can first be calculated for each voxel separately, and a final estimate can then be found by pooling over all voxels:

$$\tilde{\sigma}^2 = \frac{1}{|\Omega| - 1} \sum_{\mathbf{x}_i \in \Omega} \left[\frac{1}{4 - 1} \sum_{k=\{0,x,y,z\}} (I^k(\mathbf{x}_i) - \tilde{A}(\mathbf{x}_i))^2 \right]. \quad (10)$$

In this expression, $|\Omega|$ denotes the number of voxels with high SNR. With $\tilde{A}(\mathbf{x})$ and $\tilde{\sigma}^2$, the covariance matrix for the flow vector in Eq. 8 is fully specified.

3 Probabilistic Flow Mapping

In this section, the distribution of possible flow trajectories is addressed. Although there is no closed form expression of this distribution, it is possible to draw random samples using Monte Carlo methods and then reconstruct and visualize the distribution using a histogram technique.

3.1 Probabilistic Streamlines and Particle Traces

A 3D streamline is a trajectory described by train of vectors $\{\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_{k-1}\}$ starting in a seed point $\mathbf{x}_0 \in \mathbb{R}^3$, see Fig. 2. A 4D particle trace can be described similarly, although with the additional temporal dimension. Each vector \mathbf{s}_k is a function of the current trajectory position \mathbf{x}_k , the flow vector field \mathbb{V} , and a step length parameter T :

$$\mathbf{s}_k = \mathbf{f}(\mathbf{x}_k, \mathbb{V}; T). \tag{11}$$

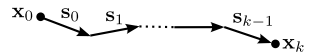


Fig. 2. A streamline

For example, an Euler sampling scheme uses $\mathbf{f}(\mathbf{x}_k, \mathbb{V}; T) = T\mathbf{v}(\mathbf{x}_k)$, but more accurate multistage functions such as Heun’s or 4th-order Runge-Kutta schemes are generally used in practise [10]. The spatial position of the streamline trajectory after k steps is

$$\mathbf{x}_k = \mathbf{x}_0 + \sum_{j=0}^{k-1} \mathbf{s}_j. \tag{12}$$

In a conventional streamline algorithm, the vectors \mathbf{s}_k are treated as deterministic variables. However, in PC MRI, these vectors are based on the estimated vector field \mathbb{V} , which is contaminated with random noise. Consequently, the vectors \mathbf{s}_k as well as the positions \mathbf{x}_k may be seen as random variables. The key question addressed in this work is the form of the statistical distribution $p(\mathbf{x}_k) = p(\mathbf{s}_0, \dots, \mathbf{s}_{k-1})$ of the streamline under the influence of noise in \mathbb{V} . A sample from $p(\mathbf{s}_0, \dots, \mathbf{s}_{n-1})$ is referred to here as a *probabilistic streamline* or a *probabilistic particle trace*. A theoretical derivation of this distribution is complicated by both the high dimensionality as well as the recursive nature of the streamline, i.e., the vector \mathbf{s}_k depends on all previous vectors $\mathbf{s}_{k-1}, \dots, \mathbf{s}_0$, which is clearly seen by combining Eq. 11 and Eq. 12. An alternative approach to investigate the form of a mathematically intractable probability distribution is to draw a large number of samples using computational Monte Carlo methods. An important observation is that the distribution above can be factorized into conditionally independent parts:

$$p(\mathbf{s}_0, \dots, \mathbf{s}_{k-1}) = \prod_{j=0}^{k-1} p_j(\mathbf{s}_j | \mathbf{x}_j). \tag{13}$$

Samples from such a distribution can be generated using so-called sequential Monte Carlo sampling [11]. In this sampling approach, a probabilistic streamline is generated iteratively by first drawing a sample $\tilde{\mathbf{s}}_0$ from $p_0(\mathbf{s}_0 | \mathbf{x}_0)$, which

gives $\tilde{\mathbf{x}}_1 = \mathbf{x}_0 + \tilde{\mathbf{s}}_0$. Next, given $\tilde{\mathbf{x}}_1$, a sample $\tilde{\mathbf{s}}_1$ can be drawn from $p_1(\mathbf{s}_1|\tilde{\mathbf{x}}_1)$, which gives $\tilde{\mathbf{x}}_2$, and so on. $p_j(\mathbf{s}_j|\mathbf{x}_j)$ remains to be determined in Eq. 13. Section 2.2 showed that the estimated flow vectors in \mathbb{V} are Gaussian distributed with covariance matrix \mathbf{C} in Eq. 8. Therefore, $p_j(\mathbf{s}_j|\mathbf{x}_j)$ also has a Gaussian form:

$$p_j(\mathbf{s}_j|\mathbf{x}_j) = N[\mathbf{f}(\mathbf{x}_j, \mathbb{V}; T), T^2\mathbf{C}]. \quad (14)$$

Expressed in words, the random vector \mathbf{s}_j has as mean the deterministic vector described by Eq. 11 obtained using a regular streamline algorithm, e.g., an Euler, Heun or Runge-Kutta scheme. This vector is perturbed with the covariance matrix \mathbf{C} in Eq. 8 scaled by the step length T .

3.2 Probabilistic Flow Map

Each probabilistic streamline or particle trace is a sample from the distribution $p(\mathbf{s}_0, \dots, \mathbf{s}_{k-1})$ that describes the probability of all possible paths a virtual zero-mass particle may take from the seed point \mathbf{x}_0 . To facilitate the visualization of this distribution, the probabilistic streamlines can be converted into a 3D spatial map $\Psi(\mathbf{x})$ or a 4D spatiotemporal map $\Psi(\mathbf{x}, t)$ for particle traces. To this end, a large number $N > 1000$ of probabilistic streamlines must be generated. Let $\mu(\mathbf{x}, N)$ be the number of occasions that each voxel \mathbf{x}_k is passed by a streamline. The flow map is then calculated as a histogram where each voxel is a bin:

$$\Psi(\mathbf{x}) = \frac{\mu(\mathbf{x}, N)}{N}. \quad (15)$$

A similar mapping technique has been used in Diffusion-Tensor MRI [12].

4 Image Data

Due to space limitations, a single 4D PC MRI data set of the blood flow in the aorta is used to demonstrate the proposed probabilistic blood flow mapping technique. The following acquisition parameters were used: TE = 3.67 ms, TR = 6.1 ms, flip angle 15, $v_{enc} = 1.5$ m/s, spatial resolution $1.7 \times 1.7 \times 3.5$ mm³ and temporal resolution 48.8 ms. A phased array coil with $K = 12$ coil elements was used to obtain the necessary spatial coverage. The acquired 4D image volumes were of size $120 \times 192 \times 24$ voxels and 14 temporal frames covering a heart beat.

5 Results

To determine the covariance matrix of the flow vectors in Eq. 8, estimates of $A(\mathbf{x})$ and σ^2 were first calculated using the procedure described in Section 2.3. A voxelwise signal-to-noise image, where $\text{SNR} = A(\mathbf{x})/\sigma$, is shown in Fig. 3. The average SNR in the aorta is about 10, resulting in a standard deviation of 68 mm/s in each velocity component when inserted into Eq. 7. As a reference, the peak flow velocity in the aorta is approximately 1500 mm/s and significantly lower in smaller vessels. Moreover, the SNR is high enough to approximate the

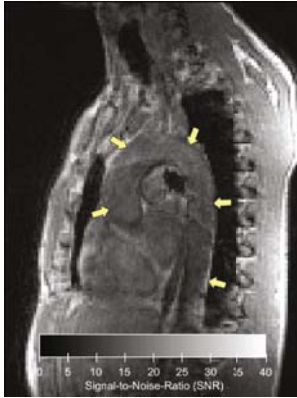


Fig. 3. SNR for each voxel in a slice estimated with the method presented in Section 2.3. The aorta is designated by the arrows.

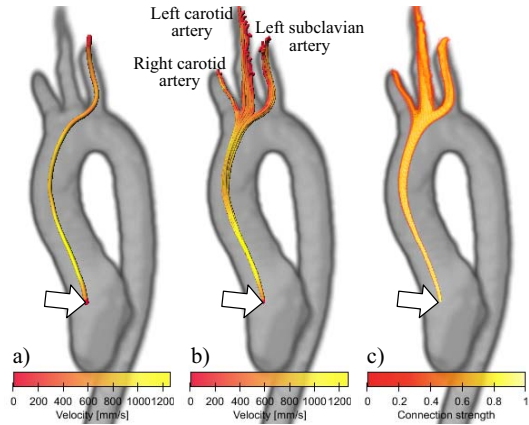


Fig. 4. a) A conventional 3D streamline visualizing the flow pattern from one point in the aorta. b) Probabilistic streamlines. c) Flow map $\Psi(\mathbf{x})$ calculated from 10,000 probabilistic streamlines.

Rician distribution with a Gaussian, as is done in Section 2. In Fig. 4a, a conventional streamline is shown, visualizing the flow pattern from the aorta into the left subclavian artery. In Fig. 4b, 50 probabilistic streamlines are shown, which were generated according to Section 3.1 and initiated in the same point. These probabilistic streamlines illustrate the uncertainty due to measurement noise that is not evident from the conventional streamline, i.e., when the noise is considered, the flow pattern may exit in any of the left carotid, right carotid, or left subclavian arteries. The distribution of possible flow trajectories, calculated as described in Section 3.2 using 10,000 probabilistic streamlines, is shown in Fig. 4c. Figure 5 shows a 4D flow map generated from 10,000 probabilistic particle traces emitted from a region instead of from a single point. The color scale

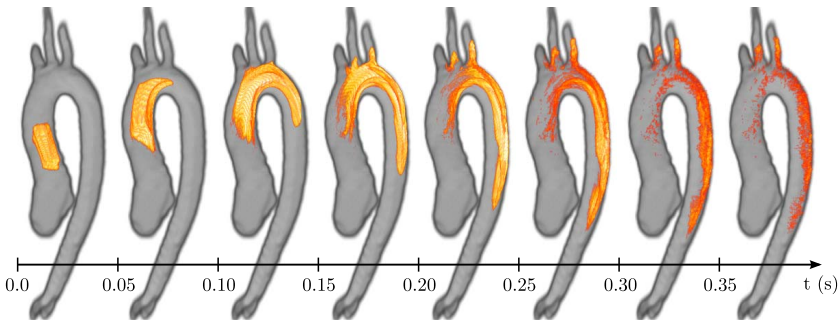


Fig. 5. 4D spatiotemporal probabilistic flow map $\Psi(\mathbf{x}, t)$ illustrating the statistical distribution of particle traces during a heart beat. The map is based on 10,000 probabilistic particle traces.

represents the probability of finding a conventional particle trace in a specific spatiotemporal box when the measurement noise is considered.

6 Discussion

In this work, the statistical properties of PC MRI velocity measurements have been derived and used for mapping the uncertainty associated with blood flow streamlines and particle traces. This is done by drawing samples from the distribution of possible flow trajectories using a sequential Monte Carlo method. The primary application demonstrated here is a visualization of the uncertainty, i.e., the flow distribution. The next step is to employ the method for flow pattern quantification. For example, stroke embolization pathway probabilities can be calculated with a probabilistic flow connection map from a plaque position to critical vessels supplying the brain. Another concrete clinical application is congenital heart disease, where blood mixing ratios at positions where two vessels merge are of interest.

References

1. Pelc, N., Sommer, F., Li, K., Brosnan, T., Herfkens, R., Enzmann, D.: Quantitative magnetic resonance flow imaging. *Magn. Reson. Q.* 10(3), 125–147 (1994)
2. Markl, M., Harloff, A., Bley, T., Zaitsev, M., Jung, B., et al.: Time-resolved 3D MR velocity mapping at 3T: Improved navigator-gated assessment of vascular anatomy and blood flow. *J. Magn. Reson. Imaging* 25(4), 824–831 (2007)
3. Napel, S., Lee, D., Frayne, R., Rutt, B.: Visualizing three-dimensional flow with simulated streamlines and three-dimensional phase-contrast MR imaging. *J. Magn. Reson. Imaging* 2(2), 143–153 (1992)
4. Wigström, L., Ebberts, T., Fyrenius, A., Karlsson, M., Engvall, J., Wranne, B., Bolger, A.: Particle trace visualization of intracardiac flow using time-resolved 3D phase contrast MRI. *Magn. Reson. Med.* 41(4), 793–799 (1999)
5. Roemer, P., Edelstein, W., Hayes, C., Souza, S., Mueller, O.: The NMR phased array. *Magn. Reson. Med.* 16(2), 192–225 (1990)
6. Conturo, T., Smith, G.: Signal-to-noise in phase angle reconstruction: dynamic range extension using phase reference offsets. *Magn. Reson. Med.* 15(3), 420–437 (1990)
7. Gudbjartsson, H., Patz, S.: The Rician distribution of noisy MRI data. *Magn. Reson. Med.* 34(6), 910–914 (1995)
8. Andersen, A.H., Kirsch, J.E.: Analysis of noise in phase contrast MR imaging. *Med. Phys.* 23(6), 857–869 (1996)
9. Henkelman, R.: Measurement of signal intensities in the presence of noise in MR images. *Med. Phys.* 12(2), 232–233 (1985)
10. Darmofal, D., Haimes, R.: An analysis of 3D particle path integration algorithms. *J. Comput. Physics* 123(1), 182–195 (1996)
11. Doucet, A., de Freitas, N., Gordon, N. (eds.): *Sequential Monte Carlo Methods in Practice*. Springer, New York (2001)
12. Koch, M., Norris, D., Hund-Georgiadis, M.: An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* 16(1), 241–250 (2002)

Rotational Encoding of C-arm Fluoroscope with Tilt Sensing Accelerometer

Victor Grzeda and Gabor Fichtinger

Queen's University, Kingston, ON, Canada
v.grzeda@queensu.ca, gabor@cs.queensu.ca

Abstract. *Purpose:* Accurate, practical, and affordable joint encoding on legacy C-arm fluoroscopes is a major technical challenge. Conventional pose tracking methods, like optical cameras and radiographic fiducials, are hampered by significant shortcomings. *Methods:* We propose to retrofit legacy C-arms with a tilt sensing accelerometer for rotation encoding. Our experimental setup consists of affixing an accelerometer to a full scale C-arm with a webcam as an alternative to X-ray imaging for this feasibility research. Ground-truth C-arm poses were obtained from the webcam that tracked a checkerboard plate. From these we constructed a series of angle and structural correction equations that can properly relate the accelerometer angle readings to C-arm pose during surgery and compensate for systematic structural C-arm deformations, such as sagging and bending. *Results:* Real-time tracking of the primary and secondary angle rotations of the C-arm showed an accuracy and precision of less than 0.5 degrees in the entire range of interest.

Keywords: fluoroscopy, C-arm, tracking, encoding, accelerometer.

1 Introduction

C-arm fluoroscopes are ubiquitous in computer-assisted interventions. They are versatile, compact, and mobile real-time X-ray imaging devices. The basic use of a C-arm is to acquire 2D X-ray images that can be reconstructed into three-dimensional representation. The reconstruction process requires that the relative poses of the 2D projection images must be known. Accurate, practical, and affordable C-arm pose tracking is a major technical challenge in using manually operated and un-encoded conventional C-arm fluoroscopes.

The motivation for this project is intra-operative implant reconstruction in prostate cancer brachytherapy [3]. C-arm images are acquired of the implanted prostate, followed by a 3D reconstruction of implants relative to the prostate gland and other surrounding structures observed in ultrasound. Then dosimetry is computed and the remainder of the implant procedure is re-optimized. For a successful implant reconstruction, the C-arm pose needs to be recovered with ideally less than 1° error [3,4,6]. Various constraints imposed by potential collisions with patient, operating table, and standard brachytherapy instrumentation limit the usable range of the C-arm to about a 30° cone in the canonical vertical position (Figure 1, left).

Conventionally, C-arm pose tracking takes one of two forms. External tracking uses a dynamic reference body that is attached to the C-arm and sensed by optical cameras to recover the pose [5]. Optical cameras can recover the C-arm pose accurately and are resistant to common problems that plague C-arms such as wheel motion and structural deformation. However, shortcomings lie in the need for direct line of sight between the reference body and the camera, leading to a complex, labor-intensive, and ultimately costly setup for most care facilities. Image-based tracking uses a radio-opaque fiducial placed in the field of imaging [3,4,7]. The precision-machined fiducial has known geometry allowing for the C-arm pose to be computed relative to the fiducial in each individual fluoroscopic image. While fiducial tracking is inexpensive, potentially accurate, and universally usable for any C-arm, it also has disadvantages. The fiducial occupies valuable real estate in the image, forcing the prostate to be positioned toward the edge of the detector, where image distortion is larger and may demand online distortion correction. Equally as important, the image needs careful processing and the fiducial must be segmented, which has been a singular point of failure in the clinical procedure [3,4]. Accelerometers have been considered for full position tracking, but those studies revealed accumulating drift error in measuring the pose [8], which thus far has prohibited the application of accelerometers in surgical guidance.

In the operating room, the brachytherapy crew is under incessant pressure to complete the image acquisition in minimum time. In such haste, more often than not the C-arm is still in motion when the X-ray image is acquired. Even if the joints look stationary, the C-arm keeps rocking on the rubber wheels and the gantry is swinging. Accelerometers detect these motions and thus enable us to gait the image acquisition accordingly. If one can guarantee stationary wheel and gantry positions, then the full pose the C-arm image can be recovered from joint encoders alone, assuming that the C-arm's kinematics had been obtained *a priori* in a one-time calibration.

We revisited the use of accelerometers that are well-suited to accurately indicate transient motion while they can be configured as tilt sensors. Preliminary bench work explored the feasibility of this approach on a crudely built and downscaled C-arm analogue [1]. A plethora of technical obstacles arises in translating this approach to a full scale clinical C-arm. This paper explains these barriers and having overcome them to achieve affordable and clinically accurate rotational encoding on a full scale C-arm fluoroscope, with the use of an accelerometer configured as tilt sensor.

2 Methodology

2.1 Calibration

The accelerometer angle encoding technique requires a one-time calibration step to relate the raw accelerometer angle to the C-arm pose. The linear workflow consists of several steps: (1) Mount the accelerometer on the C-arm; (2) Place C-arm tracking fiducial in the field of view; (3) Acquire test images while logging the initial rotation pose from the accelerometer; (4) Compute reference poses using the fiducial; (5) Compute the offset for both primary and secondary C-arm angles; (6) Confirm the accuracy of calibration on subset of independent measurements left out from the calibration computations. Preliminary investigation of this technique on a down-scaled

C-arm analogue was promising [1], but it revealed major uncertainties in translating this approach to a full scale C-arm. One such uncertainty is the effect of wheel motion. Mobile C-arms use wheel locks to secure the device in place, but these locks are not perfect. During rotational movement the C-arm tends to experience rocking and swinging motions that affect the pose of the image. In our case, since an accelerometer is used, real-time analysis of acceleration readings can detect acceleration spikes. These spikes indicate rapid motion changes (rocking, swaying) of the C-arm. By continuously monitoring for the spike to subside (steady state) a *go-ahead* signal can be flagged to indicate when images can be captured avoiding added error to the image pose.

The main area of concern is the deformation of the C-arm structure due to the force of gravity acting on the heavy X-ray source and detector. This deformation could cause an unaccounted for offset that would lead to inaccurate tracking of the C-arm pose. We mounted a webcam on a full scale C-arm unit, shown in Figure 1, *right*. Using webcam allows for conducting tests without exposure to ionizing radiation.

For the technique to be applied properly requires two key parameters that are determined from the C-arm and accelerometer combination. The first are Angle Correction Equations (ACEs) that account for the initial offset between the raw accelerometer angle readings to the C-arm pose. The second is a series of Structural Compensation Equations (SCEs) that account for the deformation of the C-arm during different rotational poses. This technique requires a one-time calibration of the C-arm to determine the ACEs and SCEs. These equations remain valid for continued use of the C-arm given that the accelerometer remains fixed on the gantry.

Our methodology of acquiring a series of test images and modeling the structural deformation of the C-arm is similar to Gorges et al [2]. Their work involved modeling the deformation as changes to the intrinsic/extrinsic parameters of the C-arm. Then they used cost functions to optimize the parameters to account for the deformations.

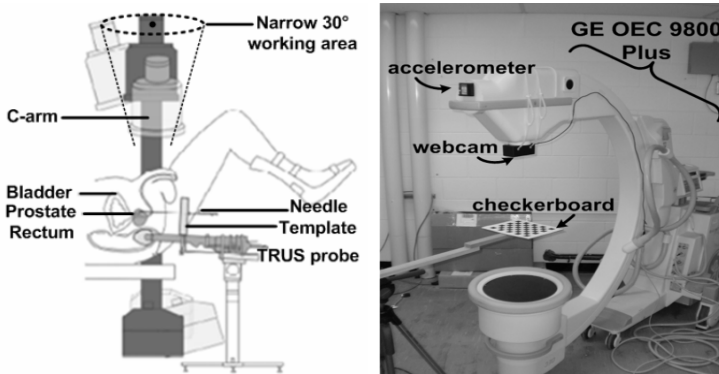


Fig. 1. (*Left*) Typical ultrasound guided brachytherapy setup with a C-arm operating in a narrow range. (*Right*) Full scale experimental setup with the accelerometer mounted on top and the webcam underneath the X-ray source. More details of devices can be found in [1].

2.2 Angle Correction Equations

The Angle Correction Equations (ACEs) quantify the offset between the webcam C-arm pose and the accelerometer angle readings. Generating the ACEs involves analyzing the differences between the initial accelerometer angle reading and ground-truth pose angle. In an actual clinical application, the ground truth pose will be computed using a radio-opaque fluoroscope tracking fiducial such as the one developed by Jain et al [3]. In our experimental setup, we applied optical imaging with a webcam mounted externally on the gantry over the exit point of the X-ray source. The webcam tracked a checkerboard pattern that provided the ground-truth C-arm poses of the 2D images through standard camera calibration techniques.

To determine the initial offset a set of test images were taken, in step-and-shoot mode, at the angles $[0^\circ \pm 5^\circ \pm 10^\circ \pm 15^\circ \pm 20^\circ \pm 25^\circ \pm 30^\circ]$ using the initial accelerometer angle reading for the primary angle (PA) and secondary angle (SA) independently. These were chosen to give enough separation between angles with a wide range to account for changes in the C-arm motion, as well as encapsulating the clinical angle limitation motivating this work. It should be noted that acquiring more angles could help to improve modeling, but when this will be used in X-ray imaging that would increase radiation exposure. Further testing will try to reduce the angle set and examine if proper tracking can be still achieved. Next, the actual pose angles of the images were computed and the differences calculated. By graphically visualizing the differences (Figure 2), best fit methods were applied creating the ACEs.

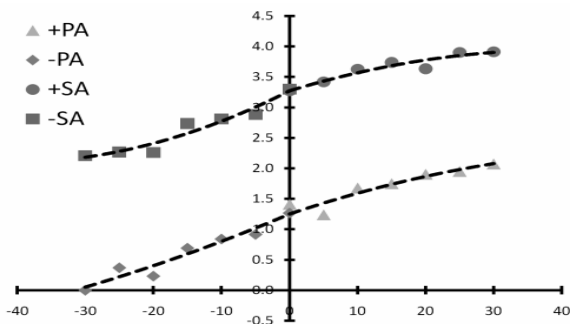


Fig. 2. A plot of the differences and fitted lines creating the Angle Correction Equations

The *Eq. 1* and *Eq. 2* are the resulting ACEs. Interestingly, the structure of the C-arm creates non-linear trends that required piecewise functions for modeling. The intercept values in the functions were set the same to avoid sudden shifts in the output during sign changes of the angles when the accelerometer is operating.

Let x represent the initial *PA* then the new output angle

$$PA^* = \begin{cases} -0.0003 \cdot x^2 + 1.037 \cdot x + 1.250 & | x \geq 0 \\ 0.0002 \cdot x^2 + 1.047 \cdot x + 1.250 & | x < 0 \end{cases}, \quad (1)$$

Let z represent the initial SA then the new output angle

$$SA^* = \begin{cases} -0.0004 \cdot z^2 + 1.034 \cdot z + 3.270 & | x \geq 0 \\ 0.0007 \cdot z^2 + 1.051 \cdot z + 3.270 & | x < 0 \end{cases} \quad (2)$$

2.3 Structural Compensation Equations

The full scale C-arm presented a major challenge due to the significant structural deformation. An initial test of the ACEs (Eq. 1 and Eq. 2) showed that proper pose tracking was achieved, but only for independent angle rotations. Figure 3, displays a chart that illustrates the problem.

ROTATION		PRIMARY ANGLE													
		DEG°	30	25	20	15	10	5	0	-5	-10	-15	-20	-25	-30
SECONDARY ANGLE	30								-1.29						
	25		24.07						29.99						-25.94
	20		27.68	18.91					-1.19					26.86	
	15			21.38	14.08				-1.12					-20.86	
	10				15.58	9.14			20.05					21.19	
	5					10.15	4.60		-0.66					-15.67	
	0						4.98	4.60	15.09					15.54	
	-5							4.98	-0.51					-10.70	
	-10								10.05					10.68	
	-15									-0.38				-5.64	
	-20									5.18				5.53	
	-25										-5.64				
	-30											-5.64			
0		29.75	24.59	19.77	14.77	9.58	4.68	-0.23	-5.13	-10.14	-14.88	-20.06	-25.09	-29.95	
-5		-0.66	-0.72	-0.66	-0.69	-0.65	-0.62	0.12	0.44	0.44	0.00	-0.31	-0.77	-1.22	
-10							4.96	0.02	-5.07						
-15							-4.95	-4.71	-4.28						
-20								10.05	0.26						
-25								-9.66	-9.68						
-30									0.55						
									-9.85						
									-9.68						
									0.46						
									-14.62						
										-9.59					
											-14.90				
											-15.29				
												-19.99			
												-21.75			
													-24.71		
													-28.40		
														-30.23	

Fig. 3. [inside cells – actual C-arm pose angles: PA (shaded) | SA below || outside border – C-arm pose positions using ACEs accelerometer readings] The chart portrays that PA and SA are accurately tracked independently, but not very well simultaneously

To correct this problem required a different approach than a simple offset correction, like the ACEs. The fact is the ACEs did accurately track the PA and SA independently. What is posited here is that the structural deformation of the C-arm has dependency on PA and SA. As the rotation angle of the source or detector moves further from vertical the effects of gravity become greater. This dependence on the pose angle of the C-arm required us to compensate for the structural deformation, thus creating the Structural Compensation Equations.

The Structural Compensation Equations (SCEs) were formulated by reprocessing the test images that created the ACEs, with additional diagonal combinations of PA

and SA, shown in Figure 3. The errors were computed and zeroed to visualize the deformation for both the PA and SA separately. Line fitting was applied to the differences to create a series of equations, shown in Figure 4.

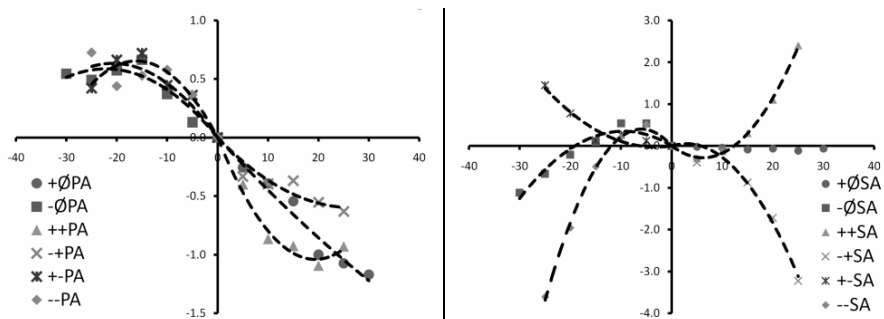


Fig. 4. The visualization of the C-arm deformations with accompanying best fit lines for PA (*left*) and SA (*right*). The horizontal axis represents the rotational angle, while the vertical indicates the amount of deformation.

The plots show that a significant amount of deformation exists and by zeroing the error creates a single starting position to make compensation easier to handle. A threshold value of τ helped to control when to activate compensating depending on the current angle values for PA and SA. The equations are piecewise functions separated into $\pm PA$ and $\pm SA$ quadrants shown here:

Let $w = PA^*$ and $x = SA^*$ then the compensated angle

$$PA^\dagger = \begin{cases} w + (0.0002 \cdot x^2 - 0.047 \cdot x) & x \geq 0 \\ w + (-0.0012 \cdot x^2 - 0.053 \cdot x) & x < 0 \\ w + (0.0028 \cdot x^2 - 0.107 \cdot x) & w > \tau, x > \tau \\ w + (0.0008 \cdot x^2 - 0.045 \cdot x) & w < -\tau, x > \tau \\ w + (-0.0025 \cdot x^2 - 0.081 \cdot x) & w > \tau, x < -\tau \\ w + (-0.0015 \cdot x^2 - 0.061 \cdot x) & w < -\tau, x < -\tau \end{cases}, \quad (3)$$

$$SA^\dagger = \begin{cases} x & w \geq 0 \\ x + (-0.0038 \cdot w^2 - 0.073 \cdot w) & w < 0 \\ x + (0.0075 \cdot w^2 - 0.093 \cdot w) & w > \tau, x > \tau \\ x + (0.0030 \cdot w^2 + 0.019 \cdot w) & w < -\tau, x > \tau \\ x + (-0.0065 \cdot w^2 - 0.038 \cdot w) & w > \tau, x < -\tau \\ x + (-0.0113 \cdot w^2 - 0.135 \cdot w) & w < -\tau, x < -\tau \end{cases}. \quad (4)$$

3 Results and Discussion

In order to characterize how accurately the ACEs and SCEs tracked the webcam C-arm pose we performed a series of tests. For the first test, the test image angles were changed for PA and SA to $[0^\circ \pm 2^\circ \pm 6^\circ \pm 10^\circ \pm 14^\circ \pm 18^\circ \pm 22^\circ \pm 26^\circ]$ for both independent and combinational rotation of PA and SA. This change was made to contain different angles than the ones used to create the equations, so that they can be independently evaluated. The results, in chart form, are shown Figure 5.

ROTATIONS		PRIMARY ANGLE															
DEG ^o		26	22	18	14	10	6	2	0	-2	-6	-10	-14	-18	-22	-26	
SECONDARY ANGLE	26								0.04								
	22	22.34							26.26								
	18	21.45	18.34						0.11							-21.74	
	14		17.74	14.47					22.20							21.55	
	10			13.91	10.35				0.17							-17.99	
	6				9.86	6.25			18.99							17.91	
	2					5.61	2.05		0.30							-13.89	
	0	26.04	21.87	18.04	14.02	10.12	6.02	2.31	0.04							13.99	
	-2	-0.30	-0.30	-0.43	-0.68	-0.56	-0.77	-0.61	0.29	-1.69	-5.87	-9.64	-13.69	-17.17	-21.62	-25.70	PA
	-6								0.20	0.25	0.42	0.27	0.22	0.04	0.02	0.07	SA
	-10								2.08	-2.04							
	-14								-1.91	-1.84							
	-18								0.15	-6.08							
-22								-5.72	-5.71								
-26								0.06	-9.79								
								-9.49	-9.70								
								0.17	-14.03								
								-13.87	-13.56								
								0.15	-18.22								
								-18.07	-17.33								
								0.10	-22.07								
								-21.88	-21.20								
								0.02									
								-25.99									

Fig. 5. Results from using the ACEs and SCEs

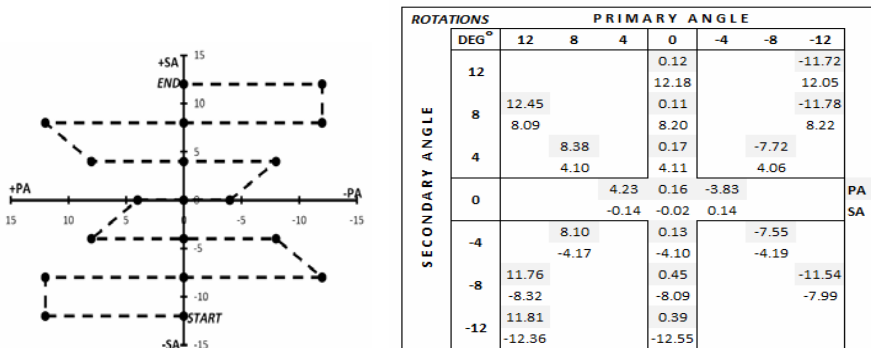


Fig. 6. The path of C-arm where the dots are the locations that the images acquired

The chart in Figure 5 suggests that, using the ACEs and SCEs the C-arm pose was tracked with an average error of $\mu = 0.11^\circ$ and standard deviation of $\sigma = 0.21^\circ$ for PA.

While for SA an average error of $\mu = 0.08^\circ$ and standard deviation of $\sigma = 0.36^\circ$. A second test was performed to show if the C-arm pose can be tracked when subjected to sequential rotation path with results shown in Figure 6.

We achieved C-arm rotation encoding by using accelerometer as tilt sensor, with an accuracy and precision of less than 0.5° . This was made possible by introducing the Structural Compensation Equations (SCEs). These equations worked in tandem with the Angle Correction Equations (ACEs) to compensate for the inherent structural deformation of the C-arm experienced at different rotational poses. We received excellent rotation encoding for the entire clinically relevant C-arm working range in prostate brachytherapy, our driving clinical application.

For other clinical application areas where the working range is larger, it is evident from Figure 5 that accelerometer tracking is not perfect where accuracy starts to waiver at the larger combined rotational poses. This effect could be attributed to the threshold values for when the SCEs need to start compensating for the structural deformation. The values control the activation of these equations, but it is not clear on when they should start. In brachytherapy, the SCEs were created by zeroing the offset, but it may be possible to start monitoring for changes farther away from the origin. The reasoning behind this is that the lower angles about the vertical axis would have less deformation caused by gravity and could still be within the ACEs capability of tracking. The next step in continuing research will be to determine a threshold location rather than a singular conditional value.

In conclusion, C-arm rotation tracking was successful with using the tilt-sensing capabilities of accelerometers. Work continues to implement this concept on fully functional X-ray imaging C-arms, with the ultimate goal of clinical deployment of an accurate, inexpensive and easy-to-use rotation technique in prostate brachytherapy.

References

- [1] Grzeda, V., Fichtinger, G.: C-arm rotation encoding with accelerometers. *International Journal of Computer Assisted Radiation and Surgery* 5(4), 385–391 (2010)
- [2] Gorges, S., Kerrien, E., Berger, M.-O., et al.: Model of a Vascular C-Arm for 3D Augmented Fluoroscopy in Interventional Radiology. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3750, pp. 214–222. Springer, Heidelberg (2005)
- [3] Jain, A.K., Deguet, A., Iordachita, I., et al.: Intra-operative Guidance in Prostate Brachytherapy Using an Average C-arm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II*. LNCS, vol. 4792, pp. 9–16. Springer, Heidelberg (2007)
- [4] Lee, J., Liu, X., Jain, A.K., et al.: Prostate brachytherapy seed reconstruction with Gaussian blurring and optimal coverage cost. *IEEE Trans Med. Imaging* 28(12), 1955–1968 (2009)
- [5] Peters, T., Cleary, K. (eds.): *Image-Guided Interventions: Technology and Applications*. Springer, Heidelberg (2008)
- [6] Tubic, D., Zaccarin, A., Beaulieu, L., Pouliot, J.: Automated seed detection and three-dimensional reconstruction. II. Reconstruction of permanent prostate implants using simulated annealing. *Medical Physics* 28(11), 2272–2279 (2002)
- [7] Yao, J., Taylor, R.H., Goldberg, R.P., et al.: A C-arm fluoroscopy-guided progressive cut refinement strategy using a surgical robot. *Comput. Aided Surg.* 5(6), 373–390 (2000)
- [8] Nikbakht, S., Mazlom, M., Khayatian,.: Evaluation of solid-state accelerometer for positioning of vehicle. In: *IEEE Int'l. Conference on Industrial Technology*, pp. 729–733 (2005)

Robotic Hand-Held Surgical Device: Evaluation of End-Effector's Kinematics and Development of Proof-of-Concept Prototypes

Ali Hassan Zahraee¹, Jérôme Szewczyk¹,
Jamie Kyujin Paik², and Guillaume Morel^{1,*}

¹ Institut des Systemes Intelligents et de Robotique
Université Pierre et Marie Curie - Paris VI, 4, place Jussieu, 75005, Paris, France

² Harvard Microrobotics Lab

Abstract. We are working towards the development of a robotic hand-held surgical device for laparoscopic interventions that enhances the surgeons' dexterity. In this paper, the kinematics of the end effector is studied. Different choices of kinematics are compared during an evaluation campaign using a virtual reality simulator to find the optimal one: the Yaw-Roll (YR) kinematics. A proof of concept prototype is made based on the results.

1 Introduction

Minimally invasive surgery (MIS) causes less operative trauma and leaves patients with less pain and scarring, speeds recovery, and reduces the incidence of post-surgical complications. Conventional instruments used in MIS are hand-held instruments with long shafts, an end effector (needle holder, dissector etc.) at one end and a handle at the other. The instrument motion is constrained at the pivot point to 4 degrees of freedom (DOF): (1) translation along the shaft of the instrument, (2) rotation around the translational axis and (3) and (4) limited inclination of the shaft pivoted through the incision [1]. Some gestures are very difficult or impossible to make using the non-dexterous conventional instruments. One could imagine a more dextrous device with a jointed end effector adding one or more DOF. The end effector must have at least 6 DOF to allow the surgeon choose the orientation and position of the end effector arbitrarily. The DOF added to the end-effector could be actuated manually, pneumatically or electrically. The latter gives a mechatronic (robotic) hand-held instrument. But making such a miniature-scale instrument with a mechanical force transmission system that can provide for the requirements in MIS, is difficult and costly. So, choosing the simplest kinematics for the added DOF that allows performing all needed movements is critical. In [2] for example, the end-effector can yaw or pitch while the surgeon can roll the instrument's shaft using his thumb. [3] has

* This work was supported by Agence Nationale de la Recherche (ANR) fund ANR-09-CORD-020.

an end-effector that can yaw, pitch and roll, but its shaft's rotation is manual and thus limited. In [4], the end-effector can yaw and roll. The *da Vinci* surgical system [5] has end-effectors that can yaw, pitch and roll. All these instruments however, have 10 mm thick shafts. Surgeons on the other hand, demand 5 mm instruments for better integration in Single-access or NOTES operations. In this paper we explain our efforts towards the development of a robotic dexterous hand-held instrument for laparoscopy with a 6 mm shaft.

2 Simulation

To evaluate and compare different handles, control modes and kinematics, we made a virtual reality (VR) simulator. Fig. 1(a) show the simulator.

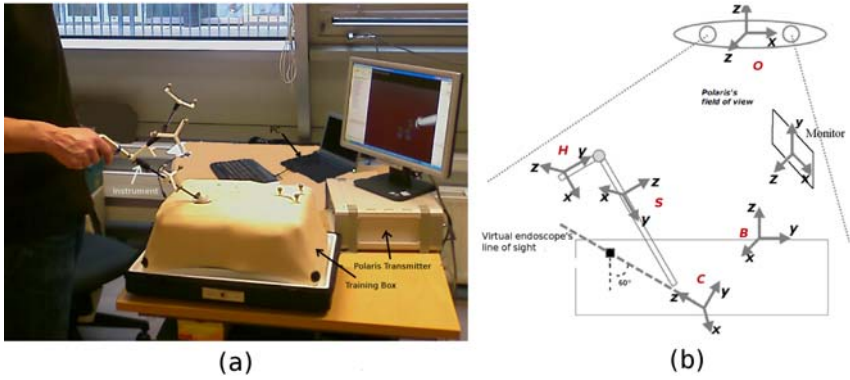


Fig. 1. (a) Simulator in use, (b) Local coordinate systems of the simulator

Polaris, a motion tracking system, keeps track of the position and the orientation of each rigid body in the scene: the training box and the instrument (a Polaris target with a local coordinate system is attached to each of them). Fig. 1(b) shows different coordinate systems present in the scene. The surgical instrument consists of a shaft and a finger-operated handle, (we used a NunchuckTM from a Wii video game console). The handle has a 2 DOF joystick under the thumb and 2 buttons under the index finger. The kinematics of the instrument's virtual end-effector can be programmed in the simulator. A 19" LCD monitor is used to show the simulated endoscopic image of the inside of the training box. It is positioned 1m away from the operator and deviated 45° from his line of sight to resemble the situation in an operation room. The control program runs on a PC. The pose of each body is calculated in the virtual endoscope's coordinate system. This virtual endoscope's position on the training box, its line of sight and its scope can be chosen arbitrarily. We chose a triangle with 10 cm sides to place the instrument and endoscope, a line of sight inclined 60° from vertical and a 75° scope. The virtual image of the inside of the box is finally rendered

using the OpenGL 2.0 library. There is no force feedback, but a visual feedback that indicates collisions between the needle and the working surface. The image shows the instrument with its end effector holding a needle, and a working surface with a grid and suturing points identified by different colors. Fig. 2 shows suturing in the simulated scene. Evaluations are based on subject performance in making sutures. One of the main advantages of a dexterous surgical device is considered to be its ability to make sutures in difficult angles e.g. sagittal sutures. Our simulated suturing task includes putting the needle in the right orientation so as to insert it in the working surface with the right angle, reaching the suture point and turning the needle to bring it out of the exit point. This is considered the end of the gesture. This series of motions (orient, reach, orient) was chosen based on previous studies done on decomposition of laparoscopic tasks [6], [7].

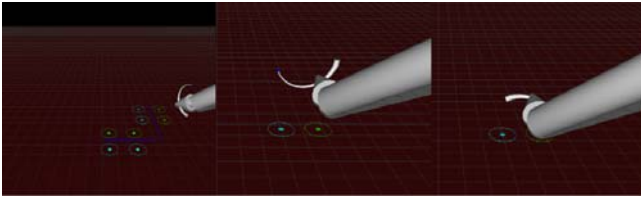


Fig. 2. From left to right: snapshots from the simulated scene showing a suture

2.1 Evaluation of Different Kinematics for the End-Effector

Reduced Set of Kinematics. Our objective was to compare different kinematics for the end effector and choose the optimal one. Surgeons need to be able to suture in different angles (frontal/sagittal). As a result they need 6 DOF needle holders. Using more than 6 DOF makes the task of visumotor control more difficult for the operator. It would not be possible to control the end-effector in its working space either, as 4 of the DOF are exclusively controlled manually. The question we are studying here is which 6 DOF kinematics is the best one for a hand-held laparoscopic instrument, knowing that 4 DOF are already defined as a result of the instrument being constrained by the pivot point. Keeping the already existing 4 DOF, we should add 2 more DOF to the end effector. The 2 DOF we want to add, are those of a 2 DOF wrist added to the end-effector. Without loss of generality, we suppose that the 2 revolute joints of the wrist have concurrent axes. For 2 revolute joints with concurrent axes, 6 combinations of rotational axes are possible. These combinations are: pitch-yaw, pitch-roll, yaw-pitch, yaw-roll, roll-pitch and roll-yaw. But pitch-yaw is the same as yaw-pitch if we only turn the shaft 90° . Pitch-roll and yaw-roll are also the same. Roll-pitch and roll-yaw are singular combinations. This leaves us with 2 possible combinations: yaw-roll (YR) and yaw-pitch (YP). The rotation of the instrument around its longitudinal axis is normally manual. Surgeons have to rotate their whole arm to rotate the instrument and still, the rotation is limited. So we decided to make this rotation automatic as well, giving subjects the ability to rotate the shaft

clockwise and counter clockwise using 2 buttons. Another possibility we thought about was to add another DOF to the wrist to make it a 3 DOF wrist (like [3] and [5]). Again, there are 6 possible combinations of 3 concurrent rotational axes. 2 of them are singular (RYP and RPY). The other 4 are become the same from the operator's point of view just by rotating the device and are equivalent to a YPR kinematics. A 3 DOF wrist makes the total number of DOF 7, with 3 of them controlled by fingers. The rotation of shaft stays manual. So the total possible kinematics for 2 and 3 DOF wrists with concurrent axes are reduced to 3 kinematics: YP, YR and YPR, shown in Fig. 3.

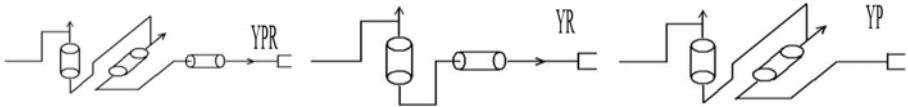


Fig. 3. 3 kinematics for end effector wrist

Evaluation Methodology and Metrics. For each evaluation, we asked test subjects to do frontal and sagittal sutures on a horizontal virtual working surface inside the training box. The suture points were identified by different colors. The subjects were engineering students with no experience in laparoscopy. Literature suggests that expert laparoscopic surgeons are significantly different from surgical novices in terms of applied forces and torques [8], [9], patterns of movement [10], task completion times [8], [9], [11], [10], trajectory length [10] and number of errors [12]. However, these studies that are mostly done for the purpose of modeling surgical gestures in laparoscopy and providing metrics for objective assessment of skills in virtual reality simulators, are done using either conventional laparoscopic instruments with 4 DOF or the da Vinci surgical system. When it comes to instruments with novel human-robot interfaces and different kinematics, expert surgeons are probably not greatly different from novices. In fact, the additional DOF and the method of controlling them may be as new to them as it is to the novices. As a result, surgeons and novices will both use their basic visumotor skills to execute the new tasks. This strongly suggests that the results of our studies would be the same, had we used expert surgeons as subjects. This has of course to be proven with experimental data in a separate study using expert surgeons as subjects. Besides, it is not even sure that expert surgeons do better than novices with these novel instruments as they do with conventional instruments or the da Vinci. For example, there is evidence that playing video games could improve surgical skills in minimally invasive surgery [13], [14]. Younger subjects though surgically novice, have generally more experience with video games and the joysticks used to play them than middle age expert surgeons.

The metric used in the evaluations is the time to completion of task (TCT). [15] states that the TCT is a practical, easy and valid objective tool for assessing acquired technical skills of urology trainees in a laparoscopic simulated environment. It is also used for comparing different surgical instruments for

laparoscopy [16]. Each user made 5 frontal and 5 sagittal sutures using each of the kinematics and his average TCT for 1 suture was calculated.

Results and Discussion. Fig. 4 shows the average TCT for each of the 15 subjects. Table 1 shows the mean TCT of all subjects for each of the 3 kinematics tested. The results show that the YR kinematics is slightly better than the YPR kinematics, and both of them are largely better than the YP kinematics, in terms of TCT. Only 3 out of 15 subjects were able to make sutures with the YP kinematics in the 3 minute per suture time limit. From the evaluation results of different kinematics for the end effector, we could see that an end effector able to yaw and roll results in the least TCT for suturing. Technologically, it is much more affordable to make a 2 DOF mesoscale wrist than a 3 DOF one. At the same time, it is dexterous enough to allow suturing in different angles. The YR kinematics of the end-effector plus rotation of the shaft, give surgeons 6 DOF instruments.

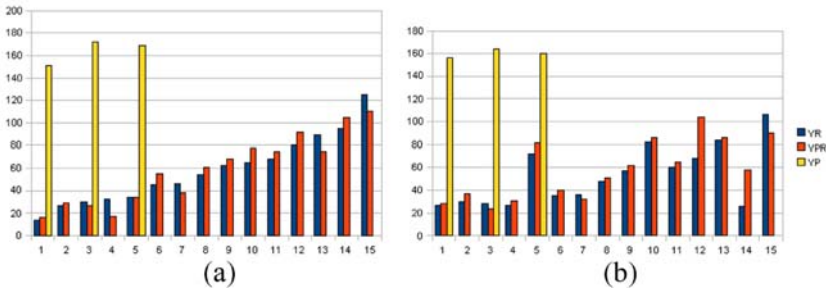


Fig. 4. Average TCT in seconds for 15 subjects using 3 different end effector wrist kinematics

Table 1. Mean time to completion for YR, YP and YPR kinematics in seconds

	Frontal	Sagittal
YR	57.73	58.67
YPR	52.4	58.33
YP	>> YR, YPR	>> YR, YPR

3 Mechanical Design

3.1 Mechanical Characteristics of the Developed System

Our second objective was to design the force transmission system and make a proof-of-concept prototype pincer for the chosen type of kinematics. The essential DOF required for the instrument pincer tip during a suturing procedure is in two independent rotational axis movements: roll and yaw. While it is crucial

to execute the full range rotations in required orientation, it is also important to maintain the position in the proposed configuration with a high stiffness and rigidity. The mechanical design challenge lies in developing an instrument in thin (5mm) cylindrical shape. Here we introduce two prototypes that can produce two independent and simultaneous orthogonal rotations amid the constraints imposed by practical usage in operating rooms.

Metallic Bellow Model. In order to transmit the rotational movement of the pincer tip (roll), while the body is in motion (yaw), a metallic bellow is used. This inner bellow tubing is actively bended by wedged sleeves and is completely independent from the outer rings which have chiseled slopes for making the curvature which are 90 deg = fold and 180 deg = tension. Such actuation is carried out by two cables (0.3mm dia. 5kg load multi-stranded steel cable) attached on the sides. A single cable (multi stranded stainless cable 0.75mm dia.) actuated pincer is affixed to the bellow joint. The cable is flexible yet robust to take the shape of the outer structure that controls the yaw direction of the end pincer. This pincer assembly rotates freely from the outer shell while maintaining its longitudinal position by a polymer bearing. The wedged sleeve links are formed that the instrument tip operates in either 0 or 90 degrees positions: The rigidity of the instrument is guarded and controlled by the cable tension on the side. Also, here it can be noted that the assembly configuration can be reversed to have the bellow either on the inside or outside.

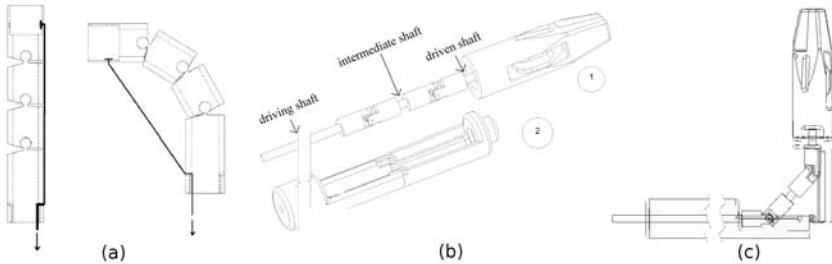


Fig. 5. (a) Metallic bellow model’s wedged shell assembly with cable transmission for bending, (b) Universal joint and pincer assembly: (1) double universal joint and pincer assembly (2) shell unit, (c) Instrument in 90 degrees yaw position

Universal Joint Model. In this model, the bending in yaw direction is actuated by the linear translation of the outer shell (Fig. 5(b2)) with respect to the two universal joints and the pincer assembly (Fig. 5(b1)). Depending on the advancement of the shell unit that can bend maximum 90 degrees from its 0 degree straight position, the universal joint unit can transmit the rotation at the pincer tip. Due to the nature of the universal joint (gimbal adjoining unit fixes two rotating shafts), the rotary transmission experiences jerks and sinusoidal rotational velocity at the driven shaft: at 45 degrees off set of rotating axes, sinusoidal rotational velocity variance is about 40% at the driven shaft. The use of double universal joints minimizes this effect.

Fig. 6 shows shows the designed and manufactured prototypes. They are both 6mm- our primary objective was 5mm- in the diameter and use the same needle holder pincer tips and are fabricated in stainless steel. Current prototypes do not have the force/torque requirements for suturing and need to be improved for an animal experiment.



Fig. 6. (a) The two developed end-effectors, (b) Photograph of one of the developed prototypes

4 Conclusion and Future Work

The results of our evaluations suggest that the YR kinematics is the optimal kinematics for the end-effector. 2 proof of concept prototypes of such an end-effector were made introducing novel designs. The device needs a handle with an ergonomic design and a more robust force transmission system enabling the surgeon to suture different types of tissue. A new version of the device is under development.

References

1. Lai, F., Howe, R.: Evaluating control modes for constrained robotic surgery. In: Proceedings of ICRA 2000, vol. 1, pp. 603–609 (2000)
2. Danitz, D.J.: Articulating mechanisms with joint assembly and manual handle for remote manipulation of instruments and tools. United States Patent Application Publication, no. US 2006/0201130 A1 (2006)
3. Lee, W., Chamorro, A.: Surgical instrument. United States Patent, no. US 7,338,513 B2. Cambridge Endoscopic Devices, Inc., Framingham (2008)
4. Jinno, M., Sunaoshi, T., Omori, S.: Development of a Master Slave Combined Manipulator for Laparoscopic Surgery Functional Model and Its Evaluation. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 52–59. Springer, Heidelberg (2002)
5. Guthart, G.S., Salisbury, J.K.: The Intuitive telesurgery system: overview and application. In: Proceedings of ICRA 2000, vol. 1, pp. 618–621 (2000)
6. Payandeh, S., Lomax, A.J., Dill, J., MacKenzie, C.L., Cao, C.: On defining metrics for assessing laparoscopic surgical skills in a virtual training environment. In: MMVR (2002)

7. MacKenzie, C.L., Ibbotson, J.A., Cao, C., Lomax, A.J.: Hierarchical decomposition of laparoscopic surgery: a human factors approach to investigating the operating room environment. *Min. Invas. Ther. & Allied Technol.* 10(3), 121–127 (2001)
8. Rosen, J., Brown, J.D., Barreca, M., Chang, L., Hannaford, B., Sinanan, B.: The blue dragon - a system for monitoring the kinematics and the dynamics of endoscopic tools in minimally invasive surgery for objective laparoscopic skill assessment. *Studies in Health Technology and Informatics - Medicine Meets Virtual Reality*, vol. 85, pp. 412–418. IOS Press, Amsterdam (2002)
9. Rosen, J., Brown, J., Chang, L., Hannaford, B.: Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. *IEEE Transactions in Biomedical Engineering* 53(3), 399–413 (2006)
10. Gallagher, A.G., Richie, K., McClure, N., McGuigan, J.: Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. *World Journal of Surgery* 25, 1478–1483 (2001)
11. Oleynikov, D., Solomon, B., Hallbeck, M.S.: Effect of visual feedback on surgical performance using the da vinci surgical system. *Journal of Laparoendoscopic & Advanced Surgical Techniques* 16(5) (2006)
12. Law, B., Atkins, M.S., Kirkpatrick, A.E., Lomax, A.J.: Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In: *ETRA 2004: Proceedings of the 2004 symposium on Eye tracking research & applications*, pp. 41–48. ACM, New York (2004)
13. Rosser, J.C., Lynch, P.J., Cuddihy, L., Gentile, J., Klonsky, D.A., Merrell, R.: The impact of video games on training surgeons in the 21st century. *Archives of Surgery* 142(2), 181–186 (2007)
14. Reilly, M.: A wii warm-up hones surgical skills. *The New Scientist* 197, 24 (2008)
15. Mishra, S.K., Ganpule, A., Kurien, A., Muthu, V., Desai, M.R.: Task completion time: Objective tool for assessment of technical skills in laparoscopic simulator for urology trainees. *Indian J. Urol.* 24, 35–38 (2008)
16. Dakin, G., Gagner, M.: Comparison of laparoscopic skills performance between standard instruments and two surgical robotic systems. *Surgical Endoscopy* 17(4), 574–579 (2003)
17. Kode, V., Cavusoglu, M.: Design and characterization of a novel hybrid actuator using shape memory alloy and dc micromotor for minimally invasive surgery applications. *IEEE/ASME Transactions on Mechatronics* 12, 455–464 (2007)

Guide-Wire Extraction through Perceptual Organization of Local Segments in Fluoroscopic Images^{*}

Nicolas Honnorat^{1,2,3}, Régis Vaillant³, and Nikos Paragios^{1,2}

¹ Laboratoire MAS, Ecole Centrale Paris, Châtenay-Malabry, France

² Equipe GALEN, INRIA Saclay-Ile-de-France, Orsay, France

³ General Electric Healthcare, Buc, France

Abstract. Segmentation of surgical devices in fluoroscopic images and in particular of guide-wires is a valuable element during surgery. In cardiac angioplasty, the problem is particularly challenging due to the following reasons: (i) low signal to noise ratio, (ii) the use of 2D images that accumulate information from the whole volume, and (iii) the similarity between the structure of interest and adjacent anatomical structures. In this paper we propose a novel approach to address these challenges, that combines efficiently low-level detection using machine learning techniques, local unsupervised clustering detections and finally high-level perceptual organization of these segments towards its complete reconstruction. The latter handles miss-detections and is based on a local search algorithm. Very promising results were obtained.

1 Introduction

The detection and the segmentation of the guidewires (GW) used in cardiac angioplasty is a challenging problem in biomedical image analysis. It is often addressed into two steps: first, interest points are detected and then linked (or grouped) together into a curvilinear structure corresponding to the GW. Both steps inherit severe technical challenges.

Detection is required due to the low signal to noise ratio. The aim is either to improve contrast between the guidewire and the background/other related anatomical structures or detect feature points corresponding to it. The first scenario is often addressed using low level operators, like dedicated filtering [1] and coherence enhancing diffusion [2]. State of the art for detection includes dedicated edge-detection methods [1], method based on the Hessian eigenvalues such in [2] and such as the Vesselness measure [3,4], steerable filters [5], and phase congruency [6]. The second class of methods aims to address detection directly using either a voting schema [7], or machine learning methods [8]. The central idea is to use patterns of appearance corresponding to vessels and learn a classifier that is able to separate them from the background. Pose and scale parameters are

^{*} This work was supported by ANRT (grant 1062/2008) and GE Healthcare. The authors thank N. Komodakis for providing the clustering method.

the most critical aspects to be handled in this process. Augmenting the training set towards encoding all possible variations of appearance is not feasible since it deteriorates the performance. However, compared to filtering methods this concept is quite promising.

Chaining of detections towards complete recovery of the guidewire is often considered afterwards [1,7,4]. The use of snake-splines on the feature images has also been investigated [2,6]. These methods are very sensitive to the presence of outliers as well as miss-detections. Such shortcomings can be addressed through the use of hierarchical grouping guided by a classification [8]. In all cases, sensitivity to the presence of outliers as well as miss-detections are challenging issues to be addressed.

In this paper we introduce a novel approach combining detection using boosting [9], with a mid-level grouping scheme based on clustering and a complete reconstruction through the minimization of a global criterion that encodes geometric and detection consistencies - through local search based on an inlier/outlier permutation model. Contrary to the other methods which can fail due to noise or low contrast [2,6], or do not recover undetected parts [1] our approach has been designed to be as robust as possible and to recover the GW as global optimum.

The most closely related work with our approach can be found in [8,10]. The first approach is purely learning based, is not invariant to the guide-wire pose parameters and cannot handle well miss-detections. The second approach shares the feature detection concept (the considered features are less efficient than the ones considered in this paper). However, [10] does not perform ordering, and is only able to provide a local grouping of segments into parts using linear programming. The method is unable to deal well enough with miss-detections as well as outliers because the final result is not necessarily perceptually meaningful.

The remainder of this paper is organized as follows: in section 2 we discuss the robust extraction of line segments that we perform and we present our approach for organizing them in section 3. The next section presents the experimental validation. Discussion concludes the paper.

2 Robust Local Segments Extraction

Our algorithm consists of two steps: (i) a low-level detection of GW pixels, that are then grouped towards extracting line segments (ii) an ordering and removal of erroneous segments. The first task is addressed using boosting and unsupervised clustering, and the second through local permutation search.

2.1 Low-Level Detection

Boosting [9] refers to a powerful classification method that combines weak classifiers towards the creation of a strong one. This classification process is guided from the miss-detection/miss-classification error. Samples are weighted according to the classification error and these weights are continuously updated. Since in our data set, one expects important discrepancy between positive and negative samples (GW pixels represent less than 0.15% of whole image) as well as important presence of outliers, we adopt the variant called Gentle AdaBoost [11] and

we replaced its re-weighting formula driving the choice of the next weak classifier with the following (where w_j is the weight of a data x_j of label $y_j \in \{-1, +1\}$ and h^m is the weak classifier chosen during the current AdaBoost step) in order to make it asymmetric like [12]:

$$w_j \leftarrow w_j e^{-y_j h^m(x_j)} e^{y_j (\text{asymmetry})}$$

When training classifiers, we have considered around 500 000 samples, almost 87% of them being negative ones, and chosen an asymmetry of 0.1. The input to the classification process was features derived from the image after appropriate filtering and in particular with separable steerable filters [5] that are optimal in the sense of Canny criteria.

In addition to the steerable ridge detectors of fourth order of [5] for $\mu = 0.25$ and scales $\sigma = \{0.8, 1.2, 1.6, 2.0, 3.0\}$ and to the edge detector of third order of scale 5.0, we introduced three features specific to our problem: (i) Difference of Gaussian (DoG) computed with the raw, the log and the next images, (ii) variances of these features around the pixel considered (in a square of size 7), both for the couples of scales: $\{(0.8, 1.2)(1.4, 1.8)(2.0, 3.0)(3.0, 4.0)(4.0, 5.0)\}$; and (iii) variances computed on the 'subtraction image' (the current image minus the next one) in squares of size 5 and 9.

Log images have indeed a physical meaning and allow to better discriminate the structures from their background, but their level of noise has been locally modified, which produces numerous false detections. Considering both images allows to combine both advantages to avoid more false detections , and considering the next frame allows to better reject the confusing structures, because they are most of the time static (and variances provides a second order description).

Clustering these points towards extracting line segments both allows to reject remaining false detections due to noise (because most of them are isolated) and provides primitives we just need to link together (with a proper handling of outliers) to delineate the GW. The fact that clustering provides a global optimum guarantees far more robustness with respect to noise. The main challenge is that neither the number of segments/clusters nor the membership function for every detected point are known in advance.

2.2 Grouping

Let us consider the N candidate pixels $p_i = (x_i, y_i)$.

Clustering of these pixels can be formulated as follows: find a set of cluster centers $c_j = (cx_j, cy_j)$ (the unknown cardinality of this set is denoted as $|C|$) and an individual membership function \mathcal{L} (labels in $\{1, \dots |C|\}$) such that all detected points p_i with the same label form a local line segment with minimal dispersion. We can express this problem within an optimization framework as follows:

$$\min_{\{c_j, |C|, \mathcal{L}\}} \sum_{j=1}^{|C|} f(c_j) + \beta \sum_{i=1}^N \delta(\mathcal{L}(i) - j) g(c_j, p_i)$$

where f is the penalty for a cluster center, $\mathcal{L}(i)$ is the label indicating the cluster for the point i and $g(c_j, p_i)$ is the cost of attributing p_i to the cluster c_j . $\delta(\mathcal{L}(i) - j)$

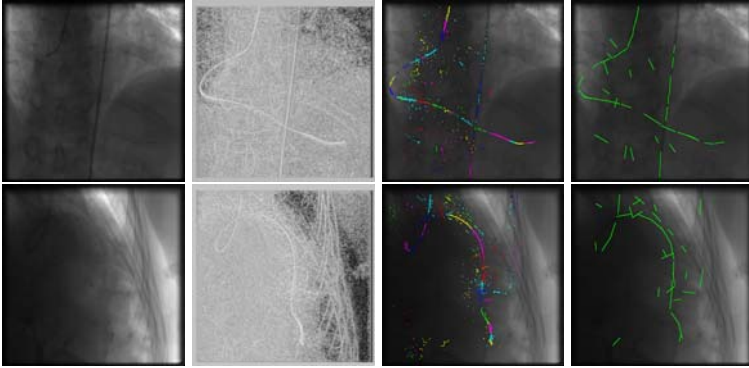


Fig. 1. 1) input 2) classification scores 3) detected points and clusters 4) detected segments (colors help to differentiate clusters in 3)

measures only the cluster cost with respect to the optimal attribution label. The definition of the functions f and g , as well as the optimization of the objective function are the main challenges to be addressed.

We consider f to be a constant (in our experiments, around 90000.0) and no-particular preference is given among the set of detected points. The definition of g is more challenging and for simplicity and clarity, let us introduce the following geometric elements: (i) (c_j, p_i) form a line segment l_{cp} , (ii) c_j and the axis of orientation given by the optimal steerable filter computed in c_j form a line l_c , and (iii) p_i and the optimal steerable filter axis computed in p_i also form a line l_p . In the ideal case, if point p_i is a member of c_j , then the three line segments will coincide.

In order to quantify this hypothesis, we consider two criteria: (i) The actual geometric distance between the two points, $d(c_i, p_j)$ since we would expect capturing local straight line segments of curvilinear structures. (ii) The distances $d_1(c_i, p_j) = d(c_i, p_j) \min(\tan(\theta_1), t_{max})$ and $d_2(c_i, p_j) = d(c_i, p_j) \min(\tan(\theta_2), t_{max})$ where θ_1 and θ_2 are the angles between the lines being formed from the two points and the ones given by steerable filters orientations, and t_{max} allows to avoid infinite distances when directions are orthogonal. These distances are combined to a single metric as follows (in our experiments, we took $\alpha = 0.3$):

$$g(c_i, p_j) = \alpha d^2(c_i, p_j) + (1 - \alpha)(d_1^2(c_i, p_j) + d_2^2(c_i, p_j))$$

The optimization of this objective function is done using [13], that finds the number of clusters and the pixels memberships simultaneously. Some clustering results are shown in: [Fig. (1)]. In order to extract the local line segments, we used a variant of RANSAC called MSAC [14] that is less sensitive to the geometric error introduced from the presence of outliers. There remains only to link these segments properly to delineate the GW - with handling outliers. Once again, we performed this task through a global optimization process, which guarantees robustness with respect to local minima created by noise.

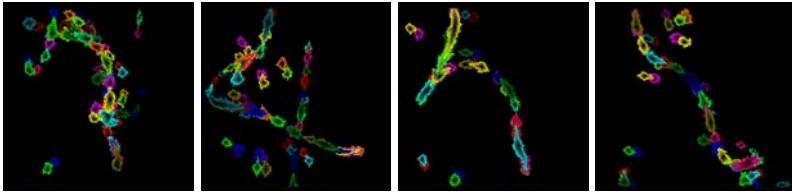


Fig. 2. Minimal geodesic costs determined using the fast marching algorithm. Colors help to differentiate and visualize costs (the brighter the color the higher the cost).

3 Perceptual Organization of Segments

Once we have obtained the line segments, we need to find which ones have to be linked and in which order for delineating the GW.

Let $|S|$ be the number of line segments and let us consider for every segment s_i a label $\{\theta_i, \omega_i, \phi_i\} \in \{1 \dots |S|\} \times \{0, 1\} \times \{0, 1\}$. θ_i denotes the rank of the segment s_i , ω_i accounts for the orientation of this segment (the conventional one assumes that the first extremity corresponds to the small horizontal coordinate) with the 0 label corresponding to the conventional one and the 1 to the extremities being reversed. Introducing such a variable is a necessity since it modifies the cost of linking two successive segments. Last, let ϕ_i be the state, that is either part of the guide-wire (1) or an outlier (0). Given such notation, one can now proceed to the definition of the local cost $C(a, b)$ corresponding to the price to be paid towards linking two segments tips a, b .

3.1 Local Ordering Cost

In order to define the cost $C(a, b)$ we consider a linear combination of the scale invariant Elastica criterion defined in [15] and a linking cost equal to the minus log-likelihood of the most likely path linking a and b . The idea is that a and b should both be linked by a path of low curvature (i.e. low Elastica) and linked by a likely path. If we assume that pixels along the path are independent, then the last criterion can be computed as a sum over the pixels of the better path, and can therefore be computed by fast marching. Denoting with $h(x)$ the detection score at location x and using a sigmoid function to compute pixel likelihood, we finally used the following cost map $c(x)$ as fast marching input:

$$c(x) = -h(x) + \log \left(e^{h(x)} + e^{-h(x)} \right)$$

Linking costs above an arbitrary threshold were not computed and approximated with the product between the Euclidean distance between the tips and the maximal linking cost normalized according to the euclidian distance between tips found by fast marching. [Fig. (2)] shows several cost maps computed with the algorithm presented in [16].

3.2 Global Ordering

Finding the optimal sequence of segments can be considered as an optimization problem. In simple words one would like to minimize:

$$\min_{\theta_i, \omega_i, \phi_i} \sum_{m=1}^{|S|} \sum_{n=1}^{|S|} \phi_m \phi_n \delta(\theta_n - \theta_m - 1) C_{\omega_m, \omega_n}(m, n) + \beta \sum_{m=1}^{|S|} (1 - \phi_m)$$

where product $\phi_m \phi_n$ guarantees that both s_m, s_n segments are retained and the term $\delta(\theta_n - \theta_m - 1)$ guarantees that the segment s_n succeeds s_m . The second term penalizes the attribution of the outlier label to a segment. Last, the cost for connecting two (conventionally oriented) segments $s_m = (a_m, b_m)$ and $s_n = (a_n, b_n)$, denoted with $C_{\omega_m, \omega_n}(m, n)$, is derived from the one earlier presented:

$$C_{\omega_m, \omega_n}(m, n) = \begin{cases} C(b_m, a_n) & \text{if } \omega_m = 0 \text{ and } \omega_n = 0 \\ C(b_m, b_n) & \text{if } \omega_m = 0 \text{ and } \omega_n = 1 \\ C(a_m, a_n) & \text{if } \omega_m = 1 \text{ and } \omega_n = 0 \\ C(a_m, b_n) & \text{if } \omega_m = 1 \text{ and } \omega_n = 1 \end{cases}$$

Starting from a configuration where all the segments are considered to be inliers and have been linked greedily (one segment has been arbitrary chosen to be the first one, and for all n , the segment chosen to be at place $n + 1$ is the nearest from the segment at place n that had not been chosen previously) our algorithm reduces the objective function iteratively until convergence. At each step, inliers and outliers are ordered, and all the possible permutations of a subsequence of inliers with a subsequence of outliers that reduce the objective function are applied.

The ordering algorithm we use for the first task tries to find by local search the order minimizing the sum of the links made between successive segments. In other words, it tries to find the 'Shortest Spanning Path' among the segments. The ordering of the outliers provides in this manner interesting subsequences for the permutation step without affecting the objective function, whereas the ordering of the inliers reduces the objective function. Given that all tasks reduce the objective function, convergence is guaranteed.

We remind that local search is a heuristic that minimizes an energy depending on a configuration by passing from the current configuration to a neighbouring one while it allows reducing the energy. Particular attention is to be paid when choosing the neighborhood definition. We chose the neighborhood defined by all the configurations reached (from the current one) when applying one of the following operations: (i) reversion of a subsequence of segments (ii) shift of a subsequence (iii) reversion followed by a shift of a subsequence.

Because local search is a meta heuristic, we cannot claim getting the global optimum, but the great size of neighborhood for both the orderings ($O(n^3)$ if n is the number of inliers) and the subsequence exchanges ($O(n^2 m^2)$ if m is the number of outliers) substantially increases the probability of finding the global optimum. In practice this was very frequently observed.

The last step of the method consists of fitting a B-spline approximation to the retained ordered segmentations towards completing the missing content.

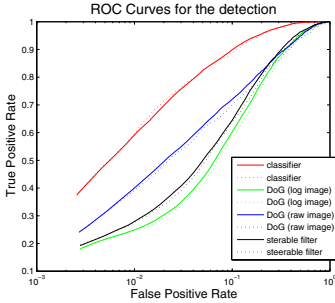


Fig. 3. Detection performances (for 5 million points from the 45 test images; dashed lines: for 1 million)

stages	MD	FD
detection	15.4%	53.2%
segments extraction	19.3%	53.0%
segments linking	23.8%	14.3%
splines fitting	25.8%	13.2%

Fig. 4. missed detections(MD) and false detections(FD) after each step. The state of the art [8] reports 22 ± 3% MD and 10 ± 3% FD.

4 Experimental Validation

Experiments were carried out on a database of 15 sequences of 10 images 1000x1000 acquired during interventions on 13 patients with a frame rate of 15 images per second, where clinical experts have manually marked GW. We performed a cross-validation: we built 15 classifiers using 75 images from 14 sequences only and tested each classifier using 3 images from the remaining sequence, using the same features and parameters for training and testing (a scale of 3 for MSAC and $\beta = 20.0$ for the ordering, in addition to constants given in preceding sections).

[Fig. (3)] compares the ROC curves of one of the classifiers with the ROC curves for three of the features that it uses: DoG of scales (1.4, 1.8) computed on raw or on log images and the ridge detector for $\mu = 0.25, \sigma = 1.6$ of [5].

Given that even features designed for curvilinear structure detection perform poorly, machine learning methods like ours are necessary. Our method produces for example 4 times less false detections than the better single feature when detector is thresholded to recover 60% of GW.

We validated our algorithm with the same metrics as in [8]: we measured the proportion of pixels of GW at more than 5 pixels (1.0 mm) of the detected structures (*missed detections*,MD) and the proportion of pixels of the detected structures at more than 5 pixels of the GW (*false detections*,FD). Because GW parts lying in catheters are difficult to detect and not always of interest, but often help delineating the other parts, we decided not to take these parts into account for the computation of *missed detection* ratio.

[Fig. (4)] presents our results. Most of the MD are due to long parts of GW hardly visible producing prohibitive linking costs. There finally remains few FD, either due to rib borders or due to linking of line segments after the GW tip. A GW tip detector like the one used in [8] could therefore help removing even more FD. This point is however out of the scope of this article.

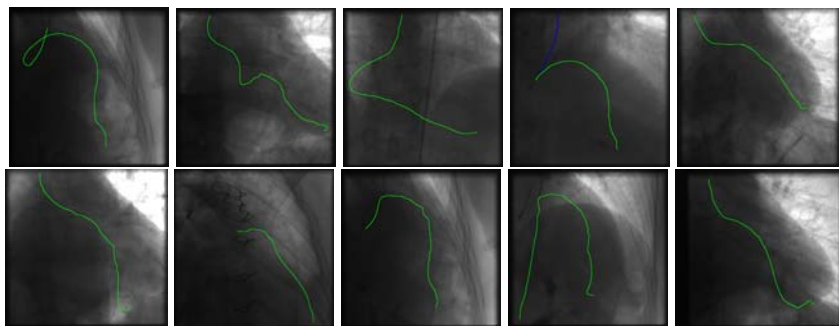


Fig. 5. Several final results (from 10 different sequences)

5 Discussion

In this paper, we have presented a global approach for GW segmentation designed to achieve high robustness. We used classifiers built by a variant of Boosting [9] and combining rotation invariant features such as improved steerable filters [5] to detect GW pixels and clustering to group them towards extracting line segments. We finally performed the delineation of the GW from the set of segments by solving an ordering problem with outlier rejection. Future work will focus on tracking the GW. Knowing the GW movement is indeed of great interest for many applications, such as cardiac motion monitoring.

References

1. Palti-Wasserman, D., Bruckstein, A.M., Beyar, R.P.: Identifying and tracking a guide wire in the coronary arteries during angioplasty from x-ray images. *IEEE Tr. on Bio. Eng.* 44(2), 152–164 (1997)
2. Baert, S.A.M., Viergever, M.A., Niessen, W.J.: Guide-wire tracking during endovascular interventions. *IEEE Tr. on Medical Imaging* 22(8), 965–972 (2003)
3. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998*. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
4. Hauke Heibel, T., Glocker, B., Groher, M., Paragios, N., Komodakis, N., Navab, N.: Discrete tracking of parametrized curves. In: *CVPR*, pp. 1754–1761 (2009)
5. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny-like criteria. *IEEE PAMI* 26(8), 1007–1019 (2004)
6. Slabaugh, G., Kong, K., Unal, G., Fang, T.: Variational guidewire tracking using phase congruency. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II*. LNCS, vol. 4792, pp. 612–619. Springer, Heidelberg (2007)
7. Franken, E., Rongen, P., van Almsick, M., ter Haar Romeny, B.M.: Detection of electrophysiology catheters in noisy fluoroscopy images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 25–32. Springer, Heidelberg (2006)

8. Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: CVPR, pp. 1–8 (2007)
9. Freund, Y., Shapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: ICML (1996)
10. Honnorat, N., Vaillant, R., Paragios, N.: Robust guidewire segmentation through boosting, clustering and linear programming. In: ISBI, pp. 924–927 (2010)
11. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. *The annals of statistics* 28(2), 337–407 (1998)
12. Viola, P., Jones, M.: Fast and robust classification using asymmetric adaboost and a detector cascade. In: NIPS (2001)
13. Komodakis, N., Paragios, N., Tziritas, G.: Clustering via lp-based stabilities. In: NIPS (2008)
14. Torr, P., Zisserman, A.: Robust computation and parametrization of multiple view relations. In: ICCV, pp. 727–732 (1998)
15. Sharon, E., Brandt, A., Basri, R.: Completion energies and scale. *IEEE PAMI* 22(10), 1117–1131 (2000)
16. Yatziv, L., Bartesaghi, A., Sapiro, G.: $O(n)$ implementation of the fast marching algorithm. *Journal of Computational Physics* 212, 393–399 (2006)

Single-Projection Based Volumetric Image Reconstruction and 3D Tumor Localization in Real Time for Lung Cancer Radiotherapy

Ruijiang Li, Xun Jia, John H. Lewis, Xuejun Gu, Michael Folkerts, Chunhua Men, and Steve B. Jiang

Department of Radiation Oncology, University of California San Diego
3855 Health Sciences Dr. 0843, La Jolla, CA 92037, USA
{rul002, xunjia, jlewis, x2gu, mfolkert, cmen, sbjiang}@ucsd.edu

Abstract. We have developed an algorithm for real-time volumetric image reconstruction and 3D tumor localization based on a single x-ray projection image. We first parameterize the deformation vector fields (DVF) of lung motion by principal component analysis (PCA). Then we optimize the DVF applied to a reference image by adapting the PCA coefficients such that the simulated projection of the reconstructed image matches the measured projection. The algorithm was tested on a digital phantom as well as patient data. The average relative image reconstruction error and 3D tumor localization error for the phantom is 7.5% and 0.9 mm, respectively. The tumor localization error for patient is ~2 mm. The computation time of reconstructing one volumetric image from each projection is around 0.2 and 0.3 seconds for phantom and patient, respectively, on an NVIDIA C1060 GPU. Clinical application can potentially lead to accurate 3D tumor tracking from a single imager.

Keywords: image reconstruction, tumor localization, lung motion, GPU, lung cancer radiotherapy.

1 Introduction

Management of tumor motion is a challenging and important problem for modern highly conformal lung cancer radiotherapy. Poorly managed tumor motion can lead to poor target coverage and an unnecessarily high dose to normal tissues [1]. Therefore, precise knowledge of real-time lung tumor motion during the treatment delivery is essential for the effectiveness of lung cancer radiotherapy [2-8]. In [9-11] 3D respiratory motion was estimated from cone beam projections based on a generic B-spline motion model. However, because of the large number of parameters in the model, many projections over a large range of angles have to be used. Therefore, the estimation process is retrospective and cannot be done in real time.

The goal of this work is to demonstrate the feasibility of extracting lung motion information from a single x-ray projection in real time. This is achieved with effective use of the prior information provided by 4DCT or 4DCBCT, where the deformation of the entire lung at different phases is efficiently represented by principal component

analysis (PCA). The inherent regularization imposed by the PCA lung motion model allows us to obtain the volumetric image of the patient from a single projection. Real-time efficiency is achieved by implementing the algorithm on a graphics processing unit (GPU).

We organize the paper as follows. In section 2, we briefly introduce the PCA lung motion model and furnish the reason behind its suitability for our goal. Then we use the PCA model to reconstruct volumetric images and localize tumors with an appropriate cost function. In section 3, we provide some details on the phantom and patient data which the algorithm was tested on. In section 4, we demonstrate the results. We conclude the paper in section 5.

2 Methods

The basic idea of our work is to first obtain a lung motion model parameterized by a few PCA coefficients, and then adapt the deformation vector field (DVF) applied to a reference CT such that its projection matches the x-ray projection images acquired during the treatment. In the following, whenever we mention CT or 4DCT, the same principle can be applied to CBCT or 4DCBCT too.

2.1 PCA Lung Motion Model

Our method starts by parameterizing lung motion with PCA. The PCA lung motion model was first proposed by [12] and was recently shown in [13] to bear a close relationship with the physiological 5D lung motion model [14] on a theoretical basis. In the PCA model, lung motion (*i.e.*, motion of each voxel in the lung along each of the 3 spatial coordinates) denoted as a vector function $\mathbf{x}(t)$, can be approximated by a linear combination of the eigenvectors corresponding to the largest eigenvalues:

$$\mathbf{x}(t) \approx \bar{\mathbf{x}} + \sum_{k=1}^K \mathbf{u}_k w_k(t) \quad (1)$$

where \mathbf{u}_k are the eigenvectors obtained from PCA and are functions of space only. The scalars $w_k(t)$ are PCA coefficients and are functions of time only. It is worth mentioning that the eigenvectors are fixed after PCA and it is the evolution of the PCA coefficients that drives the dynamic lung motion.

There are primarily two reasons why the PCA lung motion model is suitable for this work. First, PCA provides the best linear representation of the data in the least mean-square-error sense. Second, the PCA motion model imposes inherent regularization on its representation. One can show that $\|\Delta \mathbf{u}\|^2 \leq \sum_i (1/\sigma_i^2) \cdot \|\Delta \mathbf{x}\|^2$, where σ_i are the eigenvalues from PCA. This means that if two voxels move similarly, their motion represented by PCA will also be similar [13]. The combined effect is that a few scalar variables (PCA coefficients) are sufficient to dynamically deform the lung in a reasonably accurate way.

In order to build a PCA lung motion model, a representative lung motion state for one breathing cycle needs to be available for training purposes. In practice, this can be obtained from deformable image registration (DIR) between a reference CT phase and all other phases in a 4DCT (or 4DCBCT) data set, which is available from the treatment simulation (or setup).

2.2 Image Reconstruction Using the PCA Model

After we have obtained a parameterized PCA lung motion model, we seek a set of optimal PCA coefficients such that the simulated projection of the reconstructed CT matches well with the measured x-ray projection. But the simulated and measured projection may have very different intensity levels. Here we assume there exists a linear relationship between them. The cost function is:

$$\begin{aligned} \min J(\mathbf{w}, a, b) &= \left\| \mathbf{P} \cdot \mathbf{f}(\mathbf{x}, \mathbf{f}_0) - a \cdot \mathbf{y} - b \cdot \mathbf{1} \right\|_2^2 \\ \text{s.t. } \mathbf{x} &= \bar{\mathbf{x}} + \mathbf{U} \cdot \mathbf{w} \end{aligned} \quad (2)$$

where, \mathbf{U} and \mathbf{w} are comprised of a set of eigenvectors and PCA coefficients, \mathbf{x} is the parameterized DVF, \mathbf{f}_0 is the reference CT, \mathbf{f} is the reconstructed CT, \mathbf{y} is the projection image, and \mathbf{P} is a projection matrix which calculates the simulated projection. For simplicity, we have suppressed the time index under \mathbf{w} , \mathbf{x} , and \mathbf{y} .

To find the optimal values for \mathbf{w}, a, b , the algorithm alternates between the following 2 steps:

$$\text{step 1: } \mathbf{w}_{n+1} = \mathbf{w}_n - \mu_n \cdot \frac{\partial J}{\partial \mathbf{w}_n} \quad (3)$$

$$\text{step 2: } (a_{n+1}, b_{n+1})^T = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{P} \mathbf{f}_{n+1} \quad (4)$$

$$\text{where, } \mathbf{Y} = [\mathbf{y}, \mathbf{1}], \text{ and } \frac{\partial J}{\partial \mathbf{w}} = \frac{\partial \mathbf{x}}{\partial \mathbf{w}} \cdot \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \cdot \frac{\partial J}{\partial \mathbf{f}} = \mathbf{U}^T \cdot \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \cdot \mathbf{P}^T \cdot (\mathbf{P} \cdot \mathbf{f} - a \cdot \mathbf{y} - b \cdot \mathbf{1}).$$

Given the new DVF at each iteration, the reconstructed CT \mathbf{f} is found through trilinear interpolation. Accordingly, $\partial \mathbf{f} / \partial \mathbf{x}$ has to be consistent with the interpolation process in order to get the correct gradient. It turns out that $\partial \mathbf{f} / \partial \mathbf{x}$ is a linear combination of the spatial gradients of the image evaluated at the neighboring 8 grid points, weighted by the appropriate fractional part of the DVF. The step size μ_n in step 1 is found by Armijo's rule for line search. In step 2, the update for a, b is the unique minimizer of the cost function with fixed \mathbf{w} . Therefore, the cost function always decreases at each step. Note that the cost function is lower bounded by zero. The above alternating algorithm is guaranteed to converge for all practical purposes. The algorithm stops whenever the norm of the gradient is sufficiently small or the maximum number of iterations (10 in this paper) is reached.

2.3 Tumor Localization by Deformation Inversion

In order to get the correct tumor position, we distinguish between 2 different kinds of DVFs: push forward DVF and pull back DVF. The DVF found by (2) is a pull back DVF. It cannot be used directly to calculate the new tumor position. To do that, we need its inverse, *i.e.*, the push forward DVF. Here, we adopt a fixed-point approach for deformation inversion [15] and calculate tumor position.

3 Materials

3.1 Digital Phantom

The algorithm was tested using a non-uniform rational B-spline (NURBS) based cardiac-torso (NCAT) phantom [16]. This mathematical phantom has a high level of anatomical realism (*e.g.*, a beating heart, detailed bronchial trees). The respiratory motion was developed based on basic knowledge of respiratory mechanics. We generated a dynamic NCAT phantom composed of 10 phases as our simulated 4DCT. The NCAT phantom also outputs the 3D tumor position, which is used as ground truth. The CT data dimension is: $256 \times 256 \times 120$ (voxel size: $2 \times 2 \times 2.5 \text{ mm}^3$).

We used the end of exhale (EOE) phase as the reference image and did DIR between EOE phase and all other phases. The DIR algorithm is a fast demons algorithm implemented on GPU [17]. Then PCA was performed on the 9 DVFs from DIR and 3 PCA coefficients and eigenvectors were kept in the PCA model. We simulated x-ray images at different phases from those for training with different breathing amplitudes as well as at different gantry angles using Siddon's algorithm [18]. The imager dimension is down-sampled to 200×150 (pixel size: $2 \times 2 \text{ mm}^2$).

3.2 Patient Data

The algorithm was also evaluated with a patient data set. 4DCT of the patient was acquired using a GE four-slice LightSpeed CT scanner (GE Medical Systems, Milwaukee, WI, USA) and the RPM system. The cone beam projections were taken with the Varian on-board imaging system (Varian Medical Systems, Palo Alto, CA, USA) in half-fan mode with 110 kVp, 20 mA and 20 ms exposure time. For the patient in this study, there were no implanted fiducial makers and real-time 3D location of the tumor was not available to evaluate our algorithm. Instead, we projected the estimated 3D tumor location onto the 2D imager and compared with that manually defined by a clinician. From the clinician-defined contour, the tumor centroid position was calculated for each projection and used as the ground truth to evaluate the algorithm. We calculated the tumor localization error along the axial and tangential directions, both scaled back to the mean tumor position.

4 Results

4.1 Phantom Results

We tested the extrapolation performance of our algorithm by generating a new CT volume where there is an *increase* of 50% in breathing amplitude (tumor motion

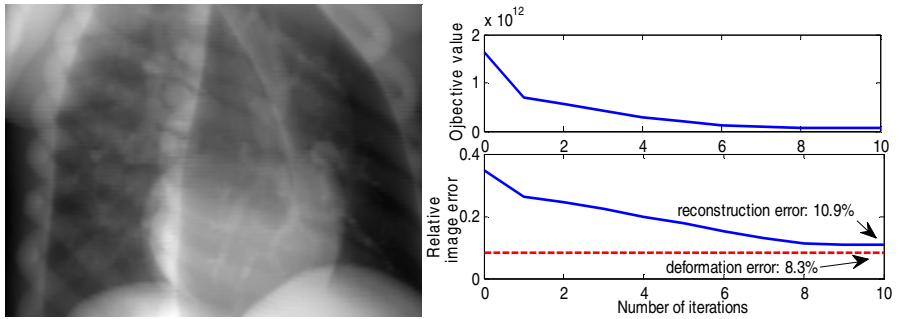


Fig. 1. a): "measured" projection of the test image at a right posterior oblique (RPO) angle; b): objective value (top) and the relative image reconstruction error (bottom) at each iteration

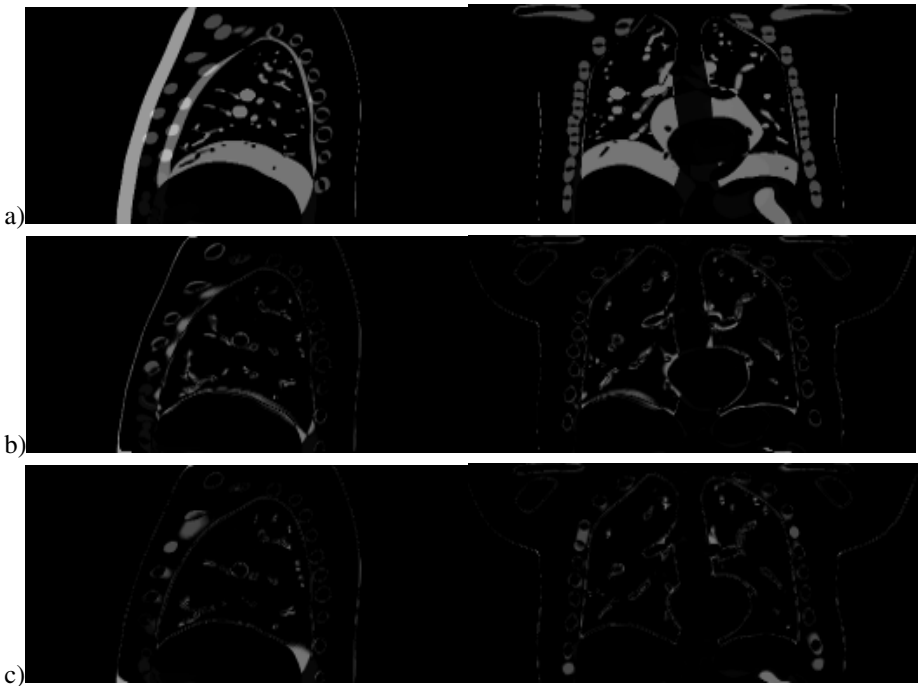


Fig. 2. Left column: sagittal view of the absolute difference image between a), test and reference CT; b), test CT and CT reconstructed using proposed algorithm; c), test CT and deformed CT using demons. Tumor is a round object near the center of the slice. Right column: same as left column, except for coronal view. Tumor is a round object in the right lung.

magnitude: 16 mm in training versus 24 mm in testing). The simulated projection image (see Fig. 1) was generated from the end of inhale phase so that it is maximally different from the reference CT. The algorithm converges within 10 iterations. Figure 2 shows the sagittal and coronal views of the reference CT and the difference images. The relative 3D RMS image reconstruction error is initially 35% and approaches to that obtained by 3D DIR with demons (11% compared with 8%). Note

that demons uses the 3D test CT instead of 2D projection image used by our algorithm. The 3D RMS tumor localization error is 0.9 mm.

We then did a more systematic evaluation by generating a dynamic phantom consisting of 60 phases with a 50% increase in breathing amplitude and 4-sec period. We simulated cone beam x-ray projections at all angles with spacing of 1° resulting in 360 projections in 1 minute (15 breathing cycles). The average relative 3D image reconstruction error is $7.5\% \pm 2.4\%$. The average 3D tumor localization error is $0.9 \text{ mm} \pm 0.5 \text{ mm}$ and is not affected by projection angles (see Fig. 3).

In order to speed up the computation, we have implemented our algorithm on an NVIDIA C1060 GPU. We initialized the PCA coefficients as those from previous frame in order to get further reduction in computation time. The image reconstruction and tumor localization for each projection was achieved between 0.2 and 0.3 seconds.

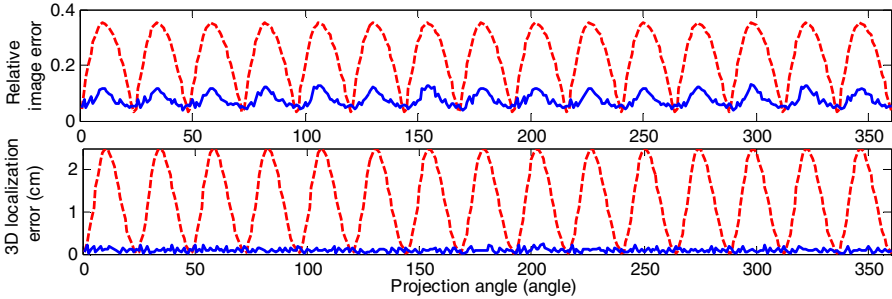


Fig. 3. Top row: relative image error between the ground truth test image under regular breathing and: reference image (*dashed line*); image reconstructed using the proposed algorithm (*solid line*) as a function of cone beam projection angle. Bottom row: same as top row, except for 3D localization error.

4.2 Patient Results

For the patient, approximately 650 projections were acquired over a full gantry rotation with a frequency of about 10.7 Hz. However, since the cone beam scans were performed in the half-fan mode (with the imager shifted about 14.8 cm laterally), and iso-center of the scan was not placed at the tumor, the tumor is only visible in a subset of these projections. The tumor was marked by the clinician in the largest continuous set of projections in which the tumor was visible. For this patient, 281 projections were used.

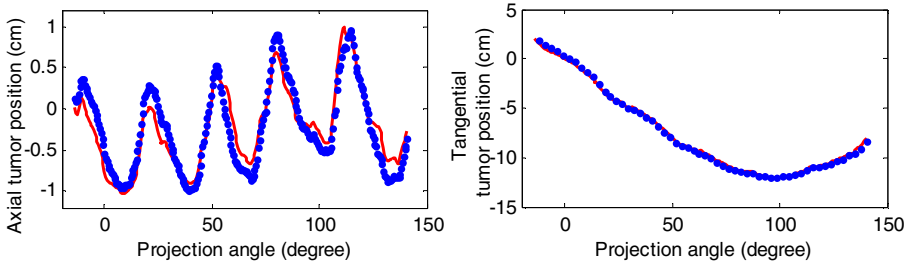


Fig. 4. Tumor localization results for the patient. Dots represent the algorithm output, and the solid lines represent the clinician marked tumor position.

The patient had quite irregular breathing during the CBCT scan, with both baseline drift and amplitude changes. Figure 4 shows the tumor localization results. For the axial direction the average localization error is 1.9 ± 0.9 mm. In the tangential direction the average error is 1.8 ± 1.0 mm. The average computation time for image reconstruction and tumor localization from each projection was around 0.3 seconds on an NVIDIA Tesla C1060 GPU card.

5 Conclusions

We have shown it is feasible to extract lung motion information from a single x-ray projection within half a second. To further speed up the computation, we can predict the current PCA coefficients using previous histories and use that as the starting point. We plan to comprehensively evaluate the accuracy of the algorithm on more clinical data. It is worth mentioning that careful quality assurance (*e.g.*, patient breathing coaching during 4DCT acquisition), better DIR algorithms will help improve the algorithm's accuracy on patient data. Also, it is not clear if the PCA motion model would be able to capture dramatic changes in breathing patterns between treatment simulation and different treatment fractions. Should that occur, 4D CBCT acquired on the day of treatment may be more appropriate to build the PCA motion model.

References

1. Keall, P.J., Mageras, G.S., Balter, J.M., Emery, R.S., Forster, K.M., Jiang, S.B., Kapatoes, J.M., Low, D.A., Murphy, M.J., Murray, B.R., Ramsey, C.R., Van Herk, M.B., Vedam, S.S., Wong, J.W., Yorke, E.: The management of respiratory motion in radiation oncology report of AAPM Task Group 76. *Med. Phys.* 33, 3874–3900 (2006)
2. Cervino, L.I., Chao, A.K., Sandhu, A., Jiang, S.B.: The diaphragm as an anatomic surrogate for lung tumor motion. *Phys. Med. Biol.* 54, 3529–3541 (2009)
3. Berbeco, R.I., Nishioka, S., Shirato, H., Jiang, S.B.: Residual motion of lung tumors in end-of-inhale respiratory gated radiotherapy based on external surrogates. *Med. Phys.* 33, 4149–4156 (2006)
4. Hoisak, J.D., Sixel, K.E., Tirona, R., Cheung, P.C., Pignol, J.P.: Correlation of lung tumor motion with external surrogate indicators of respiration. *Int. J. Radiat. Oncol. Biol. Phys.* 60, 1298–1306 (2004)
5. Wu, H., Zhao, Q., Berbeco, R.I., Nishioka, S., Shirato, H., Jiang, S.B.: Gating based on internal/external signals with dynamic correlation updates. *Phys. Med. Biol.* 53, 7137–7150 (2008)
6. Cui, Y., Dy, J.G., Sharp, G.C., Alexander, B., Jiang, S.B.: Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers. *Phys. Med. Biol.* 52, 6229–6242 (2007)
7. Xu, Q., Hamilton, R.J., Schowengerdt, R.A., Alexander, B., Jiang, S.B.: Lung Tumor Tracking in Fluoroscopic Video Based on Optical Flow. *Medical Physics* 35, 5351–5359 (2008)
8. Lin, T., Cervino, L.I., Tang, X., Vasconcelos, N., Jiang, S.B.: Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy. *Phys. Med. Biol.* 54, 981–992 (2009)
9. Zeng, R., Fessler, J.A., Balter, J.M.: Estimating 3-D respiratory motion from orbiting views by tomographic image registration. *IEEE Trans. Med. Imaging* 26, 153–163 (2007)

10. Vandemeulebroucke, J., Kybic, J., Clarysse, P., Sarrut, D.: Respiratory Motion Estimation from Cone-Beam Projections using a Prior Model. In: *Medical Image Computing and Computer-Assisted Intervention*, vol. 5762, pp. 365–372. Springer, London (2009)
11. Fluck, O., Aharon, S., Khamene, A.: Efficient Framework for Deformable 2D-3D Registration. *Medical Imaging: Visualization, Image-guided Procedures, and Modeling* (2008)
12. Zhang, Q., Pevsner, A., Hertanto, A., Hu, Y.C., Rosenzweig, K.E., Ling, C.C., Mageras, G.S.: A patient-specific respiratory model of anatomical motion for radiation treatment planning. *Med. Phys.* 34, 4772–4781 (2007)
13. Li, R., Lewis, J.H., Jia, X., Zhao, T., Lamb, J., Yang, D., Low, D.A., Jiang, S.B.: PCA-based lung motion model. In: *16th International Conference on the Use of Computers in Radiation Therapy*, Amsterdam, Netherlands (2010) (accepted)
14. Low, D.A., Parikh, P.J., Lu, W., Dempsey, J.F., Wahab, S.H., Hubenschmidt, J.P., Nystrom, M.M., Handoko, M., Bradley, J.D.: Novel breathing motion model for radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* 63, 921–929 (2005)
15. Chen, M., Lu, W., Chen, Q., Ruchala, K.J., Olivera, G.H.: A simple fixed-point approach to invert a deformation field. *Med. Phys.* 35, 81–88 (2008)
16. Segars, W.P., Lalush, D.S., Tsui, B.M.W.: Modeling respiratory mechanics in the MCAT and spline-based MCAT phantoms. *IEEE Transactions on Nuclear Science* 48, 89–97 (2001)
17. Gu, X., Pan, H., Liang, Y., Castillo, R., Yang, D., Choi, D., Castillo, E., Majumdar, A., Guerrero, T., Jiang, S.B.: Implementation and evaluation of various demons deformable image registration algorithms on a GPU. *Phys. Med. Biol.* 55, 207–219 (2010)
18. Siddon, R.L.: Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics* 12, 252–255 (1985)

A Method for Planning Safe Trajectories in Image-Guided Keyhole Neurosurgery

Reuben R. Shamir¹, Idit Tamir², Elad Dabool¹, Leo Joskowicz¹, and Yigal Shoshan²

¹ School of Engineering and Computer Science, The Hebrew University, Jerusalem, Israel

² Dept. of Neurosurgery, Hebrew University Hadassah Medical Center, Jerusalem, Israel
rubke@cs.huji.ac.il

Abstract. We present a new preoperative planning method for reducing the risk associated with insertion of straight tools in image-guided keyhole neurosurgery. The method quantifies the risks of multiple candidate trajectories and presents them on the outer head surface to assist the neurosurgeon in selecting the safest path. The surgeon can then define and/or revise the trajectory, add a new one using interactive 3D visualization, and obtain a quantitative risk measures. The trajectory risk is evaluated based on the tool placement uncertainty, on the proximity of critical brain structures, and on a predefined table of quantitative geometric risk measures. Our results on five targets show a significant reduction in trajectory risk and a shortening of the preoperative planning time as compared to the current routine method.

1 Introduction

Many image-guided keyhole neurosurgery procedures require the precise targeting of tumors and anatomical structures with a surgical tool inside the brain based on pre-operative CT/MRI images. A misplacement of the surgical tool from the planned trajectory may result in non-diagnostic tissue samples and/or severe neurological complications [1-2]. Consequently, it is desired to select a trajectory that is at a safe distance from critical structures such as blood vessels and motor and functional areas.

In current practice, trajectory planning is performed manually and may be sub-optimal, as it requires the surgeon to mentally reconstruct complex 3D brain structures and their relations based on 2D cross-sections of the patient pre-operative CT/MRI head images. The treatment risk and implications evaluation is thus a complex and time-consuming task. While volume visualization and spatial segmentation of critical brain structures are sometimes used to help the neurosurgeon with spatial perception and planning, the insertion trajectory is currently determined manually. Furthermore, it does not include any quantitative measures or trajectory-specific visualization of nearby critical structures. The resulting trajectory is thus surgeon-dependent and may not be optimal.

Several studies have proposed methods to better assess and reduce risk in image guided neurosurgery [3-10]. Some of them are aimed at tasks that are significantly different from insertion of a straight surgical tool. They include methods for identifying targets and trajectories in Deep Brain Stimulation (DBS) based on a statistical atlas [5], for optimizing a path within intracranial blood vessels [7], for path optimization

for optimal tumor evacuation [8], and for planning of non-straight trajectories with non-interactive, time-consuming optimization methods [4].

In recent work, Lee et al. [9] propose a method to fuse patient MRI head images with a registered atlas to support the manual selection of a trajectory with a visualization of the 3D atlas structures. Its main disadvantage is that the trajectory is selected manually without any quantitative information regarding nearby critical structures. Vaillant et al. [10] computes the risk of a candidate trajectory with a weighted sum and based on the trajectory intersected intracranial structures and their associated importance. The weakness of this method is that it does not consider the distance of a structure from the trajectory; thus, the damage that can be caused by surgical tool misplacement is not incorporated in the function. Tirelli et al. [3] assign each candidate trajectory with a risk value that is based on a weighted sum; the drawbacks are that no risk visualization or quantitative feedback is provided. Brunenberg et al. [6] show that computing the risk with a weighted sum can be misleading and suggest computing the maximum risk value instead. The Euclidean distance of the trajectory from critical brain structures is used to compute the risk of each voxel. Their method outputs tens to hundreds of trajectories associated with distances above a predefined threshold. Although the method significantly reduces the number of possible trajectories, it still leaves a considerable amount of manual work without quantitative feedback. Moreover, none of the studies evaluated their contribution to the actual reduction of the risk.

We present a novel preoperative straight trajectory planning method for image-guided keyhole neurosurgery. Our method quantifies the risks of multiple candidate trajectories and presents them on the outer head surface to assist the neurosurgeon in selecting the safest path. For visualization, we color-code all the trajectories according to their associated risk level and present them all at once on the relevant parts of the outer head surface. The surgeon can then select and revise the trajectory, and add or edit trajectories with visual 3D feedback and updated risk information. Our method incorporates interactive 3D visualization of critical structures and of surgical tool placement uncertainty. The computed trajectory is presented to the physician along with a ‘risk card’ that includes quantitative risk measures such as the length of the trajectory and distance between the trajectory and closest blood vessels. We observe that reporting only the maximum risk value may be partial and incomplete and compute both: the maximum and the summation of risks along trajectory. Moreover, we conduct a clinical comparative study on MRI head images to rate our method vs. routine planning on five targets.

2 Method Overview

We propose the following eight-step preoperative planning workflow for image guided keyhole neurosurgery (Fig. 1). Initially, (1) the neurosurgeon selects the target location on the CT/MRI preoperative image. Then, (2) the head outer surface is computed automatically and the neurosurgeon defines (3) the surface region on which the entry point should be located. Afterwards, (4, 5) the anatomical structures of interest for the surgery, e.g. blood vessels and ventricles, are segmented and assigned a risk value based on the potential damage of penetrating them with a surgical tool (this is done once for all patients). The input segmentations and the assigned risk values are

automatically combined into a single volume (6), called the *risk volume*, in which voxel intensity is associated with a value representing the level of damage that may be caused by a surgical tool passing through it. A trajectory risk value is then computed (7) for each trajectory automatically and the candidate trajectories risks are color-coded and superimposed on the defined entry points surface to form risk maps (Fig. 2). Finally, (8) the neurosurgeon interactively selects and refines a trajectory with visual and quantitative feedback (Figs. 2 and 3).

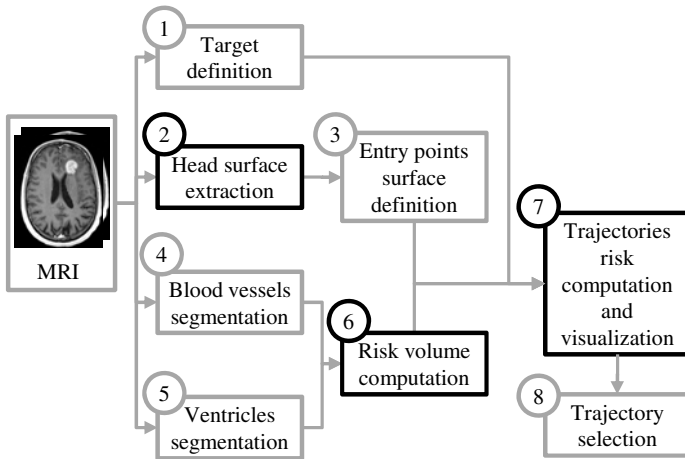


Fig. 1. Preoperative planning workflow for image guided keyhole neurosurgery

We define a risk card as a summary of geometrical parameters regarding critical structures and planned trajectories. The risk card provides the neurosurgeon with valuable information regarding the assessment of an intervention's risk and enables the direct quantitative comparison between candidate trajectories. The relevant risk parameters were identified by a senior neurosurgeon. They include measures for assessing trajectory's risk such as trajectory length, and distances of trajectory, target and entry to closest blood vessels and ventricles.

We use graphical illustrations to assist the surgeon in understanding the geometrical meaning of the measures. For example, when the neurosurgeon points with the cursor on the risk card trajectory length column, it is illustrated graphically (Fig. 3). Trajectories can be added and modified based on the 3D visualization of their localization uncertainty and predefined critical structures. The visualization of tool localization uncertainty eases the identification of cases where a planned line trajectory does not cross a critical structure, including the placement uncertainty, and indicates the possible damage to critical structures.

We describe next our method for the computation of the risk volume, the trajectory risk, and the visualization of multiple trajectories risks.

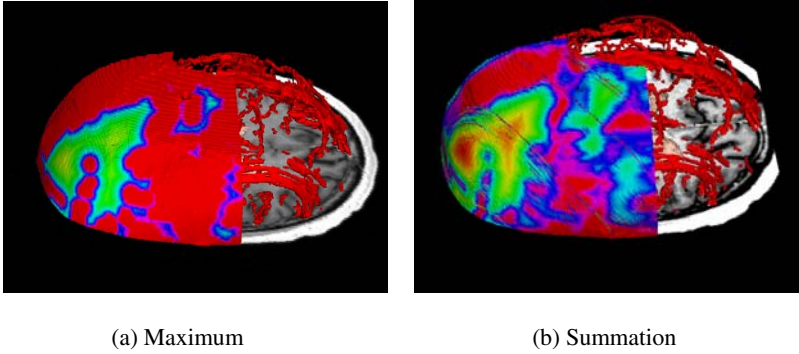


Fig. 2. Risk maps of candidate trajectories with two computation methods: (a) maximum, and; (b) summation of voxels in the risk volume that are intersected by the trajectory. Red zones are associated with high-risk trajectories. Green and yellow zones indicate safer entry points.

2.1 Computation of Risk Volume

Each voxel in the risk volume is assigned with the estimated cost incurred by penetrating the corresponding region with a surgical tool. The risk volume is generated based on two key guidelines: 1) the risk value is directly related to the estimated consequences and severity of the damage to the corresponding brain tissue or organ, severe complications and high morbidity regions are assigned with a higher risk values than tissues with minor and reversible complications, and; 2) voxels near critical structures are assigned with high risk values to reflect the intrinsic localization error of the procedure, be it freehand, frame-based stereotaxy, or with an image-guided surgery system. Therefore, voxels that are closer to a critical structure are associated with a higher risk value than those that reside further from them.

The input is a set of critical structures for which insertion of a surgical tool is forbidden or undesired, $S = \{S_1, S_2, \dots, S_p\}$, and their associated risk values $R = \{r_1, r_2, \dots, r_p\}$. The structure S_i is a segmented image. The risk value r_i is a non-negative scalar. We define each voxel in the risk volume as:

$$riskVolume(\bar{x}) = \max \left\{ \frac{r_k}{dist(\bar{x}, S_k) + \alpha} \right\}_k \quad (1)$$

where \bar{x} is the voxel center location and α is a non-negative scaling constant. Eq. 1 assigns to each voxel the maximal expected risk computed with the above cost function and with respect to the input structures and risk values. For $\alpha = 1$ and distance $dist(\bar{x}, S_k) = 0$ (e.g. voxel is located on the structure) the voxel value is the same as the input risk value r_k . It decreases as the voxel is further from the structure.

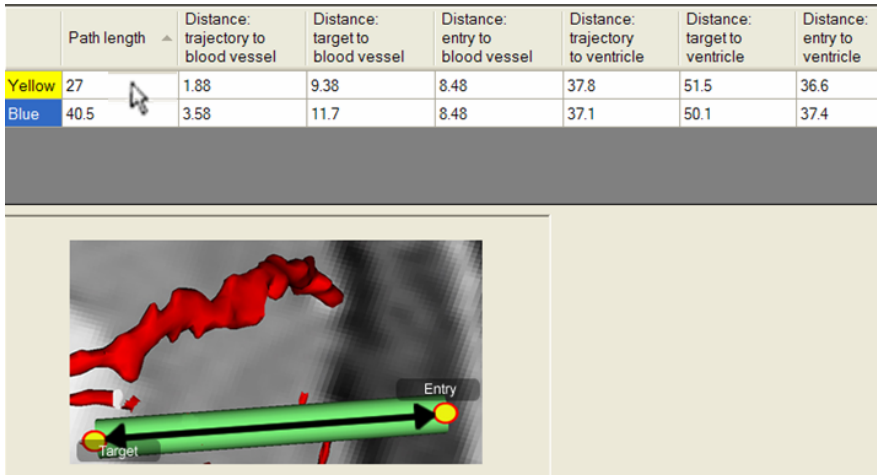


Fig. 3. Trajectory risk card showing seven parameters (all values in mm) and 3D illustration image of the “path length” risk parameter

In practice, we expect the neurosurgeon to define only a handful of risk levels, i.e., risk values $R = \{r_1, r_2, \dots, r_p\}$ are selected from a small group of numbers $r_i \in \{0, 1, 2, \dots, c\}$. In cases where two or more structures are associated with the same risk level, the voxel risk value is associated with the closest structure distance (Eq. 1). This allows computing one distance map for multiple structures that are associated with same risk level. With this approach, few distance maps can cover a large set of segmented structures.

2.2 Trajectory Risk Computation and Visualization

We assign to a given trajectory two risk values: 1) the maximal value, and 2) the sum of voxels in the risk volume that are intersected by the trajectory. The input is a target location, t , a set of candidate entry points $\{e_1, e_2, \dots, e_n\}$, and the risk volume *riskVolume*. Each target and candidate entry point pair defines a trajectory $tr_i = [e_i; t]$. The maximal trajectory risk is:

$$risk_{max}(tr_i, riskVolume) = \max \left\{ riskVolume(\bar{x}) \right\}_{\bar{x} \cap tr_i \neq \emptyset} \tag{2}$$

Note that this definition of trajectory risk does not incorporate some important risk factors such as path length and thus provides only partial information regarding the trajectory risks. For example, this definition cannot differentiate between cases where many blood vessels surround the trajectory and those where only one blood vessel is

within its proximity. Therefore, it is also desired to compute the sum of the voxels risk values along the trajectory. This measure better reflects path length, and incorporates all risks occur along the path. Note that this measure is also incomplete and should be considered with the maximal risk [6].

$$risk_{sum}(tr_i, riskVolume) = \sum_{\bar{x} \cap tr_i \neq \emptyset} riskVolume(\bar{x}) \quad (3)$$

To assist the neurosurgeon in selecting the safest paths, we propose a visualization of the relevant candidate entry-points zones on the outer head surface. Each candidate entry point on the relevant outer head surface zone is colored with respect to the risk value that was computed on its trajectory (Fig. 2). The user can change the risk computation method (Eqs. 2 or 3), the color-map, and can change the position and orientation of the 3D surface.

3 Experimental Results

We compared our method to the current routine manual approach for trajectory planning on five targets selected at various locations on four clinical MRI head images. The images are $512 \times 512 \times 122$ voxels³ with voxel size of $0.47 \times 0.47 \times 1.0$ mm³. For each target, a specialist neurosurgeon selected two trajectories: one with the conventional method based on the axial, sagittal, and coronal 2D views of the original MRI images, and the second trajectory was selected with our method. The planning protocol was as in Fig 1. A target was initially defined on the MRI image. Then, the outer surface of the head was automatically segmented and extracted [12], and sampled with $\sim 40K$ points. For each target, the user defined surface areas on the outer head surface from which the entry point can be chosen. Each candidate entry point defines a candidate trajectory with the predefined target. The blood vessels and ventricles were semi-automatically segmented and their surfaces were reconstructed. Distance maps are computed with the method of Danielsson et al. [11]. The blood vessels were associated with a risk level of $r_1 = 255$ and the ventricles were associated with a risk level of $r_2 = 100$. The risk volume was computed using Eq. 1 with r_1 and r_2 as above and $\alpha = 1$.

For each possible trajectory, the risk volume voxels that were intersected by the trajectory were identified, and the trajectory risk was computed using Eq. 2 and 3. Next, the trajectories risks were color-coded and superimposed on the relevant part of the head surface. The neurosurgeon then interactively selected an entry point. The corresponding risk card was automatically computed and a 3D visualization of the localization uncertainty and blood vessels was generated and displayed for further refinement. The method was implemented with the Visualization ToolKit (VTK) [13] and the Insight segmentation and registration ToolKit (ITK) [14] and was integrated as a set of modules in Slicer [15] on a standard PC running Windows XP OS.

Table 1. Comparison of the proposed and routine methods. Target codes are: **LF** – Left and(delete) Frontal, **LFTI** – Left Fronto-Tempo-Insular, **RPO** – Right Parieto-Occipital, **LFP** – Left Frontal-Periventricular, and **MP** - Medial Perichiasmatic and anterior to the Midbrain.

Target number	Target location	Method	Trajectory length(mm)	Distance: Trajectory to Blood Vessels(mm)	Distance: Trajectory to Ventricles (mm)	Time (min)
1	LF	Routine	35.2	5.15	9.99	-
		Proposed	31.6	9.39	9.52	-
2	LFTI	Routine	27.0	1.88	37.8	19
		Proposed	40.5	3.58	37.1	6
3	RPO	Routine	25.6	12.0	26.3	9
		Proposed	29.1	12.0	26.3	9
4	LFP	Routine	56.2	7.16	12.0	9
		Proposed	57.8	7.99	13.1	6
5	MP	Routine	94.4	0.0	2.74	6
		Proposed	96.3	1.0	4.9	10
Average		Routine	47.7	5.2	17.8	10.8
		Proposed	51.0	6.8	18.1	7.7

Table 1 summarizes the results. The mean trajectory planning time using the routine method was 10.8 min (range 6-19 min) compared to a mean of 7.7 min (range 6-10 min) using our method. Using the routine method, the mean distance of a planned trajectory to closest Blood Vessel (BV) and closest ventricle are 5.2mm (range 0.0-12.0 mm) and 17.8 mm (range 2.7-37.8 mm), respectively. Our method yielded mean distances of 6.8 mm (range 1.0-12.0 mm) from a blood vessel and 18.1mm (range 4.9-37.1 mm) from closest ventricle. The mean trajectory length with the routine method was 47.7mm (range 25.6-94.4 mm) compared to 51.0mm (range 29.1-96.3 mm) using our method.

4 Discussion

In cases 1, 2, and 5 our method resulted in significantly larger distances of up-to 4mm between trajectories and their closest blood vessel or ventricle. In cases 3 and 4, no significant difference was observed. The reason for the lack of improvement in case 3 is that the target point is located near the cranial surface where it was easier for the neurosurgeon to evaluate the risks and define a trajectory with the routine method. In case 4, the target location was the closest point along the trajectory to nearest blood vessel and the ventricle, and therefore no improvement was recorded.

The neurosurgeon that evaluated our method reports that it increased the control and confidence levels, and improved the risk assessment. The colorization of outer head surface facilitated the entry point selection. The 3D visualization of blood vessels greatly helped in understanding their complex structure and their spatial relations with respect to the planned trajectory. The risk card assisted in the trajectory selection.

5 Conclusions

We have presented a novel method to enhance the conventional trajectory planning method by a visualization of trajectories risks and by providing quantitative risk information and interactive 3D visualization of localization uncertainty and structures associated with a high risk for better assessment of the possible risks in image guided keyhole neurosurgery. Our experimental results suggest that our method produces safer trajectories in which a misplacement of a surgical tool is less likely to damage a critical structure.

Acknowledgements

This work was supported by FP7 ERC ROBOCAST Grant No. 21590.

References

1. Shamir, R.R., Joskowicz, L., Spektor, S., Shoshan, Y.: Localization and registration accuracy in image guided neurosurgery: a clinical study. *Int. J. Comput. Assist. Radiol. Surg.* 4(1), 45–52 (2009)
2. Mascott, C.R.: In vivo accuracy of image guidance performed using optical tracking and optimized registration. *J. Neurosurgery* 105(4), 561–567 (2006)
3. Tirelli, P., de Momi, E., Borghese, N.A., Ferrigno, G.: An intelligent atlas-based planning system for keyhole neurosurgery. In: *Computer Assisted Radiology and Surgery (CARS 2009) supplemental*, pp. S85–S86 (2009)
4. Popovic, A., Trovato, K.: Path planning for reducing tissue damage in minimally invasive brain access. In: *Computer Assisted Radiology and Surgery (CARS 2009) supplemental*, pp. S132–S133 (2009)
5. Guo, T., Parrent, A.G., Peters, T.M.: Automatic target and trajectory identification for deep brain stimulation (DBS) procedures. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 483–490. Springer, Heidelberg (2007)
6. Brunenberg, E.J., Vilanova, A., Visser-Vandewalle, V., Temel, Y., Ackermans, L., Platel, B., et al.: Automatic trajectory planning for deep brain stimulation: a feasibility study. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 584–592. Springer, Heidelberg (2007)
7. Fujii, T., Emoto, H., Sugou, N., Mito, T., Shibata, I.: NeuroPath planner—automatic path searching for neurosurgery. In: *Computer Assisted Radiology and Surgery (CARS)*, pp. 587–596 (2003)
8. Bourbakis, N.G., Awad, M.: A 3-D visualization method for image-guided brain surgery. *IEEE Trans. Syst. Man Cybern., Part B: Cybern.* 33(5), 766–781 (2003)
9. Lee, J.D., Huang, C.H., Lee, S.T.: Improving stereotactic surgery using 3-D reconstruction. *IEEE Eng. Med. Biol. Mag.* 21(6), 109–116 (2002)
10. Vaillant, M., Davatzikos, C., Taylor, R.H., Bryan, R.N.: A path-planning algorithm for image-guided neurosurgery. In: *First Joint Conference Computer Vision, Virtual Reality and Robotics in Med. and Med. Robotics and Comp. Assisted Surgery*, pp. 467–476 (1997)
11. Danielsson, P.E.: Euclidean distance mapping. *Computer Vision, Graphics, and Image Processing*, 227–248 (1980)
12. Joskowicz, L., Shamir, R., Freiman, M., Shoham, M., Zehavi, E., Umansky, F., et al.: Image-guided system with miniature robot for precise positioning and targeting in keyhole neurosurgery. *Comput. Aided Surg.* 11(4), 181–193 (2006)
13. The Visualization Toolkit (VTK), <http://www.vtk.org/>
14. The Insight Segmentation and Registration Toolkit (ITK), <http://www.itk.org/>
15. 3D Slicer, <http://www.slicer.org/>

Adaptive Multispectral Illumination for Retinal Microsurgery

Raphael Sznitman, Diego Rother, Jim Handa, Peter Gehlbach,
Gregory D. Hager, and Russell Taylor

Johns Hopkins University
{sznitman, jthanda, pgehlbach, hager, rht}@jhu.edu

Abstract. It has been shown that excessive white light exposure during retinal microsurgery can induce retinal damage. To address this problem, one can illuminate the retina with a device that alternates between white, and less damaging limited-spectrum light. The surgeon is then presented with a fully colored video by recoloring the limited-spectrum light frames, using information from the white-light frames. To obtain accurately colored images, while reducing phototoxicity, we have developed a novel algorithm that monitors the quality of the recolored images and determines when white light may be substituted by limited-spectrum light. We show qualitatively and quantitatively that our system can provide reliable images using a significantly smaller light dose as compared to other state-of-the-art coloring schemes.

1 Introduction

Retinal microsurgery is one of the most demanding types of surgery. The difficulty stems from the micron scale dimensions of tissue planes and blood vessels in the eye; the delicate nature of the neurosensory retina and the poor recovery of retinal function following significant injury. For many retinal surgeons the operative time is lengthy, increasing the risk of light-induced toxicity to the retina that can limit visual success despite achievement of all surgical objectives. Therefore, a central issue for the surgeon is balancing the need for adequate illumination of retinal structures with the risk of iatrogenic phototoxicity.

One of the principal sources for retinal light toxicity stems from the use of fiberoptic endoilluminators. Endoillumination is the primary means of directly illuminating surgical targets during posterior segment intraocular procedures and the primary source of retinal light toxicity during retinal surgery. Retinal phototoxicity from an operating microscope was first reported in 1983 in patients who had undergone cataract surgery with intraocular lens implantation [1] and has then been recognized as a potential complication of any intraocular surgical procedure. The frequency of these complications is reported to occur from 7% to 28% of patients undergoing cataract surgery [2,3]. Blue wavelength and ultraviolet light induce the greatest degree of retinal injury. In fact, in [4,5] it was found that commercially available light sources for endoillumination exceeded the International Commission on Non-Ionizing Radiation Protection guidelines

for retinal damage by visible light. In vitrectomy for macular hole repair, up to 7% of the patients have been reported to have experienced visually significant phototoxicity [6,7,8].

In [9], this issue was addressed by presenting a visualization system that can be used to significantly reduce the emission of highly toxic wavelengths compared to current practices. This was done by periodically illuminating the retina with white light and less damaging non-white light using a computer controlled rapidly switching multiple LED light source. A consequence of this is that images acquired are either fully colored (which we will refer to as white-light images) or monochromatic (or red-light images). To avoid visually straining the surgeon, monochromatic images were then recolored by using colors from white-light images.

In general, computer colorization schemes have existed since the 70's [10], for applications such as recoloring movies, and have since been further developed [11,12]. In general, however, such systems rely on a user to pre-select regions of the image that correspond to specific colors, making them ill-suited for this application. More recently, a time series analysis was proposed to model the retinal image scene [13]. This method however relies on having all visual cues (*e.g.* color and texture) available at all times to maintain an accurate retina model. In [9], a coloring scheme (ASR) was specifically developed for this setting. This algorithm estimates the different forms of motion which appear in the scene, taking this information into account to recolor the monochromatic images. An important limitation of ASR is that white illumination is used at regular intervals, regardless of the degree of change in the scene. This results in the use of excessive white light when no significant changes have happened in the scene, or conversely, poor recolorization (and image artifacts) when much has changed in the scene.

In order to further reduce phototoxicity and yet provide accurately colored images to the surgeon, we present a novel image acquisition and recoloring scheme for this setting. At each time frame, our algorithm determines what type of illumination should be used based on the estimated phototoxicity levels and the recolored image quality. Our method is simple and requires little parameter tuning, making it easy to use and flexible for surgical applications. We have experimentally shown that our method provides a quantitative improvement in coloring accuracy over state-of-the-art methods.

The remainder of this article is organized as follows: in section 2 our framework and colorization algorithm is presented; section 3 compares our novel approach with other methods; and finally, in section 4, we discuss some of the pitfalls of our system and future works.

2 Multispectral Illumination and Adaptive Color Fusion

In order to provide the surgeon with accurately colored images when using the light source from [9], we present an algorithm that *dynamically* chooses which illumination type to use at each time step, depending on estimates of the rendered image quality and phototoxicity levels induced. That is, the quality of the

recolorization and phototoxicity levels are continuously monitored, allowing us to estimate when it is appropriate to use white light illumination. In general, this occurs when the scene changes cannot be adequately “predicted” with the current available information.

The system we use consists of a device (as in [9]) capable of illuminating the retina using either white light, or less phototoxic red light. We define the sequence of images provided by the system as $I = \{I_1, \dots, I_N\}$ for N discrete time steps. Each image I_t is associated with a particular illumination L_t , where $L_t = 1$ means that white light was used at time t , and $L_t = 0$ means that red light was used. Consequently, when $L_t = 1$ all three color channels are available, $I_t = \{I_t^R, I_t^G, I_t^B\}$, whereas when $L_t = 0$ only the red channel I_t^R is available. We define the illumination history as $\mathbf{L}_t = \{L_1, \dots, L_t\}$. As in [9], the overall rate at which white light is flashed can be defined as $\varphi(\mathbf{L}_t) = \frac{\sum_{i=1}^t L_i}{t}$. We denote by F_t the final fully colored image rendered by our algorithm. To recolor the monochromatic images we maintain a color model of the scene for each time t , $M_t = \{M_t^R, M_t^G, M_t^B\}$.

Our goal then is to choose which illumination type, L_{t+1} , to use for the next time step. To do this, our criterion is to maximize a quantitative estimate of the patient’s wellbeing. This criterion combines the two costs incurred by the patient at time t : the *surgeon impairment cost* and the *phototoxicity cost*. The surgeon impairment cost, $S(\epsilon_t)$, is the cost of being accidentally harmed by the surgeon because of the error levels present in the recolored images, ϵ_t . The phototoxicity cost, $T(\mathbf{L}_t)$, is given by the damage to the patient produced by the illumination. In the next section we describe these costs in more detail. In section 2.2 we show how these costs are combined to select which illumination type to use at each time step.

2.1 Modeling the Cost Functions

As described in the previous section, there are two different costs incurred by the patient at time t during the procedure. The first cost is the *surgeon impairment cost*, $S(\epsilon_t)$. This is the cost (for the patient) of being accidentally harmed by the surgeon at time t . Clearly this risk (and hence the cost) increases as the recolorization error, ϵ_t (defined below) increases, since the surgeon is relying on poorer images to perform his job. The exact relationship between this cost and the error is unknown and depends, among many things, on the particular surgeon using the system. However we expect $S(\epsilon)$ to be an increasing function that levels off at a certain error, ϵ^* , at which stage the quality of the image is so poor that further deterioration does not result in additional risk. In practice, we will make sure that the system remains in the linear part of S , far from the critical value ϵ^* , where the surgeon is critically impaired. Based results from [9], we will model this relationship with the following function,

$$S(\epsilon) = \begin{cases} 1 & \text{if } \epsilon > \epsilon^* \\ \frac{\epsilon}{\epsilon^*} & \text{otherwise} \end{cases} \quad (1)$$

The recolorization error, ϵ , is due to the fact that the color model at any given time is not perfect, since the background scene changes due to the manipulations performed by the surgeon. In order to compute this error, we note that errors are only committed in the green and blue channels, since the red channel is observed at all times. We assume that the error committed in the green and blue channels at time t , $\epsilon_t^{G,B}$, is approximately equal to the error that would be obtained in the red channel, ϵ_t^R , if it were treated as the green and blue channels ($\epsilon_t^{G,B} \approx \epsilon_t^R$). Since the red channel is available at all times irrespective of the illumination type, ϵ_t^R can be directly computed as,

$$\epsilon_t^R = \|M_t^R - M_{t_w}^R\|_2 \tag{2}$$

where t_w is the last time step in which $L_{t_w} = 1$. Assuming further that the error does not change significantly in one time step, we approximate the error at time $t + 1$ by the error at time t , hence $\hat{\epsilon}_{t+1}^{G,B} \approx \hat{\epsilon}_t^{G,B} \approx \epsilon_t^R$.

The second cost, the *phototoxicity cost*, $T(\mathbf{L}_t)$, is the estimated damage at time t suffered by the patient because of the illumination used up to this point in time \mathbf{L}_t . It seems reasonable from the current literature [14] to relate the amount of phototoxic damage, T , to a function of the recent light exposure $\varphi(\mathbf{L}_t)$, where $\varphi(\mathbf{L}_t)$ is a function that models how the illumination history \mathbf{L}_t affects a cell at time t . We chose to define $\varphi(\mathbf{L}_t)$ as an exponential loss (approximated from [14]). That is, as time goes on, the influence of the past decreases exponentially fast. Hence, we approximate the phototoxicity cost by,

$$T(\mathbf{L}_t) = \begin{cases} 1 & \text{if } \varphi(\mathbf{L}_t) > L^* \\ e^{-\frac{(\varphi(\mathbf{L}_t) - L^*)^2}{2}} & \text{otherwise} \end{cases} \tag{3}$$

where L^* is some level of illumination at which irreversible damage to the patient (cell death) is produced.

It must be noted that while the choice of these functions is based on reasonable assumptions, these functions ultimately need to be empirically determined.

2.2 Choosing the Next Illumination Type

We can then formally define the estimated total cost for the patient at time $t + 1$ as the sum of the two costs described in the previous section,

$$E(\mathbf{L}_{t+1}, \hat{\epsilon}_{t+1}) = (1 - \lambda)S(\hat{\epsilon}_{t+1}) + \lambda T(\mathbf{L}_{t+1}). \tag{4}$$

where, $\hat{\epsilon}_{t+1}$, is the measure of the recolorization error defined in Eq. 2, \mathbf{L}_{t+1} is the history of illuminations at time $(t + 1)$ and λ is a tuning parameter which can be adjusted by the user (*i.e.* surgeon) to specify a bias for either image quality or phototoxicity. We select the next illumination type, by minimizing the patient wellbeing cost,

$$L_{t+1} = \arg \min_L E(\mathbf{L}_{t+1}, \hat{\epsilon}_{t+1}) = \arg \min_L \{(1 - \lambda)S(\hat{\epsilon}_{t+1}) + \lambda T(\mathbf{L}_{t+1})\}. \tag{5}$$

Notice that L can take only two values (0 or 1). Hence, this optimization reduces to

$$(1 - \lambda)S(\hat{\epsilon}_{t+1}) + \lambda(T([L_t; 1]) - T([L_t; 0])) \geq 0 \tag{6}$$

Since $\hat{\epsilon}_{t+1} = 0$ when $L_{t+1} = 1$, and $\hat{\epsilon}_{t+1} \approx \hat{\epsilon}_t$ when $L_{t+1} = 0$, all the quantities in Eq. 6 are known and choosing the next illumination type simply reduces to determining whether or not equation Eq. 6 is true.

2.3 Adaptive Active Scene Rendering

We now present the outline of our algorithm: Adaptive Active Scene Rendering (AASR). Fig. 1 provides a visual outline of AASR and associated images, respectively.

First for each image I_t , (Fig. 1 (Left)(a) and (Right)(a)) we detect and segment the tool in the image by using a 3D tool model (see 9 for more details). This provides us with a mask region for the tool, T_t (Fig. 1(Left)(b)). Then, in order to compute the new color model: if $L_t = 1$, M_t is computed by keeping pixel regions of M_{t-1} which appear where the tool is located and using I_t for regions where the tool is not present (Fig. 1(Right)(c)). This is done by using T_t to mask regions of the tool and allows for regions displaying the retina to be updated, keeping tool regions unchanged (similar to the work in 13). If $L_t = 0$, then $M_t = M_{t-1}$ (Fig. 1(Left)(c)). Rendering the recolored image, F_t , is then done by combining $M_t^{G,B}$ and I_t^R on regions outside the tool, and using a tool color model to fill in the tool ((Fig. 1(Right)(b) and (Left)(d)). Having computed these, we can then estimate the error, ϵ_t , using M_t^R and $M_{t_w}^R$, as described by Eq. 2 (Fig. 1(Right)(d)), and choosing the following illumination type can be computed as in Eq. 6 (Fig. 1(Right)(e)).

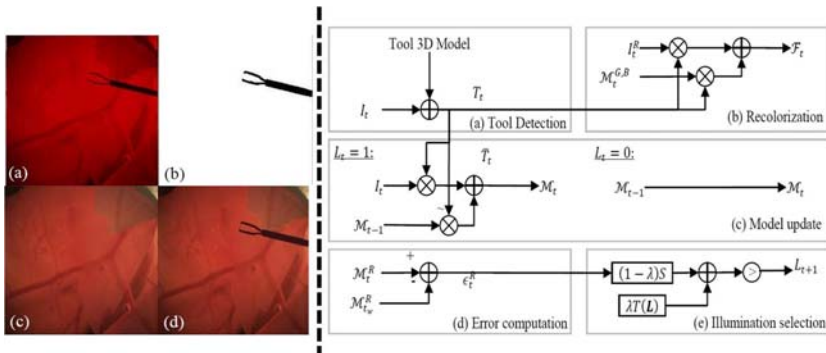


Fig. 1. (Left) Intermediate steps of the AASR algorithm: (a) non-white image provided by the device when $L_t = 0$, (b) tool segmentation, (c) representation of M_t and (d) recolored image by AASR. (Right) Block diagram of the proposed system. See section 2.3 for details.

3 Experiments

We now show how our system performs on image sequences from phantoms and from chorioallatonic chicken embryos. First, a quantitative comparison of AASR and a state-of-the-art method is presented, where it is shown that AASR surpasses ASR in a setting where ground truth is known. This is shown by measuring both image recoloring quality and quantity of white light used. We then show qualitative results of our algorithm on image sequences.

To validate the approach described in section 2, we recorded 5 image sequences of membrane peelings on phantom eyes using only white light. Each sequence consists of approximately 300 frames in similar visual settings. Doing so allows us to synthetically generate limited-spectrum images at any given time, by using only the red channel of white light images. This provides us with a way to quantitatively compare AASR and ASR, as ground truth is available.

For each image sequence we then ran AASR with three different settings: $\lambda = \{0.25, 0.5, 0.75\}$. This allows us to see results for cases where the surgeon applies a bias towards image quality, phototoxic levels, or no bias at all. For each image sequence, we also generated 4 recolored sequences using ASR, with different values of $\phi = \{1/2, 1/4, 1/8, 1/16\}$. As in [9], the L_2 (or mean squared error) norm is chosen to measure the error between the ground truth and the rendered images. In order to estimate phototoxicity levels, we observe the proportion of white-light images used.

In Fig. 2(Left) we show the results of this experiment by plotting the average recolorization error against the average estimated phototoxicity level. The dotted line (4 vertices; 1 for each value of ϕ) shows how ASR performs while the full line describes the performance of AASR (3 vertices; 1 for each value of λ). In general, we can notice that both methods displays a trade-off in accuracy: reducing one type of error induces the other and vice versa. We can also see that the AASR

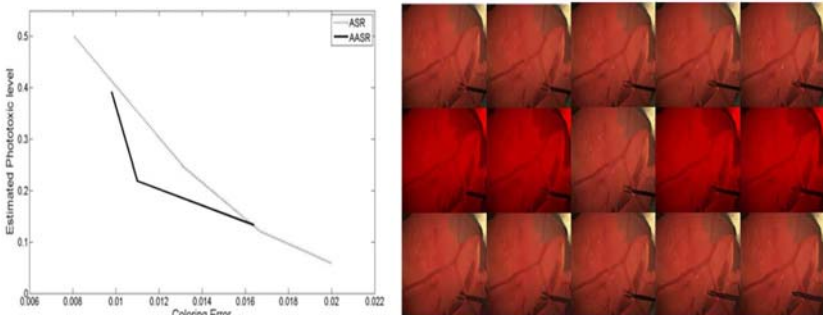


Fig. 2. (Left) Plot of estimated phototoxicity levels and recoloring error for both ASR and AASR. Notice that AASR is less phototoxic than ASR for every recoloring error level. (Right) Example image sequence of membrane peel. (top) Ground Truth, (middle) white and non-white illumination images triggered by AASR and (bottom) AASR image recolorization.

curve lies below that of ASR for every recolorization error level, hence achieving smaller total costs for the patient. In general, from our current experimental setup, AASR significantly outperform ASR, for the values of λ specified. Also, note that if all incoming images were registered to a reference frame (as in [9]) an additional reduction in colorization error would be expected.

Having observed that AASR provides a better way to model retinal-type scenes, we now present results on a typical image sequence of a chorioallatonic membrane peel from an 12 day old chicken embryo. In Fig. Fig. 2 (Right) we show a small set of images from this sequence and the resulting recolorization using AASR ($\lambda = 0.5$). The original and recolored video sequence can be seen in a video included in the supplementary materials. In the video, the same peeling sequence is visible and each row corresponds to a different value for $\lambda = \{0.25, 0.5, 0.75\}$. The first column shows the original images. The second column displays the images provided by the device, while the third column shows the images recolored by AASR. The last column displays the retina color model over time. Other similar video sequences are provided in the supplementary materials.

Notice that in general, in image sequences which contain little membrane manipulations, few white light images are used. Since in this scenario our prediction model is capable of correctly estimating the colors of the retina, few white light images are necessary. Conversely, frames which show membrane peeling require more frequent white light illumination, in order to correctly render the colors. This indicates that the framework is able to choose which illumination type to use depending on the surgeon's actions.

4 Discussion and Conclusion

In this paper we have presented a novel algorithm that can be used to reduce toxic light exposure during retinal microsurgery. When used with the LED light source from [9], our recoloring scheme can dynamically choose the illumination based on the circumstances, reducing potential light induced retinal toxicity. Our algorithm balances the risks of phototoxic retinal damage with the illumination requirements of the operating surgeon to perform the surgical tasks. In this report we provide qualitative and quantitative evidence that this novel method reduces the dose of light, and hence retinal damage, while maintaining sufficient illumination to execute required surgical maneuvers safely.

While the results we have presented are in part dependent on the modeling choices of the cost functions, our framework is generic enough to accommodate a large number of functions. This being said, a natural future direction to improve the present work is to empirically determine the specific forms of the cost functions to use. Determination of these relationships would permit a truthful quantitative evaluation of the harm reduction. In ongoing and future work, we will be exploring these issues.

Acknowledgements

Funding for this research was provided in part by NIH Grant R01 EB 007969-01, a research subcontract from Equinox Corporation and by NSF Cooperative Agreement EEC9731478.

References

1. McDonald, H., Irvine, A.: Light-induced maculopathy from the operating microscope in extracapsular cataract extraction and intraocular lens implantation. *Ophthalmology* 90, 945–951 (1983)
2. Khwarg, S., Linstone, F., Daniels, S., Isenberg, S., Hanscom, T., Geoghegan, M., Straatsma, B.: Incidence, risk factors, and morphology in operating microscope light retinopathy. I. *Am. J. Ophthalmol.* 103, 255–263 (1987)
3. Byrnes, G., Antoszyk, A., Mazur, D., Kao, T., Miller, S.: Photic maculopathy after extracapsular cataract surgery. a prospective study. *Ophthalmology*, 731–737 (1992)
4. International Commission on Non-Ionizing Radiation Protection: Guidelines on limits of exposure to broad-band incoherent optical radiation (0.38 to 3). *Health Phys.* 73, 539–554 (1997)
5. van den Biesen, R., Berenschot, T., Verdaasdonk, R., van Weelden, H., van Norren, D.: Endoillumination during vitrectomy and phototoxicity thresholds. *Br. J. Ophthalmol.* 84, 1372–1375 (2000)
6. Poliner, L., Tornambe, P.: Retinal pigment epitheliopathy after macular hole surgery. *Ophthalmology* 99, 1671–1677 (1992)
7. Michels, M., Lewis, H., Abrams, G., Han, D., Mieler, W., Neitz, J.: Macular phototoxicity caused by fiberoptic endoillumination during pars plana vitrectomy. *Am. J. Ophthalmol.* 114, 287–292 (1992)
8. Banker, A., Freeman, W., Kim, J., Munguia, D., Azen, S.: Vision-threatening complications of surgery for full-thickness macular holes. *Ophthalmology* 104, 1442–1453 (1997)
9. Sznitman, R., Billings, S., Rother, D., Mirota, D., Yang, Y., Handa, J., Gehlbach, P., Kang, J., Hager, G., Taylor, R.: Active multispectral illumination and image fusion for retinal microsurgery. In: Navab, N., Jannin, P. (eds.) *IPCAI 2010*. LNCS, vol. 6135, pp. 12–22. Springer, Heidelberg (2010)
10. Museum of Broadcast Communication: Encyclopedia of Television, O.: <http://www.museum.tv/archives/etv/c/htmlc/colorization/colorization.htm>
11. Yatziv, L., Sapiro, G.: Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing* 15(5), 1120 (2006)
12. Skora, D., Burinek, J., Zra, J.: Unsupervised colorization of black and white cartoons. In: *Int. Symp. NPAR, Annecy*, pp. 121–127 (2004)
13. Sznitman, R., Lin, H., Manaswi, G., Hager, G.: Active background modeling: Actors on a stage. In: *ICCV, Workshop on Visual Surveillance*, pp. 1222–1228 (2009)
14. Ham, W.J., Mueller, H., Ruffolo, J.J., Guerry, D., Guerry, R.: Action spectrum for retinal injury from near-ultraviolet radiation in the aphakic monkey. *Am. J. Ophthalmol.* 93, 299–306 (1982)

Motion Artifact Correction of Multi-Photon Imaging of Awake Mice Models Using Speed Embedded HMM

Taoyi Chen^{1,2}, Zhong Xue¹, Changhong Wang², Zhenshen Qu², Kelvin K. Wong¹,
and Stephen T.C. Wong¹

¹ Center for Bioengineering and Informatics, Methodist Hospital Research Institute and Department of Radiology, The Methodist Hospital, Weill Cornell Medical College, Houston, TX

² Department of Control Science and Engineering, Harbin Institute of Technology, China
zxue@tmhs.org

Abstract. Multi-photon fluorescence microscopy (MFM) captures high-resolution anatomical and functional fluorescence image sequences and can be used for the intact brain imaging of small animals. Recently, it has been extended from imaging anesthetized and head-stabilized animals to awake and head-restrained ones for *in vivo* neurological study. In these applications, motion correction is an important pre-processing step since brain pulsation and tiny body movement can cause motion artifacts and prevent stable serial image acquisition at such a high spatial resolution. This paper proposes a speed embedded hidden Markov model (SEHMM) for motion correction in MFM imaging of awake head-restrained mice. The algorithm extends the traditional HMM method by embedding a motion prediction model to better estimate the state transition probability. SEHMM is a line-by-line motion correction algorithm, which is implemented within the in-focal-plane 2-D videos and can operate directly on the motion-distorted imaging data without external signal measurements such as the movement, heartbeat, respiration, or muscular tension. In experiments, we demonstrate that SEHMM is more accurate than traditional HMM using both simulated and real MFM image sequences.

1 Introduction

Mammalian brain microscopy plays an important role in studying the electrical and biochemical processes in neurons. Previously, CCD imaging had been used to study mouse brain [1], in which a single fiber-optic bundle is placed in direct contact in the barrel cortex of an awake mouse to image cortical activities in conjunction with voltage sensitive dye. Since fiber-optic bundle have intrinsic low spatial resolution limited by the size of the fiber core, such kind of imaging system does not reach cellular-resolution but allows researchers to obtain cortical activation map with high temporal resolution. Recently, multi-photon fluorescence microscopy (MFM) imaging has been used to capture neural images from anesthetized mice at higher spatial and temporal resolution [2-5]. In [2] and [5], Ca^{2+} transients were quantified using MFM in the mice models of Alzheimer's disease (AD). Unfortunately, animals under anesthesia cannot demonstrate the neural dynamics sufficiently, and overall brain activities are suppressed. Recently, more MFM imaging studies of awake mice were carried out

[6-9], and two-photon microscope (TPM), a special variant of MFM, was demonstrated as a superior alternative due to its deeper tissue penetration, efficient light detection, and reduced phototoxicity [3].

Head-restrained TPM [8, 9] uses a standard microscope system to visualize simultaneous Ca^{2+} dynamics of a large number of individual cells, while the head-restrained mouse is free to walk or run on a stationary exercise ball [10]. Because of the mice's random motion, the relative motion between the microscope and the brain leads to line shifting artifacts, which affects image reconstruction and quantification of the functional dynamics. Thus a motion correction algorithm is needed for quantifying brain motion and aiding in stabilizing cell displacements within a region of interest (ROI) in order to reduce the number of brain-motion-related fluorescence transients.

In the literature, two methods were proposed for the motion correction of TPM in awake mice [8, 9]. Greenberg and Kerr [9] proposed a method based on the Lucas-Kanade algorithm [11, 12] to estimate the offset sequences by minimizing the squared differences between a reference frame and the subsequent frames using gradient descent method. In [8], an HMM-based method was proposed to calculate the optimal offset using a probabilistic framework. Compared to [11, 12], HMM algorithm generates more temporally stable results due to the implicit modeling of longitudinal state transition. Basically, the HMM algorithm assumes that the motion of the current line to the next line remains the same or is most likely not changing. A standard exponential model (or distribution) is used to describe this assumption. Essentially, this assumption is only realistic when the mice are still or relatively still (the relative position for scanning beam and mice's local neural region of interest stays constant), but this situation seems not very common because mice's motion is disordered and random and includes sudden changes of speed all the time during the procedure. Thus, in HMM, although the state transition model works effectively for the resting stage, it might fail and give a wrong estimation during the running stage.

In this paper, we extend the HMM by incorporating an estimated speed into the state transition model for more accurate estimation, called speed embedded HMM (SEHMM) algorithm. The major contributions for SEHMM are as follows. (1) Initial motion estimation is achieved by using a simple line-by-line searching method; (2) a grouping algorithm is used to divide the whole imaging period into resting and running stages for speed estimation; (3) SEHMM is then adopted for motion correction. In experiments, a quantitative validation was performed to compare HMM with SEHMM based on both simulated data and real data. First, simulated image sequences were generated to mimic various real motion situations, and we applied different dynamic amplitudes to make the validation more realistic and reasonable. For real data, we showed the comparative results to demonstrate the performance of SEHMM. The results illustrated that SEHMM can achieve higher estimation accuracy of image alignment as compared with HMM, especially in the running stages of the image sequences.

2 Method

2.1 Problem Formulation

A two-photon microscope captures a series of images by passing a focus of laser excitation repeatedly over a rectangular or square region of fluorescently labeled

tissue and collecting the resulting photons via a photon multiplier tube (PMT) [13]. Although the mice's motion is in 3-D, Z-axis motion shift is less than $1\mu m$ for a scanning speed of 2 ms/line , much lower than the motion along the X (medial-lateral direction along the raster scan line) and Y (rostral-caudal direction across the raster scan line) directions. Therefore, our task is to estimate the motion in X- and Y-directions. Due to the mice's motion during the raster scan progression, the relative motion can be written as,

$$\begin{cases} X_i^k = X_i^k + \delta_x^k \\ Y_i^k = Y_i^k + \delta_y^k \end{cases}, \quad (1)$$

where $X_i^k(t) = \{t / (\tau / N)\} \cdot N$ and $X_i^k(t) = [t / (\tau / N)]$ give the actual location of an object point, and (δ_x^k, δ_y^k) is its offset due to motion. $N \times N$ is the size of each frame. $[\cdot]$ represents the integer operation, and $\{\cdot\}$ denotes the fractional operation. τ is the scanning time for a frame. It can be seen that the laser moves in a zigzag pattern in X-direction and a step function pattern in Y-direction. Our goal is to estimate the offsets (δ_x^k, δ_y^k) from the serial images. Eq. (1) assumes each line has the same relative displacement, so we can choose a line-by-line motion correction algorithm to solve this problem. The reason is that the shifts for all the pixels within a line do not get beyond one-pixel in Y-direction and are very tiny in X direction. Although pixel-by-pixel correction can yield more accurate results, one has to trade off between computational speed and the gain in accuracy.

2.2 The Speed Embedded HMM (SEHMM) Algorithm

The speed embedded HMM (SEHMM) algorithm is an extension of the traditional HMM method by using a motion prediction model with HMM to better estimate the state transition probability. Denoting the displacement state for line k as (δ_x^k, δ_y^k) , the state observation probability is represented as $\pi_k^{\delta_x^k, \delta_y^k}$, and the state transition probability is denoted as $T[(\delta_x^{k-1}, \delta_y^{k-1}) \rightarrow (\delta_x^k, \delta_y^k)]$. The idea of SEHMM is that the transition probability of the displacement state is estimated based on the estimated moving speed. Thus, SEHMM can match motion more accurately not only during the resting stage but also during the moving stage.

In SEHMM, first, the transition probability is defined as,

$$T(\delta_x^{k-1}, \delta_y^{k-1}) \rightarrow (\delta_x^k, \delta_y^k) = \frac{1}{2\pi\lambda} e^{-r/\lambda}, \quad (2)$$

where r is defined as:

$$r = \sqrt{(\delta_x^k - (\delta_x^{k-1} + v_x^{k-1,k} \tau_{line}))^2 + (\delta_y^k - (\delta_y^{k-1} + v_y^{k-1,k} \tau_{line}))^2}, \quad (3)$$

where $v_x^{k-1,k}$ and $v_y^{k-1,k}$ are the estimated speed of the motion from line $k-1$ to k for X- and Y-directions, respectively. τ_{line} is the scanning time for a line. It can be seen that using Eq. (3), if the moving speed is estimated at line $k-1$, the offset at line k can be estimated. Therefore, in SEHMM, we no longer assume that the state transition probability is the highest when the mouse does not move. On the contrary, this

probability gets its peak at a linearly estimated offset value. Since the goal for motion correction is to estimate δ_x^k, δ_y^k for each line k from a given reference frame R and the current image I , we implement it by maximizing a posteriori,

$$P=(\delta_x^k, \delta_y^k | I_i^k, R) = P(I_i^k | R(X_i^k, Y_i^k)) \cdot P(\delta_x^k, \delta_y^k | \delta_x^{k-1}, \delta_y^{k-1}) = \pi_k^{\delta_x, \delta_y} \cdot T(\delta_x^{k-1}, \delta_y^{k-1}) \rightarrow (\delta_x^k, \delta_y^k). \quad (4)$$

Expressing P as a logarithm probability, one gets,

$$\ln(P) = \ln(\pi_k^{\delta_x, \delta_y}) + \ln(T). \quad (5)$$

The first term is the state observation probability $\pi_k^{\delta_x, \delta_y}$ and reflects the goodness of matching between a line in the current frame I and the corresponding line in the reference frame R . Denoting the intensity of the i th pixel in the k th line of the current frame as I_i^k , and that of the corresponding pixel in the reference frame as (x_i^k, y_i^k) , the observation probability can be modeled as a discrete Poisson distribution explicitly, $\pi_{k,i}^{\delta_x, \delta_y} = (\gamma R)^{\gamma I} \cdot e^{-\gamma R} / (\gamma I)!$, where γ is the calibration factor of the photon number, which is an inherent factor in an imaging system. Taking the logarithm transformation of $\pi_{k,i}^{\delta_x, \delta_y}$ we get,

$$\ln(\pi_{k,i}^{\delta_x, \delta_y}) = \gamma I \ln(\gamma) + \gamma I \ln(R) - \ln((\gamma I)!) - \gamma R, \quad (6)$$

where γ and I_i^k are independent of the changing offsets, and R is indeed a function of the offsets. Thus the equation can be simplified as,

$$\ln(\pi_{k,i}^{\delta_x, \delta_y}) \propto \gamma(I \ln(R) - R). \quad (7)$$

Then, the observation probabilities for all the pixels within line k can be calculated by,

$$\ln(\pi_k^{\delta_x, \delta_y}) = \sum_{i=1}^N \ln(\pi_{k,i}^{\delta_x, \delta_y}). \quad (8)$$

To illustrate how this new formulation works, Fig. 1 shows the state transition probabilities for HMM (Fig. 1(a)) and SEHMM (Fig. 1(b)) models, respectively. The center of each distribution plot is set to $(\delta_x^{k-1}, \delta_y^{k-1})$, and it can be seen that the peak of the state transition probability is right in the middle for Fig. 1 (a), and this peak can be shifted in Fig. 1 (b) according to the motion estimation. A shifting of the peak of the probability function indicates that a motion is assumed from one line to another. This compensation for adding the estimated speed value from the current offset to the following offset can remarkably improve the estimation accuracy because it satisfies the prediction requirement in all the time points but not only in the still time points.

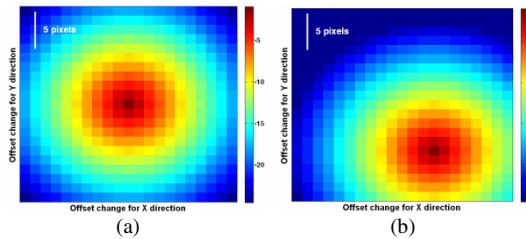


Fig. 1. The state transition probability functions without (a) and with (b) motion estimation

Once the state observation probability $\pi_k^{\delta_x, \delta_y}$ and the state transition probability $T[(\delta_x^{k-1}, \delta_y^{k-1}) \rightarrow (\delta_x^k, \delta_y^k)]$ are defined, the best displacement signal can be calculated through maximizing Eq. (5), by following two iterative steps: first, the Expectation Maximization (EM) algorithm [14] can be used to calculate the best λ that leads to the highest probability for Eq. (5). Our algorithm for doing this involves systematically scanning values of λ in a prescribed range. We scan uniformly in the logarithm space to get constant percentage sampling. For each value of λ , we run the SEHMM procedure to find the most optimal offset sequence and calculate its total probability. The value of λ we will choose is the one with the most probable offset sequence. Then, we can adopt the Viterbi algorithm to find the most likely sequence of the hidden states (offsets). This can be accomplished in two steps. First, we determine the most probable offset sequence for every state at time (line) k from any of the states at time (line) $k-1$ by marching forward through the time domain. Then, we can backtrack along the path of the most probable offset sequence to record the results of optimal offset sequence. In the implementation, a line-by-line search is first used to estimate the initial offset, and the speed of the motion can be estimated by applying the smoothness filter temporally on the estimated offsets before running SEHMM.

3 Results

3.1 Results on Simulated Datasets

Simulated serial images are used to quantitatively validate the SEHMM algorithm by comparing with the HMM method [8]. Temporal displacements are first generated to transfer a real 2D image (see Fig. 2(left)) to produce the longitudinal image sequences. The displacement signal can be generated through a pixel-by-pixel basis using a first-order differential equation with a characteristic time constant driven by a combination of the random and deterministic processes to simulate the effect of acquiring laser-scanning data from a moving sample:

$$\begin{cases} \bar{\phi}_x^{t+1} = (1 - \frac{1}{\tau})\bar{\phi}_x^t + \frac{1}{\tau}(\bar{\zeta}_{\sigma_x} + \bar{D}_x^t) \\ \bar{\phi}_y^{t+1} = (1 - \frac{1}{\tau})\bar{\phi}_y^t + \frac{1}{\tau}(\bar{\zeta}_{\sigma_y} + \bar{D}_y^t) \end{cases}, \quad (9)$$

where $\bar{\phi}_x^t$ and $\bar{\phi}_y^t$ denote the simulated offsets for X- and Y-directions, respectively, and t expresses the time-point, $t = (k-1) \cdot N + i$. τ is set to 500. σ_x and σ_y ($\sigma_x = 4, \sigma_y = 40$) are the standard deviations of the Gaussian random variables $\bar{\zeta}_{\sigma_x}$ and $\bar{\zeta}_{\sigma_y}$. \bar{D}_x^t and \bar{D}_y^t are the step functions at the time t . Then, the Poisson noise is added on a pixel-by-pixel basis to simulate the photon counting statistics because the values during scanning are Poisson distributed in terms of photon numbers but not in units of pixel intensity. σ_x , σ_y and τ control the amplitudes of temporal dynamic displacements. One simulated dataset (T1) was generated with these default parameters. In addition, we also changed the values of σ_x , σ_y and τ and kept the other parameters the same, and the second dataset (T2) was simulated by setting $\sigma_x = 10$, $\sigma_y = 30$ and

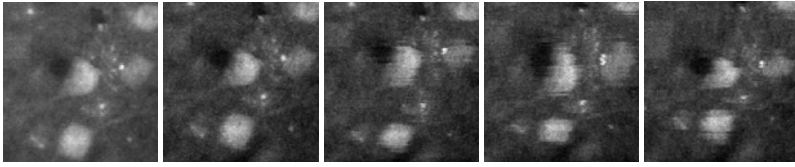


Fig. 2. The reference image (left) and some examples of the simulated images

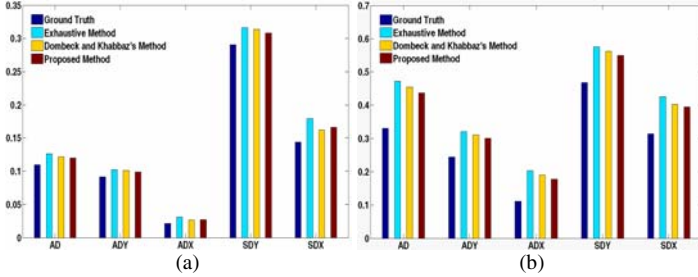


Fig. 3. Quantitative comparisons for four simulated datasets. (a-b) results of datasets T1 and T2.

$\tau = 200$. The number of frames for T1 and T2 are 65 and 51, respectively. The image size is 129×129 , and resolution is $0.39 \mu\text{m}$ /pixel.

Denoting the results by HMM as (δ_x^k, δ_y^k) , and those of SEHMM as $(\hat{\delta}_x^k, \hat{\delta}_y^k)$, the simulated offsets as (ϕ_x^k, ϕ_y^k) , and the exhaustive search results as $(\hat{\phi}_x^k, \hat{\phi}_y^k)$, we compared their motion correction accuracy. As all the three methods tested are line-by-line motion correction and the ground truth is stored in pixel resolution, we take the mean value of all ground truth offsets within each line and then round the mean value to the nearest integral, i.e., $\phi_x^k, \phi_y^k \rightarrow \phi_x^k, \phi_y^k$. The difference between these two data was also compared, which indicates how accurate we can achieve using line-by-line based methods. We used the following five performance measures: average distance (AD), average distance for Y-direction (ADY), average distance for X-direction (ADX), standard deviation for Y-direction (SDY) and standard deviation for X-direction (SDX), to evaluate the performance of the algorithms. Fig. 3 shows the comparison results between the four line-by-line results and the ground truth for data T1 and T2, respectively. The blue bars show differences between the rounded line-by-line ground truth and the simulated pixel-by-pixel ground truth, and other bars show the results of the three methods. It can be seen that for the performance measures, SEHMM yields the smallest errors (some with slightly larger SDX).

We also focused on the Y-direction because the offsets in this direction have relatively large movement than in X-direction. Fig. 4 illustrates the results for simulated dataset T2. In the figure, we show the line-by-line rounded ground truth of and the results of SEHMM. Since these two curves almost overlap each other and it is difficult to visually see the differences, we thus plot the inconsistency between them. The inconsistency is calculated as the number of lines within a frame that the results of SEHMM and HMM are different with the ground truth. It can be seen that during the resting stage, such inconsistency is small, and both SEHMM and HMM yielded good results (about 10 lines are different), and during the running stage, SHEMM can reduce the inconsistency as compared with the results of HMM.

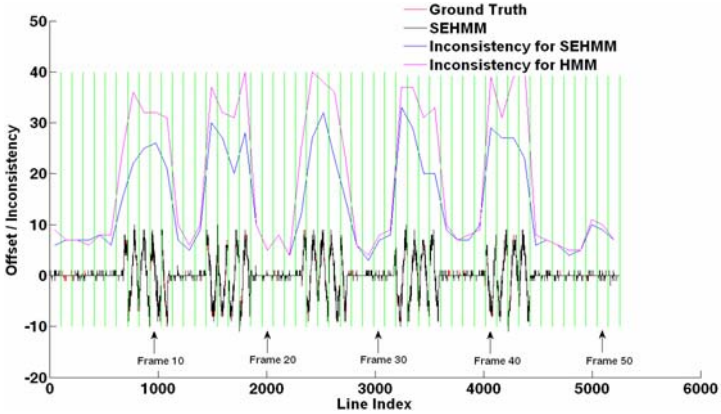


Fig. 4. Comparison results for simulated data T2. Green lines indicate frame separation.

3.2 Experiments on Real Data

For real data, we used the mean absolute intensity difference as the measure to show the performance of the methods. The mean absolute intensity difference across longitudinally corresponding pixels reflects the goodness of matching for the image sequences. Table 1 shows the comparison results for data T3 and T4 (provided by Dr. David Tank), respectively, using different methods. The number of frames are 1200 and 200, image size is 128x128 and 256x256, respectively, and the resolution is 0.39 μm /pixel for both. The improvement can be seen from the comparison results when we adopt the proposed method.

Table 1. Average of absolute intensity differences for real datasets T3 and T4

Method	Original Data	Exhaustive Method	HMM	SEHMM
T3	7.05	5.89	5.48	5.03
T4	9.54	7.74	7.15	6.56

Regarding computational speed, the exhaustive method is the fastest because limited number of possible motion is considered, and the SEHMM method is slightly more complex than the HMM method because of the embedded exhaustive method in the preprocessing step. However, during the actual motion estimation, HMM and SEHMM have essentially the same speed. Regarding motion correction accuracy, the exhaustive method is not the optimal solution for motion correction because there exist strong line-to-line correlations for offsets of successive lines, and exhaustive method does not consider this point. SEHMM outperforms HMM because the estimated speed of motion is embedded in the model itself.

4 Conclusion

We proposed a speed embedded hidden Markov model (SEHMM) for motion correction in MFM imaging. The key idea of SEHMM is that a motion prediction model is utilized to better estimate the state transition probability. Compared to the traditional HMM that assumes no motion always has the highest probability, the proposed

SEHMM can model the motion more accurately. SEHMM operates directly on the motion-distorted imaging data, without any external signal measurement such as the mice's movements, heartbeat, respiration, or muscular tension. Using simulated and real images, we demonstrated that SEHMM is more accurate than HMM using both simulated and real MFM image sequences.

Acknowledgement

The authors would like to thank Dr. David Tank from Princeton University for providing the data.

References

1. Ferezou, I., Bolea, S., Petersen, C.C.: Visualizing the Cortical Representation of Whisker Touch: Voltage-Sensitive Dye Imaging in Freely Moving Mice. *Neuron* 50, 617–629 (2006)
2. Kuchibhotla, K.V., Lattarulo, C.R., Hyman, B.T., Bacskaï, B.J.: Synchronous Hyperactivity and Intercellular Calcium Waves in Astrocytes in Alzheimer Mice. *Science* 323, 1211 (2009)
3. Denk, W., Strickler, J.H., Webb, W.W.: Two-Photon Laser Scanning Fluorescence Microscopy. *Science* 248, 73–76 (1990)
4. Denk, W., Svoboda, K.: Photon Upmanship: Techreview Why Multiphoton Imaging Is More Than a Gimmick. *Neuron* 18, 351–357 (1997)
5. Busche, M.A., Eichhoff, G., Adelsberger, H., Abramowski, D., Wiederhold, K.H., Haass, C., Staufenbiel, M., Konnerth, A., Garaschuk, O.: Clusters of Hyperactive Neurons near Amyloid Plaques in a Mouse Model of Alzheimer's Disease. *Science* 321, 1686 (2008)
6. Helmchen, F., Fee, M.S., Tank, D.W., Denk, W.: A Miniature Head-Mounted Two-Photon Microscope High-Resolution Brain Imaging in Freely Moving Animals. *Neuron* 31, 903–912 (2001)
7. Flusberg, B.A., Jung, J.C., Cocker, E.D., Anderson, E.P., Schnitzer, M.J.: In Vivo Brain Imaging Using a Portable 3.9 Gram Two-Photon Fluorescence Microendoscope. *Optics letters* 30, 2272–2274 (2005)
8. Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L., Tank, D.W.: Imaging Large-Scale Neural Activity with Cellular Resolution in Awake, Mobile Mice. *Neuron* 56, 43–57 (2007)
9. Greenberg, D.S., Kerr, J.N.D.: Automated Correction of Fast Motion Artifacts for Two-Photon Imaging of Awake Animals. *Journal of Neuroscience Methods* 176, 1–15 (2009)
10. Wilt, B.A., Burns, L.D., Ho, E.T.W., Ghosh, K.K., Mukamel, E.A., Schnitzer, M.J.: Advances in Light Microscopy for Neuroscience. *Annual review of neuroscience* 32, 435–506 (2009)
11. Lucas, B.D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: *Proceedings of the 1981 DARPA Image Understanding Workshop*, vol. 3, pp. 121–130 (1981)
12. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision* 56, 221–255 (2004)
13. Foord, R., Jones, R., Oliver, C.J., Pike, E.R.: The Use of Photomultiplier Tubes for Photon Counting. *Applied Optics* 8, 1975–1989 (1969)
14. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data Via the Em Algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, 1–38 (1977)

Diagnostic Radiograph Based 3D Bone Reconstruction Framework: Application to Osteotomy Surgical Planning

Pavan Gamage¹, Sheng Quan Xie¹, Patrice Delmas², and Wei Liang Xu³

¹ Department of Mechanical Engineering, The University of Auckland, New Zealand

² Department of Computer Science, The University of Auckland, New Zealand

³ School of Engineering & Advanced Technology, Massey University, New Zealand

Abstract. Pre-operative planning in orthopedic surgery is essential to identify the optimal surgical considerations for each patient-specific case. The planning for osteotomy is presently conducted through two-dimensional (2D) radiographs, where the surgeon has to mentally visualize the bone deformity. This is due to direct three-dimensional (3D) imaging modalities such as Computed Tomography (CT) still being restricted to a minority of complex orthopedic procedures. This paper presents a novel 3D bone reconstruct technique, through bi-planar 2D radiographic images. The reconstruction will be pertinent to osteotomy surgical diagnostics and planning. The framework utilizes a generic 3D model of the bone of interest to obtain the anatomical topology information. A 2D non-rigid registration is performed between the projected contours of this generic 3D model and extracted edges of the X-ray image to identify the planar customization required. Subsequently a free-form deformation based manipulation is conducted to customize the overall 3D bone shape.

Keywords: Osteotomy Surgical Planning, Patient-Specific Bone Customization.

1 Introduction

Pre-operative planning has been identified as an essential requirement for successful surgical outcomes [1]. The AO (Arbeitsgemeinschaft für Osteosynthesefragen) foundation for orthopedic research instructs on the importance of pre-operative planning through its medical training [1]. The surgeon should emerge from the planning phase with a clear idea of the patient's bone dimensions, the overall surgical approach, and if applicable, a detailed plan of the fixation. In current orthopedic cases most pre-operative diagnostics and planning is conducted through 2D X-ray imaging. Two-dimensional images lack significant spatial information that is present in 3D modalities. Imaging modalities such as CT have the ability to provide direct 3D volumetric images. However the use of such imaging is restricted to a minority of complex procedures due to constraints placed by cost, availability and risks posed by unwarranted detailed imaging. Thus an alternative to direct 3D imaging must be developed to augment procedures that currently rely on pure 2D radiographs. The work presented in this paper is motivated by this requirement, and proposes a 3D bone reconstruction framework from 2D radiographic images. In this paper femoral osteotomy has been

used as the application of interest, however this work can also be applied to several other fracture reduction techniques [2].

Femoral osteotomy is a procedure used for the treatment of congenital and acquired (e.g. fracture malunion) bone deformities. Long bone deformity may consist of a combination of coronal, sagittal and oblique plane distortion. The most common of which are coronal and sagittal plane angular deformities [3]. For these cases, the planar angular and torsional misalignment of the bone axis can be identified through the anterior and lateral radiographs. However several key angles including acetabular anteversion, are not measurable through 2D imaging. Moreover, the plane on which the surgical cut has to be performed and the magnitude of the angular correction can only be accurately defined by measurements viable through 3D spatial visualization [3, 4].

The rest of the paper is structured as follows. Section two describes in detail the framework proposed by the authors and documents the current state-of-the-art research in pre-operative 3D anatomical reconstruction methodologies. Section three details the verifications performed on the developed algorithms and the overall results obtained.

2 Methods

The proposed framework to achieve 3D reconstruction is specified below. The 2D X-ray images are initially processed to extract the edge points that potentially form the femur boundary. A non-rigid registration is then performed between the edges identified in the X-ray image and the projected contour points of the generic model. The identified point correspondences will next be interpolated to create a 2D planar translational field in both the anterior and lateral viewpoints. This translational field will identify the deformations required by the 3D anatomical model in the equivalent viewpoint. Finally a full 3D translational field will be created through a free-form deformation based interpolation and the 3D generic anatomical data deformed accordingly.

2.1 Anatomical Generic Model Initialization

The anatomical generic surface model utilized for our testing was segmented and extracted from CT scan data. Typically femur X-ray images are acquired in only the anterior and lateral viewpoints. Thus the pose (position/orientation) of the bone when imaged is standard on many femur X-ray images. The manual pose initialization will only be required if a different angle of acquisition is employed during the X-ray imaging. This initialization involves a six degree of freedom movement of the 3D model, with three translational and three rotational parameters.

Moreover, the generic model required certain automated pre-processing to identify the surface edge points that would be clearly identifiable on radiographic images. The outer contours in the anterior and lateral directions were identified through projection ray-tracing. As an example, the femoral generic model is shown in Figure 1(a) along with the outer contours. Accurate identification of as many of these edges as possible on radiographic images is vital as the customization process is driven by them, as discussed in the proceeding edge extraction section (Figure 1(b)).

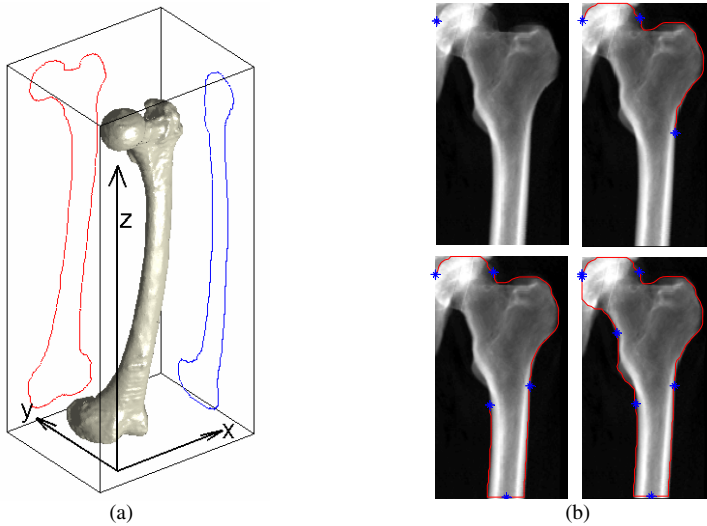


Fig. 1. Example of the clearly identifiable edges on femur radiographic images. From left to right: (a) The femoral generic model with the associated anterior and lateral projected contours; (b) Edge extraction based on an interactive LiveWire segmentation technique (detailed in section 2.2).

2.2 Edge Extraction

The main objective of the proposed edge extraction technique is to identify objects boundaries with sufficient continuity to be successfully employed in the proceeding shape based point correspondence estimation. Radiographic image edge extraction is hindered by poor contrast, ill defined boundaries, noise, and acquisition artifacts. Thus an interactive segmentation technique was utilized to achieve robust and accurate segmentation. The LiveWire technique proposed in this paper is based on [5] with modifications to suit grayscale X-ray image. The interactive segmentation requires the user to initially specify a seed point on the object boundary utilizing an input device (mouse). The input device must subsequently be moved to advance the cursor to a point further along the object boundary. A globally optimum path from the initial seed point to the current point is computed and displayed in real time. The optimal path is determined by assigning cost functions to boundary elements (edge strength/edge orientation), and then finding the minimum cost path. As the user moves the cursor slightly, different paths are computed and displayed. If the cursor moves close to the boundary, the LiveWire snaps to the edge. If the user is satisfied with the computed boundary segment, the cursor point can be set. This point becomes the new seed point and the recursive process continues.

The minimum cost path should correspond to an image object boundary that exhibit strong edge features. Thus, the local cost matrix is created from two edge features: gradient magnitude and gradient direction. The overall local cost matrix is computed as a weighted sum of these components. $C(m,n)$ represents the local cost of the link from pixel m to a neighboring pixel n , where f_G and f_D are the gradient magnitude and gradient direction cost functions respectively and w_G and w_D are the corresponding scalar weights.

$$C(m, n) = w_G f_G + w_D f_D \quad (1)$$

Gradient magnitude provides a direct correlation between edge strength and local cost. If F is the grayscale X-ray image, then the gradient magnitude G is approximated with,

$$G = \|\nabla F\| = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2} \quad (2)$$

The gradient is next scaled and inverted to ensure high gradients produce low costs to facilitate minimal cost path optimization.

$$f_G = 1 - \frac{G}{\max(G)} \quad (3)$$

Gradient direction adds a smoothness constraint to the boundary by associating a high cost for sharp changes in boundary direction. The gradient direction is the unit vector defined by the partials of an image I in x and y (I_x and I_y). The gradient direction feature cost can be defined as f_d , where $D(m)$ is the unit vector perpendicular (rotated 90 degrees clockwise) to the gradient direction at point m , and $n-m$ is the bidirectional link or edge vector between neighboring pixels m and n pointing towards n . Links are either horizontal, vertical, or diagonal (8-connectivity).

$$f_d(m, n) = \frac{1}{\pi} [\arccos(c \cdot D(m) \cdot (n-m)) + \arccos(c \cdot D(n) \cdot (n-m))] \quad (4)$$

Where,

$$D(m) = (I_y(m), -I_x(m)), \quad D(n) = (I_y(n), -I_x(n)) \quad \text{and} \quad (5)$$

$$c = \text{sign}(D(m) \cdot (n-m))$$

The link direction forces high cost to an edge between two pixels that have similar gradient directions but are near perpendicular to the link between them. Thus the direction feature cost is low when the gradient directions of the two pixels are similar to each other and the link between them.

Finally, 2D graph searching is employed for an optimal minimal cost path selection. This paper utilizes the optimal graph search presented by [6]. Dijkstra's algorithm is a graph search algorithm that solves the shortest path (path with the lowest cost) problem for a graph with nonnegative edge path costs. The sequence of images in Figure 1(b) shows the interactive LiveWire segmentation technique applied to an anterior femur X-ray image.

2.3 Shape Customization – Deformation

A non-rigid registration between the 2D projective contours of the 3D generic model and the extracted edges of the patient-specific X-ray images is performed to customize the shape of the generic bone to that of the patient. The registration is performed through a point correspondence estimation between the two point sets.

The similarity measure proposed in this paper to provide a measure of “correspondence” between two pairs of points or features in different images is based on shape histograms. The shape descriptor (Shape Context) describes the distribution of a series of points with respect to a given point on a shape [7]. The shape context based cost function (C_{shape}) to match a point P_m on the model contour to a point P_t on the target contour can be expressed as (6), where $h_m(n)$ and $h_t(n)$ denote the N-bin histogram (normalized) at P_m and P_t . Due to the chi-squared test used as the matching cost between the two shapes the C_{shape} similarity measure is intrinsically bounded in [0,1]. The primary benefit of utilizing this measure is that it is invariant to translation, rotation, scale, and shape deformation of the two shapes under consideration.

$$C_{shape}(P_m, P_t) = \frac{1}{2} \sum_{n=1}^N \frac{(h_m(n) - h_t(n))^2}{h_m(n) + h_t(n)} \quad (6)$$

The total cost of matching the point sets can be minimized through a Bi-partite graph searching methodology. The minimization of the total cost matrix C_{total} between point sets, P_m and P_t , is subject to the constraint that the matching is one-to-one. This square assignment problem is solved through the Hungarian algorithm with a time complexity of $O(N^3)$ (Figure 2(a)) [8].

The correspondences identified through the Hungarian algorithm are subsequently processed and filtered to remove any outliers (misidentified correspondences). The filtering is performed on a moving window of the identified translational values (of the correspondences) and by ensuring that they are within ± 3 standard deviations from the local mean.

Subsequent to the planar interpolation, a 3D translational field to deform the generic model is created. The full 3D deformation is based on a cubic B-spline free-form deformation (FFD) model. FFD, introduced by [9], deforms an object by manipulating a regularly subdivided 3D parallelepiped lattice containing the object. By manipulating a mesh of control points a deformation function that specifies a new position for each point on the object is calculated. In this way the deformations of the FFD lattice are passed onto the object.

The result of FFD can be obtained from the 3D tensor product cubic B-splines, along each of the three directions, U, V and W, as below,

$$T(u, v, w) = \sum_{i,j,k=0}^3 B_i(u) B_j(v) B_k(w) P_{ijk} \quad (7)$$

Where, $0 \leq u, v, w \leq 1$ and $B_i(u)$, $B_j(v)$ and $B_k(w)$ are defined as the uniform cubic B-spline basis functions evaluated at u , v and w , respectively. The main objective of the FFD based deformation technique is to utilize the scattered object point translations and to configure the control grid points such that the deformed location of the selected point matches the target point location (Figure 2(b)). This is achieved in a least squares sense, based on the manipulation algorithm proposed by [10], with additions to enable multiple point movements.

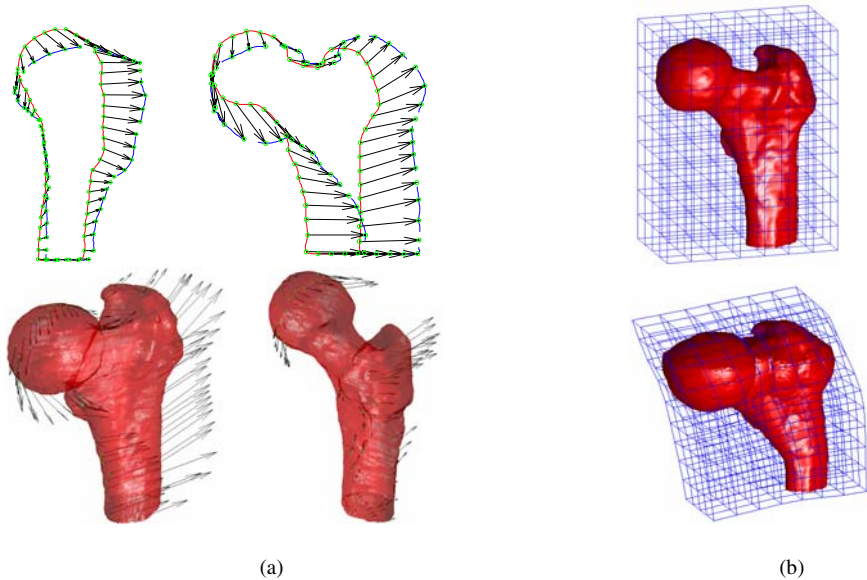


Fig. 2. Non-rigid registration performed on the edge contours and the 3D translational field. From left to right: (a) Non-rigid registration on the sub-sampled edge contour data points for the anterior/lateral views (top) and the associated translational field displayed on the generic model (bottom); (b) FFD based interpolation of the transformation field (bottom) to deform the generic model (top).

3 Results

To validate the reconstruction framework, three cases of bone deformity were tested, femoral shaft deformity (Bowling), femoral head deformity (Shepherd’s Crook), femoral condyles deformity (Erlenmeyer Flask). Two CT scans of each type of deformity and the associated X-ray images in the anterior and lateral view points were utilized in the tests. The X-ray images were used for the reconstruction and the corresponding CT scan data for accuracy assessment. Table 1 shows the quantitative results obtained from the six tests performed. Figure 3 shows several examples of the testing performed. The key accuracy measurement used during the testing is the Euclidean distance between the closest points of the reconstructed data and the ground truth CT scan data set. A discussion of the results is presented below.

Table 1. Osteotomy surgical planning 3D model reconstruction results

Deformity Type	Maximum Absolute Error (mm)	Average Absolute Error (mm)	Standard Deviation (mm)
Femoral Shaft	1.31	0.92	0.093
Femoral Shaft	1.69	1.12	0.089
Femoral Head	1.86	0.99	0.093
Femoral Head	2.36	1.14	0.094
Femoral Condyles	2.05	1.05	0.089
Femoral Condyles	2.56	1.21	0.091

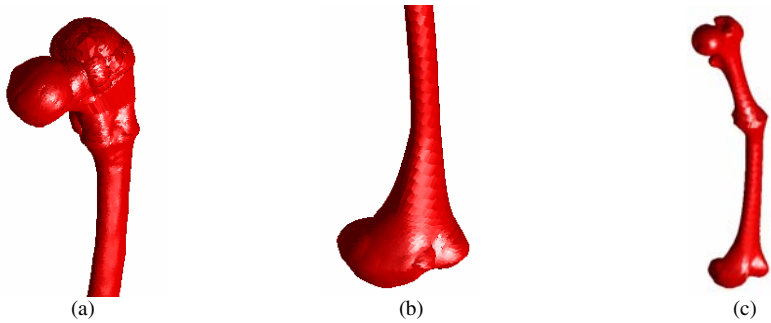


Fig. 3. Several examples of the reconstruction framework being applied. From left to right: (a) Femoral head deformity; (b) Femoral condyles deformity; (c) Femoral shaft deformity.

The results in Table 1 are within the accuracy requirements set by the authors and several other customization studies [11, 12]. During the testing the maximum errors were clustered around the femoral condyles and head (greater and lesser trochanters). This was due to the anatomical variability in these areas.

Apart from the aforementioned areas, errors seen elsewhere are attributed to several sources.

1) The utilization of orthogonal 2D radiographic images for 3D model reconstruction is ill-posed mathematically. Making use of a-priori generic models provide cross-sectional shape constraints. However there are still certain patient-specific cross-sectional variations that cannot be integrated with the use of generic models. These variations cause a majority of the reconstruction errors.

2) Bone segmentation and the defining of bone contours is also a difficult task leading to reconstruction errors. Bone diameters are typically underestimated as femur X-ray edges are blunt due to bone roundness. Femoral diameter (cortical boundary) estimation errors are typically between 0.4 to 1.0 mm [13]. These erroneous contours that are extracted cause inaccuracies when used for shape customization.

4 Conclusions

This paper introduces a bi-planar reconstruction framework which can be used for 3D osteotomy surgical planning. The authors have validated the proposed reconstruction framework, through a series of tests conducted with the aid of CT scan data. The aforementioned results exhibit the proposed method's performance in reconstruction accuracy and adaptability to several deformation types.

References

1. Schatzker, J., Tile, M.: *The Rationale of Operative Fracture Care* (1996)
2. Gamage, P., Xie, S.Q., Delmas, P., Xu, P.: Computer assisted 3D pre-operative planning tool for femur fracture orthopedic surgery. In: Kenneth, H.W., Michael, I.M. (eds.) *SPIE*, vol. 7625, p. 76253D (2010)

3. Paley, D.: Principles of Deformity Correction. Springer, New York (2002)
4. Gunay, M.: Three-dimensional bone geometry reconstruction from X-ray images using hierarchical Free-Form Deformation and non-linear optimization. Vol. PhD dissertation. Carnegie Mellon University, Pittsburgh (2003)
5. Mortensen, E.N., Barrett, W.A.: Intelligent scissors for image composition. In: Proceedings of Computer Graphics (SIGGRAPH 1995), pp. 191–198 (1995)
6. Dijkstra, E.W.: A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik* 1, 269–270 (1959)
7. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition Using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 509–522 (2002)
8. Kuhn, H.W.: The Hungarian Method for the assignment problem. *Naval Research Logistic Quarterly* 2, 83–97 (1955)
9. Sederberg, T.W., Parry, S.R.: Free-Form Deformation of Solid Geometric Models. *SIGGRAPH* 20 (1986)
10. Hsu, W.H., Hughes, J.F., Kaufman, H.: Direct Manipulation of Free-Form Deformations. *SIGGRAPH* 26 (1992)
11. Schmutz, B., Reynolds, K.J., Slavotinek, J.P.: Customization of a generic 3D model of the distal femur using diagnostic radiographs. *Journal of Medical Engineering & Technology* 32, 156–161 (2008)
12. Zheng, G., Nolte, L.P.: Surface Reconstruction of Bone from X-ray Images and Point Distribution Model Incorporating a Novel Method for 2D-3D Correspondence. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2006)
13. Zanetti, E.M., Crupi, V., Bignardi, C., Calderale, P.M.: Radiograph-based femur morphing method. *Medical and Biological Engineering and Computing* 43, 181–188 (2005)

Comparative Analysis of Quasi-Conformal Deformations in Shape Space

Vahid Taimouri¹, Huiguang He², and Jing Hua¹

¹ Computer Science, Wayne State University, Detroit, MI 48202, USA

² Institute of Automation, CAS, Beijing, 100190, China

jinghua@wayne.edu

Abstract. A novel approach based on the shape space concept is proposed to classify quasi-conformal deformations of 3D models. A new metric on the quotient space of meshes is introduced to capture changes of the curvature at each vertex of a simplicial complex during deformation. Then, the deformation curve is obtained by calculating the geodesic curve connecting two shapes in the shape space manifold. In order to compare the deformations, the deformation curves are first transferred to the same part of shape space. And then, the Multi-Dimensional Scaling method is employed to eliminate the redundant dimensions facilitating easy comparison of the deformations. To evaluate our method, some synthetic datasets and 23 datasets of gated images of the left heart ventricle during one heartbeat have been examined. Our experiments show that the algorithm can effectively classify normal and abnormal left heart ventricle deformations in shape space.

Keywords: Deformation classification, Shape space, Riemannian metric, Quasi-conformal deformation.

1 Introduction

Appropriate deformable shape analysis techniques are of utmost importance for time-varying shape comparison and classification. From a geometric point of view, in order to support efficient shape characterization, a higher level of shape abstraction and information reduction is necessary. The traditional geometry-based approaches towards the shape classification are mainly based on the techniques in which global or local geometrical features are extracted from shapes, and then, a metric is constructed based on bending energy or difference in curvature. Finally, shapes are classified by minimizing the energy functional.

The modern geometry introduces shape space, where coordinates of points in this space represent some generalized properties related to various geometrical properties. In the other words, a shape space is established such that each surface group relates to the same point in shape space, where each deformation sequence is shown by a curve. The length of the curve joining two points determines the similarity between two shapes according to the property which the shape space preserves. The desired characteristics to which shape space is invariant can be induced by choosing an appropriate geometric structure.

Invariant to the conformal transformations, a conformal structure can be computed based on the period matrix for manifolds with arbitrary topologies [1]. To induce conformal mappings in 2D and quasi-conformal mappings in 3D, the Green Coordinates [2] are used in the cage-based space deformation estimation. Continuous Ricci flow [3] conformally deforms a Riemannian metric on a smooth surface such that the Gaussian curvature evolves like a heat diffusion process. Eventually, the Gaussian curvature becomes constant and the limiting Riemannian metric is conformal to the original one. In discrete case, the discrete Ricci flow [5] conformably deforms the circle packing metrics [4] with respect to the Gaussian curvatures. In [6], the geodesic lengths of homotopy classes, measured by Hyperbolic Uniformization metric, are used to determine coordinates of each conformal class in the Teichmüller shape space to classify shapes with negative Euler number. The curvature at each point of a manifold changes according to the deformation characteristics during the deformation.

In contrast, our work is based on the geometric structure presented in [7], which computes all smooth groups of diffeomorphisms mapping two objects together. A quasi-conformal metric based on changes of the curvature at each vertex and an algorithm to classify deformations based on the quasi-conformal metric are proposed.

2 Deformation Analysis in Shape Space

Shape Space: Let \mathbf{G} be the space of all immersions with the same connectivity. The deformation of the immersion \mathbf{M} constitutes a curve in this space. Therefore, the tangent vector of X , which belongs to the tangent plane of \mathbf{G} at point \mathbf{M} , assigns a vector X_p to each point p on \mathbf{M} . Given a smooth deformation of all points on \mathbf{M} , the vector field $X(t)$ is $\partial p(t)/\partial t$ for all points p . Depending on the deformation characteristics in which we are interested, various shape spaces having different intrinsic geometries may be defined. Our algorithm to classify deformations is as follows,

1. Mesh registration to find corresponding points on different meshes.
2. Determination of deformation paths in shape space.
3. Transfer of deformation paths in shape space, and dimension reduction.
4. Construction of the 3D *spatiotemporal* space.

2.1 Quasi-Conformal Deformation Estimation

In the Riemannian geometry, a metric is defined as the inner product of two vector fields. Suppose two different deformation fields as X and Y , the distance between them are measured by $\langle\langle X, Y \rangle\rangle$ in shape space. Some geometric structures have been established to provide the Isometric and Rigid metrics based on the Riemannian metric definition [7]. Here, we propose a quasi-conformal metric within this structure. Let Σ be a simplicial complex, and a mapping $f : \Sigma \rightarrow \mathbb{R}^3$ embed Σ to the Euclidean space, then $\mathbf{M} = (\Sigma, f)$ is a triangular mesh. Also, let $\{X_p, X_q, X_r\}$ be corresponding deformations of three vertices of the face Δpqr on \mathbf{M} .

2.2 Quasi-Conformal Metric

Let Δpqr and $\Delta \acute{p}\acute{q}\acute{r}$ be two triangles such that $\acute{p}(t) = p(t + \Delta t) = p(t) + X_p(t)$, $\acute{q}(t) = q(t + \Delta t) = q(t) + X_q(t)$ and $\acute{r}(t) = r(t + \Delta t) = r(t) + X_r(t)$. In the Euclidean geometry, if the edge lengths of a triangle are multiplied by the same coefficient, the resulting triangle has the same angles as the original one; therefore,

$$\frac{\|\acute{p} - \acute{q}\|^2}{\|p - q\|^2} = \frac{\|\acute{p} - \acute{r}\|^2}{\|p - r\|^2}. \tag{1}$$

Eq 1 yields the following relation,

$$\frac{\|\acute{p}(t) - \acute{q}(t)\|^2 - \|p(t) - q(t)\|^2}{\|p(t) - q(t)\|^2} = \frac{\|\acute{p}(t) - \acute{r}(t)\|^2 - \|p(t) - r(t)\|^2}{\|p(t) - r(t)\|^2}. \tag{2}$$

The numerator shows the changes of the squared edge length within Δt . If both sides of the equal sign are divided by Δt , and also $\Delta t \rightarrow 0$; then, the numerator would be equal to the derivative of the squared edge length. By differentiating $\|p(t) - r(t)\|^2$, Eq 2 will be derived,

$$\frac{\langle p(t) - q(t), X_p(t) - X_q(t) \rangle}{\|p(t) - q(t)\|^2} = \frac{\langle p(t) - r(t), X_p(t) - X_r(t) \rangle}{\|p(t) - r(t)\|^2}, \tag{3}$$

from which $T_p(X)$ is achieved as follows,

$$T_p(X) = \langle p - q, X_p - X_q \rangle \cdot \|p - r\|^2 - \langle p - r, X_p - X_r \rangle \cdot \|p - q\|^2. \tag{4}$$

Likewise, $T_q(X)$ and $T_r(X)$ are calculated. To preserve angles of each triangle Δpqr , a similarity transformation requires that all $T_p(X)$, $T_q(X)$ and $T_r(X)$ vanish during deformation. Thus a *conformal* metric is obtained as follows,

$$T_{\Delta pqr}(X) = |T_p(X)| + |T_q(X)| + |T_r(X)|, \tag{5}$$

$$\langle \langle X, Y \rangle \rangle_M^C = \sum T_{\Delta pqr}(X) \cdot T_{\Delta pqr}(Y). \tag{6}$$

Although this metric preserves characteristics of conformal deformations, the shape space geometry of these conformal deformations cannot be reconciled to the characteristics of some non-conformal deformations. For instance, while the heart beats; one part of the heart expands while other portions contract, meaning the heart deformation is not conformal in the strict sense. This necessitates a quasi-conformal shape space which better preserves these non-conformal deformation characteristics. Let ΔK_p be the changes of curvature at the vertex p during deformation, thus $T_p(X)$ is redefined as follows,

$$T_p(X) = \Delta K_p \cdot \left[\langle p - q, X_p - X_q \rangle \cdot \|p - r\|^2 - \langle p - r, X_p - X_r \rangle \cdot \|p - q\|^2 \right], \tag{7}$$

and by substituting the new $T_p(X)$ from Eq 7 in Eq 5 and Eq 6 the final quasi-conformal metric is achieved. In this quasi-conformal metric, changes in the curvature at each vertex are considered such that each vertex with lower curvature changes has less effect in the metric. In order to find the geodesic on the shape space manifold, the energy functional $E(\mathbf{M}) = \int \langle \langle X, X \rangle \rangle_M dt$ of the curve connecting each pair of meshes is minimized.

2.3 Deformation Transfer and Dimension Reduction

Two deformations are comparable, if their deformation curves are transferred to the same part of shape space. A method based on the parallel transport approach is used for the deformation transfer. Let \mathbf{M} and \mathbf{N} be two immersions, and X and Y be their corresponding deformations, respectively. We sample each deformation path at equidistant points \mathbf{M}_i and \mathbf{N}_i , thus the line segments connecting each pair of sequential samples will be $X_i = \Delta\mathbf{M}_i$ and $Y_i = \Delta\mathbf{N}_i$, respectively. Assume the geodesic line γ_i connects each pair of meshes \mathbf{M}_i and \mathbf{N}_i in the shape space manifold. The idea is to transfer the deformation X_i to Y_i along the geodesic curve γ_i connecting two corresponding immersions.

First, the vector X_0 is transferred along the geodesic line γ_0 using the parallel transport method to get \dot{Y}_0 and $\dot{\mathbf{N}}_1 = \mathbf{N}_0 + \dot{Y}_0$; then, the geodesic line γ_1 , connecting $\dot{\mathbf{N}}_1$ and \mathbf{M}_1 is achieved and X_1 is transferred along it to get \dot{Y}_1 and $\dot{\mathbf{N}}_2 = \dot{\mathbf{N}}_1 + \dot{Y}_1$. Likewise, all the segments are transferred so that the final transferred deformation \dot{Y} is achieved [7].

At this point, all deformation paths have been transferred to the same part in shape space. However, high dimensional shape space prohibits an intuitive perception of deformation comparison. As a result, redundant dimensions of shape space should be eliminated such that the projected deformation paths are as similar as possible to the original deformation paths. To accomplish this, a distance matrix demonstrating distance of each pair of meshes is built based on the defined metric. Then, the Multi-Dimensional Scaling (MDS) method is used to reduce dimensions of shape space and project deformation paths onto a 2D space, called the 2D *spatial* space.

2.4 Deformation Classification

Although deformations take place in time, we do not have any consideration about time in this deformation structure. To address this issue, time axis is added to the 2D spatial space, as the third dimension. We call this new 3D shape space, *spatiotemporal* space. Finally, the Locally Linear Embedding (LLE) method is used to mapped 3D deformation curves onto a 2D plane such that each point in this 2D plane corresponds to one deformation curve in the spatiotemporal space. The more similar two deformations, the closer their corresponding points are in the 2D plane. By adding time axis as the third dimension to the 2D spatial space rather than to the original shape space, the time dimension will have more influence on the final deformation projection process as a naturally different dimension.

3 Experiments on Synthetic Datasets

In order to evaluate performance of our metric, some objects are deformed in accordance with different deformation patterns; then, the deformations are transferred to other objects. The more accurately the non-conformal deformations are followed by the second object without distortion, the better the metric can

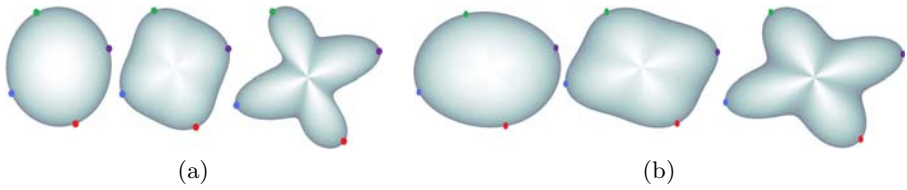


Fig. 1. a) The non-conformal deformation of a sphere within three time points from left to right. b) The ellipsoid follows the deformation of the sphere.

estimate the intrinsic geometry of shape space and preserve the invariant characteristics of the objects during deformation. Fig. 1a demonstrates a non-conformal deformation on a sphere, which is followed by an ellipsoid (Fig. 1b).

Although the shape deforms non-conformally, the ellipsoid can accurately follow the deformation, and the quasi-conformal metric can circumvent the local minima during optimization of the energy functional by learning the geometry of shape space. We applied the deformation transfer method based on our metric on 60 deformations, all of which were successfully followed by different objects.

In the next step, some different deformations are applied on a torus and classified by the classification algorithm. Since we classify the synthetic datasets, the corresponding points on different meshes are known. Therefore, the mesh registration step can be ignored. Fig. 2 illustrates classification of different deformations of a torus, in which only the representative deformations are shown next to solid dots, and the deformations corresponding to the soft dots are not depicted due to the lack of space. The similar deformations will be close together in the 2D plane.

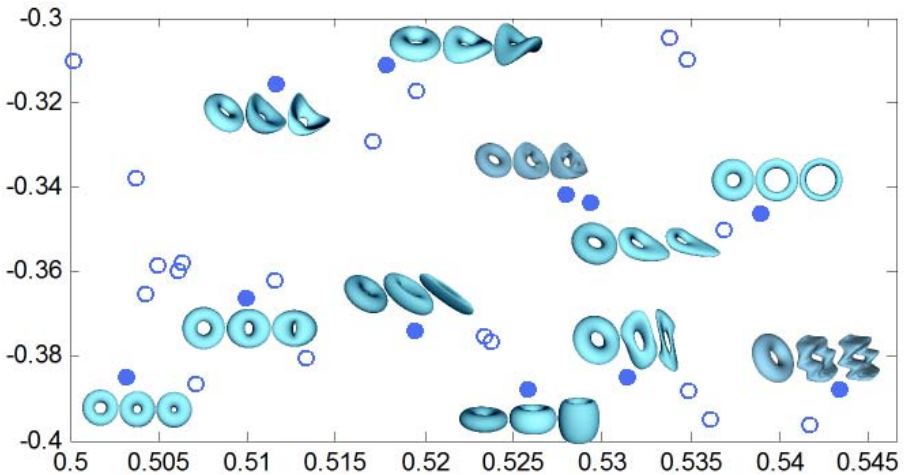


Fig. 2. The classification of different deformations applied on a torus such that each dot corresponds to one deformation (Two axes are first and second coordinates of LLE)

Due to the acquisition procedures, datasets might contain some noise, thus one aspect of classification is about how accurately a classifier can classify noisy datasets. To evaluate this aspect, the points corresponding to different deformations in the 2D plane are classified into some classes by the k -means clustering algorithm. Let \mathbf{K} be the set of all points corresponding to all deformations and \mathbf{k}_i be the set of points classified in the group i^{th} by the k -means method. In addition, let k_i belong to the set \mathbf{k}_i , and l_i be its assigned label. Eq.8 calculates the statistic τ quantifying the performance of the algorithm as follows,

$$\tau = \frac{\sum_i |\{k_i \in \mathbf{k}_i \mid l_i = i\}|}{|\mathbf{K}|}. \quad (8)$$

Table.1 shows the algorithm performance against additive noises with different distributions, which were added to the synthetic genus zero or one meshes with zero, one or two boundaries. As can be seen, the percentage of misclassified deformations is acceptable up to 10% of additive noise, meaning the pattern of deformations of topologically different meshes can be detected well. Note that, the algorithm shows better performance against the uniform noise compared with the noise with either Gaussian or Poisson distribution.

Table 1. Performance of the algorithm against noisy datasets with different percentages of additive noise variance

Noise Variance	Uniform Noise	Gaussian Noise	Poisson Noise
1%	99.6%	99.4%	99.2%
2.5%	99.3%	97.3%	97.2%
5%	96.0%	94.4%	95.1%
10%	90.1%	88.8%	89.0%

4 Application on Heart Motion Analysis

The proposed algorithm was applied on gated images of the left heart ventricle during one heartbeat. A total of 23 datasets with resolution of $128 \times 128 \times 47mm^3$ were acquired from some normal and abnormal subjects. In a preprocessing step, all the datasets are normalized within a bounding box. Subsequently, the inter-subject and intra-subject registrations are preformed to find the corresponding points among different meshes. To accomplish this, a total of 16 landmarks are chosen on meshes manually (Fig.3), and then the constrained thin plate spline method is used to perform the registration [8,9].

Although the noise distribution of the gated images is approximately Poisson, it might be estimated as Gaussian as well, depending on the number of channels used during the image acquisition. Either of the cases, the proposed method resists against the additive noises well. As can be seen in Fig.3.a, the area connecting the left ventricle to the right one is cut to make a *boundary*, which is used to set the landmarks. Then, 23 normal and abnormal registered datasets, 9 out

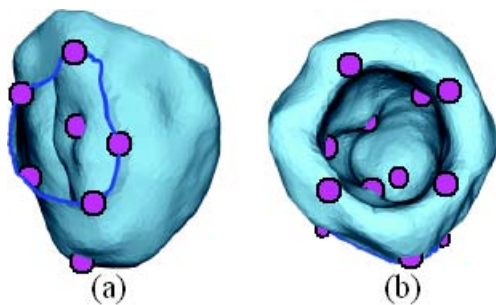


Fig. 3. The landmarks are chosen, a) on the boundary, b) inside the ventricle

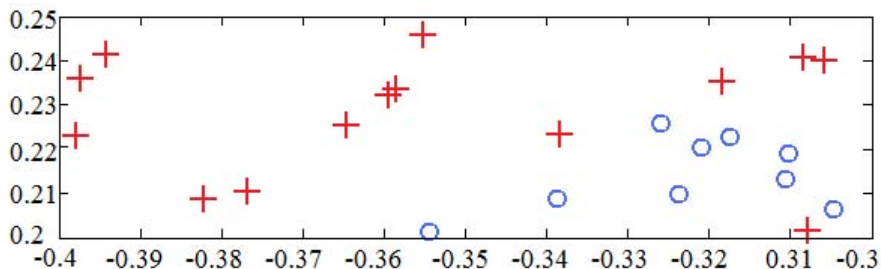


Fig. 4. The classification of normal (*circles*) and abnormal (*crosses*) subjects, in which the two axes are the first two coordinates of LLE

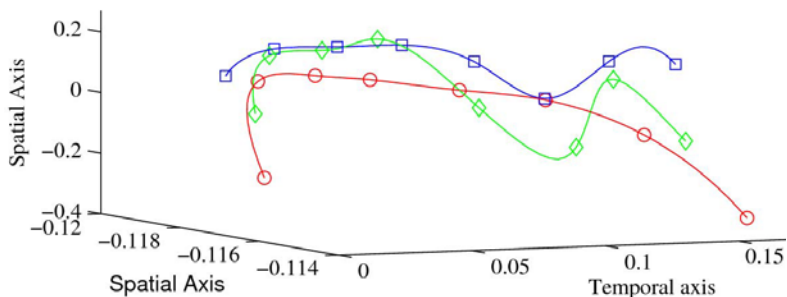


Fig. 5. Two normal subjects (*squares and diamonds*) and one abnormal subject (*circles*) in the 3D spatiotemporal space

of which are normal, are classified by the algorithm. Shown by circles in Fig. 4, the normal datasets are almost located in the same part of the 2D plane. Fig. 5 demonstrates the normal and abnormal deformations in the 3D spatiotemporal space, in which the normal deformations follow the similar pattern, which makes their corresponding points in the 2D plane close to each other. In fact, adding the extra temporal dimension is essential, as it takes into account the speed of the deformation which is of utmost importance in heart diagnosis.

5 Conclusion

In this paper, we have presented an algorithm to effectively classify and compare deformations based on a novel quasi-conformal metric in shape space. To make deformation paths intuitively comparable in shape space, the deformations are transferred and projected into a 2D plane, where each point corresponds to one deformation. The method has been successfully applied to heart data analysis on differentiating normal and abnormal heart motion.

Acknowledgements

The authors gratefully acknowledge the support of K. C. Wong Education Foundation, Hong Kong, and the National Science Foundation under the grants NSF IIS-0915933, NSF IIS-0937586 and NSF IIS-0713315.

References

1. Gu, X., Wang, Y., Yau, S.T.: Computing Conformal Invariants: Period Matrices. *Communications in Information and Systems* 3, 153–197 (2004)
2. Lipman, Y., Levin, D., Cohen, D.: Green Coordinates. In: *ACM SIGGRAPH*, vol. 27 (2008)
3. Hamilton, R.S.: Three Manifolds with Positive Ricci Curvature. *J. Differential Geometry* 17, 255–306 (1982)
4. Collins, C., Stephenson, K.: A Circle Packing Algorithm. *Computational Geometry: Theory and Applications* 25, 233–256 (2003)
5. Jin, M., Kim, J., Luo, F., Gu, X.: Discrete Surface Ricci Flow. *IEEE Trans. on Visualization and Computer Graphics (TVCG)* 14, 1030–1043 (2008)
6. Jin, M., Zeng, W., Luo, F., Gu, X.: Computing Teichmüller Shape Space. *IEEE Trans. on Visualization and Computer Graphics (TVCG)* 15, 504–517 (2009)
7. Kilian, M., Mitra, N.J., Pottmann, H.: Geometric Modeling in Shape Space. In: *ACM SIGGRAPH*, vol. 26 (2007)
8. Zou, G., Hua, J., Muzik, O.: Non-rigid Surface Registration Using Spherical Thin-plate Splines. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 367–374. Springer, Heidelberg (2007)
9. Duan, Y., Hua, J., Qin, H.: Interactive Shape Modeling Using Lagrangian Surface Flow. *The Visual Compute.* 21(5), 279–288 (2005)

Establishing Spatial Correspondence between the Inner Colon Surfaces from Prone and Supine CT Colonography

Holger Roth¹, Jamie McClelland¹, Marc Modat¹, Darren Boone²,
Mingxing Hu¹, Sebastien Ourselin¹, Greg Slabaugh³,
Steve Halligan², and David Hawkes¹

¹ Centre for Medical Image Computing, University College London, UK

² Department of Specialist Radiology, University College Hospital, London, UK

³ Medicsight PLC, London, UK

Abstract. Colonography is an important screening tool for colorectal lesions. This paper presents a method for establishing spatial correspondence between prone and supine inner colon surfaces reconstructed from CT colonography. The method is able to account for the large deformations and torsions of the colon occurring through movement between the two positions. Therefore, we parameterised the two surfaces in order to provide a 2D indexing system over the full length of the colon using the Ricci flow method. This provides the input to a non-rigid B-spline registration in 2D space which establishes a correspondence for each surface point of the colon in prone and supine position. The method was validated on twelve clinical cases and demonstrated promising registration results over the majority of the colon surface.

1 Introduction

Colorectal cancer is one of the main types of cancer, leading to more than 630,000 deaths worldwide each year [1]. Traditional colonoscopy using a video endoscope can produce miss rates of up to 27 % for adenomas smaller or equal to 5 mm [2]. Furthermore, it can cause significant discomfort and is not without risk of perforating the colon. These drawbacks have led to the development of alternative screening methods such as computed tomography colonography, which is becoming established as a standard screening tool for the detection and diagnosis of colorectal lesions in the USA, Europe and Japan. The bowel is cleansed before the procedure by administering a powerful laxative and inflated. CT images are taken in both prone and supine positions, resulting in a large deformation of the colon. Remaining faecal material and fluids can be tagged with contrast agent and removed digitally. However, faecal remnants or folds of the colonic wall can still mimic the appearances of polyps, leading to false positives in the diagnosis. Routinely, the radiologist establishes the spatial correspondence between the two views by eye which is a difficult task for even the most experienced radiologist and can thus introduce delays and errors. An automatic method of establishing correspondence between the colon surfaces visualised in prone and supine

CT images would have the potential to save time and improve accuracy during the diagnostic process. Furthermore, it could improve computer aided detection (CAD) algorithms.

Registration of prone and supine CT images of the colon has been attempted by aligning extracted centrelines of both views [3] as an index of location. However, this method essentially provides only one degree of freedom of location within the colonic surface and does not account for torsion. Another approach is to define several anatomical landmarks [4], like anus, cecum and flexures, in order to align both colons. However, flexures are difficult to locate accurately and the identification of only a small number of points is insufficient to describe the complex folding and deformation of the colon between prone and supine positions. A voxel-based method was proposed by Suh et al. [5], which also uses the centrelines to generate an initial deformation field and then treats the registration task as an optical flow process. However, it is clear from our own studies that the changes in shape and location of anatomy between prone and supine CT colonography mean that conventional image intensity-based non-rigid registration algorithms are not sufficiently robust or accurate. Lamy et al. [6] use the teniae coli as an additional feature producing a deformation field with a rotational component. However, teniae coli are difficult to extract over the full length of the colon and cannot describe the deformation of the whole colonic surface.

We propose a new method based on a 2D manifold representing the internal colon lumen surface which is able to account for the large deformations and torsion of the colon in both positions. The colon is a tube and the internal surface can be mapped onto a cylinder with two indices describing any surface location. Each location corresponds to a 3D point in the CT scan and can act as an index to a rich set of both surface and volume features. Our method uses these features from the 3D data to drive the registration, but finds a transformation between the 2D manifolds representing the colon surface in the prone and supine scans. Hence, this greatly eases the registration in comparison to full 3D registrations.

To the best of our knowledge nobody else has proposed performing non-rigid registration on the 2D manifold in order to establish a full correspondence between all points on the 3D colonic surface.

2 Methods

A number of groups have proposed methods for unfolding the colon in order to produce 2D images of the inside of the colon. These images are usually produced in order to enable better examination of the colon surface and to aid detection of polyps [7]. Conformal mappings can provide a one-to-one mapping between the 3D surface and the 2D image while minimising the necessary deformation [8,9]. A Surface S_1 in R^3 can be represented using a one-to-one mapping ϕ_1 to D_1 in R^2 . S_2 is mapped to D_2 through ϕ_2 respectively. Hence, the transformation function f between the three-dimensional surfaces S_1 and S_2 can be derived as shown in Fig. 1, where the principle is illustrated. \tilde{f} is the registration $\tilde{f}: D_1 \rightarrow D_2$ between the two flat surfaces.

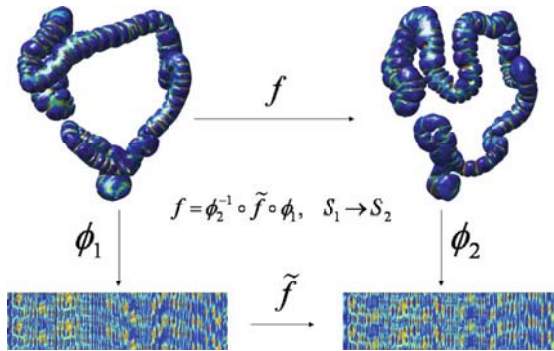


Fig. 1. The principle of surface registration using a 2D manifold, where the colour scale indicates the shape index (see section 2.3) at each coordinate of the surface computed from the 3D inner colonic surfaces S_i

2.1 Colon Unfolding with Discrete Ricci Flow

One recently developed technique for discrete surface parameterisation is based on Ricci. This method deforms a surface proportionally to its local curvature, where the curvature values evolve similarly to a heat diffusion process. It was first introduced by Hamilton [10] for Riemannian geometry. Qiu et al. [9] were the first to use this method to unfold the colon. The Ricci flow is defined as $\frac{du_i(t)}{dt} = \bar{K}_i - K_i$, where K_i is the Gaussian curvature at vertex v_i , \bar{K}_i is the desired Gaussian curvature and u_i is computed from a circle packing metric [8]. It can be shown that the Ricci flow is essentially the gradient flow of an energy function [11] which can be minimized using the gradient descent method. For the purpose of parameterisation of the colon surface, the target curvature should be zero for all vertices.

2.2 Extraction of the Inner Colonic Surface

We extracted triangulated meshes of the inner colonic surfaces S_i using the segmentations of the air inside the prone and supine colons computed by the method described in [12]. We ensured that the segmentations of the large intestine were topologically correct. In order to achieve that, some of the segmentations were manually edited. The segmentation is the input for a marching cubes algorithm with subsequent smoothing and decimation. This results in a closed and simply connected mesh along the air-to-tissue border in the CT-image.

To apply the Ricci flow method to the colon, each original genus-zero surface S_i has to be converted to a genus-one surface SD_i [8]. Therefore, the inner colonic surface S_i (which is topologically equal to a sphere) has to be converted to a torus-like surface. Hence, we remove faces of the surface mesh from cecum and rectum at user-identified positions. The remaining faces are duplicated with inverse orientation and joined at the boundaries of the previously produced holes. The resulting mesh SD_i provides the input for the Ricci flow computation which provides the two-dimensional coordinates of each location within the surface.

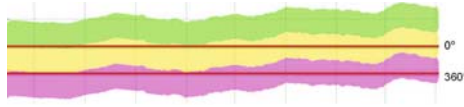


Fig. 2. Sampling the unfolded mesh (where the green and purple bands represent the repeat of the 2D coordinates of the surface) for a further 360 degrees clockwise and anticlockwise. The horizontal lines represent the re-sampled complete colon surface in a form suitable for registration

2.3 Generating the 2D Coordinates of the Surface

The Ricci flow algorithm converges to a planar surface with free border and local Gaussian curvature tending to zero everywhere by iteratively updating the edge lengths of the triangles. The optimisation carries on until the maximum difference between all K_i and \bar{K}_i is close enough to zero in order to produce a suitable parameterisation D_i which can be embedded into planar R^2 space. This is computed in a similar manner to [8] where each planar triangle is computed based on its final edge length. As D_i is not rectangular, we repeat the planar mesh so that a rectangular raster-image I_i fully samples all points around the colon (as illustrated in Fig. 2). The top (0°) and bottom (360°) edges of the image correspond to the same point on the colon surface, thus representing the inner colonic surface as a cylinder. The horizontal axis (x) corresponds to position along the colon from cecum to rectum and the vertical axis (y) corresponds to rotation around the circumference of the colon. Each pixel of I_i has an interpolated value of the corresponding shape index SI from the three-dimensional surface S_i . SI is defined as

$$SI \equiv \frac{1}{2} - \frac{1}{\pi} \arctan \left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2} \right), \quad (1)$$

where κ_1 and κ_2 are the principal curvatures extracted from S_i . Furthermore, any other measure of local curvature or other voxel grey value statistics derived from the original CT scan could be associated with a pixel in I_i . The interpolated value is always computed based on the three corner values of each triangle in D_i which correspond to the 3D vertices of S_i . The resulting supine and prone raster-images with a resolution of $n_x = 3000$ and $n_y = 150$ are shown in Fig. 3 (top) and Fig. 3 (middle).

2.4 Establishing Spatial Correspondence between Prone and Supine

The 2D manifolds are used to generate shape index images which are first aligned in the y -direction to account for differences in the 0° position arbitrarily assigned by the planar embedding. This is performed automatically by applying a circular shift in y -direction to I_1 that minimises the Sum of Squared Differences (SSD) between I_1 and I_2 . The B-spline registration is currently performed on the flat Euclidean plane and not in a cylindrical framework. Therefore I_1 and I_2 are

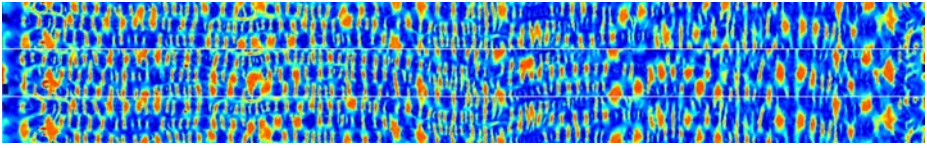


Fig. 3. Supine (top), prone (middle) and deformed supine (bottom) repeated raster-images where each pixel has the value of the corresponding shape index from the inner colonic surface

repeated in the y -direction, resulting in images with a resolution of $N_x \times N_y = n_x \times 2 \cdot n_y$, in order to simulate the cylindrical images during the registration.

A 2D B-spline registration is then performed with the shifted I_1 as target and I_2 as source. We use the fast implementation provided by [13]. The registration is performed in two stages, the first to recover the larger deformations, and the second to recover the finer deformations. The first stage consists of five resolution levels. The second stage consists of three resolution levels and uses the result from the first stage as the starting transformation for the coarsest level. Both the image and B-spline control point grid resolutions are doubled at each level. The final resolution level uses images with 3000×300 ($n_x \times 2 \cdot n_y$) pixels and control points spaced every 12.5 pixels in both directions. SSD is used as the similarity measure. The gradient of the cost function is smoothed at each iteration using a Gaussian kernel with a standard deviation of three control points for the first stage and one control point for the second. No additional penalty term is used for the first stage but bending energy is used for the second. Gaussian smoothing of the 2D images is applied at each resolution level during the first stage of registration but is not used for the second.

The central half of the B-spline registration result, from 25% to 75% of the y -axis, covers the whole inner colon surface (0° to 360°) and should have a similar displacement at $y = 0.25 \cdot N_y$ and $y = 0.75 \cdot N_y$ thanks to the duplication of image data in the y -direction prior to registration. To force the result to be fully cylindrical, such that the transformation is continuous from $y = 0.25 \cdot N_y$ to $y = 0.75 \cdot N_y$, the control points displacements at $y = 0.25 \cdot N_y$ and $y = 0.75 \cdot N_y$ are averaged together, and the control point displacements before $y = 0.25 \cdot N_y$ and after $y = 0.75 \cdot N_y$ are replaced with the corresponding control point displacements from the central section. This results in a continuous transformation around the entire inner colon surface and allows the mapping between S_1 and S_2 . From this mapping it is straightforward to determine the full 3D mapping f , as shown in Fig. 1.

3 Clinical Evaluation

Ethical permission and informed consent to utilise anonymised CT colonography datasets was obtained. Colonic cleansing and insufflations were undertaken for all

twelve subjects used in this study in accordance with current recommendations [14]. We selected patients where the colon was fully distended in the prone and supine scan with either fluid tagging or little remaining fluid in order to provide complete segmentations. The radiologist (with experience in over 500 validated colonography studies) matched pairs of reference points from the original prone and supine CT slices. Using separate multiplanar reformats, a combination of polyps, normal anatomical structures and diverticula were identified from multiple colonic segments, resulting in approximately ten pairs of coordinates per patient.

3.1 Human Observer Error Estimation

The error of establishing the correspondence points in both images was determined by repeating the validation following an interval of seven days for patients 1 to 6 (in order to reduce recall bias). The repeated validation was performed using the original coordinates from the supine datasets, and the radiologist was asked to re-identify the corresponding points in the prone dataset. The results suggest a significant difficulty in finding correct correspondences in the prone and supine CT-images, which provides further justification for a method of establishing prone-supine correspondence. We removed outliers based on the maximum likelihood estimate [15, p. 202] using the median $\sigma = 1.4828(1+5/(n-3))\text{med}|E|$, where the inliers are defined by $E_i \leq \gamma$ and $\gamma = 1.96\sigma$. This gives a threshold γ of 14.6 mm in order to get reliable landmarks to validate the registration and reduces the human observer error from (8.2 ± 12.5) mm to (3.8 ± 2.9) mm. The remaining prone landmarks are averaged and used to validate the registration for the first six patients.

3.2 Registration Error

In order to determine the registration error we find the closest surface points on S_1 to the landmarks in the prone image. We then find the closest points on S_2 to the corresponding landmarks in the supine image, transform them using the 3D mapping f , and calculate the distance to the points on S_1 .

In order to access the accuracy of the method, we had to remove 11 reference points which were outliers (determined for the first six patients) and 3 points which clearly failed to align correctly. Just using the direct mapping between D_1 and D_2 after initial alignment in the y -direction resulted in an error for the remaining 120 data points of (23 ± 15) mm. After establishing the spatial correspondence using the B-spline registration, the error was reduced to (9.0 ± 11) mm. The histogram of the registration error shows that the majority (91 %) lie below 20 mm. Successful registration could be achieved for 8 out of 12 polyps (distributed over 8 patients) with an accuracy of (4.7 ± 4.1) mm. These are promising results and have potential clinical impact to the screening process.

Four patients have clearly higher errors which correspond to the data points larger than 20 mm in Fig. 4. It appears that some of these errors may be due to some of the corresponding landmarks being assigned wrongly, e.g. to a neighboring fold. The excluded reference points were one haustral fold and the two

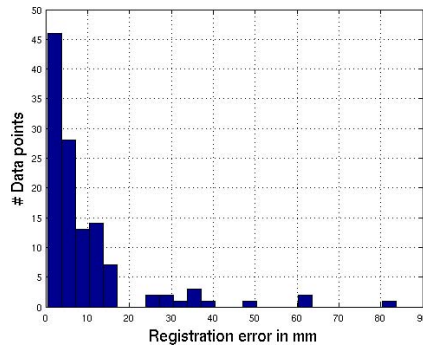


Fig. 4. Histogram of the registration error in mm

unsuccessfully registered polyps which lay in the rectum or cecum and had a large error due to the deformations in the parameterisation in that region.

4 Summary and Future Work

We propose a novel registration method for establishing spatial correspondence for the inner colon surface extracted from prone and supine CT colon images. Our registration process works by using an appropriate indexing system which reduces the registration task from a 3D- to a 2D-problem. This is done with a one-to-one conformal mapping of the entire inner colon surfaces to a parameterisation where one dimension corresponds to distance along the colon and the other to its angular orientation. From that parameterisation we compute images which correspond to 3D positions and their shape indices computed on 3D surfaces. This allows a non-rigid registration of the prone and supine colon surfaces which can handle the large deformations occurring between both positions. Furthermore, the provided framework could easily be extended to include any statistic or set of statistics derived from the original CT-images.

Future work will include phantom studies and other synthetic data for a more reliable gold standard to be used in the validation of the method. The 2D B-spline registration will be implemented in a truly cylindrical way such that repeating the images will not be necessary. We will experiment with additional statistics derived from the original CT volumes with the aim of increasing the accuracy of the method. Furthermore, we will investigate automated topological correction of the colon segmentation. This is in order to increase robustness and deal with insufficiently distended colons where the colon surface could be collapsed.

It is clear that a major limitation of this study is the lack of a sufficiently accurate gold standard. The reported accuracy of our results is almost certainly limited by observer error in picking corresponding landmarks in prone and supine CT. To address this we are building a physical phantom from porcine bowel with embedded CT visible markers to provide an accurate gold standard.

References

1. WHO: Cancer (2006), <http://www.who.int/mediacentre/factsheets/fs297/en/> (retrieved on July 2010)
2. Rex, D., Cutler, C., Lemmel, G., Rahmani, E., Clark, D., Helper, D., Lehman, G., Mark, D.: Colonoscopic miss rates of adenomas determined by back-to-back colonoscopies. *Gastroenterology* 112(1), 24–28 (1997)
3. Acar, B., Napel, S., Paik, D., Li, P., Yee, J., Beaulieu, C., Jeffrey, R.: Registration of supine and prone ct colonography data: Method and evaluation. *Radiology* 221, 332 (2001)
4. Näppi, J., Okamura, A., Frimmel, H., Dachman, A., Yoshida, H.: Region-based supine-prone correspondence for reduction of false positive cad poly candidates in ct colonography. *Academic Radiology* 12, 695–707 (2005)
5. Suh, J., Wyatt, C.: Deformable Registration of Supine and Prone Colons for Computed Tomographic Colonography. *Journal of computer assisted tomography* 33(6), 902 (2009)
6. Lamy, J., Summers, R.: Intra-patient colon surface registration based on teniæ coli. In: *SPIE Medical Imaging*, pp. 116–123 (2007)
7. Johnson K., Johnson C., F.J.M.R.S.R.: CT colonography using 360-degree virtual dissection: a feasibility study. *AJR Am J Roentgenol.* **186** (2006) 90–95
8. Jin, M., Kim, J., Luo, F., Gu, X.: Discrete surface ricci flow. *IEEE Transactions on Visualization and Computer Graphics* 14(5), 1030–1043 (2008)
9. Qiu, F., Fan, Z., Yin, X., Kaufman, A., Gu, X.: Colon Flattening with Discrete Ricci Flow. In: *MICCAI workshop*, pp. 97–102 (2008)
10. Hamilton, R.: Three-manifolds with positive Ricci curvature. *J. Differential Geom.* 17(2), 255–306 (1982)
11. Chow, B., Luo, F.: Combinatorial Ricci flows on surfaces. *J. Differential Geom.* 63(1), 97–129 (2003)
12. Slabaugh, G., Yang, X., Ye, X., Boyes, R., Beddoe, G.: A Robust and Fast System for CTC Computer-Aided Detection of Colorectal Lesions. *Algorithms* 3(1), 21–43 (2010)
13. Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S.: Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine* (2009)
14. Taylor, S., Laghi, A., Lefere, P., Halligan, S., Stoker, J.: European Society of Gastrointestinal and Abdominal Radiology (ESGAR): consensus statement on CT colonography. *European Radiology* 17(2), 575–579 (2007)
15. Rousseeuw, P., Leroy, A.: Robust regression and outlier detection. Wiley-IEEE (2003)

Heat Kernel Smoothing Using Laplace-Beltrami Eigenfunctions

Seongho Seo¹, Moo K. Chung^{1,2,3}, and Hourii K. Vorperian⁴

¹ Department of Brain and Cognitive Sciences
Seoul National University, Korea

² Department of Biostatistics and Medical Informatics

³ Waisman Laboratory for Brain Imaging and Behavior

⁴ Vocal Tract Development Laboratory, Waisman Center

University of Wisconsin, Madison, WI 53706, USA

`mkchung@wisc.edu`

Abstract. We present a novel surface smoothing framework using the Laplace-Beltrami eigenfunctions. The Green's function of an isotropic diffusion equation on a manifold is constructed as a linear combination of the Laplace-Beltrami operator. The Green's function is then used in constructing heat kernel smoothing. Unlike many previous approaches, diffusion is analytically represented as a series expansion avoiding numerical instability and inaccuracy issues. This proposed framework is illustrated with mandible surfaces, and is compared to a widely used iterative kernel smoothing technique in computational anatomy. The MATLAB source code is freely available at <http://brainimaging.waisman.wisc.edu/~chung/lb>.

1 Introduction

In medical image processing and analysis, anatomical manifolds are commonly represented as triangular meshes. Procedures such as image segmentation, mesh construction and geometric computations on the meshes introduce geometric noise to the mesh coordinates. Therefore, it is imperative to reduce the mesh noise while preserving the geometric details of the object for various applications.

Many approaches have been proposed for smoothing surface data and coordinates. Probably the most widely used method is to numerically solve diffusion equations on anatomical surfaces [1,8]. However, these approaches use discretization schemes which tend to suffer numerical instability and inaccuracy. Iterated kernel smoothing is also a widely used method in computational anatomy [3,6] and computer vision [13]. Kernel weights are spatially adapted to follow the shape of the heat kernel in a discrete fashion along a manifold. In the tangent space, the heat kernel is approximated linearly using the Gaussian kernel. However, this process compounds the linearization error at each iteration.

In this paper, we propose to construct the heat kernel analytically using the eigenfunctions of the Laplace-Beltrami operator, avoiding the need for the linear approximation used in [3,6]. Although solving for the eigenfunctions requires

the finite element method, the proposed method is analytic in a sense that heat kernel smoothing is formulated as a series expansion explicitly. Thus, the proposed method avoids the numerical instability associated with solving the diffusion equations numerically [18]. Although there are many papers on solving diffusion equations on arbitrary triangular meshes [8,14], this is the first paper that explicitly and correctly constructed heat kernel for an arbitrary surface and solved heat diffusion using the eigenfunctions of the Laplace-Beltrami operator.

2 Heat Kernel Smoothing

Consider a real-valued C^2 measure Y defined on a manifold $\mathcal{M} \subset \mathbb{R}^3$. We assume the following additive model on Y :

$$Y(p) = \theta(p) + \epsilon(p), \tag{1}$$

where $\theta(p)$ is the unknown mean signal and $\epsilon(p)$ is a zero-mean Gaussian random field. We may assume further $Y \in L^2(\mathcal{M})$, the space of square integrable functions on \mathcal{M} with the inner product $\langle f, g \rangle = \int_{\mathcal{M}} f(p)g(p)d\mu(p)$, where μ is the Lebesgue measure such that $\mu(\mathcal{M})$ is the volume of \mathcal{M} . Solving

$$\Delta\psi_j = -\lambda\psi_j, \tag{2}$$

for the Laplace-Beltrami operator Δ on \mathcal{M} , we order eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$ and corresponding eigenfunctions $\psi_0, \psi_1, \psi_2, \dots$. Then, the eigenfunctions ψ_j form an orthonormal basis in $L^2(\mathcal{M})$ [10].

Using the eigenfunctions, *heat kernel* $K_\sigma(p, q)$ is analytically written as

$$K_\sigma(p, q) = \sum_{j=0}^{\infty} e^{-\lambda_j\sigma} \psi_j(p)\psi_j(q), \tag{3}$$

where σ is the bandwidth of the kernel [3,12]. Then *heat kernel smoothing* of Y is given analytically as

$$K_\sigma * Y(p) = \sum_{j=0}^{\infty} e^{-\lambda_j\sigma} \beta_j\psi_j(p), \tag{4}$$

where $\beta_j = \langle Y, \psi_j \rangle$ are Fourier coefficients. This is taken as the estimate for θ . Since the expansion (4) is a unique solution to isotropic heat diffusion [3,12], we can avoid the need to solve the diffusion equation using less stable numerical schemes such as the finite difference method [8,14].

In this new analytic framework, we need to compute the terms in (4). We first solve for the eigensystem (2) and obtain λ_j and ψ_j . To estimate β_j , we let $\sigma = 0$ in (4). Then K_σ becomes the Dirac-delta function and we obtain

$$Y(p) = \sum_{j=0}^{\infty} \beta_j\psi_j(p) . \tag{5}$$

This is the usual Fourier expansion of Y . The finite series expansion of (5) will be then used in estimating the Fourier coefficients in the least squares fashion.

3 Numerical Implementation

Generalized Eigenvalue Problem. Since the closed form expression for the eigenfunctions of the Laplace-Beltrami operator on an arbitrary curved surface is unknown, the eigenfunctions are numerically calculated by discretizing the Laplace-Beltrami operator. To solve the eigensystem (2), we need to discretize it on a triangular mesh using the Cotan formulation [41]. In particular, Qiu et al. [11] presented a similar Cotan discretization and constructed splines on a manifold using the eigenfunctions; however, there is no direct mathematical relation between splines and heat kernel smoothing.

Using the Cotan formulation, (2) is simplified as the generalized eigenvalue problem:

$$\mathbf{C}\psi = \lambda\mathbf{A}\psi, \quad (6)$$

where \mathbf{C} is the stiffness matrix, \mathbf{A} is the mass matrix, and ψ is the unknown eigenfunction evaluated at mesh vertices. Because \mathbf{C} and \mathbf{A} are large sparse matrices, we have solved the problem using the *Implicitly Restarted Arnoldi Method* [79] without consuming large amount of memory and time for zero entries. The MATLAB code is given at <http://brainimaging.waisman.wisc.edu/~chung/lb>. Fig. 1 shows the first few eigenfunctions for a mandible surface.

Finite Eigenfunction Expansion. Let \mathcal{H}_k be the subspace spanned by up to k -th degree basis. Then an arbitrary measurement Y is estimated in the subspace \mathcal{H}_k by minimizing the sum of squared residual:

$$\arg \min_{g \in \mathcal{H}_k} \|g - Y(p)\|^2 = \sum_{j=0}^k \beta_j \psi_j(p) . \quad (7)$$

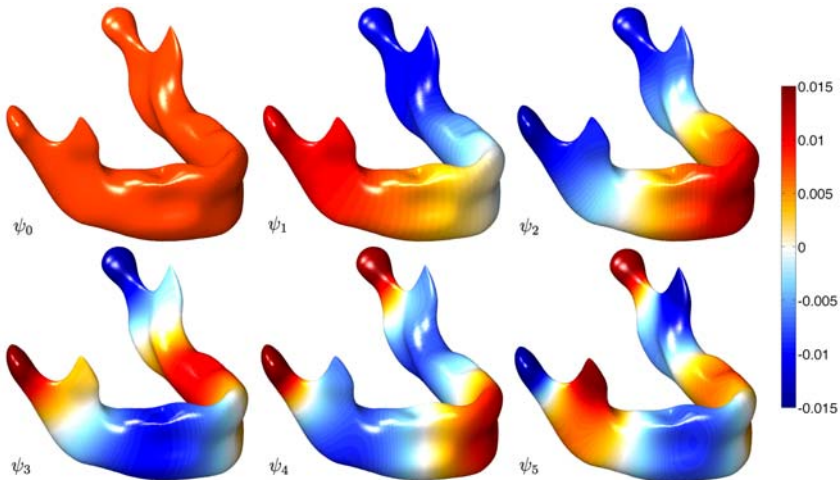


Fig. 1. The eigenfunctions are projected on the surface smoothed by the proposed heat kernel smoothing with $\sigma = 1$ and $k = 139$. The first eigenfunction is simply $\psi_0 = 1/\sqrt{\mu(\mathcal{M})}$. The color scale is thresholded at ± 0.015 for better visualization.

Consider the triangular mesh for \mathcal{M} with N_v nodes, and let $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)'$ and $\mathbf{Y} = (Y(p_1), \dots, Y(p_{N_v}))'$, for $k \leq N_v$. Then, we can represent (7) as the normal equation,

$$\mathbf{Y} = \boldsymbol{\beta}\boldsymbol{\Psi}, \quad (8)$$

where $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_0, \dots, \boldsymbol{\Psi}_k)$ and $\boldsymbol{\Psi}_j = (\psi_j(p_1), \dots, \psi_j(p_{N_v}))'$, and $\boldsymbol{\beta}$ are estimated in the least squares fashion, $\hat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{Y}$. Since the size of matrix $\boldsymbol{\Psi}'\boldsymbol{\Psi}$ can become fairly large when there is a need to obtain large number of basis, it may be difficult to directly invert the matrix. So we have adopted a more general iterative strategy to overcome possible computational bottleneck for large k .

Iterative Residual Fitting Algorithm. The Fourier coefficients are estimated based on an iterative procedure that utilizes the orthonormality of the eigenfunctions [2]. Decompose the subspace \mathcal{H}_k into smaller subspaces as the direct sum $\mathcal{H}_k = \mathcal{I}_0 \oplus \mathcal{I}_1 \cdots \oplus \mathcal{I}_k$, where each subspace \mathcal{I}_j is the projection of \mathcal{H}_k along the j -th eigenfunction. Instead of directly solving the normal equation (8), we project the normal equations into a smaller subspace \mathcal{I}_j and find the corresponding β_j in an iterative fashion from increasing the degree from 0 to k .

At degree $k = 0$, we write $\mathbf{Y} = \boldsymbol{\Psi}_0\beta_0 + \mathbf{r}_0$, where \mathbf{r}_0 is the residual of estimating \mathbf{Y} in subspace \mathcal{I}_0 . Then, we estimate β_0 by minimizing the residual in the least squares fashion:

$$\hat{\beta}_0 = (\boldsymbol{\Psi}'_0\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}'_0\mathbf{Y} \quad (9)$$

At degree j , we have

$$\mathbf{r}_{j-1} = \boldsymbol{\Psi}_j\beta_j + \mathbf{r}_j, \quad (10)$$

where the previous residual $\mathbf{r}_{j-1} = \mathbf{Y} - \boldsymbol{\Psi}_0\hat{\beta}_0 - \dots - \boldsymbol{\Psi}_{j-1}\hat{\beta}_{j-1}$. The next residual \mathbf{r}_j is then estimated as $\hat{\beta}_j = (\boldsymbol{\Psi}'_j\boldsymbol{\Psi}_j)^{-1}\boldsymbol{\Psi}'_j\mathbf{r}_{j-1}$. The optimal stopping rule is determined if the decrease of the root mean squared errors (RMSE), $\sqrt{\sum_{i=1}^{N_v} \mathbf{r}_k^2(p_i)/N_v}$, is statistically significant using the F -test (Fig. 2). We compute the F statistic at each degree, and find the degree of expansion where corresponding p -value first becomes bigger than the pre-specified significance 0.01.

4 Experimental Results

We applied the proposed smoothing method to mandible surfaces obtained from CT and further compared it against iterative kernel smoothing method [3,6].

Image Acquisition and Preprocessing. The CT images were obtained using several different models of GE multi-slice helical CT scanners. The CT scans were acquired directly in the axial plane with a 1.25 mm slice thickness, matrix size of 512×512 and 15–30 cm FOV. Image resolution varied and was in the range of 0.29 to 0.59 mm as determined by the ratio of field of view (FOV) divided by the matrix. CT scans were converted to DICOM format and subsequently Analyze 8.1 software package (AnalyzeDirect, Inc., Overland Park, KS) was used

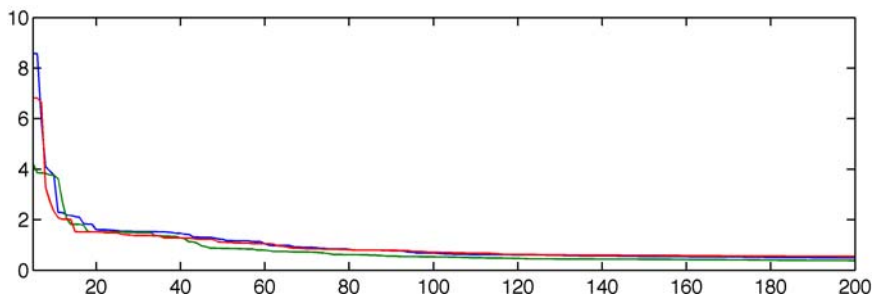


Fig. 2. The plot of the root mean squared errors (RMSE) for coordinates x (blue), y (red) and z (green) for a sample mandible surface, varying degree k from 5 to 200. The optimal degree for the surface is determined as 139.

in segmenting binary mandibular structure based on histogram thresholding. By checking the Euler characteristic, holes in mandible images were automatically filled up using morphological operations. This process was necessary to make the mandible binary volume to be topologically equivalent to a solid sphere and produces the surface mesh that is topologically equivalent to a sphere.

Results. We applied the proposed method in smoothing a mandibular surface. The optimal eigenfunction expansion was determined using the F -test at $\alpha = 0.01$. Fig. 2 shows the plot of the RMSE for varying degrees between 5 to 200. As the degree k increases, the RMSE for each coordinate rapidly decreases and starts to flatten out at a certain degree.

The numerical implementation was done with MATLAB 7.9 in 2×2.66 GHz Quad-Core Intel Xeon processor MAC PRO with 32 GB memory. For a mesh with 22050 vertices, the entire process took approximately 105 seconds: 85 seconds for setting up the generalized eigenvalue problem (6), 10 seconds for actually solving (6), 0.1 seconds for the IRF, and 9 seconds for finding the optimal degree.

Comparison. The proposed heat kernel smoothing was compared against widely used iterated kernel smoothing [3,6] of which the MATLAB code is given in <http://www.stat.wisc.edu/mchung/software/hk/hk.html>. In iterated kernel smoothing, the weights of the kernel are spatially adapted to follow the shape of heat kernel in discrete fashion along a surface mesh. Smoothing with large bandwidth is broken into iterated smoothing with smaller bandwidths:

$$K_{m\sigma} * Y = \underbrace{K_{\sigma} * \dots * K_{\sigma}}_{m \text{ times}} * Y . \quad (11)$$

For small σ , using the *parametrix expansion* [12], we can approximate heat kernel locally using the Gaussian kernel for small bandwidth:

$$K_{\sigma}(p, q) = \frac{1}{\sqrt{4\pi\sigma}} \exp\left[-\frac{d^2(p, q)}{4\sigma}\right][1 + O(\sigma^2)], \quad (12)$$

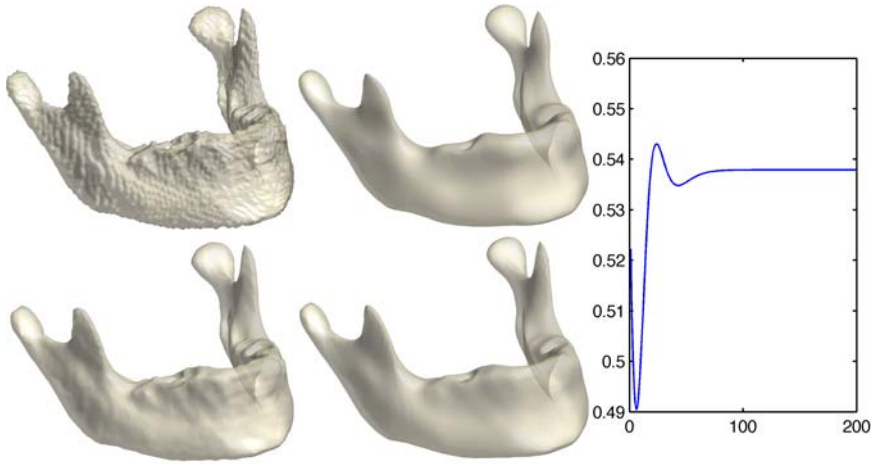


Fig. 3. Top left: original mandible surface. Top right: heat kernel smoothing with $\sigma = 1$ and $k = 139$. This is taken as the ground truth and iterative kernel smoothing is compared. Bottom left: iterative kernel smoothing with $\sigma = 1/6$ and $m = 6$ when the reconstruction error is minimum. Bottom right: iterative kernel smoothing with $\sigma = 1/6$ and $m = 30$. Right: RMSE for x -coordinate over the number of iterations m .

where $d(p, q)$ is the geodesic distance between p and q . For sufficiently small bandwidth, all the kernel weights are concentrated near the center, so the first neighbors of a given vertex in a mesh is used in the approximation.

By taking the proposed heat kernel smoothing as the ground truth, we were able to determine the performance of iterated kernel smoothing. Due to the lack of the ground truth, there are no studies in the literature validating the performance of iterated kernel smoothing except [5]. For heat kernel smoothing, we used the bandwidth $\sigma = 1$ and eigenfunctions up to $k = 139$ degree. For iterated kernel smoothing, we varied the number of iterations $1 \leq m \leq 200$ with the correspondingly smaller bandwidth $1/m$ to have the effective bandwidth of 1. The performance of the iterated kernel smoothing depended on the number of iterations, as shown in the plot of RMSE of x -coordinate over the number of iterations (Fig. 3 right). The RMSE was up to 0.5430 and it did not decrease even when we increase the number of iterations. This comparison directly demonstrates for the first time, the limitation of iterated heat kernel smoothing which does not converge to heat diffusion.

In another comparison (Fig. 4), we numerically constructed a heat kernel with a small bandwidth 0.2 as a sample data. Then we performed the additional iterated kernel smoothing with $\sigma = 0.2$ and $m = 49$ on the sample data to obtain the effective smoothing bandwidth of 10. The result was compared with the proposed heat kernel smoothing with the bandwidth 9.8 on the sample data making the effective bandwidth of 10. From Fig. 4 we immediately see that the shapes of two kernels are different. This visually demonstrates iterated kernel smoothing substantially diverges from heat kernel smoothing.

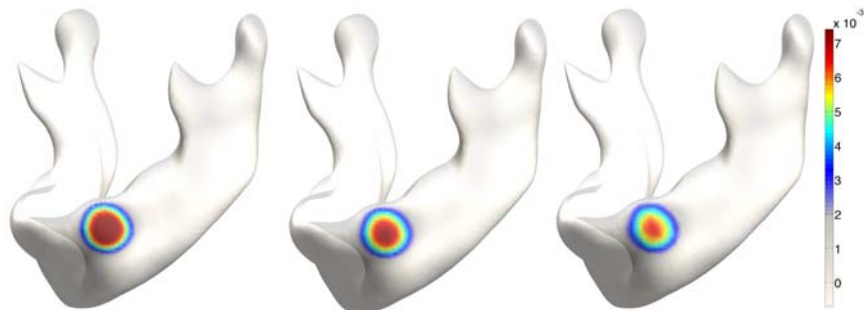


Fig. 4. Comparison of the shape of kernels. Left: heat kernel with 0.2 used as a sample data. Middle: The sample data is smoothed with heat kernel smoothing method with bandwidth $\sigma = 9.8$ making the effective bandwidth of 10. Right: iterated kernel smoothing with bandwidth $\sigma = 0.2$ and $m = 49$ iterations applied to the sample data.

5 Conclusions

We present a novel surface data smoothing framework where a smoothed surface measure is represented as a weighted linear combination of Laplace-Beltrami eigenfunctions. The expansion analytically solves isotropic heat diffusion. Taking the expansion as the ground truth, the proposed method is compared against widely used iterated kernel smoothing. The proposed method outperforms iterated kernel smoothing in accuracy. The divergence of iterated kernel smoothing is visually represented as kernel shapes on the mandible confirming the superiority of this proposed framework.

Acknowledgement. This work was supported by NIH-grant R01 DC6282 from the National Institute of Deafness and other Communicative Disorders, a core grant P-30 HD03352 to the Waisman Center from the National Institute of Child Health and Human Development, WCU-grant from the government of Korea to the Department of Brain and Cognitive Sciences, Seoul National University. We thank Dongjun Chung, Lindell R. Gentry, Mike S. Schimek, Katelyn J. Kassulke and Reid B. Durtschi for assistance with image acquisition and segmentation.

References

1. Andrade, A., Kherif, F., Mangin, J., Worsley, K.J., Paradis, A., Simon, O., Dehaene, S., Le Bihan, D., Poline, J.-B.: Detection of fmri activation using cortical surface mapping. *Human Brain Mapping* 12, 79–93 (2001)
2. Chung, M.K., Dalton, K.M., Shen, L., Evans, A.C., Davidson, R.J.: Weighted Fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging* 26, 566–581 (2007)
3. Chung, M.K., Robbins, S., 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)

4. Chung, M.K., Taylor, J.: Diffusion smoothing on brain surface via finite element method. In: Proceedings of IEEE International Symposium on Biomedical Imaging, ISBI (2004)
5. Hagler Jr., A.P., Saygin, D.J., Sereno, M.I.: Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage* 33, 1093–1103 (2006)
6. Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., et al.: Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage* 32, 180–194 (2006)
7. Hernandez, V., Roman, J.E., Tomas, A., Vidal, V.: A survey of software for sparse eigenvalue problems. Technical Report STR-6, Universidad Politécnic de Valencia (2006), <http://www.grycap.upv.es/slepc>
8. Joshi, A.A., Shattuck, D.W., Thompson, P.M., Leahy, R.M.: A parameterization-based numerical method for isotropic and anisotropic diffusion smoothing on non-flat surfaces. *IEEE Transactions on Image Processing* 18(6), 1358–1365 (2009)
9. Lehoucq, R.B., Sorensen, D.C., Yang, C.: ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. SIAM Publications, Philadelphia (1998)
10. Lévy, B.: Laplace-Beltrami Eigenfunctions: Towards an Algorithm that Understands Geometry. In: IEEE International Conference on Shape Modeling and Applications, p. 13. IEEE, Los Alamitos (2006)
11. Qiu, A., Bitouk, D., Miller, M.I.: Smooth functional and structural maps on the neocortex via orthonormal bases of the laplace-beltrami operator. *IEEE Transactions on Medical Imaging* 25, 1296–1396 (2006)
12. Rosenberg, S.: The Laplacian on a Riemannian Manifold. Cambridge University Press, Cambridge (1997)
13. Spira, A., Kimmel, R., Sochen, N.: A short-time Beltrami kernel for smoothing images and manifolds. *IEEE Transactions on Image Processing* 16, 1628–1636 (2007)
14. Tasdizen, T., Whitaker, R., Burchard, P., Osher, S.: Geometric surface smoothing via anisotropic diffusion of normals. In: Geometric Modeling and Processing, pp. 687–693 (2006)

Under-Determined Non-cartesian MR Reconstruction with Non-convex Sparsity Promoting Analysis Prior

Angshul Majumdar and Rabab K. Ward

Department of Electrical and Computer Engineering, University of British Columbia
{angshu.lm, rababw}@ece.ubc.ca

Abstract. This work explores the problem of solving the MR reconstruction problem when the number of K-space samples acquired in a non-Cartesian grid is considerably less than the resolution (number of pixels) of the image. Mathematically this leads to the solution of an under-determined and ill-posed inverse problem. The inverse problem can only be solved when certain additional/prior assumption is made about the solution. In this case, the prior is the sparsity of the MR image in the wavelet domain. The non-convex l_p -norm ($p < 1$) of the wavelet coefficient is a suitable metric for sparsity. Such a prior can appear in two forms – in the synthesis prior formulation, the wavelet coefficients of the image is solved for while in the analysis prior formulation the actual image is solved for. Traditionally the synthesis prior formulation is more popular. However, in this work we will show that the analysis prior formulation on redundant wavelet transform provides better MR reconstruction results compared to the synthesis prior formulation.

Keywords: MRI, image reconstruction, non-convex optimization.

1 Introduction

In Magnetic Resonance Imaging (MRI) the K-space (Fourier frequency) samples are collected. The problem is to reconstruct the image from the K-space samples. The reconstruction is trivial (inverse Discrete Fourier Transform) when the K-space is densely sampled. But such dense sampling is time consuming. In this work, we investigate the possibility to reduce MR acquisition times by reducing the number of K-space samples.

We are particularly interested in non-Cartesian K-space sampling like the radial, spiral or rosetta. For such sampling, the relationship between the image space and the K-space is given by [1],

$$y_{m \times 1} = F_{m \times n} x_{n \times 1} + \eta_{n \times 1}, m < n \quad (1)$$

where, y is the collected K-space samples, F is the Non Uniform Fast Fourier Transform (NUFFT), x is the image and η is the noise, assumed to be Gaussian with mean zero.

The inverse problem (1) is under-determined, and does not have a unique solution. However, if additional information is available regarding the character solution, one

may be able to pick up a unique solution satisfying that character. In [2] it is argued that the MR image is sparse in Wavelet domain. Therefore, a solution to (1) was seek which had a sparse representation in the wavelet domain. It was shown experimentally in [2] that the said assumption is valid and leads to impressive image reconstruction results from under-sampled K-space.

Incorporating the wavelet transform into the inverse problem (1) leads to the following equation,

$$y = FW^T \alpha + \eta \tag{2}$$

Where W is the forward wavelet transform (W^T is the inverse) and α is the transform coefficients (assumed to be sparse for MR images).

Assume that the vector α is k-sparse (i.e. k non-zeroes and n-k zeroes). Ideally one would solve the following optimization problem in order to search for the sparsest solution.

$$\min \|\alpha\|_0 \text{ subject to } \|y - FW^T \alpha\|_2 \leq \sigma \tag{3}$$

where $\|\cdot\|_0$ is the number of non-zeroes in the vector and σ is the standard deviation of noise.

Since the wavelet coefficient of the MR image is assumed to be k-sparse (i.e. k non-zeroes and n-k zeroes), only $m \approx 2k$ samples are needed for obtaining the solution of (7) via (8). Unfortunately (8) is an NP hard problem. But theoretical study [3] proves that – (i) if a linear system has a sparse solution, the solution is typically unique; and (ii) the solution can be obtained by the following convex optimization (4) instead of an NP hard problem.

$$\min \|\alpha\|_1 \text{ subject to } \|y - FW^T \alpha\|_2 \leq \sigma \tag{4}$$

This is a very strong result which guarantees recovery of the image by a tractable algorithm, but on the other hand, the number of samples required is larger $m \approx Ck \log(n/k)$. In [2], the MR reconstruction was based on (4).

Our aim is to reduce the MR acquisition time as much as possible, which in turn is proportional to the number of K-space samples. Therefore to reduce MR acquisition time, we should be able to reconstruct the images with as few samples as possible. Between the two extremes (3) (very few samples but NP hard reconstruction algorithm) and (4) (considerably larger number of samples but easy reconstruction algorithm) there exists reconstruction algorithms based on non-convex optimization.

$$\min \|\alpha\|_p \text{ subject to } \|y - FW^T \alpha\|_2 \leq \sigma, 0 < p < 1 \tag{5}$$

Such non-convex methods require samples that are intermediate between the requirements of (3) and (4), $m \approx C_1 k + pC_2 k \log(n/k)$ [4]. As the value of ‘p’ is reduced, the number of samples required for reconstruction reduces smoothly.

Algorithms for solving non-convex optimization problems (5) are not more complex than solving convex optimization problems, but are not guaranteed to reach a global minima. However, (5) is not any non-convex problem, rather it is quasi-convex. Convergence proofs for quasi-convex formulations can be found in [5].

The formulation discussed so far (3-5) is called the synthesis prior formulation. In this formulation, the sparse wavelet coefficients of the MR image are solved for. Once the wavelet domain representation is obtained, the image is reconstructed via the wavelet synthesis equation ($x=W\alpha$). The synthesis prior formulation (for the convex case, i.e. $p = 1$) has been previously used [2] for MR image reconstruction. However, a recent work [6] has pointed out that an alternate analysis prior (for the convex case, $p=1$) formulation (6) yields better reconstruction results when redundant wavelets are used instead of orthogonal wavelets (for the orthogonal case, the two formulations are the same). The experiments in [6] were carried out on synthetically generated data.

$$\min \|Wx\|_p \quad \text{subject to} \quad \|y - Fx\|_2 \leq \sigma, 0 < p < 1 \quad (6)$$

The analysis prior formulation solves for the image directly rather than the wavelet coefficients. In this work, we show that the non-convex analysis prior when employed on redundant wavelets yield better MR reconstruction results compared to synthesis prior on orthogonal wavelets.

There are many algorithms to solve the convex optimization problem with synthesis prior (4). There is only a single general purpose algorithm [7, 8] to solve the non-convex synthesis prior problem (5). This algorithm is called the Iterative Reweighted Least Squares (IRLS) algorithm. However, IRLS is a second order method and is consequently very slow. To overcome the limitations of speed we propose a new algorithm for solving the non-convex synthesis prior problem.

There are a few algorithms for solving the convex optimization problem for analysis priors, but none for the non-convex version. Therefore, we had to develop a new algorithm to solve it. Our proposed algorithms for non-convex optimization on analysis/synthesis priors are based on the Majorization-Minimization approach.

The rest of the paper is organized into several sections. The next section contains informal derivations of the non-convex algorithms proposed in this paper. The experimental results are presented in Section 3. Finally in Section 4, the conclusions of this work are discussed.

2 Optimization Algorithms

The constrained problems (5) and (6) are hard to solve. Rather, we proceed to solve their unconstrained versions,

$$\min \|y - FW^T\alpha\|_2^2 + \lambda \|\alpha\|_p^p \quad (7)$$

$$\min \|y - Fx\|_2^2 + \lambda \|Wx\|_p^p \quad (8)$$

The parameter λ in the equations (7) and (8) are related to σ in (5) and (6), but the relationship is not analytical. For the time being we concentrate on the solution of equations (7) and (8). How to make the solutions of (7) and (8) reach those of (5) and (6) iteratively, will be discussed later.

2.1 Majorization-Minimization

The Majorization-Minimization framework is outlined in [9]. The problems (7) and (8) do not have a closed form solution and has to be solved iteratively. In the Majorization-Minimization (MM) approach, at each iteration the function to be minimized is replaced by a surrogate function which has a higher functional value at all points except at its minima; the value of the surrogate function and the original function are the same at the minima. The surrogate function is easier to solve than the original function. The idea behind this approach is to construct such a surrogate function so that its solution after each iteration, is closer to the desired solution.

MM Framework

Let $J(x)$ be the original function that is to be minimized.

Initialize: iteration counter $k = 0$; initial estimate x_0 .

Repeat the following steps until a suitable exit criterion is met.

1. Chose $G_k(x)$ such that:
 - 1.1. $G_k(x) \geq J(x), \forall x$
 - 1.2. $G_k(x_k) = J(x_k)$
2. Set: $x_{k+1} = \min G_k(x)$

Set: $k=k+1$ and return to step 1.

2.2 Landweber Iterations

Let us consider the minimization of the following optimization problem,

$$J(x) = \|y - Mx\|_2^2 \tag{9}$$

Here M is a generic matrix, for (7) it needs to be replaced by FW^T and for (8) it needs to be replaced by F .

For this minimization, the surrogate function $G_k(x)$ is chosen to be,

$$G_k(x) = \|y - Mx\|_2^2 + (x - x_k)^T (aI - M^T M)(x - x_k) \tag{10}$$

To maintain strict convexity $a > \max \text{eig}(M^T M)$.

$G_k(x)$ is minimized by Landweber iterations,

$$x_{k+1} = x_k + \frac{1}{a} M^T (y - Mx_k) \tag{11}$$

Using this update formula, $G_k(x)$ can be expressed as,

$$G_k(x) = a \|x - x_{k+1}\|_2^2 - ax_{k+1}^T x_{k+1} + y^T y + x_k^T (aI - M^T M)x_k \tag{12}$$

Note that all the terms apart from the first term are independent of x and does not play any part in the minimization.

2.3 Synthesis Prior Algorithm

The problem to be solved is,

$$J(x) = \|y - FW^T \alpha\|_2^2 + \lambda \|\alpha\|_p^p \tag{13}$$

After the discussion in the previous sub-section the choice of $G_k(x)$ is quite obvious,

$$G_k(x) = \| \alpha - \alpha_{k+1} \|_2^2 + \frac{\lambda}{a} \|\alpha\|_p^p \tag{14}$$

where $\alpha_{k+1} = \alpha_k + \frac{1}{a} WF^T (y - FW^T \alpha_k)$

The minimizer of (16) is the following (we omit the derivation owing to limitations in space),

$$\alpha = \text{signum}(\alpha_{k+1}) \max(0, |\alpha_{k+1}| - \frac{\lambda}{2a} p \cdot \text{Diag}(|\alpha_{k+1}|^{p-2})) \tag{15}$$

2.4 Analysis Prior Algorithm

We are interested in solving the following problem,

$$J(x) = \|y - Fx\|_2^2 + \lambda \|Wx\|_p^p \tag{16}$$

Following the discussion in sub-section 3.2 the choice of $G_k(x)$ is,

$$G_k(x) = \|x - x_{k+1}\|_2^2 + \frac{\lambda}{a} \|Wx\|_p^p \tag{17}$$

where $x_{k+1} = x_k + \frac{1}{a} F^T (y - Fx_k)$

Owing to limitations in space, we skip the derivations. (17 is minimized by the following updates,

$$z_{k+1} = (\frac{2a}{\lambda} D^{-1} + cI)^{-1} (cz_k + H(x_k - H^T z_k)), \text{ where } c > \max \text{eig}(W^T W)$$

$$x_{k+1} = x_k - H^T z_{k+1}$$

$$x_k = x_{k-1} + \frac{1}{a} A^T (y - Ax_{k-1})$$

2.5 Constrained Optimization via Cooling

Theoretically the constrained forms (5) and (6) and the unconstrained forms (7) and (8) are equivalent for correct choice of λ and σ . However, for all practical case, it is

not possible to find the relation between the two explicitly. Owing to the smoothness of the Pareto curve, solving the unconstrained problems for a decreasing sequence of λ is guaranteed to reach the solution desired by the constrained form [10]. Based on this idea, we use a cooling algorithm to solve (5) and (6) via iteratively solving (7) and (8).

1. Choose a high value of λ initially.
2. Solve the unconstrained optimization (7)/(8) for the given value of λ .
3. Decrease the value of λ and go back to step 1.
4. Continue steps 2 and 3, until the mismatch $(\|y - FW^T \alpha\|_2 / \|y - Fx\|_2)$ between the data and the solution is less than σ .

3 Experimental Evaluation

The experimental evaluation was carried out on two real brain slices – Brainweb and NIH. The data is assumed to be noiseless. The experiments were simulated for radial scan lines in the k-space. The number of samples in k-space was always less than the number of pixels in the reconstructed image. As mentioned earlier the mapping from the Cartesian image space to the non-Cartesian k-space is the NUFFT [1]. Haar wavelets are used as the sparsifying basis. The normalized mean squared error between the ground-truth and the reconstructed image is used as an evaluation metric.

For all experiments it was found that the best reconstruction results were obtained at $p=0.8$ for both the analysis and the synthesis prior problems. In the following tables the reconstruction results for one slice each from Brainweb (Table 1) and NIH (Table 2) are reported for different sampling ratios (ratio of number of samples in K-space to the number of pixels in image).

Table 1. Reconstruction Results for Brainweb Slice

Algorithm	Sampling Ratio (in percent)					
	19.94	23.83	27.73	31.62	35.52	39.41
Orthogonal Synthesis	0.155	0.132	0.114	0.099	0.092	0.088
Redundant Synthesis	0.331	0.227	0.197	0.278	0.247	0.207
Redundant Analysis	0.105	0.091	0.079	0.071	0.068	0.060

Table 2. Reconstruction Results for NIH Slice

Algorithm	Sampling Ratio (in percent)					
	19.94	23.83	27.73	31.62	35.52	39.41
Orthogonal Synthesis	0.258	0.233	0.212	0.196	0.183	0.170
Redundant Synthesis	0.484	0.385	0.376	0.376	0.343	0.303
Redundant Analysis	0.203	0.184	0.167	0.156	0.144	0.134

Tables 1-2 confirm that minimizing the l_p -norm on a redundant analysis prior gives the best MR reconstruction results. We also provide a qualitative comparison of the reconstruction results; owing to the limitations in space we provide results only for the NIH slice (since it is more challenging).

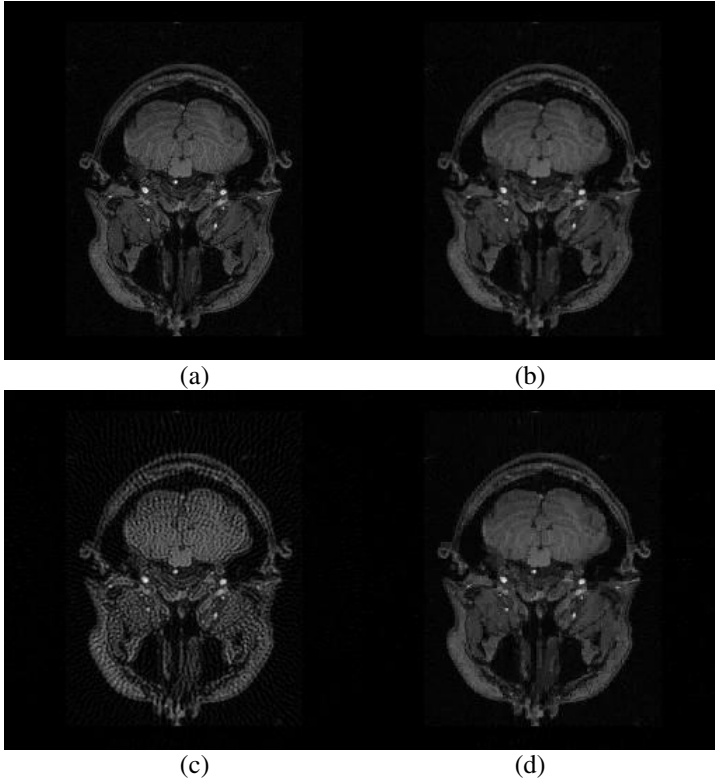


Fig. 1. NIH slice – (a) Ground truth; (b) Orthogonal Haar Synthesis; (c) Redundant Haar Synthesis and (d) Redundant Haar Analysis.

On close inspection the reconstructed images show blocky artifacts. This is because, we are using Haar wavelets for reconstruction. More sophisticated wavelets like Dualtree or Fractional Spline will reduce such artifacts.

4 Conclusion

In this work we look at the problem of reducing the MR acquisition times from a signal processing perspective. The acquisition time for the scanners are directly proportional to the number of number of samples collected in the K-space. Therefore in order to reduce the acquisition time the number of samples needs to be reduced. However, this leads to under-determinacy of the reconstruction problem. In spite of having an under-determined system of linear equations it is possible to exploit the sparsity of

the MR images in the wavelet domain to obtain a unique solution via a non-convex optimization problem.

The optimization problem can be formulated in two flavours – sparsity promoting analysis prior and sparsity promoting synthesis prior. In this work, we show that better reconstruction results can be obtained with analysis prior algorithm using redundant (Haar) wavelets.

Non-convex algorithms for sparsity promoting priors is not a well developed subject. There is only a single algorithm [7, 8] to solve the synthesis prior problem, which is computationally slow; hence we did not employ it. In this paper we propose a new fast algorithm for solving the non-convex synthesis prior problem. The analysis prior problem in the non-convex setting was never encountered earlier, therefore we developed an efficient first order algorithm to solve this problem.

References

- [1] Fessler, J.A.: On NUFFT-based gridding for non-Cartesian MRI. *Jour. Magn. Reson.* 188(2), 191–195 (2007)
- [2] Lustig, M., Donoho, D.L., Pauly, J.M.: Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. *Magnetic Resonance in Medicine* 58(6), 1182–1195 (2007)
- [3] Donoho, D.L.: For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Comm. on Pure and Applied Maths* 59(6), 797–829 (2006)
- [4] Chartrand, R.: Nonconvex compressed sensing and error correction. In: *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Honolulu, Hawaii (April 2007)
- [5] Luenberger, D.G.: Quasi-convex programming. *SIAM Journal on Applied Mathematics* 16, 1090 (1968)
- [6] Selesnick, W., Figueiredo, M.A.T.: Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In: *Proceedings of SPIE, vol. 7446, Wavelets XIII* (2009)
- [7] Majumdar, A., Ward, R.K.: Non-Convex Compressed Sensing from Noisy Measurements. *The Open Signal Processing Journal* 2, 40–45 (2009)
- [8] Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: *ICASSP* pp. 3869–3872 (2008)
- [9] <http://cnx.org/content/m32168/latest/>
- [10] Hennenfent, G., van den Berg, E., Friedlander, M.P., Herrmann, F.J.: New insights into one-norm solvers from the pareto curve. *Geophysics* 73(4) (2008)

A Statistical Approach for Achievable Dose Querying in IMRT Planning

Patricio Simari¹, Binbin Wu², Robert Jacques³, Alex King¹, Todd McNutt²,
Russell Taylor¹, and Michael Kazhdan¹

¹ Department of Computer Science, Johns Hopkins University, USA

² Department of Radiation Oncology and Molecular Radiation Science, Johns
Hopkins University, USA

³ Department of Biomedical Engineering, Johns Hopkins University, USA*

Abstract. The task of IMRT planning, particularly in head-and-neck cancer, is a difficult one, often requiring days of work from a trained dosimetrist. One of the main challenges is the prescription of achievable target doses that will be used to optimize a treatment plan. This work explores a data-driven approach in which effort spent on past plans is used to assist in the planning of new patients. Using a database of treated patients, we identify the features of patient geometry that are correlated with received dose and use these to prescribe target dose levels for new patients. We incorporate our approach in a quality-control system, identifying patients with organs that received a dose significantly higher than the one recommended by our method. For all these patients, we have found that a replan using our predicted dose results in noticeable sparing of the organ without compromising dose to other treatment volumes.

Keywords: Data-driven IMRT planning, achievable dose querying, Overlap Volume Histogram.

1 Introduction

We explore a data-driven approach for achievable dose querying in intensity-modulated radiation therapy (IMRT) planning. Our work is motivated by the time-intensive nature of the planning process, in which a dosimetrist specifies target dosage for the different organs and an optimization process is used to find the treatment plan that best meets these objectives. In general, a dosimetrist will seek doses that minimize the radiation received by organs at risk (OARs) while providing sufficient coverage of the primary treatment volume (PTV).

The pitfalls of specifying dosage objectives are two-fold: If the dosimetrist specifies too low a target dose to the OAR, the treatment plan may not be realizable and the treatment objectives will not be met. Conversely, if the dosimetrist specifies too high a dose to the OAR, the treatment plan will be realized at the

* This research was supported in part by the generosity of Paul Maritz, Philips Radiation Oncology Systems (Madison, WI), and by Johns Hopkins University internal funds.

cost of excessive irradiation of essential organs. As a result, IMRT planning often requires numerous trial-and-error iterations of computationally intensive dose simulation in order to finalize on a treatment plan.

Our goal is to facilitate the planning process by using information from previously treated patients in designing treatment plans for new patients. To this end, we use a database of previously treated patients to query for the lowest *achievable* dose to a new patient’s OARs. We do this in two steps. First, for each OAR, we identify a subset of *relevant* database patients – patients that have OARs which we expect to be at least as difficult to spare as the new patient’s. Then, using the assumption that a dose achieved by these more challenging patients should be achievable by the new patient, we set the target dose for the new patient’s OAR to the *minimum* dose achieved, over the subset of relevant patients.

Contribution. To prescribe an achievable dose for a query patient’s OAR, we need to identify the database patients that are harder to treat than the query. As with previous approaches (e.g. [4,1,7,3,6]), we use a shape descriptor that characterizes the geometry of the treatment volumes to retrieve patients from the database. However, while previous approaches have used shape descriptors primarily for measuring similarity, our research must address the more challenging task of using the geometric information to *rank* achievable dose.

We address this challenge by using the patient database to learn the correlation between features of the shape descriptors and achievable doses. Specifically, we use the recently developed Overlap Volume Histogram (OVH) [6], which characterizes the distribution of distances of points within the OAR from the surface of the PTV. For each OVH, we extract a set of features (e.g. the mean of the distribution, the standard deviation, etc.) and compute the rank correlation between the features of the OVH and the dose the associated OAR received.

Then, we predict that a database patient is more difficult to treat if a comparison of the database and query patients’ features indicate that the database patient is more difficult to plan, for *every* significantly correlated feature. That is, a database patient’s organs are said to be more difficult to spare than the query’s if they have larger feature values for every feature that is positively correlated with dose and smaller feature values for every feature that is negatively correlated.

We deployed our dose querying as a means for quality-control in a database of 91 head-and-neck patients. Using our approach we identified database patients whose OARs received a dose exceeding the dose recommended by the RTOG treatment protocol [2] but for which our method returned a dose that would meet the requirement. These patients were flagged for re-planning, with the retrieved dose used to set the target dose for optimization. For each of these patients, the re-plan lowered the dose to the OAR without compromising dose to other treatment volumes and resulted in treatments whose achieved dose were never larger than one Gy of the dose prescribed by our method.

Related Work. In the works most similar to ours, Hunt *et al.* [5] and Wu *et al.* [9] seek to address a similar problem by explicitly using the properties of the OVH to predict dose. For example, Hunt *et al.* use the overlap of the PTV with

the OAR (equivalent to the zero-crossing of the OVH) to predict dose, while Wu *et al.* use the OVH values to predict Dose Volume Histogram (DVH) values.

The use of rank correlation in our method distinguishes our approach in two ways. First, by using rank correlation to define relevant OAR instances, we are able to query for achievable dose without explicitly fitting a model that relates OVH values to dose. Second, our use of rank correlation allows us to identify predictive features on a per-OAR basis. Thus while we may identify the zero-crossing as an important feature for organs close to the PTV (where it is an indication of the proximity of the organ to the tumor, and hence a good predictor of the radiation the organ will receive), we may disregard it for organs distant from the PTV (where we expect it to always be zero, and hence have no predictive value).

Outline. The individual steps of our implementation are described in detail in the following sections.

- **Feature Definition:** We define a set of candidate OVH features and compute their correlations with OAR dose in Sec. 2.
- **Patient Relevance and Dose querying:** We formalize our notion of patient *relevance* based on feature sets and their correlations to dose in Sec. 3.
- **Feature Selection:** We select a maximally predictive subset of features based on our notion of relevance in Sec 4.
- **Dose Prediction:** We apply our achievable dose querying framework for plan quality control and present results in Sec. 5.

2 Overlap Volume Histogram Features

Recently, Kazhdan *et al* [6] introduced the Overlap Volume Histogram (OVH) as a means to characterize the spatial relationship between an OAR and a PTV. For a PTV T and OAR O , the value of the OVH of O with respect to T at distance t is defined as the fraction of volume of O that is at most a distance of t from T :

$$\text{OVH}_{O,T}(t) = \frac{|\{p \in O | d(p, T) \leq t\}|}{|O|},$$

where $d(p, T)$ is the signed distance of p from T 's boundary and $|\cdot|$ represents the volume of a set.

In this paper we will consider the OVH distributions in their normalized differential form (dOVH). Seen in this way, they can be interpreted as probability density functions. Given an OAR O , a target volume T , and a distance t , $\text{dOVH}_{O,T}(t)$ represents the relative likelihood that a uniformly randomly selected point in O will be at a distance of t from T .

Fig. 1 illustrates examples of these distributions for this paper's particular organs of interest with respect to PTV⁷⁰ for three patients in our database. We note that the different organs' dOVH curves have characteristically different shapes. Also of interest is the fact that in all cases except the mandible, the zero crossing value is zero and thus provides no information.

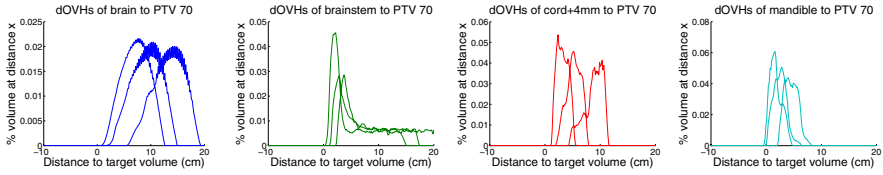


Fig. 1. dOVHs for the brain, brainstem, cord+4mm and and mandible with respect to PTV 70 for three patients in our database. While exhibiting variability within a same organ, each organ’s dOVH has a characteristically different shape.

The characteristic shapes of the dOVH curves for the different organs motivate us to develop a dose prediction model in which the different features are correlated with dose in different ways and the correlation is determined in an organ-dependent manner. In particular, we consider the following statistical features (and denote them with the parenthesized abbreviation): mean (*mu*), standard deviation (*sig*), skewness (*skw*), median (*med*), mode (*mod*), minimum (*min*), maximum (*max*), and the zero-crossing (*zc*) of the cumulative OVH. Additionally, we compute each of these for PTVs at 58.1, 63 and 70 Gy prescription values, and denote each feature with the corresponding number. For example, for a given patient and OAR, *mu70* denotes the mean of the dOVH distribution of this organ with respect to PTV⁷⁰. This results in a total of 24 features (the 8 statistical features mentioned above times the 3 PTVs).

We begin by analyzing the rank correlation of the above-mentioned features in relation to the maximum dose delivered to each of our database cases. This provides a better alternative to standard linear correlation since we wish to make minimal assumptions as to the nature of the functional relationship between the feature and the dose value. Furthermore, while some features are expressed in linear distance units, (e.g. *mu*), others are cubic (e.g. *zc*), while yet others are unitless (e.g. *skw*). As a measure of rank correlation, we use the Kendall tau coefficient, which is defined as $\tau = (n_c - n_d) / (\frac{1}{2}n(n-1))$, where n_c is the number of concordant pairs (i.e. pairs in which the ordering of the feature values agrees with that of dose), n_d is the number of discordant pairs (pairs in which the ordering is reversed), and the denominator expresses the total number of pairs in the data set of n entries. The value of τ ranges from -1 to 1 and its sign indicates if the correlation is positive or negative.

3 Achievable Dose Querying

To query for achievable dose to an organ, we define the notion of a *relevant organ*. The intuition is that a database OAR will be relevant to a query OAR if it is at least as difficult to spare when irradiating the PTV. In doing so, we query for the dose to each organ independently, implicitly assuming that organ dose is primarily dictated by its proximity to the PTV, as in [6,9], and not by the dose received by other organs.

Let us assume for the moment that we have selected a subset of the features discussed above $F = \{f_1, f_2, \dots, f_n\}$. We formalize the notion of relevance as follows. Given treatment plans for two OARs of the same type with corresponding dOVHs o_1 and o_2 , we say that o_2 is relevant to o_1 , and denote it $o_1 \prec_F o_2$, if for all i it holds that $f_i(o_1) < f_i(o_2)$ if $\tau(f_i) \geq 0$ and $f_i(o_1) > f_i(o_2)$ if $\tau(f_i) < 0$.

Using our database, we query for the achievable dose to a query OAR as the minimum of the doses delivered to all relevant database cases; that is:

$$\text{achievable}(o) = \min\{\text{maxdose}(o_i) : o \prec_F o_i\}$$

where $\text{maxdose}(o_i)$ denotes the maximum dose delivered to the organ associated to o_i . Notice that this formulation generalizes the approach of Wu *et al.* [9] reducing to it under the feature set $F = \{\text{min58}, \text{min63}, \text{min70}\}$.

4 Feature Selection

The above framework for achievable dose querying requires that we determine a set of features to use. In principal, we could use all features. However, the \prec operator induces a partial order and, as such, not all pairs of feature vectors are comparable. Specifically, given dOVHs o_1 and o_2 it can easily be the case that neither $o_1 \prec o_2$ nor $o_2 \prec o_1$ is true. As we increase the number of features, the tendency will be for fewer pairs to be comparable, exponentially fewer in the general case, thus dramatically reducing the size of instances relevant to a query. This observation provides a strong incentive to choose a small set of features under which the \prec operator is most predictive of dose relationship while still allowing for many instances to be comparable.

We can measure the predictiveness of a candidate set of features F by using a generalization of rank correlation to this multi-feature case. For all dOVH pairs o_i and o_j of a given organ type with corresponding maximum dose values d_i and d_j , we count the number of times in which the \prec relation agrees with the dose relation, that is $o_i \prec_F o_j$ and $d_i < d_j$, denoted n_c . We also count the total number of comparable pairs n_{\prec} , that is, pairs such that $o_i \prec_F o_j$ is true. We define our measure of predictiveness as n_c/n_{\prec} and our measure of comparability as $n_c/(\frac{1}{2}n(n-1))$.

To increase robustness to small perturbations of dose, we only examine pairs of organs for which $|d_i - d_j| > \frac{1}{2}\sigma$, where σ is the standard deviation of all observed dose values to the organ in question. Note that we do not count discordant pairs as in τ since the directionality of the correlation has been taken into account in the definition of \prec .

Our goal is to select the subset of features that maximizes predictiveness while retaining a minimum measure of comparability. Given that the space of all subsets is too large to explore exhaustively, we formulate a greedy approach. Beginning with an empty set $F_0 = \emptyset$, at each iteration k we select the feature f_i that maximizes the the predictiveness of $F_k \cup \{f_i\}$ and thus define F_{k+1} . We terminate the iteration when the comparability of F_k drops below 50%. This is formalized in Algorithm [11](#).

Algorithm 1. Select Features (*dOVH list O* , *max dose list D*)

Require: the i -th member of D is the max dose received by the organ instance whose dOVH is the i -th instance of O .

Feature set $F \leftarrow \emptyset$

repeat

$f \leftarrow \operatorname{argmax}_f \text{predictiveness}(F \cup \{f\}, O, D)$

$c \leftarrow \text{comparability}(F \cup \{f\}, O)$

if $c \geq .5$ **then**

$F \leftarrow F \cup \{f\}$

end if

until $c < .5$

return F

5 Experimental Results

We ran our feature selection algorithm on a clinical database consisting of 91 patients, their OAR contours, and associated DVHs for their clinical plans. We applied our method to four high priority organs: brain, brainstem, cord+4mm expansion and mandible and extracted the following features as maximally predictive of achievable maximum dose while maintaining at least 50% comparability.

Organ	F	Pred (%)	Comp (%)
Brain	$zc58, max58, mu58$	98.38	50.03
Brainstem	$min58, skw58$	75.12	64.64
Cord+4mm	$mod70, mod58, min58, min70$	68.88	53.17
Mandible	$zc70, max63, min58, min63, max70$	99.11	51.01

The way these numbers are interpreted is, for instance in the case of the mandible, of all pairs of organ instances in our database, 51.01% are comparable according to the relevance relationship. Of those instances, in 99.11% of the cases the relevance relationship was in agreement with the dose relationship. We note that in the case of the brain and brainstem, the most predictive features come from the relationship to the PTV^{58.1}. Clinicians take great care to avoid contamination from the higher dose PTV⁶³ and PTV⁷⁰ onto these organs. In the case of the other two, the higher dose PTVs come into play, especially in the mandible where overlap is frequent and the overlap volume with the PTV⁷⁰ (given by $zc70$) is selected first.

To evaluate our method, we applied it to the database. For each of the four organs we found an instance where the RTOG protocol max dose threshold [2] had been exceeded. In our database of 91 patients, there were 31 instances where the dose exceeded the recommended value for the brain, 35 for the brainstem, 66 for the cord+4mm and 12 for the mandible. We chose those instances where the dose was highest and our method predicted the threshold could be met. We re-planned each of these instances using the clinical plan as a starting point, lowering the objective of the organ in question as indicated by our query. In each instance, our query results were very nearly met or exceeded without sacrificing coverage or dose to other OARs. These results are summarized in Fig. 2.

Patient #	Organ Dose Point	Clinical Plan (Gy)	Est. Achievable (Gy)	Re-plan (Gy)
<i>a</i>	brain <i>D0</i>	60.24	52.02	52.91
<i>b</i>	brainstem <i>D0</i>	60.04	52.75	53.70
<i>c</i>	cord+4mm <i>D0</i>	50.26	46.86	42.91
<i>d</i>	mandible <i>D0</i>	77.80	71.33	71.10

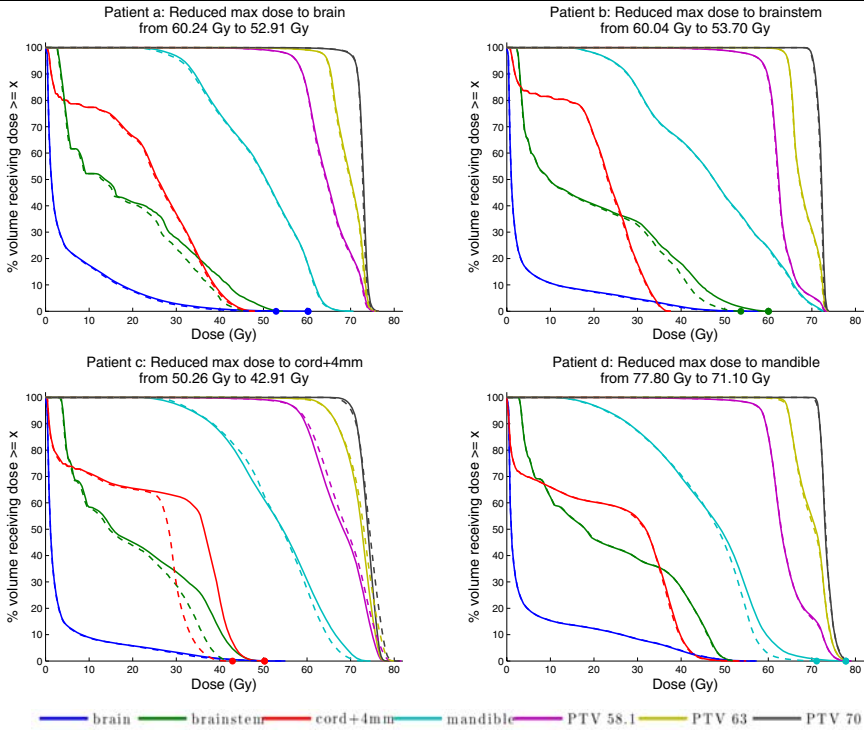


Fig. 2. DVHs resulting from re-planning patients based on our achievable dose query results. In each case, the recommended max dose threshold for the organ had been violated in the clinical plan (solid line) and was met in a re-plan (dashed line), which closely approximated our query results without sacrificing coverage or dose to other OARs.

6 Discussion and Future Work

We have presented a statistical approach for the querying of achievable dose to OARs in IMRT planning of head-and-neck cancer patients. In this presentation, our description has focused on the use of the OVH in prescribing treatment for head-and-neck patients. In future work, we will consider generalizing our approach in two ways. First, we will consider the use of other shape descriptors for defining feature sets for dose prediction. Second, we will explore the use of other OARs and other treatment sites. In concurrent work, we have used a simpler, heuristic OVH feature set to automatically generate all necessary planning constraints [8]. In future work, we will attempt to use the feature selection presented here in the same manner.

With regards to limitations, it should be noted that the quality of our query results will naturally be determined by the quality and case coverage of the database being queried. Furthermore, our approach relies on an assumption that the relative configuration of anatomic structures is static and cannot be purposefully altered by the clinician to facilitate treatment. This may not hold for other sites such as thorax and prostate respectively.

References

1. Ankerst, M., Kastenmüller, G., Kriegel, H., Seidl, T.: 3D shape histograms for similarity search and classification in spatial databases. In: Güting, R.H., Papadias, D., Lochovsky, F.H. (eds.) *SSD 1999. LNCS*, vol. 1651, pp. 207–228. Springer, Heidelberg (1999)
2. Eisbruch, A., Chao, K.C., Garden, A.: Phase I/II study of conformal and intensity modulated irradiation for oropharyngeal cancer (RTOG 0022). Radiation Therapy Oncology Group of the American College of Radiology (2004)
3. Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D.: A search engine for 3D models. *ACM Transactions on Graphics* 22(1), 83–105 (2003)
4. Horn, B.: Extended Gaussian images. *Proceedings of the IEEE* 72, 1656–1678 (1984)
5. Hunt, M.A., Jackson, A., Narayana, A., Lee, N.: Geometric factors influencing dosimetric sparing of the parotid glands using IMRT. *International journal of radiation oncology, biology, physics* 66, 296–304 (2006)
6. Kazhdan, M., Simari, P., McNutt, T., Wu, B., Jacques, R., Chuang, M., Taylor, R.: A shape relationship descriptor for radiation therapy planning. *Medical Image Computing and Computer-Assisted Intervention* 5762/2009(12), 100–108 (2009)
7. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Matching 3D models with shape distributions. In: *IEEE International Conference on Shape Modeling and Applications*, pp. 154–166 (2001)
8. Wu, B., Ricchetti, F., Sanguineti, G., Kazhdan, M., Simari, P., Jacques, R., Taylor, R., McNutt, T.: A data-driven approach to generating achievable dvh objectives in imrt treatment planning. *International Journal of Radiation Oncology, Biology, Physics* (2010) (to appear)
9. Wu, B., Ricchetti, F., Sanguineti, G., Kazhdan, M., Simari, P., Chuang, M., Taylor, R., Jacques, R., McNutt, T.: Patient geometry-driven information retrieval for IMRT treatment plan quality control. *Medical Physics* 36(12), 5497–5505 (2009)

Multivariate Statistical Analysis of Deformation Momenta Relating Anatomical Shape to Neuropsychological Measures

Nikhil Singh¹, P. Thomas Fletcher¹, J. Samuel Preston¹, Linh Ha¹, Richard King¹, J. Stephen Marron², Michael Wiener³, and Sarang Joshi¹

¹ University of Utah, Salt Lake City, UT

² University of North Carolina at Chapel Hill, Chapel Hill, NC

³ University of California, San Francisco, CA

`nikhil@cs.utah.edu`

Abstract. The purpose of this study is to characterize the neuroanatomical variations observed in neurological disorders such as dementia. We do a global statistical analysis of brain anatomy and identify the relevant shape deformation patterns that explain corresponding variations in clinical neuropsychological measures. The motivation is to model the inherent relation between anatomical shape and clinical measures and evaluate its statistical significance. We use Partial Least Squares for the multivariate statistical analysis of the deformation momenta under the Large Deformation Diffeomorphic framework. The statistical methodology extracts pertinent directions in the momenta space and the clinical response space in terms of latent variables. We report the results of this analysis on 313 subjects from the Mild Cognitive Impairment group in the Alzheimer’s Disease Neuroimaging Initiative (ADNI).

1 Introduction

In previous work, Large Deformation Diffeomorphic Metric Mapping (LDDMM) has been extensively used for the characterization of anatomical changes associated with various diseases [1], including the analysis of changes in anatomy with normative aging [2]. Most of the earlier studies on characterization of neuroanatomical changes have focused on the statistical analysis of deformation maps either using the associated Jacobian of the transformations, as in the now ubiquitous deformation-based morphometry [3], or directly by the analysis of the displacement maps. In this article we present a multivariate analysis of diffeomorphic transformations of the whole brain for relating complex anatomical changes observed in the population with neuropsychological responses, such as clinical measures of cognitive abilities, audio-verbal learning and logical memory. Contemporary studies using large deformation diffeomorphic transformations have focused on the characterization of group differences in the shape of specific substructures such as the hippocampus [4].

In this paper we consider the Mild Cognitive Impairment (MCI) subjects as a continuous class rather than a discrete class. Using Partial Least Squares (PLS),

we study the covariance of the anatomical structures in the entire brain volume without any segmentation or a priori regions of interest identification. The purpose of this study is to extract and identify shape deformation patterns in brain anatomy that relate to observed clinical scores depicting cognitive abilities. We show in our results that anatomical measures, such as cortical thickness and hippocampal volume, used in previous studies of Alzheimer’s and related dementia emerge naturally as the result of our analysis.

2 Technical Background

Fundamental to the study of the anatomical variation in a population is the construction of transformations, mapping one anatomical image to another, commonly referred to as registration maps. Conventionally, anatomical variation has been studied by generating transformation maps between each subject in the population using a pre-selected anatomical template. In contrast to this approach, we study anatomical variation by simultaneously generating transformations between the entire population and a common reference coordinate system. Several authors suggested this idea of nonlinear averaging of shapes for the construction of mean anatomical image or the ”atlas” representing the entire population. Avants et. al [5] proposed an energy minimization algorithm employing the diffeomorphic constraints on the average transformation. Cootes et. al [6] emphasized the notion of dense correspondences across group of images using diffeomorphic functions to optimize a groupwise objective function. In this paper, we follow the unbiased atlas construction approach introduced by Joshi et. al [7] under the large deformation diffeomorphic setting. As the properties of large deformation diffeomorphic transformations and atlas construction are fundamental to our analysis we briefly review it here, for a detailed analysis we refer the reader to a comprehensive review by Younes et. al [8] and references therein.

2.1 Riemannian Metric, Atlas Construction, and Momenta Maps

Riemannian Metric. Diffeomorphic transformations of the underlying coordinate space Ω are invertible transformations which are continuously differentiable with a differentiable inverse. This definition implies that the set of all diffeomorphisms of Ω has a group structure. A convenient and natural machinery for generating diffeomorphic transformations is by the integration of ordinary differential equations (ODE) on Ω defined via the smooth time indexed velocity vector fields $v(t, y) : (t \in [0, 1], y \in \Omega) \rightarrow \mathbb{R}^3$. The function $\phi^v(t, x)$ given by the solution of the ODE $\frac{dy}{dt} = v(t, y)$ with the initial condition $y(0) = x$ defines a diffeomorphism of Ω . One defines a Riemannian metric on the space of diffeomorphisms by inducing an energy via a Sobolev norm with the partial differential operator L on these velocity fields. The distance between the identity transformation and a diffeomorphism ψ is defined as the minimization

$$d(id, \psi)^2 = \min \left\{ \int_0^1 \langle Lv(t, \cdot), v(t, \cdot) \rangle dt : \phi^v(1, \cdot) = \psi(\cdot) \right\}$$

The distance between any two diffeomorphism is defined as $d(\phi, \psi) = d(id, \psi \circ \phi^{-1})$.

Atlas Construction. This Riemannian metric defined on the space of diffeomorphisms can now be used to compute a deformation that matches two images. The deformation ϕ is defined as the ‘optimal’ time-varying velocity field \hat{v} , based on the minimum energy criteria:

$$\hat{v} = \operatorname{argmin}_{v: \phi_t = v_t(\phi_t)} \int_0^1 \langle Lv(t, \cdot), v(t, \cdot) \rangle^2 dt + \frac{1}{\sigma^2} \int_{\Omega} \|I^0 \circ \phi^{-1} - I^1\|^2 dx \quad (1)$$

where the second term allows inexact matching, and σ is a free parameter controlling the tradeoff between exactness of the match and smoothness of the velocity fields. Now, given a collection of anatomical images $\{I^i, i = 1, \dots, N\}$, the minimum mean squared energy atlas construction problem is that of jointly estimating an image \hat{I} and N individual deformations:

$$\{\hat{I}, \hat{\phi}_i\} = \operatorname{argmin}_{I, \phi_i} \frac{1}{N} \sum_{i=1}^N \int_{\Omega} \|I \circ \phi_i^{-1} - I^i\|^2 dx + d(id, \phi_i)^2 \quad (2)$$

The implementation details are described in greater detail in Section 3

Momenta Maps. The joint minimizer of the atlas construction problem (2) estimates an atlas image \hat{I} while simultaneously solving the N LDDMM image matching problems. An important consequence of this is that the Euler-Lagrange equations associated with the LDDMM problem coincides with the Euler-Lagrange equations of geodesics on the group of diffeomorphisms. As shown in Younes et. al (8) the geodesic equations are completely determined via the initial momenta Lv_0 and furthermore it is in direction of the gradient of deforming image. This implies that for the atlas matching problem above, at the minimizer, for each of the N image matching problems the initial velocity is given by the equation $Lv^i(0, x) = a_0^i(x) \nabla \hat{I}(x)$. The quantity $a_0^i(x) \nabla \hat{I}(x)$ is referred to as the initial momenta. Each of the $i = 1, \dots, N$ geodesic equations evolve according to

$$Lv^i(t) = a^i(t) \nabla \hat{I}(t) \quad (3)$$

$$\frac{da^i(t, \cdot)}{dt} + \nabla \cdot (a^i(t) v^i(t)) = 0 \quad (4)$$

$$\frac{d\hat{I}(t)}{dt} = \nabla \hat{I}(t)^T v^i(t) \quad (5)$$

Equation (5) is the infinitesimal action of the velocity field v^i on the image, while (4) is the conservation of momenta.

2.2 Multivariate Analysis with Partial Least Squares

Traditionally, Partial Least Squares (PLS) has been used to characterize pertinent directions between independent variable and dependent variable in a high

dimensional multivariate regression setting. This approach to multivariate analysis using PLS was introduced in the neuroimaging community by Bookstein [9] with a detailed review by McIntosh et. al [10] and references therein. We now adapt the PLS methodology for the purpose of extracting and identifying deformation patterns in brain anatomy that relate to k observed clinical measures $y^i \in R^k$ depicting cognitive and neuropsychological responses of each of the $i = 1, \dots, N$ subjects. As described above, the anatomical variation in the collection of I^i is captured by the initial scalar momenta maps ($a^i(x)$) at the atlas \hat{I} . These momenta maps govern the deformation of the atlas along the geodesic in the group diffeomorphism towards the respective individual images I^i . Analogous to the classical PLS, we find directions \hat{a} in the momenta space, defined at the atlas in terms of deformation momenta a^i 's, and directions \hat{y} in the clinical response space, defined by y^i 's that explain their association in the sense of their common variance. We propose to extract these directions such that initial momenta when projected on to \hat{a} and the corresponding clinical responses when projected on to \hat{y} have maximum covariance. We call these projections latent variables, l_a and l_y respectively.

To find the anatomical variation that covaries maximally with clinical responses, we perform PLS analysis between the scalar momenta fields a^i and the response space y^i . The PLS problem is given by:

$$\max \text{cov}(\langle \hat{a}, a^i \rangle, \langle \hat{y}, y^i \rangle) \text{ subject to } \|\hat{a}\| = 1, \|\hat{y}\| = 1 \quad (6)$$

The subsequent directions are found by removing the component extracted (deflating the data) both in momenta space and the clinical response space as:

$$a^i \leftarrow a^i - \langle \hat{a}, a^i \rangle \hat{a} \text{ and } y^i \leftarrow y^i - \langle \hat{y}, y^i \rangle \hat{y}$$

The solution to the above maximization problem (6) is the Singular Value Decomposition (SVD) of the covariance matrix of the dependent and independent variables. The corresponding direction vectors \hat{a} 's and \hat{y} 's are the respective left and right singular vectors. The successive latent variables l_a 's and l_y 's are computed once by a single SVD.

Statistical Significance. The statistical significance of the directions extracted by PLS analysis can be assessed using the projected data (the latent variables) l_a 's and l_y 's. We use non-parametric permutation tests for calculating the significance of the regression of l_y 's on l_a 's and use the R^2 (the proportion of variance explained in l_y 's by l_a 's) as the test statistics. The distribution of the R^2 statistic under the null hypothesis is calculated by randomly reordering the momenta and clinical response association and then recalculating the new SVD and its associated R^2 each time. The significance of a particular latent variable is measured by the p-value from the empirical distribution.

3 Computational Framework

Two main challenges exist in implementing the LDDMM atlas building framework: the intensive computational cost and large memory requirements. Even

with a very low-resolution time discretization, and efficient multithreaded implementation atlas generation for a set of 84 MRI brain images with a resolution of $144 \times 192 \times 160$ takes approximately 10 hours on a high-end 32-core shared memory machine and uses approximately 50GB of memory. This makes parameter tuning and cross-validation schemes impractical, and limits the size of the population for which an atlas can reasonably be generated. The target of this study is the analysis of 313 MCI subjects from the ADNI data sets and on a single machine would require 200GB of memory and 40 hours of computation time for a single run.

Equation (2) can be efficiently solved using a parallel alternating algorithm by interleaving the updates of the optimal deformations and the estimate of the atlas image \hat{I} . For a fixed atlas image \hat{I} the N individual deformations are updated by performing a gradient step of (1) using the algorithm of Beg et. al (11). These deformations are completely independent of each other, naturally yielding to a distributed memory implementation. Further, the parallel nature of many of the image processing algorithms used in the deformation update process lends themselves to an efficient and massively parallel GPU-based implementation. An implementation of LDDMM atlas building for use on a GPU computing cluster was therefore developed based on MPI and the GPU image processing framework by Ha et. al (12). Individual deformation calculations are distributed across computing nodes, and nodes further distribute deformation calculations among GPUs. In this manner the only inter-GPU and inter-node communication required is in the atlas update step. Inter-GPU atlas computation is done in host (node) shared memory, and inter-node atlas computation is efficiently done by a parallel-reduce summation MPI call.

The GPU cluster used consists of 64 8-core computing nodes and 32 NVIDIA Tesla s1070 computing servers, each containing four GPUs. Each node controls two of the four GPUs contained in a s1070. Communication from the host to the GPU is via external x16 PCIe bus, and inter-node communication is through a 20Gbit/s 4x DDR infiniband interconnect. Using 42 nodes of the GPU cluster, the resulting implementation generated the atlas of the population of 313 brain images in under 40 minutes. This represents a speed up of an order of magnitude.

4 Results

313 patients with mild cognitive impairment (MCI) were selected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The T1 weighted, bias field corrected and N3 scaled structural Magnetic Resonance Images (MRI) and the selected neuropsychological data were downloaded from ADNI. Two global cognitive and functional assessment test were used. The first was the Clinical Dementia Rating scale, Sum of Boxes (cdr.sb). The second test was the modified Alzheimer's Disease Assessment Scale modified cognitive battery (adas-cog), which includes delayed word recall and number cancellation. Episodic memory was assessed using the Rey Auditory Verbal Learning Test (AVLT) and the Logical Memory test of the Wechsler Memory Scale-Revised. Both memory

tests had immediate recall (avlt.imm, logic.imm) and 30 minute delayed recall (avlt.del, logic.del). Note the AVLTL used the immediate recall after the 5th learning trial. Preprocessing the MRI involved skull stripping and registration to talairach coordinates using freesurfer [13]. Tissue-wise intensity normalization for white matter, gray matter and cerebrospinal fluid was performed using the expectation maximization (EM) based segmentation followed by the piecewise polynomial histogram matching algorithm.

Data Processing Details. The atlas was constructed with the 313 MCI subjects on the GPU cluster and the associated initial momenta fields a^i were computed. Each p dimensional a^i ($i = 1, \dots, 313$, $p = 144 \times 192 \times 160$) represents a row of a large $313 \times p$ matrix \mathbf{X} of momenta maps. The corresponding k dimensional clinical outcome y^i ($i = 1, \dots, 313$ and $k = 6$) populates the rows of the 313×6 matrix \mathbf{Y} of clinical outcomes. The PLS was then performed on \mathbf{X} and \mathbf{Y} data matrices. The significance tests for the extracted momenta direction and the clinical response directions was performed using 100,000 permutations.

Table 1. Significance test - 100000 permutations

latent variable	1	2	3	4	5	6
R^2	0.5010	0.5059	0.4515	0.4618	0.4356	0.5271
p-value	0.0669	0.1033	0.2289	0.3435	0.2775	0.0050
adas-cog	0.4373	0.0220	0.0973	-0.7900	0.4102	0.0800
cdr.sb	0.2069	0.9639	-0.1050	0.1077	-0.0266	-0.0688
avlt.imm	-0.4738	0.2175	0.4403	-0.2855	-0.2802	0.6118
avlt.del	-0.4484	0.1256	0.5166	0.0003	0.4574	-0.5542
logic.imm	-0.4325	0.0582	-0.5045	-0.5228	-0.3517	-0.3976
logic.del	-0.3916	0.0626	-0.5141	0.0972	0.6478	0.3865

Visualization of geodesic shooting. In our initial study we found that latent variables 1 and 6 (LV1 and LV6) had the lowest p-values with latent variable 6 being highly significant (p-value=0.005). We analyzed the extracted deformation directions for LV1 and LV6, owing to their low p-value, by evolving the atlas image \hat{I} along the geodesic in the corresponding projected momenta directions (\hat{a}) and interpreting the association with the directions extracted in their clinical response counterparts (\hat{y}).

The initial momenta direction was scaled to the maximum of the projections of all the momenta over the population of 313 images (along LV1 and LV6). The log Jacobians of the deformation, overlaid on atlas image \hat{I} , resulting from shooting \hat{I} along the geodesic with this momenta are shown in Fig. 1 and Fig. 2 respectively. The selected slices from this 3D overlay shown here capture relevant regions of the neuro-anatomical structures, such as hippocampus and ventricles, pertinent to cognitive impairment in Alzheimer's and related dementia. Fig. 3 displays the axial and sagittal slices of the 3D MRI deforming brain from the atlas along the geodesic for the equal intervals, t , both in the positive and the negative LV1 direction ($t = 0$ correspond to the atlas image, \hat{I}).

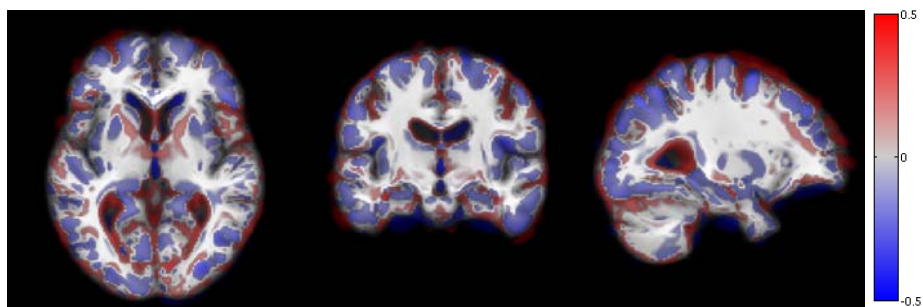


Fig. 1. LV1 log Jacobians overlaid on atlas. Red denotes regions of local expansion and blue denotes regions of local contraction (see electronic version for colored images)

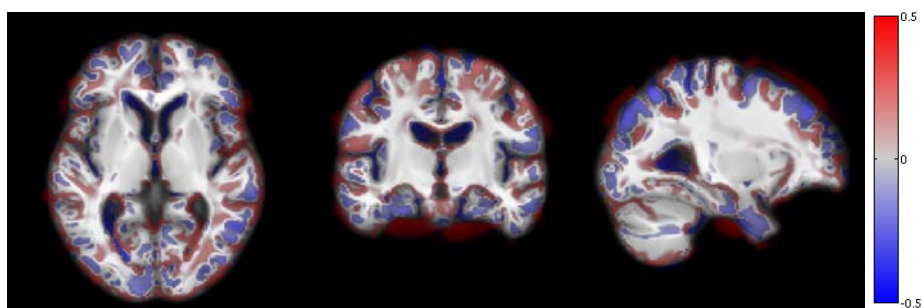


Fig. 2. LV6 log Jacobians overlaid on atlas

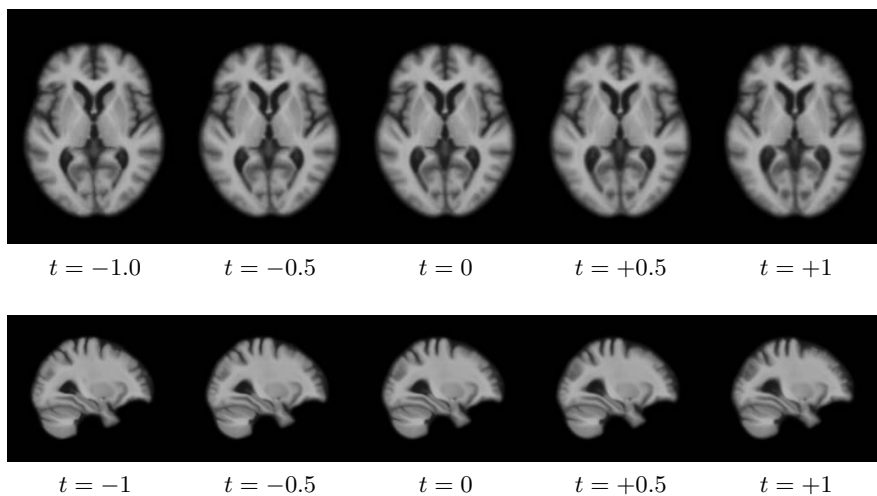


Fig. 3. Deformation of mean brain along LV1: t is the scaling parameter along LV1. The value $t = 0$ corresponds to the mean brain

5 Discussion

The major contribution of this article is that the shape deformation patterns in anatomical structures show up evidently as a result of the PLS analysis of the momenta.

LV1: We notice the expansion of lateral ventricles and CSF, together with the shrinkage of cortical surface along LV1 (Fig. 1 and Fig. 3). The corresponding neuropsychological clinical response direction is that of increasing adas-cog and cdr.sb (measures of increasing cognitive degeneration) and decreasing AVLT and logical scores (measures of audio verbal learning and logical memory). Another critical observation is the clearly evident shrinkage of the hippocampus and cortical and sub-cortical gray matter along these directions. Such patterns of atrophy are well known to characterize the disease progression in AD and related dementia.

LV6: The highly statistically significant LV6 explains an altogether independent set of anatomical deformation patterns that relate to corresponding patterns in audio-verbal learning scores and memory scores (immediate and delayed recall). The LV6 mainly explains deformations for learning and memory, owing to high absolute weights for AVLT and logic scores but very low weights to adas.cog and cdr.sb. The deformation patterns in anatomy Fig. 2 show almost invariance in hippocampal region well known to be important to memory formation. This observation has to our knowledge, never been reported before and we are studying further the medical importance of this finding.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). The research in this paper was supported by NIH grant 5R01EB007688, the University of California, San Francisco (NIH grant P41 RR023953) and the National Science Foundation (NSF grant CNS-0751152).

References

1. Ashburner, J., Csernansk, J., Davatzikos, C., Fox, N., Frisoni, G., Thompson, P.: Computer-assisted imaging to assess brain structure in healthy and diseased brains. *The Lancet Neurology* 2(2), 79–88 (2003)
2. Davis, B., Fletcher, P., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: *Proceeding of ICCV* (2007)
3. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping* 6(5-6), 348–357 (1998)
4. Wang, L., Beg, M., Ratnanather, J., Ceritoglu, C., Younes, L., Morris, J., Csernansky, J., Miller, M.: Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *IEEE Transactions on Medical Imaging* 26(4), 462 (2007)

5. Avants, B., Gee, J.C.: Shape averaging with diffeomorphic flows for atlas creation. In: IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004, pp. 595–598 (2004)
6. Cootes, T.F., Marsland, S., Twining, C.J., Smith, K., Taylor, C.J.: Groupwise diffeomorphic non-rigid registration for automatic model building. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 316–327. Springer, Heidelberg (2004)
7. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
8. Younes, L., Arrate, F., Miller, M.: Evolutions equations in computational anatomy. *NeuroImage* 45(1S1), 40–50 (2009)
9. Bookstein, F.L.: Partial least squares: a Dose-Response model for measurement in the behavioral and brain sciences (1994)
10. McIntosh, A.R., Lobaugh, N.J.: Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* 23(supp. 1), 250–263 (2004)
11. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision* 61(2), 139–157 (2005)
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. Dale, A., Fischl, B., Sereno, M.: Cortical surface-based analysis I. Segmentation and surface reconstruction. *Neuroimage* 9(2), 179–194 (1999)

Shape Analysis of Vestibular Systems in Adolescent Idiopathic Scoliosis Using Geodesic Spectra

Wei Zeng^{1,4}, Lok Ming Lui², Lin Shi³, Defeng Wang³, Winnie C.W. Chu³, Jack C.Y. Cheng³, Jing Hua¹, Shing-Tung Yau², and Xianfeng Gu⁴

¹ Wayne State University, Detroit, MI 48202, USA

² Harvard University, Cambridge, MA 02138, USA

³ The Chinese University of Hong Kong, Hong Kong, China

⁴ Stony Brook University, Stony Brook, NY 11794, USA

Abstract. Adolescent Idiopathic Scoliosis (AIS) characterized by the 3D spine deformity affects about 4% schoolchildren worldwide. One of the prominent theories of the etiopathogenesis of AIS was proposed to be the poor postural balance control due to the impaired vestibular function. Thus, the morphometry of the vestibular system (VS) is of great importance for studying AIS. The VS is a genus-3 structure situated in the inner ear and consists of three semicircular canals lying perpendicular to each other. The high-genus topology of the surface poses great challenge for shape analysis. In this work, we propose an effective method to analyze shapes of high-genus surfaces by considering their geodesic spectra. The key is to compute the canonical hyperbolic geodesic loops of the surface, using the Ricci flow method. The Fuchsian group generators are then computed which can be used to determine the geodesic spectra. The geodesic spectra effectively measure shape differences between high-genus surfaces up to the hyperbolic isometry. We applied the proposed algorithm to the VS of 12 normal and 15 AIS subjects. Experimental results show the effectiveness of our algorithm and reveal statistical shape difference in the VS between right-thoracic AIS and normal subjects.

1 Introduction

Adolescent Idiopathic Scoliosis (AIS) is a 3D spinal deformity which affects about 4% schoolchildren worldwide. It is believed to be a multi-factorial disease. The right-thoracic curve is the most typical type of AIS. The current treatment of AIS is still unsatisfactory because the etiopathogenesis remains unclear. AIS patients had a significantly higher mean body sway compared with healthy controls when their visual and somatosensory systems were simultaneously challenged [3]. It has been proposed that the altered processing of vestibular information might contribute to the etiology of AIS [11]. Based on the findings of abnormal postural balance control in AIS [3,9,11,21], one of the prominent theories of the etiopathogenesis of AIS was proposed to be the poor postural balance control

due to the impaired vestibular function [22]. The vestibular system (VS) is responsible for perception of head movements and sending postural signals to the brain. The important role of the vestibular system in the development of idiopathic scoliosis was manifested by an animal study using the frog model [14]. Recently, some preliminary results have also revealed the statistical difference in global morphology of the VS between right-thoracic AIS and normal controls [20]. The shape of the central contour of each canal for the VS surfaces has also been studied [2]. However, the complete geometry of the surface has not been fully analyzed. Thus, the morphometry of the VS is of great importance for understanding AIS. The VS is a genus-3 structure situated in the inner ear and consists of three semicircular canals lying perpendicular to each other, see Figure 1 (a). Because of the high-genus topology of the VS, it poses great challenge for shape analysis. This motivates us to propose a stable algorithm to study the VS systematically by making full use of rigorous structural features.

A lot of works have been done on shape analysis for anatomical structures. For example, Dale et al. [6] proposed a surface-based methods for studying the structural features of brain; Ashburner et al. [1] proposed a deformation-based morphometry for studying the 2D or 3D local displacement vector fields in brain volume or shape; Pizer et al. [15] proposed to measure shape differences by medial representations (M-reps); Davies et al. [7] developed an algorithm for shape discrimination in the hippocampus using the minimum description length (MDL) models; Chung et al. [5] proposed a tensor-based morphometry for examining spatial derivatives of deformation maps using spherical harmonic representation; Reuter et al. [17,16] presented Laplace-Beltrami spectra as shape DNA for statistical shape analysis. These methods provide a good way for analyzing anatomical shapes, but they generally cannot deal with high-genus structures.

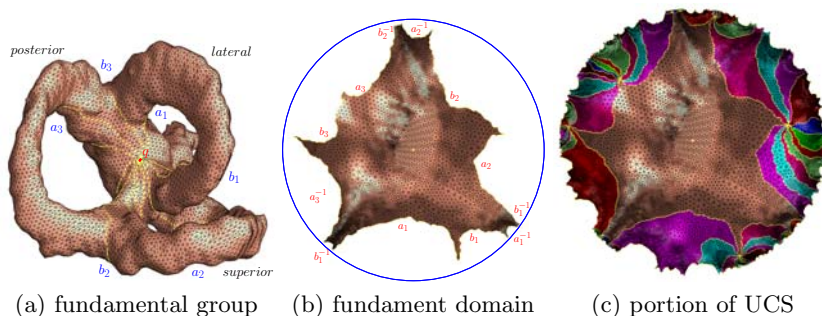


Fig. 1. Computing the hyperbolic metric for a vestibular system (VS) surface. (a) The VS is of genus-3 and consists of 3 semicircular canals lying perpendicular to each other: the lateral canal with (a_1, b_1) , the superior canal with (a_2, b_2) and the posterior canal with (a_3, b_3) , where $\{a_1, b_1, a_2, b_2, a_3, b_3\}$ are the homotopy group generators. (b) The fundamental domain embedded in the hyperbolic space, the Poincaré disk. (c) The finite portion of the universal covering space (UCS). Different fundamental domains are encoded in different colors.

In this work, we propose an effective algorithm for shape analysis of high-genus surfaces by considering their geodesic spectra. Under the hyperbolic metric, one can find the shortest geodesic loop in each homotopy class of the surface, and the lengths of such geodesic loops form the geodesic spectrum. The geodesic spectrum determines the structure of a high-genus surface up to hyperbolic isometry. To compute the geodesic spectrum, the basic idea is to compute the hyperbolic metric using the Ricci flow method. Ricci flow was first introduced by Hamilton [10] and later it was generalized to the discrete case [4]. Zeng et al. [23] applied the surface Ricci flow method to study general 3D shape matching and registration. Recently, Jin et al. [13] introduced the Teichmüller shape space to index and compare general surfaces with various topologies, geometries and resolutions. After the hyperbolic metric is computed, we can compute the Fuchsian group generators which can be used to determine the geodesic spectrum. The method is rigorous and practical. Experimental results on 12 normal and 15 AIS subjects show the effectiveness of our algorithm and reveal statistical shape difference in the VS between the normal and unhealthy groups.

2 Theoretic Background

This section briefly introduces the background knowledge of conformal geometry, and the basic concepts of algebraic topology and hyperbolic geometry. For more details, we refer readers to the classical textbooks [18] and [19].

2.1 Fundamental Group and Universal Covering Space

Let S be a topological surface with a base point q . All loops through q are classified by homotopic relations. All homotopy equivalence classes form the *homotopy group* or *fundamental group* $\pi_1(S, q)$, where the product is defined as the concatenation of two loops through their common base point. For a genus g closed surface, one can find *canonical homotopy group generators* $\{a_1, b_1, a_2, b_2, \dots, a_g, b_g\}$, such that a_k only intersects b_k once. Then the fundamental group is represented as $\langle a_1, b_1, a_2, b_2, \dots, a_g, b_g | a_1 b_1 a_1^{-1} b_1^{-1} a_2 b_2 a_2^{-1} b_2^{-1} \dots a_g b_g a_g^{-1} b_g^{-1} = e \rangle$.

A *covering space* of S is a space \tilde{S} together with a continuous surjective map $h : \tilde{S} \rightarrow S$, such that for every $p \in S$ there exists an open neighborhood U of p , $h^{-1}(U)$ is a disjoint union of open sets in \tilde{S} , each of which is mapped homeomorphically onto U by h . The map h is called the *covering map*. The *universal covering space (UCS)* is simply connected. A *deck transformation* of a cover $h : \tilde{S} \rightarrow S$ is a homeomorphism $f : \tilde{S} \rightarrow \tilde{S}$ such that $h \circ f = h$. All deck transformations form the so-called *deck transformation group*. A *fundamental domain* of S is a simply connected domain, which intersects each orbit of the deck transformation group only once. Figure 1 shows a fundamental domain and a finite portion of the universal covering space of a genus-3 surface. The canonical homotopy group basis are computed using the algorithm in [8].

2.2 Hyperbolic Uniformization and Ricci Flow

Two Riemannian metrics are conformal, if they differ by a scalar function. If any surface admits a conformal Riemannian metric of constant Gaussian curvature, such metric is called the *uniformization metric*. The uniformization metric for a genus $g > 1$ surface induces -1 Gaussian curvature [10], which is called the hyperbolic metric of the surface. The hyperbolic metric can be computed by the *Ricci flow method*. Intuitively, a surface Ricci flow is a process to deform the Riemannian metric of the surface, in proportion to Gaussian curvatures, such that the curvature evolves in the same manner as heat diffusion:

$$\frac{dg_{ij}}{dt} = -2Kg_{ij} + \frac{\chi(S)}{A},$$

where K is the Gaussian curvature induced by the metric $\mathbf{g}(t)$, A is the area of the surface. For closed surfaces with non-positive Euler numbers χ , Hamilton proved the convergence of Ricci flow in [10].

The universal covering space of S with the hyperbolic metric can be isometrically embedded in the hyperbolic space \mathbb{H}^2 , here we use *Poincaré disk model*. All the deck transformations are Möbius transformation, called *Fuchsian transformation*. The deck transformation group is called the *Fuchsian group* of S .

2.3 Geodesic Spectrum

Each homotopy class has a unique closed geodesic under the hyperbolic metric. The lengths of all such geodesics form the *geodesic spectrum* of the surface. Geodesic spectrum determines the hyperbolic metric completely.

Theorem 1 (Geodesic Spectrum). *Suppose (S_1, \mathbf{g}_1) and (S_2, \mathbf{g}_2) are surfaces with hyperbolic metrics. Let $f : S_1 \rightarrow S_2$ be a diffeomorphism, such that for each homotopy class $[\gamma] \in \pi_1(S_1, p)$, it is mapped to a homotopy class $[f(\gamma)] \in \pi_1(S_2, f(p))$. If the length of the geodesic in $[\gamma]$ equals to the length of the geodesic in $[f(\gamma)]$, for all homotopy classes, then f is an isometry.*

Each homotopy class corresponds to a unique Fuchsian transformation. The length of the unique geodesic in the class can be explicitly computed by the corresponding Fuchsian transformation. If two surfaces are conformal equivalent, $f : (S_1, \mathbf{g}_1) \rightarrow (S_2, \mathbf{g}_2)$ is the conformal mapping. Then under their uniformization hyperbolic metrics, $f : (S_1, \tilde{\mathbf{g}}_1) \rightarrow (S_2, \tilde{\mathbf{g}}_2)$ is an isometry. Therefore, the two surfaces have the same hyperbolic geodesic spectrum. Geodesic spectrum is the key to measure the similarity of the conformal structures of the high-genus surfaces.

In the above, the notations g , \mathbf{g} , and g_{ij} are used to denote the concepts of surface genus, Riemannian metric and component of metric tensor, respectively.

3 Algorithms

The main goal of our algorithm is to compute the hyperbolic geodesic spectrum of a high-genus closed surface for shape analysis. The pipeline is as follows: 1) canonical homotopy group generators [8], 2) hyperbolic metric with Ricci flow, 3) Fuchsian group generators, and 4) geodesic spectrum.

Discrete Hyperbolic Ricci Flow. The computation of the hyperbolic metric on a triangular mesh M is based on the discrete hyperbolic Ricci flow. We will briefly describe the algorithm on triangular mesh. Details can be found in [4] and [12].

1. Assign a circle at vertex v_i with radius r_i ; For each edge $[v_i, v_j]$, two circles intersect at an angle ϕ_{ij} , called edge weight.
2. The edge length l_{ij} of $[v_i, v_j]$ is determined by the hyperbolic cosine law,

$$\cosh l_{ij} = \cosh r_i \cosh r_j + \sinh r_i \sinh r_j \cos \phi_{ij}.$$

3. The angle θ_i^{jk} , related to each corner $i\angle_j^k$, is determined by the current edge lengths with the inverse hyperbolic cosine law.
4. Compute the discrete Gaussian curvature K_i of each vertex v_i :

$$K_i = \begin{cases} 2\pi - \sum_{f_{ijk} \in F} \theta_i^{jk}, & \text{interior vertex} \\ \pi - \sum_{f_{ijk} \in F} \theta_i^{jk}, & \text{boundary vertex} \end{cases}$$

where θ_i^{jk} represents the corner angle attached to vertex v_i in the face f_{ijk} .

5. Update the radius r_i of each vertex v_i : $r_i = r_i - \epsilon K_i \sinh r_i$.
6. Repeat the step 2 through 5, until $\|K_i\|$ of all vertices are less than the user-specified error tolerance.

Computing Fuchsian Group Generators. The Fuchsian group generators are obtained in the following way:

1. Slice M open along a set of canonical homology basis $a_1, b_1, a_1^{-1}, b_1^{-1}, a_2, b_2, a_2^{-1}, b_2^{-1}$ to form the fundamental domain \bar{M} . See Figure 1(a).
2. Embed the fundamental domain of M into Poincaré disk with boundary segments $\partial\bar{M} = a_1 b_1 a_1^{-1} b_1^{-1} a_3 b_3 a_3^{-1} b_3^{-1}$. See Figure 1(b).
3. Compute the unique Möbius transformation (hyperbolic rigid motion) that maps a_i to a_i^{-1} on \bar{M} , which is β_i ; compute α_j which maps b_j^{-1} to b_j .

Then we get the canonical Fuchsian group generators $\{\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3\}$. Figure 1(c) shows a finite portion of the universal covering space of a genus-3 VS surface embedded in the Poincaré disk. Different fundamental domains are encoded by different colors.

Computing Geodesic Spectrum. The fundamental group generators are $\{a_1, b_1, a_2, b_2, a_3, b_3\}$. Any homotopy class is represented as a word in $\pi_1(S, q)$, $w = \omega_1 \omega_2 \cdots \omega_n$, the corresponding Fuchsian transformation is given by $\tau = \tau_1 \tau_2 \cdots \tau_n$, where ω_j is replaced by τ_j . If ω_j equals $a_k^{\pm 1}$ (or $b_k^{\pm 1}$), then τ_j is $\alpha_k^{\pm 1}$ (or $\beta_k^{\pm 1}$). Then we use the upper half plane model of \mathbb{H}^2 . τ is the product of the matrices of τ_j . The length of the unique geodesic loop in the class of w is given by the trace of the matrix representation of τ ,

$$2 \cosh\left(\frac{l}{2}\right) = \text{trace}(\tau).$$

Therefore, the geodesic spectrum can be efficiently computed in an explicit form. Figure 2 visualizes the computing process of the geodesic loops for canonical homotopy group generators.

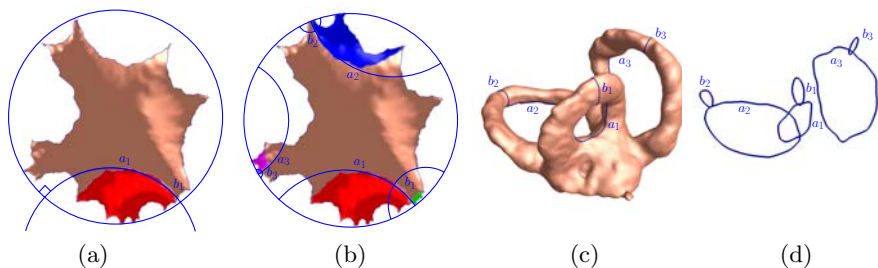


Fig. 2. Computing the geodesic loops for each homotopy class of a VS surface. (a) The unique geodesic loop in the homotopy class $[a_1]$ is the axis of the Möbius transformation α_1 on UCS, which is a circular arc orthogonal to the unit circle on the Poincaré disk. (b) All the geodesics loops form the homotopy group basis on the Poincaré disk. (c) The geodesic loops on the original surface. (d) The geodesic loops in \mathbb{R}^3 .

4 Experiments

Subjects and image acquisition. 15 girls with right-thoracic AIS (mean age 15 years old; mean Cobb's angle 27.27 degrees) and 12 age-matched healthy girls participated in this study. All participants in this study underwent a T2-weighted MRI scanning of the inner ear using 1.5T MR Scanner (Sonata, Siemens, Erlangen, Germany) with a quadrature head coil. High-spatial resolution anatomic imaging was performed with a 3D constructive interference steady state (CISS) sequence with $TR = 11.94msec$, $TE = 5.97msec$, flip angle=70 degrees, $FOV = 130mm$, slice thickness = $1mm$, no gap, matrix = 320×288 , number of excitation = 1. The sequence yielded high quality T2-weighted images with the voxel size of $0.5 \times 0.4 \times 1.0mm^3$. After on-site image interpolation, the voxel size was adjusted to $0.2 \times 0.2 \times 1.0mm^3$.

Segmentation of the vestibular system. 3D segmentation of vestibular system from T2-MRI in this study was achieved by a tailor-made pipeline proposed in [20]. Triangle mesh of the vestibular system was constructed using the marching cube algorithm. As it was found that the left-side vestibular system is abnormal in AIS compared to normal controls, in this study, our analysis is focused on the left-side vestibular system.

Computation of geodesic spectrum. The geodesic spectrum is intrinsic to the surface geometry and invariant to the triangulation resolution of the mesh. The algorithms are tested on a laptop with 2.0GHz CPU and 3.0GB RAM. The genus-3 VS surfaces are with $50k$ triangles. The computation of hyperbolic metric based on Ricci flow took 20 seconds; the computation for Fuchsian group generators and geodesic spectrum took 5 seconds. The whole pipeline is fully automatic without any human intervention.

Geodesic spectrum for shape analysis. We have done statistical analysis to study the structural difference of VS surfaces between the normal and AIS groups.

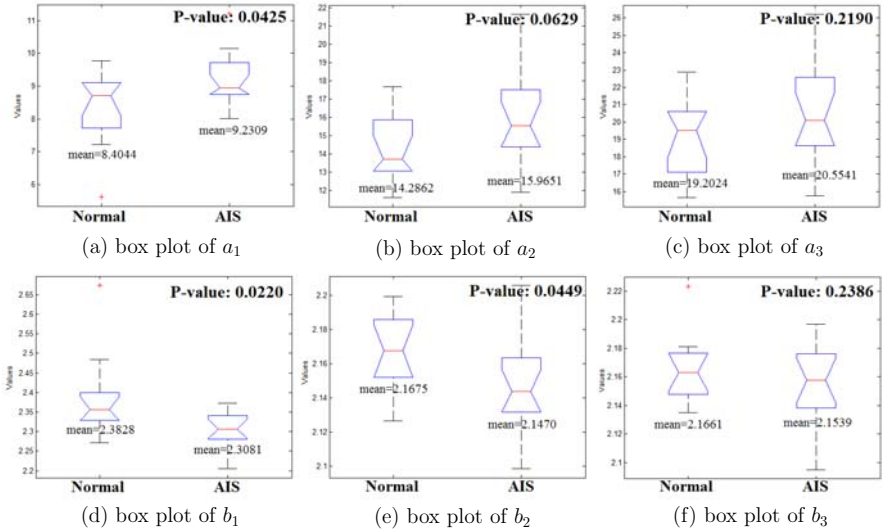


Fig. 3. Distribution of geodesic spectra for the normal and AIS groups

The geodesic spectrum $(a_i, b_i)_{i=1}^3$ on each VS surface from the normal and AIS groups is computed. The statistical difference in the geodesic spectra between groups are evaluated using t tests. The box plots of a_i 's and b_i 's are plotted in Figure 3, which shows the distribution of the geodesic spectra of each group. The AIS group tends to have higher a_i 's and smaller b_i 's. Intuitively, a higher a_i means a longer canal whereas a smaller b_i means a thinner canal. The shape difference in lateral canal (a_1, b_1) between two groups tends to be more statistically significant with P -value < 0.05 . In particular, the AIS group tends to have smaller b_1 and b_2 with high statistical significance. It means the lateral and superior canals are generally thinner for AIS subjects. We also studied the difference in shape index using different combination of a_i and b_i between groups, and their statistics are shown in Table 1. The ratio a_1/b_1 tends to be larger in the AIS group with high statistical significance ($P = 0.0338$). It means the conformal modules of the lateral canal are significantly different. This implies once again that the shape in the lateral canal tends to be significantly different between the two groups.

Table 1. Statistics on geodesic spectra between the normal and AIS groups

Combination	Mean(Normal)	Mean(AIS)	P-value
$b_1 + b_2 + b_3$	6.7164 ± 0.1103	6.6091 ± 0.0641	0.0040
$a_1 + a_2 + a_3$	41.8930 ± 4.1591	45.7500 ± 5.5327	0.0563
$a_1 + b_1 + a_2 + b_2 + a_3 + b_3$	48.6094 ± 4.0848	52.3591 ± 5.4889	0.0603
a_1/b_1	3.5514 ± 0.6145	4.0070 ± 0.4397	0.0338
a_2/b_2	6.5982 ± 0.8969	7.4461 ± 1.2299	0.0568
a_3/b_3	8.8719 ± 1.0674	9.5539 ± 1.5345	0.2036

5 Conclusion

We present a rigorous algorithm which computes the global feature, called the geodesic spectrum, for analyzing high-genus shapes. The key is to compute the canonical hyperbolic geodesic loops of the surface, using the Ricci flow method. The experiments on the morphometry study of the vestibular system (VS) of 12 normal and 15 AIS subjects show the effectiveness of our algorithm and reveal statistical shape difference in the VS between right-thoracic AIS and normal subjects. The performance demonstrates the great potential for AIS disease detection. In general, our proposed model is stable and can be applied to shape analysis for general anatomical structures of any genus.

Acknowledgments

Xianfeng Gu is supported by NIH grant 1R01EB0075300A1, NSF IIS 0916286, CCF0916235, CCF0830550, III0713145, and ONR N000140910228.

References

1. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K., Bg, L.W.: Identifying global anatomical differences: Deformation-based morphometry. *Human Brain Mapping* 6(5-6), 348–357 (1998)
2. Bradshaw, A., Curthoys, I., Todd, M., Magnussen, J., Taubman, D., Aw, S., Halmagyi, G.: A mathematical model of human semicircular canal geometry: a new basis for interpreting vestibular physiology. *J. Assoc. Res. Otolaryngol.* 11(2), 145–159 (2010)
3. Byl, N., Gray, J.: Complex balance reactions in different sensory conditions: adolescents with and without idiopathic scoliosis. *J. Orthop. Res.* 11(2), 215–227 (1993)
4. Chow, B., Luo, F.: Combinatorial Ricci flows on surfaces. *Journal of Differential Geometry* 63(1), 97–129 (2003)
5. Chung, M.K., Dalton, K.M., Davidson, R.J.: Tensor-based cortical morphometry via weighted spherical harmonic representation. *IEEE Transactions on Medical Imaging* 27(8), 1143–1151 (2008)
6. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis - I. segmentation and surface reconstruction. *Neuroimage* 9, 179–194 (1999)
7. Davies, R.H., Twining, C.J., Allen, P.D., Cootes, T.F., Taylor, C.J.: Shape discrimination in the hippocampus using an MDL model. In: Taylor, C.J., Noble, J.A. (eds.) *IPMI 2003. LNCS, vol. 2732*, pp. 38–50. Springer, Heidelberg (2003)
8. Erickson, J., Whittlesey, K.: Greedy optimal homotopy and homology generators. In: *SODA 2005: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1038–1046 (2005)
9. Guo, X., Chau, W., Hui-Chan, C., Cheung, C., Tsang, W., Cheng, J.: Balance control in adolescents with idiopathic scoliosis and disturbed somatosensory function. *Spine* 31, 437–440 (2006)
10. Hamilton, R.S.: The Ricci flow on surfaces. *Mathematics and general relativity* 71, 237–262 (1988)
11. Herman, R., Mixon, J., Fisher, A., Maulucci, R., Stuyck, J.: Idiopathic scoliosis and the central nervous system: a motor control problem. *Spine* 10(1) (1985)

12. Jin, M., Kim, J., Luo, F., Gu, X.: Discrete surface Ricci flow. *IEEE Transactions on Visualization and Computer Graphics* 14(5), 1030–1043 (2008)
13. Jin, M., Zeng, W., Luo, F., Gu, X.: Computing Teichmüller shape space. *IEEE Transactions on Visualization and Computer Graphics* 15(3), 504–517 (2009)
14. Lambert, F.M., Malinvaud, D., Glaunés, J., Bergot, C., Straka, H., Vidal, P.-P.: Vestibular asymmetry as the cause of idiopathic scoliosis: A possible answer from *Xenopus*. *J. Neuroscience* 29, 12477–12483 (2009)
15. Pizer, S., Fritsch, D., Yushkevich, P., Johnson, V., Chaney, E.: Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transaction on Medical Imaging* 18(10), 851–865 (1999)
16. Reuter, M., Wolter, F.-E., Peinecke, N.: Laplace-beltrami spectra as “shape-dna” of surfaces and solids. *CAD* 38(4), 342–366 (2006)
17. Reuter, M., Wolter, F.-E., Shenton, M., Niethammer, M.: Laplace-beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *CAD* 41(10), 739–755 (2009)
18. Munkres, J.R.: *Elements of Algebraic Topology*. Addison-Wesley Co., Reading (1984)
19. Schoen, R., Yau, S.-T.: *Lectures on Differential Geometry*. International Press, Boston (1994)
20. Shi, L., Wang, D., Chu, W., Heng, P., Burwell, R., Cheng, J.: Statistical morphometry of the vestibular system in adolescent idiopathic scoliosis. In: 6th Annual World Congress for Brain Mapping and Image Guided Therapy. Harvard Medical School, Boston (2009)
21. Simoneau, M., Lamothe, V., Hutin, E., Mercier, P., Teasdale, N., Blouin, J.: Evidence for cognitive vestibular integration impairment in idiopathic scoliosis patients. *BMC Neuroscience* 10(102) (2009)
22. Simoneau, M., Richer, N., Mercier, P., Allard, P., Teasdale, N.: Sensory deprivation and balance control in idiopathic scoliosis adolescent. *Experimental Brain Research* 170, 576–582 (2005)
23. Zeng, W., Samaras, D., Gu, X.D.: Ricci flow for 3D shape analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(4), 662–677 (2010)

Value-Based Noise Reduction for Low-Dose Dual-Energy Computed Tomography

Michael Balda¹, Björn Heismann^{1,2}, and Joachim Hornegger¹

¹ Pattern Recognition Lab, Friedrich-Alexander University, Erlangen, Germany

² Siemens Healthcare, Erlangen, Germany

Abstract. We introduce a value-based noise reduction method for Dual-Energy CT applications. It is based on joint intensity statistics estimated from high- and low-energy CT scans of the identical anatomy in order to reduce the noise level in both scans. For a given pair of measurement values, a local gradient ascension algorithm in the probability space is used to provide a noise reduced estimate. As a consequence, two noise reduced images are obtained. It was evaluated with synthetic data in terms of quantitative accuracy and contrast to noise ratio (CNR)-gain. The introduced method allows for reducing patient dose by at least 30% while maintaining the original CNR level. Additionally, the dose reduction potential was shown with a radiological evaluation on real patient data. The method can be combined with state-of-the-art filter-based noise reduction techniques, and makes low-dose Dual-Energy CT possible for the full spectrum of quantitative CT applications.

1 Introduction

Dual-Energy CT (DECT) measures two image-sets at different energy weightings, e.g. by performing two scans with tube voltages set to 80kV and 140kV respectively. Alvarez [1] and Macovski [2] introduced a reconstruction technique for multi-energy scans based on a decomposition of the spectral attenuation coefficient into basis functions. Dual-Energy applications can yield valuable information for intervention planning and diagnosis. The most popular current Dual-Energy CT diagnostic applications are bone removal [3], PET/SPECT attenuation correction [4], lung perfusion diagnostic or quantification of contrast agent concentrations, for instance in the myocardium.

It is commonly agreed that the two dual-energy scans should require about the same total X-ray dose as the corresponding single-energy scan. This calls for effective noise reduction techniques in Dual-Energy CT. The most common filtering strategy applied in medical CT scanners is modifying the high-pass reconstruction kernel used for filtered back-projection (FBP) in a way that high frequencies are less amplified or blocked. Additionally, adaptive filters are applied occasionally. This type of filters steers the filter strength according to a noise estimate. It causes less smoothing in regions where noise is low, i.e. the X-rays are weakly attenuated. More sophisticated edge preserving filters have also been investigated. Weickert [5] introduced an edge-preserving anisotropic diffusion

filter which can be adapted to CT data as shown in [6]. Bilateral filtering [7] tries to achieve a similar goal by combining frequency- and intensity-based smoothing. These filters steer the smoothing locally according to distance and similarity of neighboring intensity values. In CT, frequency based noise reduction filters are usually applied in the projection domain as the spectral noise properties in the CT-image or -volume domain can hardly be derived analytically. The image noise is inhomogeneous and non-stationary and estimating local noise properties is complex, [8] shows how to estimate local variance and analyze noise correlations for filtering CT-data in the image domain.

We introduce a noise reduction technique that is solely based on joint intensity statistics of the two Dual-Energy datasets. Most multi-energy modalities have an unequal noise distribution between images due to tube limitations and/or strong absorption in low-kVp images. The individual images from a multi-energy scan are much noisier than an image of a single energy scan, as the overall scan dose should be in an identical range. Quantitative CT (QCT) applications based on dual energy data are very sensitive to image noise. Image-based basis material decomposition (BMD) [9], for instance, causes noise amplification in the resulting basis material coefficient images. For other QCT applications, like Rho-Z projection [10], similar problems arise.

The proposed Dual-Energy noise reduction improves images from one energy weighting by using knowledge on the joint intensities from both datasets. As it is purely value-based and does not utilize any frequency information, it is compatible with the frequency-based filters explained above. This method operates on the reconstructed images, so it can be applied to any reconstructed DECT data-set. Since it is a post-reconstruction method, it may not enhance the quantitative correctness of the reconstructed images and relies on the correctness of the DECT reconstruction and beam hardening correction method.

2 Method

As input data, the noise-reduction method uses a low-kVp CT-volume $\bar{\mu}_1(\mathbf{x})$ and a high-kVp volume $\bar{\mu}_2(\mathbf{x})$. The voxel coordinate is indicated by \mathbf{x} . This method is neither restricted in terms of the number of multi-energy input datasets nor their dimension. However, for practical reasons, the description focuses on Dual-Energy data of two or three spatial dimensions. Our method estimates the most likely true object attenuation values $\tilde{\mu}_1(\mathbf{x})$ and $\tilde{\mu}_2(\mathbf{x})$ for each measured $(\bar{\mu}_1(\mathbf{x}), \bar{\mu}_2(\mathbf{x}))$ -pair by a gradient ascent in the joint histogram.

Joint histogram computation: The first step is estimating the joint probability density $P(\bar{\mu}_1, \bar{\mu}_2)$ for all $(\bar{\mu}_1(\mathbf{x}), \bar{\mu}_2(\mathbf{x}))$ -pairs. Several methods are available for this purpose, e.g. computing histograms, data clustering or Parzen windowing [11]. We use kernel density estimation using a bivariate, uncorrelated Normal Distribution as kernel function. This method offers the possibility to apply smoothing to the $P(\bar{\mu}_1, \bar{\mu}_2)$ -estimate by adjusting the bandwidth of the kernel:

$$\mathcal{N}_{(\sigma_1, \sigma_2)}(\bar{\mu}_1, \bar{\mu}_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2} \left(\frac{\bar{\mu}_1^2}{\sigma_1^2} + \frac{\bar{\mu}_2^2}{\sigma_2^2} \right) \right\}. \quad (1)$$

The standard deviation (σ_1, σ_2) is the bandwidth parameter. Slightly oversmoothed density estimates yield better noise suppression at the cost of increased bias. In the following, the selected bandwidth parameter set is denoted (b_1, b_2) . Once a set of bandwidth parameters is selected, the density estimate $P_{b_1, b_2}(\bar{\mu}_1, \bar{\mu}_2)$ can be computed by a convolution.

Noise estimation: The next step is to estimate true attenuation values for each pair of measured values by analyzing $P_{b_1, b_2}(\bar{\mu}_1, \bar{\mu}_2)$. In order to find the most probable true $\bar{\mu}_1$ -value for $(\bar{\mu}_1(\mathbf{x}), \bar{\mu}_2(\mathbf{x}))$, the $\bar{\mu}_2(\mathbf{x})$ -value is kept fixed and a gradient ascent along the $\bar{\mu}_1$ -direction within the density estimate is performed.

The gradient direction is given by

$$G_1(\bar{\mu}_1, \bar{\mu}_2) = \left(\text{sgn} \left\{ \frac{d}{d\bar{\mu}_1} (P_{(b_1, b_2)}(\bar{\mu}_1, \bar{\mu}_2)) \right\}, 0 \right). \tag{2}$$

Figure 1 shows an example for this procedure. The resulting local maximum for $\bar{\mu}_1(\mathbf{x})$ is called $\tilde{\mu}_1(\mathbf{x})$. It represents the most probable $\bar{\mu}_1$ -value for the measured intensity pair with fixed $\bar{\mu}_1(\mathbf{x})$. The distance $d_1(\mathbf{x}) = |\tilde{\mu}_1(\mathbf{x}) - \bar{\mu}_1(\mathbf{x})|$ is an estimate for the noisiness of $\bar{\mu}_1(\mathbf{x})$. If the actually measured value $\bar{\mu}_1(\mathbf{x})$ is very close to the most probable value, it is likely to be less noisy. The process of finding the $\tilde{\mu}_2$ -value works analogously: The gradient ascent is performed in $\bar{\mu}_2$ -direction. This leads to the estimates $\tilde{\mu}_2(\mathbf{x})$ and $d_2(\mathbf{x})$.

Parameter reduction: The method has two bandwidth parameters of the bivariate Gaussians. These can be reduced to b_1 as the only free parameter and automatically compute an appropriate b_2 value for the $\bar{\mu}_2$ -image by taking the relative contrast and noise of the $\bar{\mu}_2$ -image into account:

We generate a set \mathcal{X} of all soft tissue voxel positions in image 1 by thresholding. The standard deviations $\sigma_{1, \mathcal{X}}$ and $\sigma_{2, \mathcal{X}}$ are computed for \mathcal{X} in both images. The standard deviation for the second image $\sigma_{2, \mathcal{X}}$ can differ from $\sigma_{1, \mathcal{X}}$ for two reasons: A different noise level and/or a different tissue contrast in the second image. A lower noise level or less contrast require a smaller bandwidth. A higher bandwidth can be chosen if the noise level lower or the tissue contrast is higher. Consequently we set $b_2 = b_1 \cdot \frac{\sigma_{2, \mathcal{X}}}{\sigma_{1, \mathcal{X}}}$. For the example of a second image with a higher tube acceleration voltage at the same tube current, a lower noise level and reduced tissue contrast is observed. This case leads to a reduced bandwidth b_2 . If the noise level is similar in both images, the reduced tissue contrast prohibits strong noise reduction on the high-kVp image.

Noise reduction: The quality of the estimates $\tilde{\mu}_1(\mathbf{x})$ and $\tilde{\mu}_2(\mathbf{x})$ is affected by noise in the original intensity pairs $(\bar{\mu}_1(\mathbf{x}), \bar{\mu}_2(\mathbf{x}))$: High noise in $\bar{\mu}_1(\mathbf{x})$ and $\bar{\mu}_2(\mathbf{x})$ generally worsens both estimates, low noise in the $\bar{\mu}_1(\mathbf{x})$ -value increases the quality of the estimate $\tilde{\mu}_2(\mathbf{x})$ and vice versa. The noise-reduction process should use a measure that takes these properties into account and adjusts the strength of the correction process accordingly. The following measure offers these properties and allows governing the strength of the filter by a single parameter:

$$\hat{\mu}_1(\mathbf{x}) = \alpha(\mathbf{x}) \cdot \bar{\mu}_1(\mathbf{x}) + (1 - \alpha(\mathbf{x})) \cdot \tilde{\mu}_1(\mathbf{x}) \tag{3}$$

The correction strength is defined by a parameter α , the final noise-reduced image value is called $\hat{\mu}_1(\mathbf{x})$; α is computed as follows:

$$\alpha(\mathbf{x}) = \left(\frac{d_1(\mathbf{x})}{d_t(\mathbf{x})} \right)^r \quad \text{with } d_t(\mathbf{x}) = \sqrt{d_1(\mathbf{x})^2 + d_2(\mathbf{x})^2} \quad (4)$$

The parameter r determines the correction strength, values of $r = [0, 1]$ generally favor the $\tilde{\mu}_1(\mathbf{x})$ - resp. $\tilde{\mu}_2(\mathbf{x})$ -estimates whereas $r =]1, +\infty[$ directs the filter to use the estimates only when the image noise estimate is close to the overall noise estimate $d_t(\mathbf{x})$. In our experiments, we used $r = 5$.

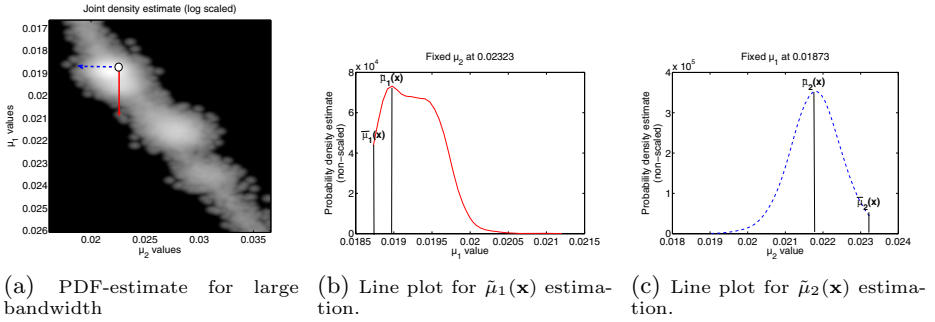


Fig. 1. Examples for the gradient ascent procedure with $\mu_1(\mathbf{x}) = 0.0187 \text{ mm}^{-1}$ and $\mu_2(\mathbf{x}) = 0.0232 \text{ mm}^{-1}$. Arrows indicate the gradient direction, the start point is located at the base of the arrows.

3 Evaluation

In order to evaluate the theoretical limits of the proposed denoising method against ground truth data, we conducted several simulations and a radiological evaluation.

1. *Contrast-to-noise ratio*: Evaluation of the achievable CNR-gain for various contrasts and dose-levels.
2. *Quantitative accuracy and precision*: Noise reduction and quantitative accuracy for energy calibration application using basis material decomposition.
3. *Radiological evaluation*: Perceived and measured noise reduction while maintaining visibility of important image details.

The simulations involve an analytic forward projection of a geometrically defined thorax phantom and a standard filtered back-projection (see Fig. 2). The tissue compositions were taken from the ICRU Report 46¹ and the elemental mass attenuation coefficients from [12]. In order to avoid beam hardening artifacts, we created dual-energy data-sets with mono-chromatic radiation at 54 keV and 73 keV which corresponds to the effective energy of 80 kVp and 140 kVp scans.

¹ International Commission on Radiation Units and Measurements: Report 46. Photon, Electron, Proton and Neutron Interaction Data for Body Tissues, 1998.

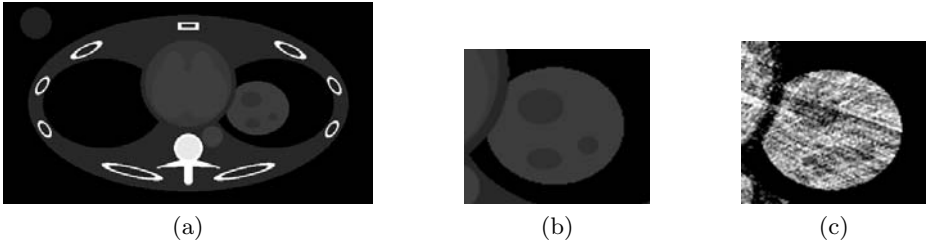


Fig. 2. (a) Thorax phantom (with a small water cylinder at the top-left corner used for water scaling); (b) Low contrast lesions for CNR evaluation. (c) Lesion example at 80 kVp with 14 HU contrast and 70000 primary photons (intensity window center: 45 HU, width: 35 HU).

Contrast-to-noise ratio: The possible improvement in contrast-to-noise ratio (CNR) of the data was evaluated by placing various small lesions inside the liver (see Fig. 2b). Four different contrasts were evaluated at five noise levels. The liver and cirrhotic liver tissue composition was chosen according to the ICRU report 46 with a density of $1.060 \frac{\text{g}}{\text{cm}^3}$ for the healthy liver tissue. The liver lesion densities were set to 1.040, 1.045, 1.050 and $1.055 \frac{\text{g}}{\text{cm}^3}$. The resulting contrasts are 29, 24, 19 and 14 HU for the effective 80 kV spectrum and 23, 18, 13 and 8 HU for the 140 kV spectrum. The ratio of quanta in the low and high energy spectra was kept fix at 3 to 1. Six different bandwidths were selected for the parameter b_1 : 0.5, 1.0, 2.0, 4.0, 7.0 and 11.0 HU. The according b_2 -values were determined automatically. Figure 3 shows an excerpt of the resulting CNR-values. The results show an improvement in all cases as long as an appropriate bandwidth parameter is chosen. The choice of a too large bandwidth can, however, decrease the resulting CNR for very low contrasts. The possible dose reduction can be deduced by comparing the number of primary photons needed to get a CNR value similar to the original one. In the evaluated cases, 70% to 40% of the original photon numbers yield similar CNR values. In general, the CNR gain is larger in the 140 kV cases, since the tissue contrast in the 140 kV images is smaller, so even small improvements with respect to noise have a large impact in terms of CNR.

Quantitative accuracy and precision: For this evaluation we performed a two-material BMD with the basis materials water and femur bone on the thorax phantom introduced above. The liver lesions were removed and the medium noise case was selected with $1.4 \cdot 10^5$ primary photons for the effective 80 kVp spectrum. Here we use the BMD to estimate the spectral attenuation coefficients $\mu(E, \mathbf{x})$ at every voxel position \mathbf{x} . These coefficients are then weighted with an effective 120 kVp spectrum ($w_{120\text{kVp}}(E)$) to create a virtual 120 kVp image from the input images by computing $\bar{\mu}_{120\text{kVp}}(\mathbf{x}) = \int_0^\infty w_{120\text{kVp}}(E)\mu(E, \mathbf{x})dE$. The resulting image is compared with the analytically computed ground truth data in terms of mean-shift and standard deviation. The value-based noise reduction may cause a minor shift of the mean attenuation values along with noise reduction. This evaluation is meant to quantify the trade-off between noise reduction and decrease in quantitative accuracy caused by the mean shift.

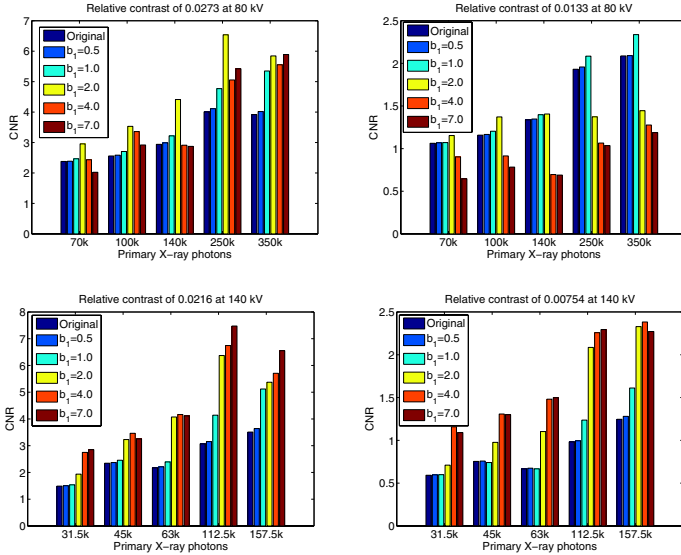


Fig. 3. CNR test results for different contrasts at 80 kVp and 140 kVp tube voltage and selected noise reduction parameters.

Table 1. Ground truth and calculated mean and standard deviation of attenuation values for virtual 120 kVp image from original and processed images. All quantities given in mm^{-1} . The values show a very small bandwidth dependence of the quantitative accuracy (mean values). The noise (standard deviation) decreases with increasing bandwidth.

Bandwidth	Average soft tissue	Healthy liver	Red Marrow
Ground Truth	2.072e-2	2.156e-2	2.050e-2
Original	2.060e-2 (1.605e-4)	2.140e-2 (1.655e-4)	2.039e-2 (3.711e-4)
1.08e-5	2.060e-2 (1.583e-4)	2.140e-2 (1.647e-4)	2.040e-2 (3.714e-4)
4.33e-5	2.059e-2 (1.286e-4)	2.140e-2 (1.526e-4)	2.044e-2 (3.643e-4)
8.66e-5	2.058e-2 (0.859e-4)	2.139e-2 (1.001e-4)	2.051e-2 (3.204e-4)
1.52e-4	2.059e-2 (1.002e-4)	2.138e-2 (1.094e-4)	2.054e-2 (2.665e-4)

Table 1 shows the results at different bandwidths for the tissues average soft tissue (large area), liver (medium area) and red marrow (small area). Noise suppression is achieved for all tissues at a tolerable mean shift. At lower bandwidths the mean shift is almost negligible. At very high bandwidths the noise standard deviation can get worse in some cases since the gradient ascent tends to the wrong direction for some voxels.

Radiological evaluation: Images from seven different Dual-Energy scans were presented to three radiologists in a typical clinical environment. We selected data from typical examinations of different body regions with and without contrast agents and varying scan parameters. The input images were all Dual-Source

images acquired with a Siemens Definition or a Siemens Definition Flash CT-scanner (Siemens AG, Forchheim, Germany). The Siemens Definition device uses 80 kVp and 140 kVp tube voltages, whereas the Definition Flash uses 100 kVp and 140 kVp with an additional tin-filter for the high-energy spectrum that improves spectral separation. The images were presented in randomly ordered pairs of original and denoised images with different bandwidth settings. The radiologists were asked to compare the image-pairs in terms of perceived noise level and visibility of important details. We evaluate to which extent the perceived and measured noise level can be reduced without important structures being visibly weakened compared to the original images. The corresponding bandwidth parameter for each dataset is called *optimal bandwidth* in the following. Table 2 lists the average measured noise reduction for the optimal bandwidth on all datasets. All radiologists agreed that using the optimal bandwidth, a noise reduction could be perceived in all Dual-Energy datasets.

Table 2. Average noise reduction for several different test scenarios with optimal bandwidth setting. First number corresponds to low-kVp image, second to high-kVp. Image noise was determined by evaluating the standard deviation in homogeneous image regions. (*) only 80kV image was evaluated for the "Foot" dataset.

Dataset	Head	Foot*	Liver	Lung
Noise reduction	-27% / -25%	-33%*	-29% / -25%	-29% / -25%
	Abdomen	Lower Abdomen	Pelvis	
	-20% / -32%	-24% / -18%	-22% / -22%	

4 Conclusion

The introduced noise-reduction technique for Dual-Energy CT data showed the potential of reducing image noise by about 20 to 30% in subjective tests on real CT-images of various body regions. The quantitative accuracy was evaluated in an energy calibration application on phantom data. It revealed a tolerable accuracy loss of approx. 1 to 2% which is in the range of the overall accuracy of a CT system. The noise standard deviation could be reduced by approx. 30% for this application. A CNR test for small, low-contrast lesions yielded a CNR-gain ranging from 10% to over 100% depending on noise level, tube voltage setting and tissue contrast. The CNR values of the original image could be reproduced with 30% up to 70% less primary photons needed. This is directly related to an accordingly reduced radiation dose.

References

1. Alvarez, R.E., Macovski, A.: Energy-selective Reconstructions in X-ray Computerized Tomography. *Phys. Med. Biol.* 21(5), 733–744 (1976)
2. Macovski, A., Alvarez, R.E., Chan, J.L.H., Stonestrom, J.P., Zatz, L.M.: Energy dependent reconstruction in X-ray computerized tomography. *Comput. Biol. Med.* 6(4), 325–334 (1976)

3. Zheng, L., Maksimov, D., Stutzmann, T., Jochum, S., Brockmann, C., Diehl, S., Hesser, J.: Removal of bone information in dual energy CT. In: Proceedings of the MICCAI-CVII Workshop, pp. 120–127 (2008)
4. Heismann, B.J., Balda, M.: Quantitative image-based spectral reconstruction for computed tomography. *Medical Physics* 36(10), 4471–4485 (2009)
5. Weickert, J.: *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, Stuttgart (1998)
6. Wang, J., Lu, H., Li, T., Liang, Z.: Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters, vol. 5747, pp. 2058–2066. SPIE, San Jose (2005)
7. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision (1998)
8. Borsdorf, A., Kappler, S., Raupach, R., Hornegger, J.: Analytic Noise Propagation for Anisotropic Denoising of CT Images. In: Sellin, P. (ed.) 2008 IEEE Nuclear Science Symposium Conference Record, pp. 5335–5338 (2008)
9. Taguchi, K., Zhang, M., Frey, E.C., Xu, J., Segars, W.P., Tsui, B.M.W.: Image-domain material decomposition using photon-counting CT, vol. 6510, p. 651008. SPIE, San Jose (2007)
10. Heismann, B.J., Leppert, J., Stierstorfer, K.: Density and atomic number measurements with spectral X-ray attenuation method. *J. Appl. Phys.* 94(3), 2074–2079 (2003)
11. Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Stat.* 33, 1065–1076 (1962)
12. Berger, M.J., Hubbell, J.H., Seltzer, S.M., Chang, J., Coursey, J.S., Sukumar, R., Zucker, D.S.: XCOM: Photon Cross Sections Database. NIST Standard Reference Database 8 (XGAM) (1998)

Automatic Detection of Anatomical Features on 3D Ear Impressions for Canonical Representation

Sajjad Baloch¹, Rupen Melkisetoglu¹, Simon Flöry², Sergei Azernikov¹,
Greg Slabaugh³, Alexander Zouhar¹, and Tong Fang¹

¹ Siemens Corporate Research, Princeton, NJ, USA

² Vienna University of Technology, Wien, Austria

³ Medicsight PLC, London, UK

Abstract. We propose a shape descriptor for 3D ear impressions, derived from a comprehensive set of anatomical features. Motivated by hearing aid (HA) manufacturing, the selection of the anatomical features is carried out according to their uniqueness and importance in HA design. This leads to a *canonical ear signature* that is highly distinctive and potentially well suited for classification. First, the anatomical features are characterized into *generic* topological and geometric features, namely *concavities*, *elbows*, *ridges*, *peaks*, and *bumps* on the surface of the ear. Fast and robust algorithms are then developed for their detection. This indirect approach ensures the generality of the algorithms with potential applications in biomedicine, biometrics, and reverse engineering.

1 Introduction

Shape analysis typically involves the abstraction of complex structures by removing redundant details. It captures the essence of geometry via morphological descriptors that emphasize more on informative and distinctive features of the underlying surface. Such descriptors may be employed for classification in addition to guiding registration [1] and segmentation. The problem is very challenging due to the variability of organic surfaces. Moreover, it is typically not possible to consistently identify key features solely from geometric and topological information, and the need of anatomical features becomes imminent.

We focus on the surfaces representing the external and outer human ear and construct a canonical ear signature (CES). The idea is to derive a descriptor for subsequent modeling of the anatomy of the human ear. This will eventually lead to a 3D digital human ear atlas, and a framework for the design of implants and prosthetic devices for the ear. This is in line with the recent trends in medical image analysis, where the imaging technology is exploited for data modeling, intervention planning, and corrective treatment. Our approach herein is to derive the CES from the anatomy itself, which is quite well known in the medical community. The first comprehensive atlas dates back to Gray in 1858 (Fig. 1(a), with a 3D reconstruction in Fig. 1(b)). As shown, an ear impression consists of a spindle shaped *canal* that sits deep in the outer ear and a base that resides

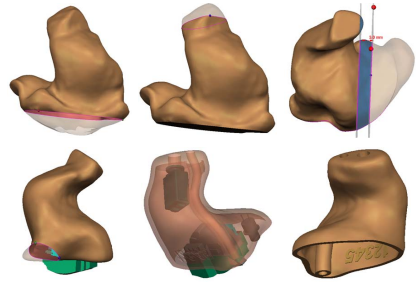
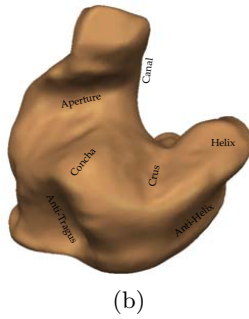
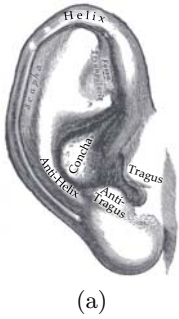


Fig. 1. The human ear anatomy (courtesy [2]) and its 3D reconstruction

Fig. 2. Shell detailing and modeling. Last item is the finished shell

in the external ear. The two are separated by a narrow opening called *aperture*. Around the bottom of the external ear, there are two clamp like structures known as *tragus* and *anti-tragus*. The deepest point in the external ear is the *concha peak* that sits on the big bowl shaped *concha*. On the top, there is a narrow *helix* structure sandwiched between the *anti-helix* and the *crus*. These anatomical parts along with others are considered for the construction of the CES. Once a 3D impression of an ear is acquired, the problem then reduces to extracting its signature representation. To this end, we adopt a generalized approach, and first characterize the anatomical features into more generic topological and geometric features, such as *peaks*, *pits*, *concavities*, *elbows*, *ridges*, and *bumps*. Fast and robust algorithms are then constructed for the detection of generic features. As a result, the application of proposed algorithms is not limited to ears. For instance, the ridge detection algorithm may readily be used for tracing the nose in face recognition, or the crest lines on human brains.

There is substantial literature on surface feature detection [3,4,7]. However, not much work has been reported on ear shape analysis. We introduce three major contributions: (a) It is the first work that comprehensively analyzes the **outer** and **external** ear shape to automatically determine the CES. (b) Due to the varied nature of the features, a collection of algorithms are proposed to produce an overall system capable of their reliable detection. These methods enable automation in the HA manufacturing (HAM) [6] and have the potential of faster and more consistent design. (c) The resulting signature will serve as a foundation to a 3D digital atlas of the human ear. In comparison to [3,7], our algorithms are fast, and guarantee protection against broken ridges. We also consider the detection of planes, which has not been attempted previously.

2 Canonical Ear Signature

First, we identify the anatomical constructs that uniquely identify the shape of an ear impression. To this end, we are inspired by the HAM application. HAs are generally custom made. Once a mesh is constructed from the 3D scan of an ear

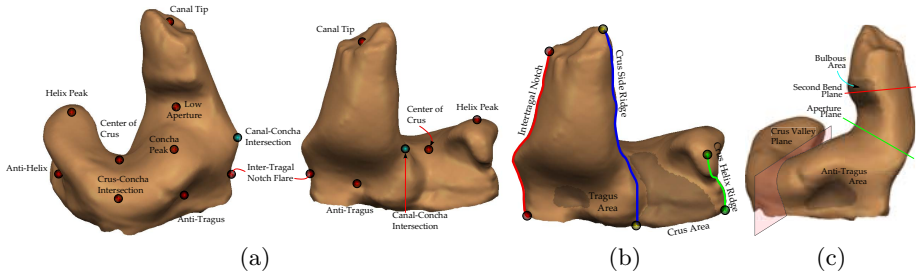


Fig. 3. Anatomical Features: (a) Points; (b) Curves and areas; (c) Planes and areas

impression, an operator carries out a sequence of surface modification operations (Fig. 2), to design an HA shell [6]. Typically, he relies on manual measurements (on sub-mm scale) implicitly using certain key anatomical features. Our hypothesis is that since an HA is designed to comfortably fit the ear(s) of a patient, the underlying features form a canonical representation of the anatomy. We have, therefore, identified a comprehensive set of features that are well known in the medical anatomy, and are implicitly used by HA designers (Fig. 3). Collectively, they capture the structure of an ear in a *Canonical Ear Signature*.

3 Generic Feature Detection

Our approach to canonical feature detection is based on generalization. We characterize various anatomical features via a set of generic features, *peaks* $\{\mathcal{F}_p\}$, *concavities* $\{\mathcal{F}_c\}$, *elbows* $\{\mathcal{F}_e\}$, *ridges* $\{\mathcal{F}_r\}$, and *bumps* $\{\mathcal{F}_b\}$. Although some anatomical features may not be represented by these geometric primitives, they may still be derived from the latter or other derived features. Algorithms are constructed for the generic features, thereby later providing a foundation for the derived features. An advantage of such a hierarchical approach is that the generic algorithms may readily be applied to other applications such as biometrics, and reverse engineering, without compromising on the complexity. The relationship between generic and anatomical features is tabulated in Table II.

3.1 Problem Formulation

Let \mathcal{M} be a 2D manifold representing a 3D ear impression embedded in \mathbb{R}^3 locally parameterized as $\phi : \Omega \rightarrow \mathcal{M}$, where $\Omega \subset \mathbb{R}^2$ represents the parameter space. The problem is to detect a set of generic features $\mathcal{F} = \{\mathcal{F}_p \cup \mathcal{F}_c \cup \mathcal{F}_e \cup \mathcal{F}_r \cup \mathcal{F}_b\}$.

3.2 Peak Detection

A peak point is a prominent topological landmark on a surface. The basic idea for its detection emanates from *Morse theory*, and involves defining a smooth real valued function $f : \mathcal{M} \rightarrow \mathbb{R}$ on \mathcal{M} to abstract its shape via critical points. A point $p = \phi(u) \in \mathcal{M}$, $u \in \Omega$ is a *critical point* of f if the gradient of $f \circ \phi$ vanishes

Table 1. Generic Characterization of Anatomical Features

Characterization	Anatomical Features
Peak	Canal Tip, Concha Peak, Helix Peak
Concavity	Tragus, Anti-Tragus, Anti-Helix, Center of Crus
Elbow	First Bend, Second Bend
Ridge	Inter-Tragal Notch, Crus-Side Ridge, Crus-Helix Ridge
Bump	Canal Bulbous

at u . A critical point $p \in \mathcal{M}$ is regarded as *non-degenerate* if the Hessian $\nabla^2 f \circ \phi(u)$ is non-singular at $\phi(u)$. f is called *Morse*, if all of its critical points are non-degenerate. The problem, therefore, reduces to finding a suitable Morse function. For peak detection, we employ a *height function*, $h : \mathcal{M} \rightarrow \mathbb{R}$, that assigns to each point $p(x, y, z) \in \mathcal{M}$ a value equal to its height, $h(p) := h(x, y, z) = z$. For a non-degenerate surface, h is a Morse function and its critical points are the peaks, passes and pits of the surface. We use it for peak detection as in [4].

The algorithm for detecting the critical points of h follows from *Morse deformation lemma*, and analyzes the level sets of h (horizontal planes) for topological changes. By gradually increasing $h \in [0, H]$ in K steps, we find the intersections of the surface with corresponding planes. Intersections are subsequently analyzed for changes in topology between two successive planes. If a change in topology is detected, we notice the existence of a critical level between them, and zoom in to analyze the surface with larger K (effective $K \leftarrow K^2$, since the interval is smaller). The process is repeated until convergence to a critical point.

3.3 Concavities

Concavities are marked by depressions on a surface. For their detection, we utilize orthogonal scans on a surface to generate a surface profile that is composed of the intersection contours. Spline representations of individual contours are analyzed for variations in signed curvature, where the negative sign identifies a concavity. First, a profile in one direction is considered, and subsections of contours with negative curvature are identified. For these subsections, the points of least curvature are found, with their average computed as a seed point. This seed point is corrected by a scan, orthogonal to the previous scan, shifting it towards the lowest curvature point. Consequently, the seed point is pushed deeper in the valley. The process is repeated iteratively to achieve the absolute local minimum similar to minimization by alternating variables. Once the center of concavity is identified, region growing based on negative curvature is used to determine the concave region.

3.4 Elbows

The detection of elbows in the presence of noise and bumps is quite challenging, which limits the use of only the curvature. Our approach, first identifies points of high curvature on a surface followed by a selection/rejection strategy to fit a plane along the elbow. This leads to a robust algorithm when compared with computations based solely on the skeleton.

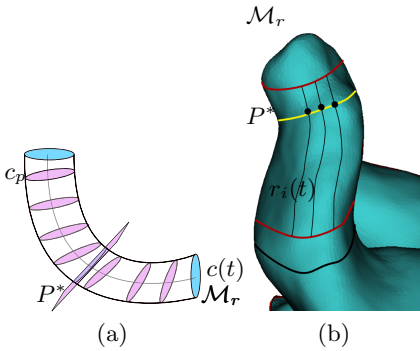


Fig. 4. Elbow Detection: (a) Scanning; (b) Two planes (shown red) define the ROI \mathcal{M}_r in the computation of t_c -sensitive elbow (yellow). Black radial lines represent the spline representations r_i .

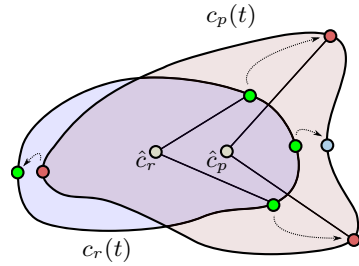


Fig. 5. Dotted lines indicate correspondence between c_p and c_r . Corresponding points (q_p, q_r) are compared via their distances from individual centroids (\hat{c}_p, \hat{c}_r). $q_p \in c_p$ is bumpy if $\|q_p - \hat{c}_p\|_2 > \|q_r - \hat{c}_r\|_2$. Red points are bumpy but not the blue.

The proposed algorithm relies on a predefined region of interest (ROI) $\mathcal{M}_r \subset \mathcal{M}$. First, \mathcal{M}_r is scanned with planes P_t oriented along the centerline $c(t) : t \in [0, 1]$, to generate its x-sectional profile $\{c_p\} = \{\mathcal{M} \cap P_t, \forall t \in [0, 1]\}$ (Fig 4(a)). The profile contours, thus, capture information about bends on the surface. Correspondences are then established among these contours along the radial direction (Fig 4(b)), and each radial contour is eventually represented by its spline parametrization $r_i(t)$ for robustness to noise. Once radial contours are parameterized, we proceed to identify points $Q = \{q_i\}$ of maximal curvature along these contours. However, not all points reliably represent an elbow, due to potential presence of bumps. The set Q is, hence, pruned via a point rejection strategy for plane fitting similar to the deterministic RANSAC to increase robustness.

3.5 Ridge Detection

We define a ridge $r : [0, 1] \rightarrow \mathcal{M}$ as a geodesic on \mathcal{M} characterized by points of high curvature, $\kappa : \mathcal{M} \rightarrow \mathbb{R}$. Note that local extrema of κ alone are not sufficient for its detection, as isolated points may exhibit high curvature due to noise. This is a major limitation of existing methods [3] in addition to their inherent inability to maintain the integrity of a ridge if not prominent, or if broken.

At times, it is also required to identify a ridge in addition to its detection. We, therefore, develop a method that first detects its starting and ending points p_s and p_e respectively, using either the geometric information and/or some a priori knowledge. A geodesic is then run between the two points. We are specifically interested in a geodesic g that minimizes the cost of going from p_s to p_e , where the cost is defined as a weighted combination of the geodesic distance and surface curvature: $\mathcal{C}(g) = \int_0^1 w(\kappa(t))g(t)dt$, where $g(0) = p_s$ and $g(1) = p_e$ and $w(\kappa(t))$ is selected as a decreasing functional of curvature. Hence, the ridge is a minimizer of \mathcal{C} , and may easily be computed for a triangulated mesh, through Dijkstra's algorithm with curvature weighted edge lengths. Consequently, the resulting

ridge does not suffer from the limitations outlined earlier, and, its accuracy depends solely on robust detection of the end points. Curvature weighting ensures that the geodesic passes through the high curvature ridge.

3.6 Bumps

We are interested in bumps on somewhat tubular subsurfaces. A bump is characterized with gentle slope, instead of pointy high curvature area. The problem is significant in CAD, for removing unwanted bumps, or in medical image analysis, for instance, for detecting soft-plaques in CT angiography.

Our approach is to analyze a tubular subsurface via x-sectional scans in an ROI, \mathcal{M}_r , to generate a set of profile contours $\{c_p\}$. The problem is, hence, reduced to (1) identifying the contours that correspond to a bump, followed by (2) determining “bumpy points” of a “bump contour”. A reference contour c_r is defined, with which all c_p are compared, starting at one end of \mathcal{M}_r and gradually moving towards the other. \mathcal{M}_r is selected to be sufficiently large to avoid the boundary of \mathcal{M}_r falling on a bump. Hence, the x-section at a boundary of \mathcal{M}_r may be selected as the initial c_r , which is later continuously updated with the last contour found without a bump.

For the identification of bump contours, the area of each contour c_p is compared to that of c_r . If it exceeds the reference area by a certain threshold t_P , c_p is considered to be on a bump. Once a bump contour is identified, the challenge lies in determining its bumpy part. A bump contour is projected to the plane of c_r . In this 2D subspace of the reference plane, a correspondence is found between the projected points and c_r . Corresponding points are then compared by way of their distances from individual centroids. If the distance of a projected point to the centroid of c_p , \hat{c}_p , exceeds the distance of the corresponding reference point to the reference centroid, \hat{c}_r , the point is marked bumpy. Such a criterion takes into account possible shifts of centroids and gives accurate estimates for bump points. After all x-sections are examined, bump points are binned into areas of connected points through recursion within 1-ring neighborhood.

4 Anatomical Feature Detection

We now detect the CES. The inter-tragal notch and the crus-side ridge require the detection of their end points. We use PCA of the x-section to identify two points on the canal as candidates for the inter-tragal notch top and the crus-side ridge top. The bottom of the inter-tragal notch is detected by considering the convexity of the bottom contour between tragus and anti-tragus. Geodesics are run from the ridge tops to the bottom point according to Section 3.5. The shorter geodesic is classified as the inter-tragal notch. The other ridge top is regarded the crus-side ridge top, and the shortest geodesic from this point to the bottom contour is the crus-side ridge. For detecting bulbous areas, and first and second bends, the canal was selected as the ROI. The canal-concha (or canal-crus) intersection is detected as an intersection of two geodesics, one running along a canal ridge, while the other traced from the concha peak (or crus).

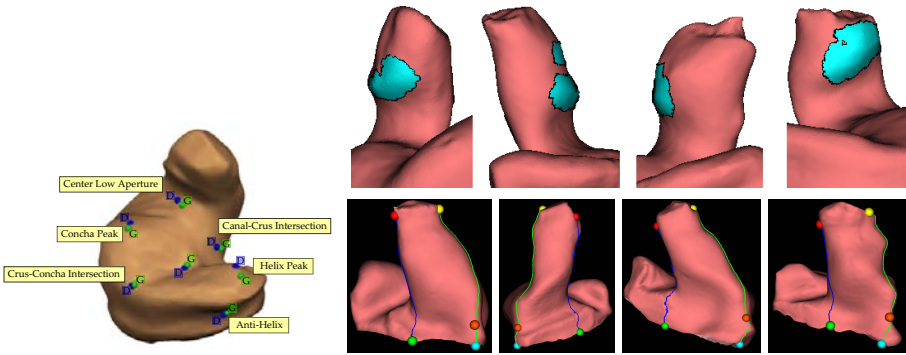


Fig. 6. Automatic detection of Point features: G = GT, D = detected features

Fig. 7. Automatic detection of: (Top) Canal Bulbous; (Bottom) Inter-tragal notch, crus-side ridge, and inter-tragal notch flare

Helix ridge is detected as the shortest curvature weighted geodesic between the helix and the shell boundary. Crus area is computed as the area enclosed by appropriately weighted geodesics run between the following feature pairs: (1) center crus-helix ridge bottom; (2) center crus-crus ridge bottom; and (3) the boundary contour. Crus-concha intersection is detected by analyzing the tangential profile of the intersection of the shell with the crus valley plane.

5 Experiments

Statistical validation of the detected features was carried out over a dataset of 198 shells, for which an expert annotated the Ground Truth (GT). Features were then detected automatically, and compared with GT (e.g., Fig. 6).

For quantitative validation, point features were compared by way of their distance from the GT. The similarity measure for plane features considered their orientation and location. Orientations were compared via inner product between plane normals. Deviation of plane locations was the average distance between the two planes. Sensitivity and specificity were used for validating area features. The mean and standard deviations of these measures are given in Fig. 8. The results indicate good agreement with the GT. For the crus area, the mean *sensitivity* was found to be 0.83 with a mean *specificity* of 0.93. Since bulbous is not always present on an impression, only its qualitative evaluation was considered (Fig. 7(Top)). The ridges were analyzed qualitatively and in more than 90% of the cases the experts were satisfied with the performance. Examples in Fig. 7(Bottom) precisely follow the actual ridges on the canal. Inter-tragal notch flare is also at the correct location, precisely where the saddle is formed.

In the second set of experiments, an expert provided acceptable error tolerances for various features (3mm for points and plane locations, 15° for plane orientations). The *success rate* was then computed as the percentage of cases that resulted in features within the acceptable range. Mean success rate was

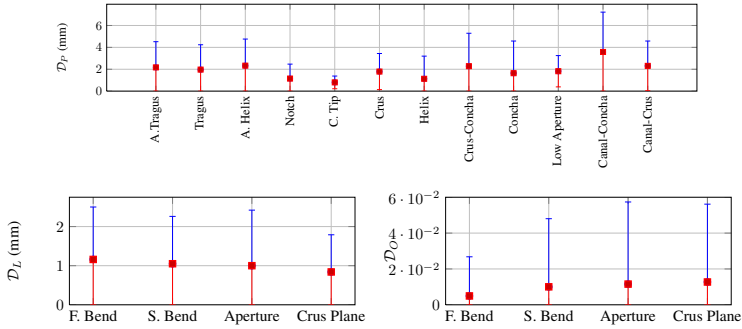


Fig. 8. Feature Validation – Comparison with GT: (Top) Point Features – Mean and standard deviation of the point deviations. (Bottom) Plane Features – Mean and standard deviation of deviations between plane (Left) locations; (Right) orientations.

87% for points (excluding Canal-Concha Intersection) and 82% for the planes. On average, all individual features exhibited acceptable performance (overall average deviation of $\sim 2\text{mm}$ and $\sim 10^0$) with the exception of the Canal-Concha Intersection and the Second Bend. Computationally, on average it takes 1.7s to compute all features for a dense mesh (79ms per feature) of $\sim 23\text{k}$ vertices.

6 Conclusions

We have proposed a comprehensive set of features for canonical representation of human ears, in addition to fast and robust algorithms for their detection. The representation may be used for automatic design of HA shells [5], feature guided registration, and classification. Due to their robustness, they have already been introduced for the automatic design of HAs by a major HA manufacturer. We emphasized that the utility of the proposed algorithms is not limited to the HAs. They have been designed with generality in mind, thereby providing a very powerful tool for medical image analysis and CAD.

References

1. Cohen-Or, D., et al.: Three-dimensional distance field metamorphosis. *ACM Trans. Graph.* 17(2), 116–141 (1998)
2. Henry, G.: *Gray’s Anatomy: Descriptive and Surgical* (1858)
3. Hildebrandt, K., et al.: Smooth feature lines on surface meshes. In: *SGP 2005* (2005)
4. Shinagawa, Y., Kunii, T.L.: Constructing a Reeb graph automatically from cross sections. *Computer Graphics and Appl.* 11(6), 44–51 (1991)
5. Sickel, K., Baloch, S., et al.: Semi-automatic manufacturing of customized hearing aids using a feature driven rule-based framework. In: *VMV 2009* (2009)
6. Slabaugh, G., et al.: 3D Shape Modeling for Hearing Aid Design. *IEEE Signal Processing Magazine* 5(25), 98–102 (2008)
7. Stylianou, G., Farin, G.: Crest lines for surface segmentation and flattening. *IEEE Visualization and Computer Graphics* 10(5), 536–544 (2004)

Probabilistic Multi-Shape Representation Using an Isometric Log-Ratio Mapping

Neda Changizi and Ghassan Hamarneh

Medical Image Analysis Lab, Simon Fraser University, Canada
{nca19,hamarneh}@cs.sfu.ca

Abstract. Several sources of uncertainties in shape boundaries in medical images have motivated the use of probabilistic labeling approaches. Although it is well-known that the sample space for the probabilistic representation of a pixel is the unit simplex, standard techniques of statistical shape analysis (e.g. principal component analysis) have been applied to probabilistic data as if they lie in the unconstrained real Euclidean space. Since these techniques are not constrained to the geometry of the simplex, the statistically feasible data produced end up representing invalid (out of the simplex) shapes. By making use of methods for dealing with what is known as compositional or closed data, we propose a new framework intrinsic to the unit simplex for statistical analysis of probabilistic multi-shape anatomy. In this framework, the isometric log-ratio (ILR) transformation is used to isometrically and bijectively map the simplex to the Euclidean real space, where data are analyzed in the same way as unconstrained data and then back-transformed to the simplex. We demonstrate favorable properties of ILR over existing mappings (e.g. LogOdds). Our results on synthetic and brain data exhibit a more accurate statistical analysis of probabilistic shapes.

1 Introduction

Numerous sources of uncertainties exist in shape boundaries including tissue heterogeneity [1], image acquisition artifacts, segmentation by multiple-raters, and image segmentation algorithms intentionally designed to output fuzzy results [2,3]. It is important not to ignore these uncertainties in subsequent analyses and decision-making [1,4]. In order to capture information concerning uncertainty in addition to the shape of multiple structures, probabilistic multi-shape representations have been proposed. Being able to perform statistical analysis on these probabilistic multi-shape representations is important in understanding normal and pathological geometrical variability of anatomical structures.

Multi-shape non-probabilistic (crisp) representations have been proposed by Tsai et al. [5], and Babalola and Cootes [6]. Leventon et al. adopted the signed distance map (SDM) representation and performed linear principal component analysis (PCA) to extract shape statistics [7]. Besides the fact that SDMs are not designed to encode uncertainty, the main disadvantage of this method is that it is not obvious how to impose a vector space structure on SDMs. This

is usually dealt with by projecting samples from the distribution given by the PCA coefficients back onto the manifold of valid SDMs [8]. Pohl et al. used the logarithm of the odds ratio (LogOdds) method to place probability atlases in a linear vector space [9]. In [9], linear PCA was done on layered SDMs (interpreted as LogOdds maps) to build a statistical shape atlas of brain structures. This means that, while the layered SDMs are exactly the logarithm-of-odds representations, results after algebraic manipulation in the logarithm-of-odds space often yield invalid SDMs (but still valid logarithm-of-odds representations). Using such results, computing probabilities as described in [9] may yield erroneous likelihoods, since vector operations are not closed under the set of SDMs. Malcolm et al. proposed a mapping of labels to vertices of a regular simplex which form the basis of a convex linear structure [10]. Hamarneh and Changizi proposed a proper inverse function for label space representation based on barycentric coordinates [11]. Performing linear PCA on the label space representation then exploring the variational modes may result in new invalid points out of the simplex (e.g. negative probabilities or sums exceeding unity). A possible remedy is to project such points onto the simplex [12]. However, this causes points along modes of variation (from some point onward) to collapse onto a single point, which gives improper results and violates the indefinite Gaussian distribution assumption of PCA. Another way to circumvent the problem is to impose some limits when exploring the modes of variation to force the points to stay within the simplex. This is undesired, however, as it renders traditional statistical exploration invalid, e.g. one may no longer be able to explore the variability within ± 3 standard deviations from the mean.

Compositional or *closed* data are multivariate data with positive values that sum up to a constant, usually chosen as 1. Compositional data has arisen in many different disciplines such as geology (mineral compositions of rocks), environmental (pollutant compositions), economics (household budget compositions), etc. [13]. We extend this list of application areas to anatomical compositions within image pixels, resulting from the aforementioned uncertainties (e.g. partial volume effect) [12,3,4].

Standard statistical techniques may lead to misleading results if they are directly applied to closed data. Statistical analysis of compositional data has been a developing area since 1986, when Aitchison introduced two transformations of compositional data to real space: the additive log-ratio (ALR) and the centered log-ratio (CLR) transformations, as well as a proper distance metric in the simplex [13]. Aitchison applied classical statistical analysis to the transformed observations, using ALR for modeling and CLR for those techniques based on a metric. The underlying reason was that ALR does not preserve distances, whereas CLR preserves distances but leads to a singular covariance matrix [14]. Egozcue et al. defined a new isometric log-ratio (ILR) transformation [15], which is an isometry between the simplex and the real space of the same dimension, avoiding the drawbacks of both ALR and CLR.

In this paper, we propose to consider the algebraic-geometrical structure (Hilbert space) of the simplex, based on operations of perturbation, power

transformation and the Aitchison inner product [13]. Through the ILR transformation, we develop techniques for statistical analysis of probabilistic multi-shape representations, which are not only intrinsic to the simplex but are bijective, isometric, and yielding non-singular covariance.

2 Methods

A sample space of compositional data, such as the probabilistic representation $\mathbf{x} = [x_1, x_2, \dots, x_n]$ of a pixel with n labels, is an $(n - 1)$ -dimensional unit simplex $S^{n-1} \subset R^n$ given as

$$S^{n-1} = \left\{ \mathbf{x} = [x_1, x_2, \dots, x_n] \in \mathbb{R}^n \mid x_i > 0, i = 1, 2, \dots, n; \sum_{i=1}^n x_i = 1 \right\}. \quad (1)$$

Two basic operations were defined on the simplex by Aitchison [13]. (i) **Perturbation**, which is analogous to displacement or translation in real space and is defined for any $\mathbf{x}, \mathbf{y} \in S^{n-1}$ as

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1y_1, x_2y_2, \dots, x_ny_n]. \quad (2)$$

(ii) **Power transformation**, which is analogous to scalar multiplication in real space and is defined for a vector $\mathbf{x} \in S^{n-1}$ and a scalar $\alpha \in \mathbb{R}$ as

$$\alpha \odot \mathbf{x} = \mathcal{C}[x_1^\alpha, x_2^\alpha, \dots, x_n^\alpha]. \quad (3)$$

In equations (2) and (3), \mathcal{C} is the closure operation defined for a vector $\mathbf{z} = [z_1, z_2, \dots, z_n] \in \mathbb{R}_+^n$ as

$$\mathcal{C}(\mathbf{z}) = \left[\frac{z_1}{\sum_{i=1}^n z_i}, \frac{z_2}{\sum_{i=1}^n z_i}, \dots, \frac{z_n}{\sum_{i=1}^n z_i} \right]. \quad (4)$$

The internal simplicial operation of perturbation, and the external operation of powering define an $(n - 1)$ -dimensional vector space (indeed a Hilbert space) on S^{n-1} [16][17][18][19]. The structure can be extended to produce a metric vector space by the introduction of the simplicial metric d_S defined by Aitchison as [13]

$$d_S(\mathbf{x}, \mathbf{y}) = \left[\frac{1}{n} \sum_{i < j} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2 \right]^{\frac{1}{2}} = \left[\sum_{i=1}^n \left(\ln \frac{x_i}{g(\mathbf{x})} - \ln \frac{y_i}{g(\mathbf{y})} \right)^2 \right]^{\frac{1}{2}}, \quad (5)$$

where $g(\mathbf{x}) = (\prod_{i=1}^n x_i)^{\frac{1}{n}}$ is the component-wise geometric mean of the composition. An inner product $\langle \mathbf{x}, \mathbf{y} \rangle_S$ and a norm $\|\mathbf{x}\|_S^2 = \langle \mathbf{x}, \mathbf{x} \rangle_S$, consistent with this metric, complete the Euclidean structure of the simplex:

$$\langle \mathbf{x}, \mathbf{y} \rangle_S = \frac{1}{D} \sum_{i < j} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} = \sum_{i=1}^D \ln \frac{x_i}{g(\mathbf{x})} \ln \frac{y_i}{g(\mathbf{y})}. \quad (6)$$

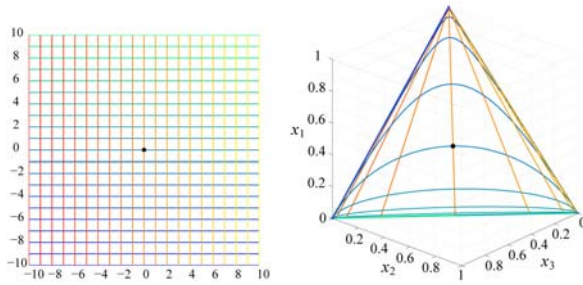


Fig. 1. Left: Orthogonal grids of lines in \mathbb{R}^2 . Right: Orthogonal compositional lines (colors correspond) in S^2 , equally spaced by 1 unit in Aitchison distance [13].

In any finite dimensional Hilbert space, a geodesic curve connecting two points is understood to be the only continuous curve whose length is minimum with respect to the distance metric of the space. Such a geodesic is a segment of a ‘straight’ line within the geometry of the space considered. To avoid confusion, straight lines with respect to the Aitchison geometry are called *compositional lines* (Figure 1). A compositional line going from \mathbf{x}_0 to $\mathbf{x}(t)$, with the leading vector \mathbf{p} , in S^{n-1} is given by

$$\mathbf{x}(t) = \mathbf{x}_0 \oplus (t \odot \mathbf{p}) \quad t \in \mathbb{R}, \mathbf{x}_0, \mathbf{p} \in S^{n-1}. \tag{7}$$

Three log-ratio transformations have been proposed so far to map the simplex to the Euclidean real space. (i) **The ALR transformation** [13] (the same as the LogOdds method [9]):

$$alr : S^{n-1} \rightarrow \mathbb{R}^{n-1}, \quad alr(\mathbf{x}) = \left[\ln \frac{x_1}{x_n}, \ln \frac{x_2}{x_n}, \dots, \ln \frac{x_{n-1}}{x_n} \right]. \tag{8}$$

This transformation is asymmetric in the parts of the composition. By changing the part in the denominator, we obtain different *alr* transformations. Although all the statistical procedures are invariant under a permutation of the compositional parts [13]. But the main drawback of ALR is the fact that it is not an isometric (but an isomorphic or bijective) transformation from the simplex, with the Aitchison metric, onto the real space, with the ordinary Euclidean metric. In fact, *alr* coefficients are coordinates in an *oblique* basis, something that affects distances if the usual Euclidean distance is computed from the *alr* coordinates. (ii) **The CLR transformation** [13]:

$$clr : S^{n-1} \rightarrow \mathbb{R}^n, \quad clr(\mathbf{x}) = \left[\ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_n}{g(\mathbf{x})} \right]. \tag{9}$$

CLR is symmetrical in the components, but since the sum of the components has to be zero, this transformation leads to a singular covariance matrix. CLR is an isometry but between the $(n - 1)$ -dimensional simplex S^{n-1} and a subspace of real space \mathbb{R}^n . In this paper, we propose to adopt the third transformation:

(iii) **The ILR transformation [15]:**

$$ilr : S^{n-1} \rightarrow \mathbb{R}^{n-1}, \quad \mathbf{y} = ilr(\mathbf{x}) = [\langle \mathbf{x}, \mathbf{e}_1 \rangle_S, \langle \mathbf{x}, \mathbf{e}_2 \rangle_S, \dots, \langle \mathbf{x}, \mathbf{e}_{n-1} \rangle_S]. \quad (10)$$

ILR is based on the choice of *orthonormal* bases $\mathbf{e}_i, i = \{1, 2, \dots, n - 1\}$ of the simplex S^{n-1} that can be calculated using the Gram-Schmidt procedure. The ILR transformation of any $\mathbf{x} \in S^{n-1}$ gives the coordinates of \mathbf{x} with respect to the basis $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{n-1}$. The *inverse* ILR transformation corresponds to the expression of \mathbf{x} in the reference basis of S^{n-1} :

$$\mathbf{x} = ilr^{-1}(\mathbf{y}) = \bigoplus_{i=1}^{n-1} (\mathbf{y}_i \odot \mathbf{e}_i), \quad \text{where } y_i = \langle \mathbf{x}, \mathbf{e}_i \rangle_S. \quad (11)$$

3 Results

We first show the shortcomings of ALR and CLR in comparison to ILR. We created 10,000 random pairs (\mathbf{x}, \mathbf{y}) of probability vectors for different number of labels $n = 2, 3, 4$, sampled from a Dirichlet distribution with parameters $\alpha_1 = \dots = \alpha_n = 1$, giving a uniform distribution within the $(n - 1)$ -dimensional open simplex and zero elsewhere [20]. We calculated the distance between the two vectors in each pair in different spaces: (a) In S^{n-1} , using the simplicial metric $d_S(\mathbf{x}, \mathbf{y})$ from equation (5), (b) Euclidean distance between the ILR-transformed vectors, i.e. $|ilr(\mathbf{x}) - ilr(\mathbf{y})|_2$, (c) $|alr(\mathbf{x}) - alr(\mathbf{y})|_2$, and (d) $|\mathbf{x} - \mathbf{y}|_2$. The scatter plots in Figure 2 show the relation between the different distances. Note how only the distances of the proposed ILR-transformed vectors remain faithful to the proper simplicial distance (diagonal line in the leftmost plot). We also show this difference in distance calculations by starting with a set of probability vectors forming a unit circle, according to the simplex geometry, in S^2 and noting the isometry (between S^2 and \mathbb{R}^2) only when transforming these vectors to ILR space, but not ALR or CLR (Figure 3).

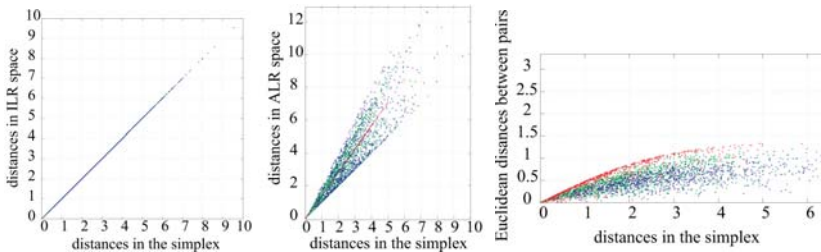


Fig. 2. Comparing distances between 10,000 random pair of probability vectors calculated in the ILR space (Left), ALR space (Center), and Euclidean distances between probability vectors (Right) vs. using the proper simplicial metric. Red, green, and blue points correspond to distances between vectors with 2, 3, and 4 labels, respectively. For ILR, all colored points collapse onto the diagonal.

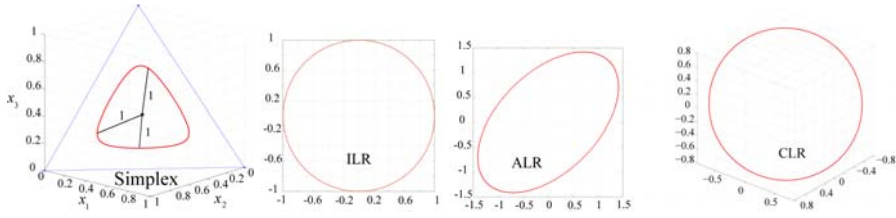


Fig. 3. Left to right: Unit circle according to the simplex (S^2) geometry, transformed isometrically to a unit circle in ILR space, distances are unpreserved in ALR space whereas an extra dimension is added in CLR space

Next, we show that performing statistical analysis (e.g. PCA) in ILR and ALR spaces lead to different results. In the first experiment, we sampled a simple curve in S^2 and transformed the samples to ILR and ALR. After performing PCA in the Euclidean spaces of ILR and ALR, we back-transformed the results to the simplex. We also performed PCA on the label space representation of the samples. It is shown that moving along the modes of variation obtained from ALR and ILR transformations result in different probabilities. It is also obvious that, in label space, it is not allowable to move freely along the modes since, at some point, the mode will leave the simplex (e.g. negative probabilities). This experiment is repeated for a random set of points in S^2 (Figure 4).

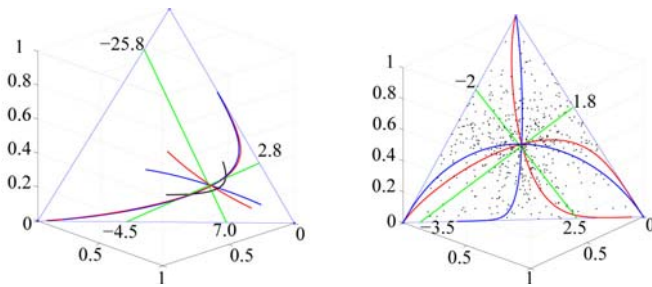


Fig. 4. For a given set of points (black) in S^2 , new points along the modes of variation (-10 to $10\sqrt{\lambda}$) are obtained using ILR (red) and ALR (blue), where λ is the variance explained by that mode. Note that when the label space method is used (green), limits are imposed (factors of $\sqrt{\lambda}$ for each mode, shown where green reaches the simplex boundary) to stay in the simplex

In the next experiment, we performed PCA on ILR and ALR-transformed probabilistic segmentation maps of 20 subjects from BrainWeb [21]. For each subject and at each pixel, 12 probability values capture the pixel’s fuzzy membership in 12 classes such as cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM). The mid-sagittal plane of all subjects are transformed to ILR and ALR space. After performing PCA in those spaces, new probability

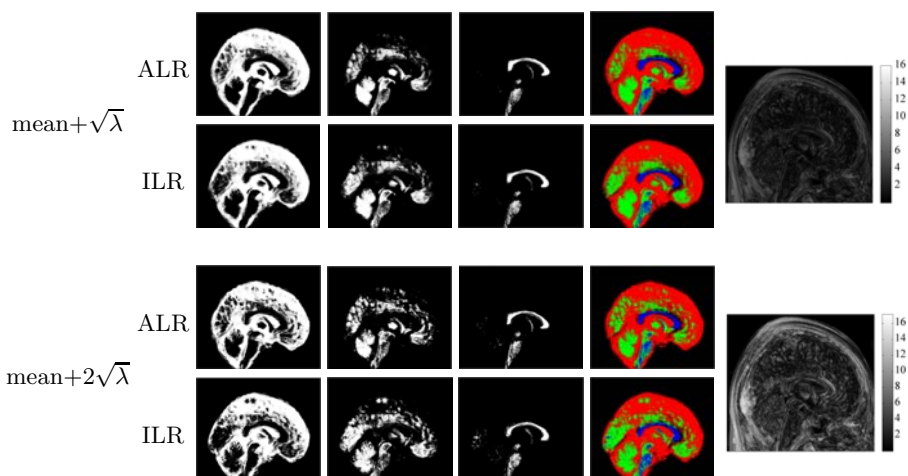


Fig. 5. Columns 1–3 show the variability in the resulting probability maps for CSF, GM, and WM. The 4th column uses the first three columns as RGB channels. λ is the variance explained by the first mode. The 5th columns shows the simplicial distances (errors when using ALR (logOdds)) between the two probability maps resulting from ILR and ALR at each pixel.

maps along the first mode of variation are obtained and shown in Figure 5. Different images obtained using ILR and ALR transformations show that the error from ALR can be avoided using the proposed method. The simplicial distances between pixels of the the new probability maps using ILR and ALR transformations are calculated and also shown in Figure 5.

4 Conclusions

We proposed an alternative probabilistic multi-shape representation: Isometric Log-Ratio. It has several desired properties: forms a vector space, isometric and thus isomorphic to the probability simplex, and results in a non-singular covariance. These properties do not exist together in any previously offered probabilistic computational anatomy work. We demonstrated how the lack of some of these properties degrades the results, e.g. statistical analysis using linear PCA. In the future, we intend to apply the method within the context of clinical applications.

References

1. Udupa, J.K., Grevera, G.J.: Go digital, go fuzzy. *Pattern Recognit. Lett.* 23(6), 743–754 (2002)
2. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20(1), 45–57 (2001)

3. Grady, L.: Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(11), 1768–1783 (2006)
4. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23(7), 903–921 (2004)
5. Tsai, A., Wells, W., Tempany, C., Grimson, E., Willsky, A.: Mutual information in coupled multi-shape model for medical image segmentation. *Med. Image Anal.* 8(4), 429–445 (2004)
6. Babalola, K., Cootes, T.: Registering richly labelled 3D images. In: *IEEE ISBI*, pp. 868–871 (2006)
7. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: *IEEE CVPR*, pp. 316–323 (2000)
8. Golland, P., Grimson, W., Shenton, M., Kikinis, R.: Detection and analysis of statistical differences in anatomical shape. *Med. Image Anal.*, 69–86 (2005)
9. Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Using the logarithm of odds to define a vector space on probabilistic atlases. *Med. Image Anal.* 11(5), 465–477 (2007)
10. Malcolm, J., Rath, Y., Shenton, M., Tannenbaum, A.: Label space: A coupled multi-shape representation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II. LNCS*, vol. 5242, pp. 416–424. Springer, Heidelberg (2008)
11. Hamarneh, G., Changizi, N.: Barycentric label space. In: *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, pp. 162–173 (2009)
12. Cremers, D., Schmidt, F.R., Barthel, F.: Shape priors in variational image segmentation: Convexity, Lipschitz continuity and globally optimal solutions. In: *IEEE CVPR*, pp. 1–6 (2008)
13. Aitchison, J.: *The statistical analysis of compositional data*. Chapman & Hall Ltd., Boca Raton (1986)
14. Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R.: *Lecture notes on compositional data analysis* (2007), <http://hdl.handle.net/10256/297>
15. Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barcel-Vidal, C.: Isometric logratio transformations for compositional data analysis. *Math. Geol.* 35(3), 279–300 (2003)
16. Pawlowsky-Glahn, V., Egozcue, J.J.: Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment* 15(5), 384–398 (2001)
17. Pawlowsky-Glahn, V., Egozcue, J.J.: BLU estimators and compositional data. *Math. Geol.* 34(3), 259–274 (2002)
18. Billheimer, D., Guttorp, P., Fagan, W.F.: *Statistical analysis and interpretation of discrete compositional data*. Technical Report 11, University of Washington, Seattle, Washington (1997)
19. Billheimer, D., Guttorp, P., Fagan, W.F.: Statistical interpretation of species composition. *J. of the American Statistical Association* 96(456), 1205–1214 (2001)
20. Connor, R.J., Mosimann, J.E.: Concepts of independence for proportions with a generalization of the dirichlet distribution. *J. of the American Statistical Association* 64(325), 194–206 (1969)
21. Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Pike, G.B., Evans, A.C.: BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage* 5, 425 (1997)

Efficient Robust Reconstruction of Dynamic PET Activity Maps with Radioisotope Decay Constraints

Fei Gao¹, Huafeng Liu², and Pengcheng Shi¹

¹ Golisano College of Computing and Information Sciences,
Rochester Institute of Technology, Rochester, NY, 14623, USA

² State Key Laboratory of Modern Optical Instrumentation, Zhejiang University,
Hangzhou, 310027, China

Abstract. Dynamic PET imaging performs sequence of data acquisition in order to provide visualization and quantification of physiological changes in specific tissues and organs. The reconstruction of activity maps is generally the first step in dynamic PET. State space H_∞ approaches have been proved to be a robust method for PET image reconstruction where, however, temporal constraints are not considered during the reconstruction process. In addition, the state space strategies for PET image reconstruction have been computationally prohibitive for practical usage because of the need for matrix inversion. In this paper, we present a minimax formulation of the dynamic PET imaging problem where a radioisotope decay model is employed as physics-based temporal constraints on the photon counts. Furthermore, a robust steady state H_∞ filter is developed to significantly improve the computational efficiency with minimal loss of accuracy. Experiments are conducted on Monte Carlo simulated image sequences for quantitative analysis and validation.

1 Introduction

Dynamic positron emission tomography (PET) is a molecular imaging technique that is used to monitor the spatiotemporal distribution of a radiotracer *in vivo* [1]. Dynamic PET performs sequence of contiguous short-interval scans, and the reconstructed activity maps reflect the real-time concentration of radiotracers in tissues. The activity maps are conventionally reconstructed frame-by-frame from the dynamic data sequence with analytical or statistical methods, without taking into consideration potential temporal relations between frames. Further, through application of *a priori* models, parameter maps can be recovered to illustrate the dynamic changes of radiotracer kinetics. Together, the activity maps and the parameter maps provide visualization and quantification of physiological changes that indicate the functional states of specific tissues that are of significant research and clinical values [2,3].

Many research efforts have concentrated on the design of iterative statistical reconstruction algorithms for PET activity reconstruction [4], such as those based

on expectation maximization (EM) and maximum a priori (MAP) criteria. These strategies assume the statistical distribution of acquired data to be known, which is not necessarily valid for *corrected data* from most clinical scanners. Liu *et al.* have first introduced a robust state space approach for PET image reconstruction [5], where they have utilized the robust H_∞ filtering without any assumptions on the system and data noise statistics. However, as all other frame-by-frame methods, it ignores temporal constraints in the dynamic reconstruction of the image sequence. Recently, Gao *et al.* have extended the work to dynamic dual-tracer PET reconstruction [6] where they introduced the temporal radioisotope decay model. However, the radio decay model is not incorporated as temporal constraints but rather used to separate two radioisotopes with drastically different half lives. Furthermore, a big shortcoming of these state space H_∞ reconstruction algorithms is their computational complexity because of the need for large-scale matrix inversion, which has hindered their practical value.

In this paper, we concentrate on the development of an efficient and robust minimax activity reconstruction framework with radioisotope physics-based temporal constraints. The radio decay model, considering the natural decay property of the radioisotope, is introduced into the objective function as the temporal guidance for multi-frame image sequence reconstruction. The optimization problem is solved by continuous steady state H_∞ filter for all image frames [7], which yields robust and accurate results with significant computation reduction compared to conventional H_∞ filter reconstruction. Experiments are conducted on Monte Carlo simulated image sequences for quantitative analysis and validation.

2 Theory

2.1 Radioisotope Decay Constrained Dynamic PET Imaging

The projection equation of dynamic PET imaging can be formulated through an affine transform between the projection data and emission object as:

$$y(t) = Dx(t) + r(t) + s(t) \quad (1)$$

where the emission sinogram data is represented by a vector y , and the activity of emission object is represented by x . D is system probability matrix, which gives the probability of a photon emitted from i^{th} voxel being detected in projection j^{th} bin. t is the time frame. r and s are the contributions of random coincidence events and scatter coincidence events. After the conventional online delayed-window random correction, the equation (1) can be rewritten as:

$$y(t) = Dx(t) + e(t) \quad (2)$$

here e is an error vector, which represents *unknown* measurement uncertainties including scatter coincidence events.

In the conventional state space reconstruction of PET imaging, the distribution of the radioisotopes in the body is assumed to be temporally stationary corresponding to the autoradiographic model, however, in the real situation, the radioisotope will decay with time, and its activity at time t should be

$$x = X_0 e^{\frac{\ln(0.5)}{T}t} \tag{3}$$

here X_0 is the initial activity distribution, and T is the half life of the radioisotope. So the real-time change of radioisotope can be represented as

$$\frac{dx}{dt} = X_0 \frac{\ln(0.5)}{T} e^{\frac{\ln(0.5)}{T}t} \tag{4}$$

then the dynamic change of radioisotope from one frame to the next can be obtained from the integral of equation(4). A general representation of state transition will be

$$x(t + 1) = H(t)x(t) + v(t) \tag{5}$$

where $x(t)$ is the radioactivity concentration at time frame t , and $H(t)$ is a coefficient matrix for state transition at time frame t . $v(t)$ represents the uncertainties during state transition. With the introduction of decay model shown as equation(5), we are able to make use of the radioisotope’s own temporal properties as constraints to guide our reconstruction.

2.2 Robust Reconstruction Criterion

Seeking solutions under uncertainties is a difficult problem. Robust discrete optimization seeks to identify solutions that will perform well under any circumstances. Although many criteria are available, one reasonable choice is the minimax criterion, which allows one to identify a robust solution as one that has the best worstcase performance. In general, a robust discrete optimization problem can be formulated as follows. Let X be the set of all solutions, E be the set of uncertainties of measurement in single time frame, M be the set of uncertainties for state transition among time frames, and performance of a solution $x \in X$ under uncertainties $e \in E$ and $v \in M$ be $F(x, e, v)$, now the problem is to find the solution that has the best worst-case performance, which is the same as minimizing (over all solutions) the maximum (over all uncertainties) performance:

$$\min_{x \in X} \max_{e \in E, v \in M} F(x, e, v) \tag{6}$$

from the description of equation(2) and (5), the estimation of activity distribution $x(t)$ at time t is not only computed based on measurement $y(t)$, but also affected by previous estimations, so we define a linear combination of $x(t)$ as

$$z(t) = \mathcal{F}x(t) = g(x(k), H(k), v(k)) \quad k = 1, 2...t \tag{7}$$

so the measurement of performance $F(x, e, v)$ is given by

$$J = \frac{\sum \|z(t) - \hat{z}(t)\|_{Q(t)}^2}{\|x(0) - \hat{x}(0)\|_{p_0^{-1}}^2 + \sum (\|v(t)\|_{V(t)}^2 + \|e(t)\|_{N(t)}^2)} \tag{8}$$

where the notation $\|x\|_G^2$ is defined as the square of the weighted (by G) L_2 norm of x (i.e. $\|x\|_G^2 = x^T G x$). $N(t)$, $V(t)$ and $Q(t)$ are the weighting matrices for the

measurement uncertainties, state transition uncertainties, and the estimation error at time t respectively. $\hat{x}(0)$ is the initial estimate of the state. p_o is the weighting matrix for the initial conditions. With the state estimate $\hat{x}(n)$ at time n , the measurement uncertainty $e(n)$ and state transition uncertainty $v(n)$ can be expressed as $e(n) = Y(n) - D\hat{x}(n)$ and $v(n) = \hat{x}(n + 1) - H(n)\hat{x}(n)$. Then both the measurement-system matrix constraint and decay model constraint are implied in J . The optimal estimate $z(t)$ among all possible $\hat{z}(t)$ should satisfy:

$$\|J\|_\infty = \sup J < \gamma^2 \tag{9}$$

where $\gamma^2 > 0$ is a prescribed level of disturbances. It is assumed that the L_2 norms of $e(t)$ and $v(t)$ exist. Then the minimax performance criterion of equation (8) where the estimator strategy $z(t)$ playing against the exogenous inputs $e(t)$, $v(t)$ and the initial state $x(0)$ becomes

$$\min_{z(t)-\hat{z}(t)} \max_{v,e,x(0)} J = \sum \|z(t) - \hat{z}(t)\|_{Q(t)}^2 - \gamma^2 \|x(0) - \hat{x}(0)\|_{p_o^{-1}}^2 - \gamma^2 \sum (\|v(t)\|_{V(t)^{-1}}^2 + \|e(t)\|_{N(t)^{-1}}^2) \tag{10}$$

Now the problem becomes to solve the above objective function, and the decay model is successfully incorporated. Robust H_∞ filter solution has already been presented in [5], so here we concentrate on a more efficient steady state H_∞ filter.

2.3 Efficient Implementation

A matrix inverse is required at every time step in conventional H_∞ filter in order to calculate the H_∞ gain. Generally, inversion of small matrices is fairly easy, but the inversion of a large matrix will require more computational costs in a practical implementation. The steady state H_∞ filter is much easier to be implemented in a system in which real-time computational effort or code size is a serious consideration [7].

The minimax objective function is given as equation (10), where the parameters (N, V and Q) are symmetric positive definite matrices based on the specific problem. Since the designed parameters of the underlying system can be treated as fixed values for input, then the steady state solution to the minimax problem can be obtained. Referring to H_∞ filter, the steady state solution will be

$$K = PSD^T N^{-1} \tag{11}$$

$$P = HPSH^T + V \tag{12}$$

$$\hat{x}(m + 1) = H\hat{x}(m) + HK(m)(y(m) - D\hat{x}(m)) \tag{13}$$

$$S = (I - \gamma^{-2}\bar{Q}P + D^T N^{-1}DP)^{-1}$$

$$\bar{Q} = \mathcal{F}^T Q \mathcal{F}$$

In order to have a solution to the problem, the following condition must be hold:

$$P^{-1} - \gamma^{-2}\bar{Q} + D^T N^{-1}D > 0 \tag{14}$$

If γ^{-2} , \mathcal{F} , N or Q is too large, or D is too small, the H_∞ estimator will have no solution. After the conditions above satisfied, equation (12) can be written as

$$P = H[P^{-1} - \gamma^{-2}\bar{Q} + D^T N^{-1}D]^{-1}H^T + V \quad (15)$$

Applying the matrix inversion lemma to the inverse of the above expression, we can get

$$\begin{aligned} P &= HP - P[P + (-\gamma^{-2}\bar{Q} + D^T N^{-1}D)^{-1}]^{-1}PH^T + V \\ &= HPH^T - HP[P + (-\gamma^{-2}\bar{Q} + D^T N^{-1}D)^{-1}]^{-1}PH^T + V \end{aligned} \quad (16)$$

Equation (16) is a discrete-time algebraic Riccati equation that can be solved by control system software or numerically iterating equation (12) until it converges to a steady state value.

The disadvantage of the steady state H_∞ filter is that theoretically it does not perform as well as the time-varying filter. However, the reduced performance that is seen in the steady state H_∞ solution is often a small fraction of the optimal performance, whereas the computational saving can be significant [7].

3 Experiments and Results

The data set used for validation in this study was acquired by Monte Carlo simulations. Monte Carlo simulations can provide a relatively accurate reference for the development and assessment of new image reconstruction algorithms. Simulations in our study are performed using toolbox GATE. The simulated PET scanner is the newly designed full 3D whole body PET scanner Hamamatsu SHR74000 from Hamamatsu Photonics K.K.. The phantom chosen here for simulations is Zubal thorax phantom.

2 experiments were performed here. In the 1st experiment, the number of coincidence events acquired in every time frame was controlled to be similar, and this is in accordance with the conditions of conventional state space H_∞ filter reconstruction. However, in real situation, it is not easy to control the number of coincidence events in every time frame, so we designed the 2nd experiment that sampled the data with the same time interval more intuitively. These 2 experiments are named as *Experiment 1* and *Experiment 2* in the following parts respectively. The initial activity per voxel was set to be low to avoid the contamination of random coincidence events. In *Experiment 1*, 20 time frames were acquired in all, and the total number of coincidence events was around 68k for every time frame; in *Experiment 2*, same 20 time frames were acquired in all, but the acquisition time was set to be 360s for every time frame. The 2 experiments correspond to 2 different protocols for dynamic PET acquisition.

The simulation outputs were stored in sinograms. Before reconstruction, random correction, normalization correction, attenuation correction and scatter correction were performed properly. The sinograms were reconstructed as images of size 128×128 by H_∞ filter without decay model, H_∞ filter with decay model and steady state H_∞ filter respectively.

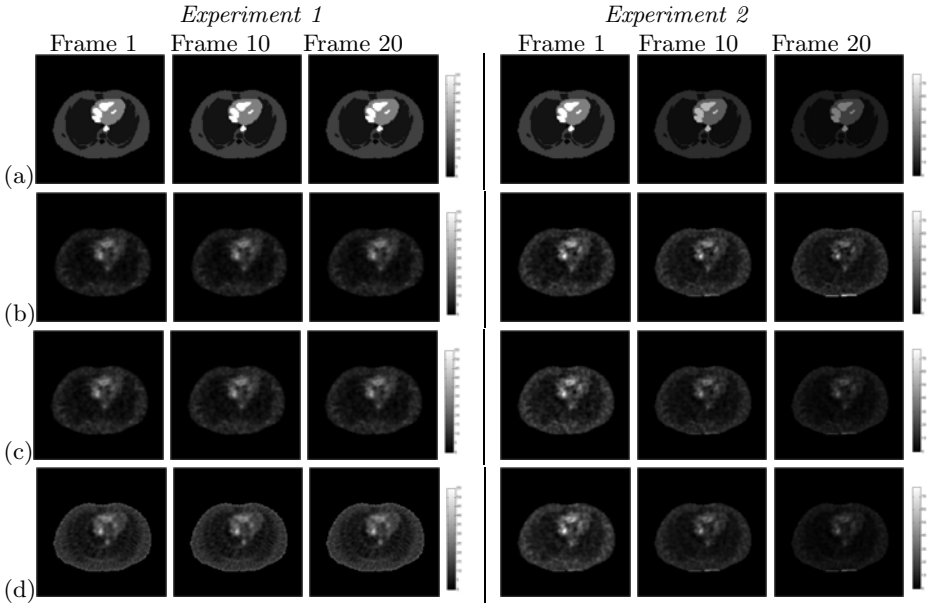


Fig. 1. True activity distributions and reconstructed images at time frame #1, #10 and #20 in 2 experiments. (a): true activity. (b): H_∞ filter reconstruction without decay model. (c): H_∞ filter reconstruction with decay model. (d): steady state H_∞ filter reconstruction.

Fig. 1 gives the true activity distributions and reconstructed images of the central slice by 3 algorithms of both experiments at time frame #1, #10 and #20. The reconstruction results from H_∞ filter with decay model show better recovery of activity distributions than H_∞ filter without decay model. Steady state H_∞ filter can also reconstruct the activity distributions properly: the regions with high activity concentration are recovered well, but the non-activity regions are little overestimated.

Since these experiments are based on Monte Carlo simulations, we can get the true activity distributions at anytime exactly. In order to analyze the reconstruction results quantitatively, we define the Variance (Var) as follow: $Var = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \hat{x}_i)^2}{\hat{x}_i^2}$, where x_i is the estimated activity value of pixel i , \hat{x}_i is the true activity value of pixel i , and n is the number of pixels.

The calculated variances between reconstruction results and true activity distributions through all time frames are plotted in Fig 2, and some results are extracted and listed in Table 1 for better understanding. From the quantitative analysis, in *Experiment 1*, due to the similar number of coincidence events acquired through all time frames, which corresponds to the assumptions of conventional H_∞ filter reconstruction without decay model, reasonable results can be obtained through all the time frames by it; but in *Experiment 2*, the experiment conditions do not agree with the assumptions, so the calculated variances

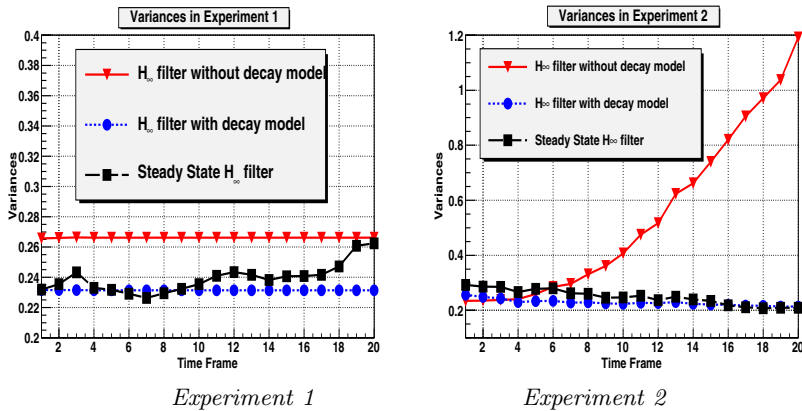


Fig. 2. Variances of reconstructed images by 3 algorithms through all time frames

Table 1. Variances of reconstructed images by 3 algorithms in *Experiment 1* and *Experiment 2* at time frame #1, #5, #10, #15 and #20

Variances of <i>Experiment 1</i>	Frame 1	Frame 5	Frame 10	Frame 15	Frame 20
H_∞ without decay model	0.2657	0.2664	0.2662	0.2662	0.2662
H_∞ with decay model	0.2316	0.2314	0.2314	0.2314	0.2314
Steady State H_∞	0.2320	0.2318	0.2355	0.2408	0.2624
Variances of <i>Experiment 2</i>	Frame 1	Frame 5	Frame 10	Frame 15	Frame 20
H_∞ without decay model	0.2342	0.2595	0.4066	0.7393	1.1914
H_∞ with decay model	0.2567	0.2331	0.2239	0.2206	0.2136
Steady State H_∞	0.2922	0.2784	0.2470	0.2345	0.2090

become larger as time increases, which means the quantitative accuracy of reconstruction results becomes worse. However, for the reconstruction results by H_∞ filter with decay model, due to the introduction of the decay model, they are already more accurate than that reconstructed by H_∞ filter without decay model in *Experiment 1*, furthermore, in *Experiment 2*, as a result of the temporal constraints of the decay model in reconstruction, the reconstruction results are much more stable through all the time frames, and the quantitative accuracy of reconstruction results are similar for all time frames and much better than that reconstructed by H_∞ filter without decay model. The variances are reduced by 11% and 57% in average for *Experiment 1* and *Experiment 2* respectively. For steady state H_∞ filter reconstruction, the results are not so stable as H_∞ filter with decay model and the quantitative accuracies are little worse, but the reconstruction results are still much better than that reconstructed by H_∞ filter without decay model, most importantly, significant computation saving is achieved by steady state H_∞ filter shown as follow.

Table 2. Time consumptions of 3 algorithms in *Experiment 1* and *Experiment 2*

Algorithm	H_∞ without decay model	H_∞ with decay model	Steady State H_∞
<i>Experiment 1</i>	5536s	6228s	127s
<i>Experiment 2</i>	30531s	39924s	540s

In this paper, all reconstructions are finished using MATLAB on a single 2.26GHz CPU, 12G RAM PC. The time consumptions are summarized in Table 2. Compared with H_∞ filter, the computation costs were reduced by 99.7% in *Experiment 1* and 98.6% in *Experiment 2* by using steady state H_∞ filter. This makes steady state H_∞ filter more efficient for practical usage, furthermore, there are cases that steady state H_∞ filter is used in embedded system, so further acceleration may be achieved by hardware. Related researches are underway.

4 Conclusion

An efficient and robust minimax activity reconstruction for dynamic PET imaging with radioisotope decay constraints is proposed in this paper. A physical decay model is introduced as temporal guidance in reconstruction. A robust steady state H_∞ filter is adopted to solve the problem which leads to significant computation reduction and makes the algorithm more efficient and practical. Reconstruction results from simulation experiments show the robustness and improvement in quantitative accuracy of the minimax reconstruction with decay model. The time consumption analysis shows the efficiency of the steady state H_∞ filter.

Acknowledgements. This work is supported in part by the National Basic Research Program of China (No: 2010CB732500).

References

1. Bailey, D.L., Townsend, D.W., Valk, P.E., Maisey, M.N.: Positron emission tomography: basic sciences. Springer, Heidelberg (2005)
2. Wernick, M., Infusino, E., Milosevic, M.: Fast spatio-temporal image reconstruction for dynamic PET. IEEE Transactions on Medical Imaging 18(3), 185–195 (1999)
3. Krestyannikov, E., Tohka, J., Ruotsalainen, U.: Joint penalized-likelihood reconstruction of time-activity curves and regions-of-interest from projection data in brain pet. Physics in Medicine and Biology 53(11), 2877–2896 (2008)
4. Qi, J., Leahy, R.: Iterative reconstruction techniques in emission computed tomography. Physics in Medicine and Biology 51, R54–R78 (2006)
5. Liu, H., Tian, Y., Shi, P.: PET image reconstruction: a robust state space approach. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 197–209. Springer, Heidelberg (2005)
6. Gao, F., Liu, H., Jian, Y., Shi, P.: Dynamic dual-tracer pet reconstruction. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 38–49. Springer, Heidelberg (2009)
7. Simon, D.: Optimal state estimation: kalman, H_∞ and nonlinear approaches. John Wiley and Sons, Chichester (2006)

Nonlinear Embedding towards Articulated Spine Shape Inference Using Higher-Order MRFs^{*}

Samuel Kadoury¹ and Nikos Paragios²

¹ Philips Research North America, Briarcliff Manor, NY, USA
samuel.kadoury@philips.com

² Ecole Centrale de Paris, Laboratoire MAS
GALEN Group, INRIA Saclay, Ile-de-France, France
nikos.paragios@ecp.fr

Abstract. In this paper we introduce a novel approach for inferring articulated spine models from images. A low-dimensional manifold embedding is created from a training set of prior mesh models to establish the patterns of global shape variations. Local appearance is captured from neighborhoods in the manifold once the overall representation converges. Inference with respect to the manifold and shape parameters is performed using a Markov Random Field (MRF). Singleton and pairwise potentials measure the support from the data and shape coherence from neighboring models respectively, while higher-order cliques encode geometrical modes of variation for local vertebra shape warping. Optimization of model parameters is achieved using efficient linear programming and duality. The resulting model is geometrically intuitive, captures the statistical distribution of the underlying manifold and respects image support in the spatial domain. Experimental results on spinal column geometry estimation from CT demonstrate the approach’s potential.

1 Introduction

Statistical models of shape variability have been successful in addressing fundamental vision tasks such as segmentation and registration in medical imaging. For example, Active Shape and Appearance Models have been used in recovering single object geometries obtained from dense collection of data points. Implicit representations is an alternative formulation to address model-based segmentation while more recently numerous methods based on point distribution models (PDM) and embedding on various geometric spaces (spherical [1]) have been proposed. However, in the case of articulated objects such as the spinal column, model-based segmentation of single objects typically leads to fitting errors when pathologies are present. The result is thereby sensitive to model initialization and therefore limited to the capture range. Simultaneous multi-object inference is often beneficial compared to the separate segmentation of individual objects. In [2], an extension of PDMs was considered for modeling relations between shape constellations using conditional probabilities between 2D-contours. Rigid

^{*} Funding provided by Stereos+ (Medicen), INRIA and FQRNT grants.

transformations were considered as statistics between part sets to create reconstructed models [3], achieve CT segmentation [4] and multi-modal inference [5].

On the other hand, the high dimensionality and complex non-linear underlying structure unfortunately makes the commonly used linear statistics inapplicable for articulated structures. A manifold learning algorithm of particular interest to this work is locally linear embedding [6]. It maps high-dimensional observation data that are presumed to lie on a nonlinear manifold, onto a single global coordinate system of lower dimensionality. Such a concept reveals the underlying structure of the data which can be used for statistical modeling. Inferring a model from the underlying manifold is a novel concept but far from being trivial. In this work, we tackle the problem of spinal deformity pathologies to model both global statistics of the articulated model and local shape variations of vertebrae based on local measures. We propose a spine inference/segmentation method from CT, where the model representation is optimized through a Markov Random Field (MRF) graph, balancing prior distribution with image data.

We introduce a deformable articulated spine instantiation through a statistical modeling of inter-object transformations. Our method is structured in two parts. The first relates to the creation of a nonlinear manifold embedding of spine articulations which can handle both small and large deformations in a given population. To this end, a novel articulated metric is introduced to create local linear patches. Second, we propose an inference framework using high-order MRF. This graph involves costs related to the imaged data, prior geometrical dependencies and global higher-order cliques. This paper is organized as follows. Section 2 presents the theoretical methodology for the manifold representation of articulated mesh models, while in Section 3 we propose the MRF-based inference framework. In Section 4 we present our evaluation results applied to images of the spinal column and the last section concludes the paper.

2 Manifold Embedding of Articulated Spine Models

The method performs an embedding of a training set of annotated vertebra shape constellations into a sub-space which dimensionality corresponds to the domain of admissible variations. Local shape is determined via analysis of variations in patch of the manifold. Fig. 1 illustrates a flowchart of the method.

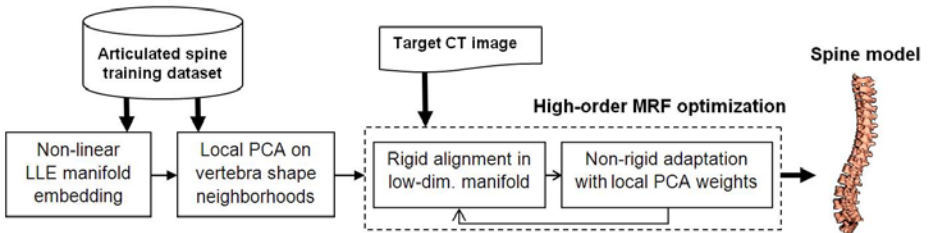


Fig. 1. Flowchart diagram of the proposed spine inference approach.

2.1 Representation of Articulated Spine Models

Our spine model $\mathbf{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_L\}$ consists of an interconnection of L vertebrae. For each vertebra \mathbf{s}_i , we recover a triangular mesh with vertices $\{\mathbf{v}_j^i | j = 1, \dots, V\}$, where the j^{th} vertex corresponds to approximately the same location from one shape to another and V the number of vertices. Additionally, every \mathbf{s}_i is annotated with landmarks on each model to rigidly register each object to its upper neighbor. Hence, an articulated deformable model (ADM) is represented by a vector of local inter-vertebral rigid transformations $A = [T_1, T_2, \dots, T_L]$. To perform global shape modeling of \mathbf{S} , we convert A to an absolute representation $\mathbf{A}_{\text{abs}} = [T_1, T_1 \circ T_2, \dots, T_1 \circ T_2 \circ \dots \circ T_L]$ using recursive compositions. The transformations are expressed in the local coordinate system (LCS) of the lower vertebra. Center of transformation is the intersection of all 3 vertebral axes, following antero-posterior, cranial-caudal and left-right directions. Rigid transformations described here are the combination of a rotation matrix R , a translation t and scaling s . We formulate the rigid transformation $T = \{s, R, t\}$ of a triangular mesh model as $y = sRx + t$ where $x, y, t \in \mathbb{R}^3$.

2.2 Nonlinear Manifold Embedding of Articulated Spine

Given a training set of N articulated spine shape models \mathbf{S} expressed by the absolute vector representation $\mathbf{A}_{\text{abs}}^i$ of dimensionality D , we seek their low-dimensional manifold \mathcal{M} with points $Y_i, Y_i \in \mathbb{R}^d$ where $d \ll D$. The set consisted of pre-reconstructed patient X-rays exhibiting a wide range of deformities (normal to severe). A free-form deformation was applied to obtain vertebral meshes with vertex correspondences. In all, 6 classes of spine deformations were included to represent the underlying population structure. In the sub-cluster corresponding to a pathological population, each individual point of the training set and its neighbours would lie within a locally linear patch on the manifold.

Nearest neighbor selection. In our approach, we adopt the intrinsic nature of the Riemannian manifold geometry allowing us to discern between articulated shape deformations in a topologically suited framework. The K closest neighbours are selected for each point using a distortion metric adapted for geodesic metrics, defined as $d_{\mathcal{M}}(\mathbf{A}_{\text{abs}}^i, \mathbf{A}_{\text{abs}}^j)$ which estimates the distance of models i, j . Distance measure can be expressed as a sum of the L articulation deviations:

$$d_{\mathcal{M}}(\mathbf{A}_{\text{abs}}^i, \mathbf{A}_{\text{abs}}^j) = \sum_{k=1}^L d_{\mathcal{M}}(T_k^i, T_k^j) = \sum_{k=1}^L \|c_k^i - c_k^j\| + d_G(R_k^i, R_k^j) \quad (1)$$

where the canonical representation encodes the intrinsic (c) and orientation (R) parameters. The first term evaluates intrinsic distances in the L_2 norm evaluating the inter-vertebral translations. The second defines a diffeomorphism d_G between rotation neighborhoods. We use the concept of geodesics to evaluate the unfolding in the manifold space \mathcal{M} of orientation vectors in order to estimate how both diverge from the tangent plane.

Forward and inverse manifold mapping. The manifold is built by locally linear embedding [6], where the low-dimensional representation of an absolute articulated vector $\mathbf{A}_{\text{abs}}^i$ is obtained from it’s neighbors $\mathbf{A}_{\text{abs}}^j$ for all data points. Hence, a new ADM can be inferred in the embedded d -space as a low-dimensional data point by finding its optimal manifold coordinates Y_i . To obtain the articulation vector for a new embedded point in the ambient space, one has to determine the representation in high-dimensional space based on its intrinsic coordinates. The inverse mapping of Y_i is performed by estimating the relationship between the D -space and manifold \mathcal{M} as a joint distribution. The manifold should follow a conditional expectation which captures the overall trend of the data in a local neighborhood of the manifold. Using *Nadaraya-Watson* regression, Gaussian kernels are used to estimate these densities in the conditional expectation setting [7]. By assuming the kernels G generalizes the expectation such that the observations are defined in terms of a metric $d_{\mathcal{M}}$ in manifold space:

$$f_{\text{NW}}(Y_i) = \underset{\mathbf{A}_{\text{abs}}^i}{\text{argmin}} \frac{\sum_{j \in \mathcal{N}(i)} G(Y_i, Y_j) d_{\mathcal{M}}(\mathbf{A}_{\text{abs}}^i, \mathbf{A}_{\text{abs}}^j)}{\sum_{j \in \mathcal{N}(i)} G(Y_i, Y_j)} \tag{2}$$

which integrates the distance metric $d_{\mathcal{M}}(\mathbf{A}_{\text{abs}}^i, \mathbf{A}_{\text{abs}}^j)$ defined in (II) and updates $f_{\text{NW}}(Y_i)$ using the closest neighbors of point Y_i in the manifold space. This constrains the regression to be valid for similar data points in its vicinity since locality around Y_i preserves locality in $\mathbf{A}_{\text{abs}}^i$.

2.3 Local Vertebra Appearances in the Manifold

The key idea of capturing vertebral shape appearance lies on the assumption that global models, represented in a local neighborhood of \mathcal{M} , will also manifest similar local geometries. The motivation stems from the fact that global shape deformation belonging to the same class will induce similar local biomechanical patterns, creating morphologically comparable vertebrae. We assume here that vertebra appearances follow a linear distribution within the low-dimensional manifold. Hence, given a data point Y_j and its K neighbors, the local shape model \mathbf{s}_i , representing the i^{th} element of the ADM, is obtained by building a particular class of shapes given the set of examples $\{\mathbf{s}_i^1, \dots, \mathbf{s}_i^K\}$. We approximate the distribution of the shape using a parameterized linear model by computing the deformation vectors formed for the $K - 1$ shape samples. We compute the n eigenvalues and corresponding eigenvectors v so that a new vertebra $\mathbf{s}_i^{\text{new}} = \bar{\mathbf{s}}_i + [v_1 \dots v_n][\omega_1 \dots \omega_n]$ can be instantiated where $\bar{\mathbf{s}}_i$ is the mean shape of the K neighboring local objects and $\mathbf{w} = [\omega_1 \dots \omega_n]$ the weight vector. This step warps individual instances to infer new local vertebra models.

3 Inference through MRF Optimization

Once an appropriate modeling of spine shape variations is determined, a successful inference between the image and manifold must be accomplished. We describe

here how a new model is deformed. We search the optimal embedded manifold point $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_d)$ of the global spine model. Such a strategy offers an ideal compromise between the prior constraints, as well as the individual shape variation described by the weight vector $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)$ in a localized sub-patch. The energy E of inferring the model \mathbf{S} in the image \mathcal{I} is a function of the set of displacement vectors Δ in the manifold space for global shape representation. This involves: (a) a data-related term expressing the image cost and (b) a global prior term measuring deformation between low-dimensional vectors with shape models. The third term represents (c) a higher-order term which is expressed by the reconstruction weights Ω for local vertebra modeling. The energy E can be expressed as the following combination of a global and local optimization:

$$E(\mathbf{S}^0, \mathcal{I}, \Delta, \Omega) = V(\mathbf{Y}^0 + \Delta, \mathcal{I}) + \alpha V(\mathbf{N}, \Delta) + \beta V(\mathbf{H}, \Delta, \Omega). \quad (3)$$

3.1 Rigid Alignment of the Spine

The global alignment of the model with the target image primarily drives the deformation of the ADM. The purpose is to estimate the set of articulations describing the global spine model by determining its optimal representation \mathbf{Y}^0 in the embedded space. This is performed by obtaining the global representation using the mapping in (2) so that: $f_{\text{NW}}(Y_i + \Delta) = f_{\text{NW}}(\{y_1 + \delta_1, \dots, y_d + \delta_d\})$. This allows to optimize the model in manifold space coordinates while retrieving the articulations in \mathcal{I} . The global cost can be expressed as:

$$V(\mathbf{Y}^0 + \Delta, \mathcal{I}) = V(f_{\text{NW}}(\{y_1 + \delta_1, \dots, y_d + \delta_d\}), \mathcal{I}). \quad (4)$$

The inverse transform allows to obtain $\mathbf{A}_{\text{abs}}^i + \mathbf{D}$, with \mathbf{D} as deformations in the image space. Since the transformations T_i are implicitly modeled in the absolute representation $\mathbf{A}_{\text{abs}}^0$, we can formally consider the singleton image-related term as a summation of costs associated with each L vertebra of the ADM:

$$V(\mathbf{A}_{\text{abs}}^0 + \mathbf{D}, \mathcal{I}) = \sum_{i=1}^L V_i(\mathbf{s}_i * (T_i^0 + \mathbf{d}_i), \mathcal{I}) \quad (5)$$

where $V_i(\mathbf{s}, \mathcal{I}) = \sum_{\mathbf{v}_i \in \mathbf{s}} \mathbf{n}_i^T(\mathbf{v}_i) \nabla \mathcal{I}(\mathbf{v}_i)$ minimizes the distance between mesh vertices of the inferred ADM and gradient image \mathcal{I} by a rigid transformation. Here, \mathbf{n}_i is the normal pointing outwards and $\nabla \mathcal{I}(\mathbf{v}_i)$ the image gradient at \mathbf{v}_i .

The prior constraint for the rigid alignment are pairwise potentials between neighboring models y_i such that the difference in manifold coordinates is minimal with regards to a prior distribution of neighboring distances P :

$$\alpha V(\mathbf{N}, \Delta) = \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^0 + \delta_i, y_j^0 + \delta_j, P). \quad (6)$$

This term represents the smoothness term of the global cost function to ensure that the deformation δ_i applied to point coordinates are regular, with $V_{ij} = (0, 1)$ a distance assigning function based on the distances to P .

3.2 Non-rigid Adaptation of Vertebral Shapes

Finally, local shape geometry for each of the ADM’s vertebrae is obtained by varying the weight parameters of the principal variations. We parameterize these potentials with a set \mathcal{C} of clique variables c , controlled by high-order potential V_c [8] which assigns a cost to a configuration of c . Each clique are assigned to weight vectors ω_c . Hence the third term of (3) is described as a high-order functional:

$$\beta V(\mathbf{H}, \Delta, \Omega) = \beta \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^0 + \omega_c) \tag{7}$$

where independent clique variables c are treated as a graph minimization problem. The prior term is represented by higher-order potentials of degree n , based on the eigenvalues of the L local vertebrae from our model \mathbf{S} . Our work is inspired from a mesh reconfiguration [9] where costs are associated to cliques c based on the positions of the morphed mesh vertices \mathbf{v}_i . A search is performed along the normal \mathbf{n}_i from \mathbf{v}_i to find the optimal compromise between boundary detection and the distance to the mean eigenvalue shape. We therefore penalize deformations which deviates from the local distribution.

One can integrate the global data and prior terms along with local shape terms parameterized as the higher-order cliques, by combining (4), (6) and (7):

$$E(\mathbf{S}^0, \mathcal{I}, \Delta, \Omega) = V(f_{\text{NW}}(\{y_1 + \delta_1, \dots, y_d + \delta_d\}, \mathcal{I})) + \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^0 + \delta_i, y_j^0 + \delta_j) + \beta \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^0 + \omega_c). \tag{8}$$

3.3 Energy Minimization

The optimization strategy of the resulting MRF (8) in the continuous domain is not a straightforward problem. The convexity of the solution domain is not guaranteed, while gradient-descent optimization approaches are prone to non-linearity and local minimums. We seek to assign the optimal labels $\mathcal{L}^\Delta = \{l_1, \dots, l_d\}$ and $\mathcal{L}^\Omega = \{l_1, \dots, l_n\}$ which are associated to the quantized space Δ of displacements and local weight parameters Ω respectively. We consider that displacing the coordinates of point y_i^0 by δ^{l_i} is equivalent to assigning label l_i to y_i^0 . An incremental approach is adopted where in each iteration t we look for the set of labels that improves the current solution s.t. $y_i^t = y_i^0 + \sum_t \delta^{l_i t}$, which is a temporal minimization problem. Then (8) can be re-written as:

$$E^t(\mathcal{L}^\Delta, \mathcal{L}^\Omega) = V(f_{\text{NW}}(\{y_1^{t-1}, l_1^\Delta, \dots, y_d^{t-1}, l_d^\Delta\}, \mathcal{I})) + \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^{t-1}, y_j^{t-1}, l_i^\Delta, l_j^\Delta) + \beta \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^{t-1}, l_c^\Omega). \tag{9}$$

We solve the minimization of the higher-order cliques in (9) by transforming them into quadratic functions [8]. We apply the FastPD method [10] which solves the problem by formulating the duality theory in linear programming.

4 Experiments and Results

In order to evaluate the performance, we considered modeling pathological spinal columns for CT inference. We used a database of $N = 711$ spine models reconstructed in 3D and exhibiting different types of deformations relative to global and local shapes. For each spine, 6 landmarks on each of the 17 vertebrae composing the spinal column were used to extrapolate the inter-object transformation (Fig. 2(a)). Optimal neighborhood size was found at $K = 10$, while intrinsic dimensions was $d = 7$ and $n = 5$, dictating the number of nodes in our global graph model. α and β balance the contribution of the energy terms and set at 0.3 and 0.5 respectively. We tested the algorithm on a subset of 20 unseen 3D reconstructed cases from the database (modifying the modular data term to vertex correspondences) and 12 CT volumes. We quantitatively compared our method to an AAM modeling based on global PCA. Dice scores and root-mean-square (RMS) landmark distances show improvement of the proposed MRF approach via a non-linear shape analysis in Fig. 2(b). Successful examples from CT inferred data are shown in Fig. 2(c). We evaluated the performance based on the density of the input sample points, affecting the global shape inference starting when only 20% of points are available. Furthermore, added Gaussian distributed noise

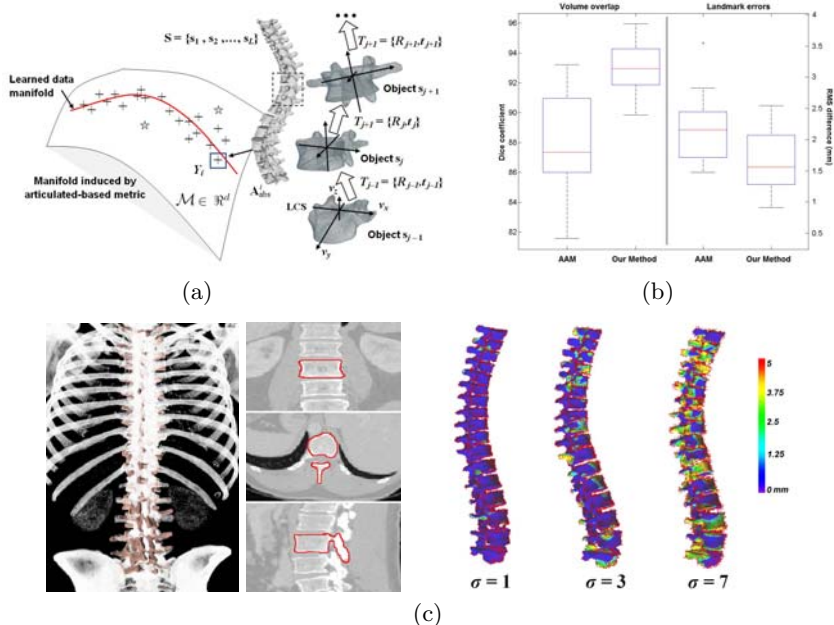


Fig. 2. (a) Representation of inter-vertebral transformations. (b) Plots of Dice coefficients and landmark errors comparing our method to AAM. (c) Spine inferences of triangulated from CT with orthogonal views. Local shape distortions with significant noise level increase σ added to target points (error-coded for ground-truth distances).

to the target data introduces distortions when $\sigma = 3$ (Fig. 2). These validations prove how our method elegantly encodes prior knowledge with image constraints in an MRF framework, and efficiently minimizes the energy term to converge towards an optimal solution. One drawback remains the computational time due to the inverse regression mapping and higher-order clique potential minimization.

5 Discussion and Future Work

Our main contribution consists in modeling complex, non-linear patterns of spine deformations in a Riemannian manifold. Point-based models are created from statistical knowledge in terms of global and local variations. To this end, an articulated distance metric based on intrinsic and orientation properties was proposed. The non-linear embedding is constructed in such a way to avoid creating shape distortions, as well as collisions between shapes. This is accomplished by constraints within the manifold parameters to restrict outlying values corrupting global shape appearance. The proposed framework can be extended in medical imaging applications to allow global/local shape modeling. While the method remains computationally expensive for the inverse map, inference of articulated models based on optimal clique decomposition can be beneficial to this end.

References

1. Nain, D., Haker, S., Bobick, A., Tannenbaum, A.: Multiscale 3-D shape representation and segmentation using spherical wavelets. *T. Med. Imag.*, 598–618 (2007)
2. de Bruijne, M., Lund, M., Tanko, L., Pettersen, P., Nielsen, M.: Quantitative vertebral morphometry using neighbor-conditional shape models. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 1–8. Springer, Heidelberg (2006)
3. Boisvert, J., Chretien, F., Pennec, X., Labelle, H., Ayache, N.: Articulated spine models for 3-D reconstruction from partial radiographic data. *T. Biomed. Eng.* 11, 2565–2574 (2008)
4. Klinder, T., Wolz, R., Lorenz, C., Franz, A., Ostermann, J.: Spine segmentation using articulated shape models. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 227–234. Springer, Heidelberg (2008)
5. Kadoury, S., Paragios, N.: Surface/volume-based articulated 3D spine inference through Markov Random Fields. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 92–99. Springer, Heidelberg (2009)
6. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326 (2000)
7. Davis, B., Fletcher, P., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: *ICCV*, pp. 1–7 (2007)
8. Rother, C., Kohli, P., Feng, W., Jia, J.: Minimizing sparse higher order energy functions of discrete variables. In: *CVPR*, pp. 1382–1389 (2009)
9. Kaus, M., Pekar, V., Lorenz, C., Truyen, R., Lobregt, S., Weese, J.: Automated 3D PDM construction from segmented images using deformable models. *IEEE T. M. Imag.* 22, 1005–1013 (2003)
10. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic mrf: Setting the state of the art with primal dual strategies. *CVIU* 112(1), 14–29 (2008)

Improved Method for Point-Based Tracking

Andrei Danilchenko¹, Andrew D. Wiles^{2,3},
Ramya Balachandran⁴, and J. Michael Fitzpatrick¹

¹ Department of Electrical Engineering and Computer Science, Vanderbilt University,
Nashville, TN, USA

² Northern Digital Inc., Waterloo, ON, Canada

³ Robarts Research Institute, The University of Western Ontario, London, ON,
Canada

⁴ Department of Otolaryngology, Vanderbilt University, Nashville, TN, USA

Abstract. Image-guided surgery systems have a wide range of applications where the level of accuracy required for each application varies from millimeters to low sub-millimeter range. In systems that use optical tracking, it is typical to use point-based registration without any weighting schemes to determine the pose of the tracked tool with very good accuracy. However, recent advancements in methods to estimate the measurement uncertainty for each tracked marker and the development of an anisotropically weighted point-based registration algorithm have allowed for the optical tracking accuracy to be improved. In this article, we demonstrate a new tracking method that improves the tracking accuracy by 20 – 45% over the traditional tracking methodology.

1 Introduction

Image-guided surgery (IGS) systems have improved the standard of care in brain, spine and orthopaedic interventions by combining pre-operative medical images and virtual reality using spatial tracking technologies [1]. Recently, real-time imaging techniques, such as ultrasound and endoscopy, and robotics technology have been integrated with traditional IGS systems by tracking the imaging and robotic devices similarly to tracked surgical tools. The real-time images and robotic devices are displayed in the virtual reality environment resulting in an augmented view of the surgical target. One such application is the robotic drilling system for a mastoidectomy [2] where a section of bone is resected from behind the ear of a patient for various otolaryngology procedures. In this application, an optical tracking system (NDI Polaris Spectra[®], Waterloo, ON, Canada) is used to track the poses of both the patient and the robot. If the patient moves during the drilling procedure, the drilling plan is automatically updated to reflect the change in patient positioning using the real-time information provided by the optical tracking system. This procedure serves as an example of applications in which a high degree of tracking accuracy is crucial to success. Such applications motivate the focus of this paper, which is the development and validation of an improved method for highly accurate optical tracking.

The method presented in this paper is the combination of two recently developed algorithms: (i) a method to estimate the anisotropic fiducial localizer error (FLE) at each of the optically tracked markers [3] and (ii) a registration procedure that accounts for the anisotropic weighting at each of the markers [4]. We show that the FLE covariance statistics can be used as an anisotropic weighting function in the registration procedure. Accuracy improvements of 20 – 45% are demonstrated and this improvement is deemed important for applications that rely on highly accurate real-time tracking such as fully automatic robotic mastoidectomy.

2 Method

The goal of this paper is to demonstrate that the anisotropic weighted point-based registration algorithm [4] provides better results than the standard isotropic point-based registration method [5,6]. The comparison is done using the target registration error (TRE) statistics for both registration methods.

2.1 Computing the Real-Time Weightings

The weighting for the anisotropic registration algorithm at each marker is determined using the FLE estimates from the algorithm presented in [3]. The FLE covariance matrix at a given marker a is found by solving a complex set of linear equations that relates the FLE statistics at each marker to the estimated FRE statistics: $\mathbf{A}_a \mathbf{x}_a = \mathbf{b}_a$ where \mathbf{A}_a is based on the geometry of the tracked rigid body, \mathbf{x}_a is the six independent FLE covariance components and \mathbf{b}_a is the six independent fiducial registration error (FRE) components estimated from the previous M frames. For an optical tool with N tracked markers, there are N sets of linear equations that can be solved in order to obtain an estimate of the FLE at each marker. However, to improve numerical stability we can solve an overdetermined set of equations for a single FLE covariance matrix by stacking the matrices and vectors such that

$$\mathbf{x}_{avg} = \left(\mathbf{A}_{stack}^T \mathbf{A}_{stack} \right) \mathbf{A}_{stack}^T \mathbf{b}_{stack} \quad (1)$$

where

$$\mathbf{A}_{stack} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_N \end{bmatrix} \quad \text{and} \quad \mathbf{b}_{stack} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_N \end{bmatrix}. \quad (2)$$

Taking the six independent FLE covariance components from \mathbf{x}_{avg} (with or without stacking) and rewriting them as a FLE covariance matrix $\mathbf{\Sigma}$, the weightings are computed as

$$\mathbf{W} = \mathbf{\Sigma}^{-1/2}. \quad (3)$$

2.2 Obtaining a Good Estimate of the True Target Location

In the work previously presented in the literature, the algorithms were tested using Monte Carlo simulations whereby the true location of the target was known exactly. However, experimentally the true target location is never known exactly and therefore a method of estimating the true target location is needed so that the TRE can be computed. Here we use two rigidly connected optically tracked tools shown in Fig. 1 to solve this problem.

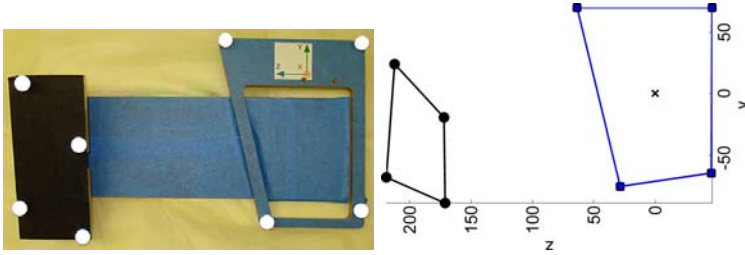


Fig. 1. Photograph and schematic of the tracked rigid bodies used in the experiment. The tracked tool under test is on the left and the tracked tool used to estimate the true location of the target is on the right. The target is the centroid of the right-hand tool, which is marked by an \times in the schematic. Measurements are in millimeters and coordinate frame is the local tool coordinate frame.

In Fig. 1, the four markers on the left represent tracking markers rigidly attached to a tool (e.g., drill or hand-held pointer tool). The four markers on the right are arranged so that, for the robotic application, the centroid (denoted by \times in the schematic) is at the tip of the drill and for the hand-held-tool application the centroid is at the tip of the probe pointer. Previous to the tracking experiments a standard ‘tool definition’ calibration was carried out. The robotic arm was held stationary in several poses while the positions of the tool markers on the left and the target markers on the right were measured repeatedly (1000s of times). The average of these measurements over the various poses for each marker provides a highly accurate standard configuration. The centroid of the four markers on the right is defined to be the ‘target’ position (e.g., drill tip, or probe pointer tip). Then, during the tracking experiments (see Step iv below), each detected configuration of the tool is registered to the standard tool. Therefore, using the right-hand rigid body to estimate the true location of the target, the target is estimated with target localization error (TLE) statistics of

$$\Sigma_{\text{tle}} = \frac{\Sigma_{\text{fle},1} + \Sigma_{\text{fle},2} + \dots + \Sigma_{\text{fle},N}}{N^2}, \quad \text{RMS}_{\text{tle}} = \sqrt{\text{trace}(\Sigma_{\text{tle}})} \quad (4)$$

where N is the number of markers on a tool ($N = 4$ for our example) and RMS is the root-mean-square statistic. If the FLE is homogeneous across the markers, then the covariance matrix and RMS reduce to $\Sigma_{\text{tle}} = \Sigma_{\text{fle}}/N$ and $\text{RMS}_{\text{tle}} = \text{RMS}_{\text{fle}}/\sqrt{N}$, respectively.

2.3 Comparing Isotropic and Anisotropic Registrations

The TRE statistics of the isotropic and anisotropic registrations algorithms are compared using the centroid of the right-hand tracked tool as the target ground truth. The two rigid bodies are rigidly attached to one another and moved together. The experimental protocol is to carry out these steps for every frame of data returned from the tracking system:

- (i) The 3D positions of each of the markers on both rigid bodies are measured with the optical tracking system at an instance in time which we will refer to as a frame of data.
- (ii) The target location, \mathbf{r}_{ref} , is obtained by taking the mean of the four markers on the right-hand tracked tool.
- (iii) The right-hand tracked tool is registered using isotropic registration and the FLE estimates are updated for these markers. The TLE statistics of the target are obtained using (4).
- (iv) The left-hand tracked tool is registered using isotropic registration and the FLE estimates are updated for these markers. The target location, \mathbf{r}_{iso} , is computed using the transform computed from isotropic registration.
- (v) The left-hand tracked tool is registered using anisotropic registration using the FLE estimates of the markers for the weighting as per (3). The target location, \mathbf{r}_{aniso} , is computed using the transform computed from anisotropic registration.
- (vi) The TRE vectors for both registration methods are computed such that $\mathbf{tre}_{iso} = \mathbf{r}_{iso} - \mathbf{r}_{ref}$ and $\mathbf{tre}_{aniso} = \mathbf{r}_{aniso} - \mathbf{r}_{ref}$.

After 1000 frames of data are collected, the results are plotted and a set of observational statistics is computed. The results section provides a comparison of the algorithms for different dynamic paths over which the tool traveled and also provide examples of the FLE estimates for the markers obtained with the algorithm in [3].

3 Results

Using the method described in Section 2, we compared our new tracking method, performed using anisotropic registration [4], with traditional tracking method, performed using isotropic registration [5,6]. The tracked tools were placed at a distance of approximately 160 cm from the tracking system. We noticed that the results of the comparisons are distinct for different types of motion, thus we collected data under the following conditions:

- Test A: translate approximately 10 cm parallel to Polaris' x-axis (up-down)
- Test B: translate approximately 10 cm parallel to Polaris' y-axis (left-right)
- Test C: translate approximately 10 cm parallel to Polaris' z-axis (front-back)
- Test D: translate in all directions
- Test E: rotate in all directions
- Test F: random path including translations and rotations

Table 1. Results of tests A-F. All statistics are computed over the usable 800 frames. The estimate of the TLE RMS is provided using (4). The TRE RMS statistics are computed for all the distance errors in a given test. The percent difference between the two methods is shown using the the isotropic RMS as the reference. Finally, noting that the TLE contributes to the measured TRE RMS statistic, see (5), we correct the TRE RMS statistics and recompute the TRE RMS percent difference.

Test	TLE RMS	TRE RMS	TRE RMS	TRE RMS	TRE RMS
		Isotropic	Anisotropic	% Diff.	% Diff Corr.
A	0.08 mm	0.22 mm	0.18 mm	-19.3%	-22.6%
B	0.07 mm	0.33 mm	0.19 mm	-42.4%	-45.2%
C	0.08 mm	0.32 mm	0.18 mm	-42.8%	-46.8%
D	0.06 mm	0.34 mm	0.24 mm	-28.7%	-29.9%
E	0.08 mm	0.60 mm	0.46 mm	-23.2%	-23.7%
F	0.06 mm	0.34 mm	0.22 mm	-34.5%	-36.0%

A total of 1000 frames of data was collected for each test. Since a sliding window of 200 frames was used to estimate the FLE, only the last 800 frames are used for statistical analysis because it takes 200 frames until the FLE estimate stabilizes.

Three key sets of results are presented in Table 1. First, an estimate of the TLE RMS is given using (4) with estimates of the FLE covariance matrices found from the same FLE estimation algorithm used to determine the weightings for the anisotropic registration algorithm.

Next, the TRE RMS computed for both the isotropic and anisotropic registrations are provided. We notice that the anisotropic TRE RMS is lower for each test and the percent difference between the two statistics is provided (4).

Finally, we note that the TRE RMS for each registration has a contribution from the TLE where the measured TRE RMS can be related to the actual TRE RMS by

$$\text{RMS}_{\text{tre,meas}}^2 = \text{RMS}_{\text{tre,actual}}^2 + \text{RMS}_{\text{tle}}^2. \quad (5)$$

Taking into consideration this relationship, we correct the percent differences between isotropic and anisotropic TRE RMS statistics by using the $\text{RMS}_{\text{tre,actual}}$ and show this new percent difference in the last column of Table 1. A small increase in the accuracy is observed with this correction.

Moving beyond the observational statistics, we provide details of the data measured during Test F. In Fig. 2, the FLE RMS estimates are provided for (i) each marker estimated individually and (ii) the average FLE estimated by solving the over-determined system of equations in (1).

In Fig. 3, the magnitudes in the three principal directions of the average FLE covariance matrix are given. Here we observe the common behavior of optical tracking systems where one of the components is much larger than the other two directions. The direction of the higher magnitude error is along the viewing direction of the optical tracking system (z -axis of the Polaris Spectra).

¹ The percent difference is computed using the isotropic registration as the reference value so that $\% \text{Diff} = 100 \times (\text{RMS}_{\text{tre,aniso}} - \text{RMS}_{\text{tre,iso}}) / \text{RMS}_{\text{tre,iso}}$.

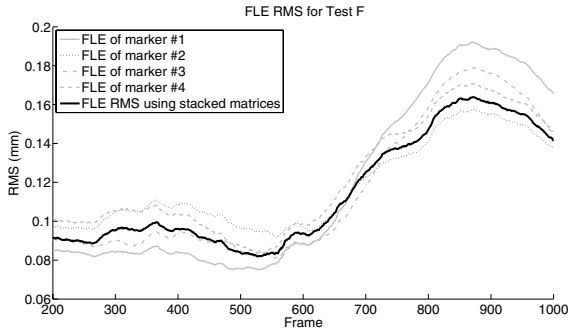


Fig. 2. Sample of the FLE RMS for Test F. The RMS is computed for (i) each marker individually and (ii) using the stacked, over-determined system of equations described in (II).

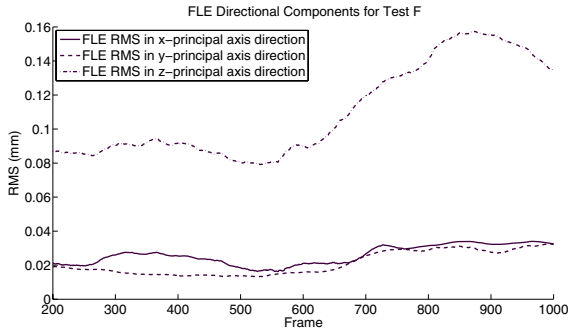


Fig. 3. Sample of the FLE directional components for Test F. The directional components are singular values of the covariance matrix where the principal axes of the covariance matrix are approximately aligned with the axes of the optical tracking system.

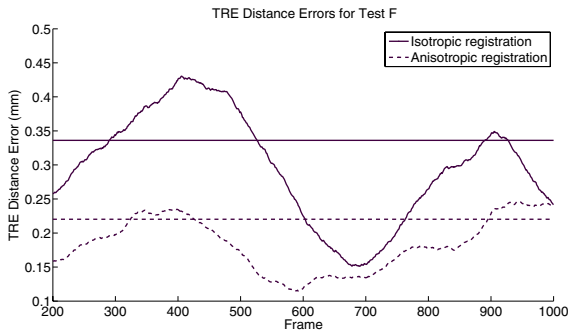


Fig. 4. TRE distance errors for Test F. The distance errors are smoothed using a moving average of 200 frames to better view the trend. Horizontal lines represent the RMS for each test.

The last plot in Fig. 4 shows the TRE distance errors at each frame in Test F. We clearly demonstrate here that the anisotropic registration provides a better estimate of the target location. The TRE RMS computed over the entire data collection is shown with the horizontal lines. We also performed a paired t-test for this case and found $p \ll 0.05$, suggesting that the difference between isotropic and anisotropic TREs is statistically significant.

4 Discussion

We have developed an improved method for point-based tracking of rigid bodies. The improvement is accomplished by replacing the isotropic point-based registration method typically employed with one that takes into consideration the anisotropy of tracking system's FLE. Our method combines a recently published algorithm [3] for estimating the covariance matrix of anisotropic FLE with a novel rigid registration algorithm [4] that accommodates anisotropic weighting. Incorporating automatically generated weighting matrices optimized from the estimated FLE covariance matrices into the anisotropic registration algorithm exhibits a surprising level of accuracy, surpassing the current state of the art. Further improvements might be made by controlling for outliers or applying additional weightings during the solution of the equations, such as maximum-likelihood methods, M-estimators, regularization, etc. [7,8].

The method was tested on data obtained with NDI Polaris Spectra[®] with various motions of the tracked tool in the region approximately 160 cm from the tracking system. For a variety of motions, including pure translation, pure rotation, and combinations of both, we measured an increase in tracking accuracy in the range of 20 – 45%, and in every case accuracy was improved. A statistical analysis confirmed that the tracking of optical rigid bodies using our new approach is more accurate than tracking using the standard method ($p \ll 0.05$).

The input to each FLE calculation is a set of previous FRE vectors collected from the most recent M frames, using the isotropic registration algorithm. We found that the size of the window, M , used for estimating the FLE statistics is an important factor in the tracking accuracy for our algorithm. The size of the window was investigated experimentally and we found that a contiguous collection of $M = 200$ frames of FRE measurements produced the best results. Larger windows gave only a very small improvement in accuracy but also increased the lag in obtaining a good estimate of the FLE statistics. With a window size of $M = 200$, one must allow the system to perform 200 data collections, before the improved tracking can commence. For the experiments described in this article, our method was implemented to run off-line in Matlab (Version 2009b, Mathworks, Inc., Natick, MA) on an Intel Core 2 Duo 2.2 GHz with 2GB of RAM while utilizing only one CPU. Each complete update, comprising FLE estimation and registration, required less than 6.4 milliseconds. For the NDI Polaris Spectra[®], whose update rate is 60 Hz (period of 17 ms), our improved tracking algorithm can run in real-time. Furthermore, at 60 Hz with the NDI Polaris Spectra[®], a delay of only four seconds is required to obtain the first set of 200

frames of FRE data to begin providing accurate estimates of the FLE statistics. In summary, the algorithm has promise to be included in imaged-guided surgery applications that use point-based registration for highly accurate optical tracking.

Acknowledgements

This work was supported by the NIH. Three authors (AD, RB, and JMF) gratefully acknowledge salary support via two grants—R01 DC008408 and R01 DC010184. The robot was purchased via R21EB002886.

References

1. Peters, T.M.: Image-guidance for surgical procedures. *Physics in Medicine and Biology* 51(14), R505–R540 (2006)
2. Danilchenko, A., Balachandran, R., Toennies, J.L., Baron, S., Munske, B., Wester III, R.J., Labadie, R.F.: Robotic mastoidectomy. In: Abstracts Selected for the American Otological Society Spring Meeting, Las Vegas, NV, USA (May 2010)
3. Wiles, A.D., Peters, T.M.: Real-time estimation of FLE statistics for 3D tracking with point-based registration. *IEEE Transactions on Medical Imaging* 28(9), 1384–1398 (2009)
4. Balachandran, R., Fitzpatrick, J.M.: Iterative solution for rigid-body point-based registration with anisotropic weighting. In: Miga, M.I., Wong, K.H. (eds.) *Proceedings of SPIE, Medical Imaging, Visualization, Image-Guided Procedures, and Modeling*, Orlando, FL, USA, vol. 7261, p. 72613D. SPIE, San Jose (February 2009)
5. Schönemann, P.H.: A generalized solution of the orthogonal procrustes problem. *Psychometrika* 31, 110 (1966)
6. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A* 4(4), 629–642 (1987)
7. Huber, P.J.: *Robust Statistical Procedures*. SIAM, Philadelphia (1996)
8. Mair, E., Strobl, K.H., Suppa, M., Burschka, D.: Efficient camera-based pose estimation for real-time applications. In: *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2696–3703 (October 2009)

A Texton-Based Approach for the Classification of Lung Parenchyma in CT Images

Mehrdad J. Gangeh¹, Lauge Sørensen², Saher B. Shaker³,
Mohamed S. Kamel¹, Marleen de Bruijne^{2,4}, and Marco Loog⁵

¹ Department of Electrical and Computer Engineering, University of Waterloo, Canada
{mgangeh, mkamel}@pami.uwaterloo.ca

² Department of Computer Science, University of Copenhagen, Denmark
{lauges, marleen}@diku.dk

³ Department of Respiratory Medicine, Gentofte University Hospital, Hellerup, Denmark

⁴ Biomedical Imaging Group Rotterdam, Erasmus MC, The Netherlands

⁵ Pattern Recognition Laboratory, Delft University of Technology, The Netherlands
m.loog@tudelft.nl

Abstract. In this paper, a texton-based classification system based on raw pixel representation along with a support vector machine with radial basis function kernel is proposed for the classification of emphysema in computed tomography images of the lung. The proposed approach is tested on 168 annotated regions of interest consisting of normal tissue, centrilobular emphysema, and paraseptal emphysema. The results show the superiority of the proposed approach to common techniques in the literature including moments of the histogram of filter responses based on Gaussian derivatives. The performance of the proposed system, with an accuracy of 96.43%, also slightly improves over a recently proposed approach based on local binary patterns.

1 Introduction

Computerized quantitative analysis in pulmonary computed tomography (CT) images is a vital tool in the analysis of chronic obstructive pulmonary disease (COPD). The disease is projected to become the fifth most burdening disease worldwide by 2020 [1]. COPD is a chronic lung disease characterized by limitation of airflow, and it comprises two components: small airway disease and emphysema, which is characterized by gradual loss of lung tissue.

Current methods for the diagnosis and quantification of COPD suffer from several limitations. Common computerized methods on CT images do not use all the information available in a CT image. For example, the relative area of emphysema below a threshold (RA) [2] considers only independent pixel intensity values and relies on a hand-picked parameter, the threshold. The primary diagnostic tool for COPD is spirometry by which various pulmonary function tests (PFTs) are performed [1]. These are cheap and fast to acquire, but they have a low sensitivity to early stages of COPD.

This work focuses on improving the assessment of emphysema in CT images. Emphysema lesions, or bullae, are visible in CT images as areas of abnormally low attenuation values close to that of air, and it is often classified into three subtypes [3].

These subtypes are the following: centrilobular emphysema (CLE), defined as multiple small low-attenuation areas; paraseptal emphysema (PSE), defined as multiple low-attenuation areas in a single layer along the pleura often surrounded by interlobular septa that is visible as thin white walls; and panlobular emphysema (PLE), defined as a low-attenuation lung with fewer and smaller pulmonary vessels.

As the texture of lung tissue is affected by the type of disease, texture analysis can be used for quantitative assessment of different subtypes of emphysema. Classifying emphysema in CT images of the lung using texture features was first introduced in [4]. Since then, various features have been used for the classification of emphysema and other disorders in lung CT images including moments of histograms computed on the outputs of filter banks consisting of Gaussian derivatives [5], measures on gray-level co-occurrence matrices (GLCM), measures on gray-level run-length matrices (GLRLM), and moments of the attenuation histogram [4, 6, 7].

Recently, it was shown that small-sized local operators like local binary patterns (LBP) [8] and the patch representation of small local neighborhood in texton-based approaches [9] yield excellent texture classification performance on standard texture databases. It should be noted here that small-sized local operators are desirable in situations where the region of interest (ROI) is rather small, which is often the case in texture analysis in medical imaging, where pathology can be localized in small areas. This is because of two reasons: first, convolution with large support filter banks suffers from boundary effects; second, more patches can be extracted using small-sized local operators that makes the estimation of image statistics more reliable [9].

In this paper, we propose to use small patch representation in texton-based approaches along with support vector machines (SVMs) for the classification of emphysema in CT images of the lung. To our knowledge, this technique has never been used for the classification of CT lung images.

The effectiveness of small-sized local operators in medical imaging is shown in [10] using LBP texture features and k -NN classifier with similar results to filter bank approaches based on Gaussian derivatives. In this work, we also show that texton-based approaches using a SVM with radial basis function (RBF) kernel produces better results than common filter bank approaches and slightly better results than LBP, which can be considered as the state of the art in emphysema classification [10].

2 Texton-Based Texture Classification

In this section, the principle of texton-based texture classification is reviewed [9, 11, 12]. This approach is independent of the representation used to describe local image information, i.e., it could be raw pixel representation, outputs of filter banks convolved with the patches, or even more complex representations. The texton-based approach can be divided into three stages: 1) construction of a codebook of textons using a clustering algorithm such as k -means, 2) learning texton histograms from the training set, and 3) classification of the test set by finding the histogram of textons in the test set and comparing to those found during stage two to find the nearest pattern. These three steps are explained in the remaining of this section.

2.1 Construction of Texton Codebook

To construct the texton codebook, small-sized local patches are randomly extracted from each image in the training set. These small patches are then aggregated over all images in a class and clustered using an unsupervised algorithm such as k -means. Obtained cluster centers form a dictionary that represents the class of textures used. It is saved as the codebook to be used in the next stage. Fig. 1 displays sample images of lung CT ROIs used in this paper and the codebook dictionary computed over all ROIs using the texton size of 7×7 pixels and $k = 40$ in k -means.

2.2 Learning the Model

The next stage is to find the features (learn the model) using the images in the training set. To this end, these steps are followed: first, extract small patches of the same size as the previous stage by sliding a window over each training image in a class. Second, find the distance between each patch to all textons in the dictionary to find the closest match. Third, update a histogram of textons accordingly for each image based on the closest match found. This yield a histogram for each image in the training set, which is used as the features representing that image after normalization. These features are used for training a classifier such as SVM. Fig. 2 illustrates the construction of the codebook and learning the model in a texton-based classification system.

2.3 Classification

To classify a test image, the same steps as in the learning stage are followed to find the features for the test image. This includes extraction of small patches from each test image in a class, finding the closest match to these patches from the dictionary, and computing the normalized histogram of obtained closest textons to define a feature vector for the image. The trained classifier in the learning stage is used to find the class of the test image.

Both k -NN and SVM are tested in this paper for the classification of texton-based features. In SVM, a RBF kernel as given in (1) is used as it is recommended as the first kernel choice in [13]. In (1), γ is the kernel width and \mathbf{x}_i and \mathbf{x}_j are two sample patterns.

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}. \quad (1)$$

3 Experimental Setup

Data Preparation. The data used for the experiments is the same as in [10], which is collected from a set of thin-slice CT images of the thorax from an exploratory study carried out at the Department of Respiratory Medicine, Gentofte University Hospital, Denmark [14]. The slices were reconstructed using a high spatial resolution (bone) algorithm. Each subject was scanned in the upper, middle, and lower lung, resulting in three 1.25 mm thick slices with a resolution of 0.78×0.78 mm per subject.

The leading pattern in 75 CT slices from 25 subjects, 8 healthy non-smokers, 4 smokers without COPD, and 13 smokers diagnosed with moderate or severe COPD



Fig. 1. Sample ROIs of size 50×50 pixels (*left*) in three classes, i.e., normal lung (*top left row*), CLE (*middle left row*), and PSE (*bottom left row*). The constructed codebook using texton sizes of 7×7 pixels and $k = 40$ in k -means (*right*).

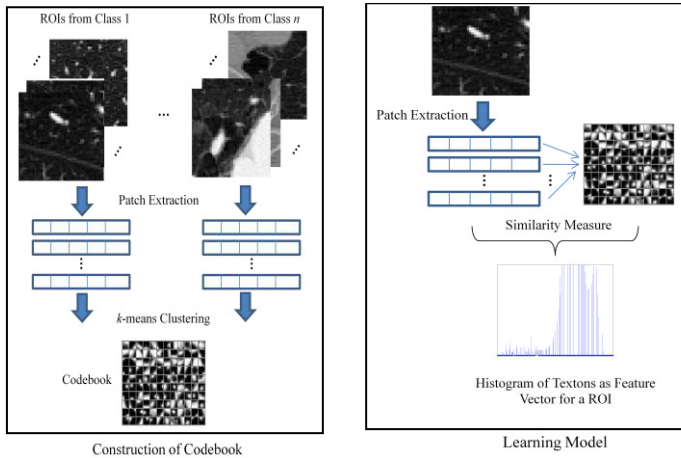


Fig. 2. The illustration of different stages of a texton-based texture classification system: the generation of texton codebooks using k -means clustering (*left*) and the generation of features by computing the texton histograms of training set (*right*).

according to PFTs [1], was visually assessed by an experienced chest radiologist and a CT experienced pulmonologist independently. The leading pattern could either be normal tissue (NT), CLE, PSE, or PLE, in each of the slices, and consensus readings were obtained in all cases of disagreement. 168 non-overlapping ROIs of size 50×50 pixels were subsequently annotated in the slices representing the three classes: NT (59 ROIs), CLE (50 ROIs), and PSE (59 ROIs). The NT ROIs were annotated in the non-smokers and the CLE and PSE ROIs were annotated in the two smokers' subject classes, within the area(s) of the leading pattern. PLE was excluded due to very few cases in the data set (only 2 out of 20 subjects diagnosed with COPD).

Computation of Texton-Based Features. For the construction of the codebook, 500 random patches are extracted from each ROI in each class. Patch sizes of 3×3 to 8×8 pixels are used in the experiments. No filter banks are applied and raw pixel representation is used instead. Since in CT images, the mean of the intensity in the images

indicate a physical property of the tissue displayed, it should not be removed. The patches extracted from different ROIs of each class are clustered using k -means to find the codebook that represents a class. Two different values of k , i.e., $k = 10$ and $k = 40$ are tested in the experiments leading to $3 \times 10 = 30$ or $3 \times 40 = 120$ (3 is the number of classes) textons in the codebook, respectively (refer to Fig. 1). After construction of texton codebook, the texton frequency histograms of the ROIs are computed to find the model. In this stage, small overlapping patches with the same size as what was used in the clustering stage are systematically extracted from each ROI. As in the clustering stage, no filter bank is used and raw pixel representation is considered. Euclidean distance between the resulting textons (collection of small patches) and the textons in the codebook is computed in order to identify the most similar texton in the codebook and the corresponding histogram of textons is updated accordingly. The histograms are normalized and used as the features.

Classifier and Evaluation. Both k -NN and SVM are used in the experiments. The crucial issue in using SVMs is finding a suitable kernel and the optimum trade-off parameter C . RBF kernel is used and the optimum kernel width, i.e., γ in (1) and C are found by a grid search on the training set at each specific texton size and k value (in k -means) used in the experiments. This grid search is performed by leave-one-subject-out on the training set. The computation of the texton codebook is performed each time on the training set, excluding the validation set. The proposed texton-based classification system using SVM as classifier with RBF kernel and optimal C and γ is evaluated using leave-one-subject-out. This means that, at each trial, all ROIs from one patient are held out as the test set and the remaining ROIs as the training set.

4 Results and Discussions

In this section, we first present the results for the proposed texton-based texture classification system using SVM as classifier with the parameters chosen as explained in previous section. Then comparison with other techniques is provided.

After finding the optimal C and γ at each texton size and k value (of k -means), the accuracy of the classification system is evaluated using leave-one-subject-out for the particular texton size and k . The results are shown in Table 1 for various texton sizes and two different values of k in k -means. It can be seen from these results that using $k = 40$ in k -means improves the performance of the classification system over $k = 10$. The best result is obtained at the texton size of 4×4 pixels with $k = 40$. It can be observed from the results that the accuracy of classification system decreases with increasing texton size. This can be because increasing the texton size leads to a higher dimensional space in k -means, requiring more data for reliable clustering. Also, fewer patches can be extracted from the ROIs at higher texton sizes that may degrade the estimation of model as explained in Section 2.2.

Comparison with Other Techniques. The first comparison is made between texton-based classification system using SVM and k -NN classifiers. The optimal parameter k of the k -NN classifier is found using a validation set in the training set in a similar way as the grid search performed for the parameters in the SVM classifier. The results are shown in Table 1, and as can be seen, the SVM classifier performs much better

The confusion matrix for the best results in [10], i.e., the results of LBP2, and our best results are provided in Table 2. The proposed approach attains performance better than LBP2 though McNemar's test does not indicate the difference to be significant ($p = 0.75$). The specificity of texton-based and LBP2 approaches are 98.31% and 93.33%, while their sensitivity are 95.41% and 97.25%, respectively (when comparing NT versus CLE and PSE).

5 Conclusion

In this paper, a texton-based texture classification system using a SVM with RBF kernel is proposed for the classification of emphysema in CT images of the lung. It is shown that the proposed approach performs significantly better than common approaches based on moments of histograms of filter bank responses using Gaussian derivatives and slightly improve the performance over a recently proposed approach based on LBPs. LBP operators are, by design, invariant to monotonic intensity transformations. While this is desirable in some applications, in the classification of Lung CT images, the mean of intensity is important and this explains the poor performance of LBP1 in Table 2 as it discards the mean of intensity in the ROIs. Due to this drawback of LBPs, in [10], the joint intensity and LBP histograms are considered (LBP2). This improves the performance of the LBPs in this application at the cost of adding to the complexity of the approach. The texton-based approach does not suffer from this problem as it is not invariant to intensity transformations. It is also shown that using SVM in the proposed approach yields higher accuracy than a k -NN classifier.

As the state of the art in examination of lung is volumetric chest CT scans, one may wonder whether texton-based approach can be extended to 3D data. As the computation of texton signatures is fast especially when raw pixel representation is used, we expect that the approach can easily be extended to 3D data. The main obstacle might be the computational cost for optimizing the SVM parameters in a grid search that can be reduced by using an m -fold cross-validation at patient level instead of leave-one-subject-out used in our experiments.

In future work, combining the classification outputs at different texton sizes using multiple classifier systems (MCS) will be investigated for possible improvement of the results. This improvement is expected if the misclassification is done on different ROIs in different texton sizes such that the MCS yield better results due to the diversity of the classifiers [16].

Acknowledgments. This work was partly funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Canada Graduate Scholarship (CGS D3-378361-2009), the Danish Council for Strategic Research under the Programme Commission for Nanoscience and Technology, Biotechnology and IT, the Netherlands Organization for Scientific Research (NWO), and AstraZeneca, Lund, Sweden. The first author would also like to thank Laurens van der Maaten from Delft University of Technology, the Netherlands for useful discussions on texton-based approach in texture classification.

References

1. Rabe, K.F., Hurd, S., Anzueto, A., Barnes, P.J., Buist, S.A., Calverley, P., Fukuchi, Y., Jenkins, C., Rodriguez-Roisin, R., van Weel, C., Zielinski, J.: Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: GOLD Executive Summary. *Am. J. Respir. Crit. Care Med.* 176(6), 532–555 (2007)
2. Müller, N.L., Staples, C.A., Miller, R.R., Abboud, R.T.: Density Mask-An Objective Method to Quantitate Emphysema Using Computed Tomography. *Chest* 94(4), 782–787 (1988)
3. Webb, W.R., Müller, N., Naidich, D.: High-Resolution CT of the Lung, 3rd edn. Lippincott Williams & Wilkins (2001)
4. Uppaluri, R., Mitsa, T., Sonka, M., Hoffman, E.A., McLennan, G.: Quantification of Pulmonary Emphysema from Lung Computed Tomography Images. *Amer. J. Respir. Crit. Care Med.* 156(1), 248–254 (1997)
5. Sluimer, I.C., Prokop, M., Hartmann, I., van Ginneken, B.: Automated Classification of Hyperlucency, Fibrosis, Ground Glass, Solid, and Focal Lesions in High-Resolution CT of the Lung. *Medical Physics* 33(7), 2610–2620 (2006)
6. Chabat, F., Yang, G.Z., Hansell, D.M.: Obstructive Lung Diseases: Texture Classification for Differentiation at CT. *Radiology* 228(3), 871–877 (2003)
7. Xu, Y., Sonka, M., McLennan, G., Guo, J., Hoffman, E.A.: MDCT-based 3-D Texture Classification of Emphysema and Early Smoking Related Lung Pathologies. *IEEE Trans. Med. Imag.* 25(4), 464–475 (2006)
8. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. on PAMI* 24(7), 971–987 (2002)
9. Varma, M., Zisserman, A.: A Statistical Approach to Material Classification Using Image Patch Exemplars. *IEEE Trans. on PAMI* 31(11), 2032–2047 (2009)
10. Sørensen, L., Shaker, S.B., de Bruijne, M.: Quantitative Analysis of Pulmonary Emphysema Using Local Binary Patterns. *IEEE Trans. Med. Imag.* 29(2), 559–569 (2010)
11. Leung, T., Malik, J.: Representing and Recognizing the Visual Appearance of Materials Using Three-Dimensional Textons. *Int'l. J. Computer Vision* 43(1), 29–44 (2001)
12. Varma, M., Zisserman, A.: A Statistical Approach to Texture Classification from Single Images. *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis* 62(1-2), 61–81 (2005)
13. Fan, R.E., Chen, P.H., Lin, C.J.: Working Set Selection Using the Second Order Information for Training SVM. *Journal of Mach. Learning Research* 6, 1889–1918 (2005)
14. Shaker, S.B., von Wachenfeldt, K.A., Larsson, S., Mile, I., Persdotter, S., Dahlbäck, M., Broberg, P., Stoel, B., Bach, K.S., Hestad, M., Fehniger, T.E., Dirksen, A.: Identification of Patients with Chronic Obstructive Pulmonary Disease (COPD) by Measurement of Plasma Biomarkers. *The Clinical Respiratory Journal* 2(1), 17–25 (2008)
15. Caputo, B., Hayman, E., Fritz, M., Eklundh, J.O.: Classifying Materials in the Real World. *Image and Vision Computing* 28(1), 150–163 (2010)
16. Kuncheva, L.I.: Combining Pattern Classifiers Methods and Algorithms. John Wiley & Sons, New Jersey (2004)

Active Learning for an Efficient Training Strategy of Computer-Aided Diagnosis Systems: Application to Diabetic Retinopathy Screening

C.I. Sánchez¹, M. Niemeijer², M.D. Abràmoff³, B. van Ginneken^{1,2}

¹ Department of Radiology, Radboud University Nijmegen Medical Centre, The Netherlands

² Image Sciences Institute, University Medical Center Utrecht, The Netherlands

³ Department of Ophthalmology and Visual Sciences, University of Iowa, USA

Abstract. The performance of computer-aided diagnosis (CAD) systems can be highly influenced by the training strategy. CAD systems are traditionally trained using available labeled data, extracted from a specific data distribution or from public databases. Due to the wide variability of medical data, these databases might not be representative enough when the CAD system is applied to data extracted from a different clinical setting, diminishing the performance or requiring more labeled samples in order to get better data generalization. In this work, we propose the incorporation of an active learning approach in the training phase of CAD systems for reducing the number of required training samples while maximizing the system performance. The benefit of this approach has been evaluated using a specific CAD system for Diabetic Retinopathy screening. The results show that 1) using a training set obtained from a different data source results in a considerable reduction of the CAD performance; and 2) using active learning the selected training set can be reduced from 1000 to 200 samples while maintaining an area under the Receiver Operating Characteristic curve of 0.856.

1 Introduction

In the last decade a variety of computer-aided diagnosis (CAD) systems has been developed for the automatic screening of diverse diseases, such as breast or lung cancer or diabetic retinopathy [1]. In general, these systems receive as input an exam consisting of one or more images from a patient and they generate as output a degree of suspicion for the disease. A typical CAD system relies on multiple stages: segmentation of normal anatomy, localization of abnormalities and fusion of different findings to obtain the final decision.

Supervised classification has been widely adopted for CAD systems as the optimal solution for the fusion and generation of the final outcome. In this type of approach, the training phase is of paramount importance for the development of these systems. During this phase, the classifier 'learns' from a group of *a priori* manually-annotated sample exams, namely training set. This training set should be representative enough to cope with the image variability and the

range of pathologies usually encountered in medical environments in order to obtain good prediction ability. Therefore, the choice of this training set has a great impact on the final performance. A non-representative training set may dramatically reduce the system accuracy, as we will show in this paper.

The common approach for the selection of training sets in CAD systems is to use public labeled databases or to extract a group of samples from the available data in a random way. This procedure does not always guarantee a representative training set, especially when retrieving data from the large imbalanced databases usually found in CAD applications. This results in the selection of large training sets in order to assure data generalization and high quality results. The annotation of a large amount of data is a tedious and time consuming task for medical experts. Additionally, the computational costs for the training phase increases drastically when the number of training samples increase.

Active learning is a machine learning approach that attempts to retrieve training data maximizing the system performance and minimizing the labeling effort [2]. In active learning approaches, only the most informative samples are dynamically selected from the unlabeled data and their correct labels are requested from an expert. Therefore, the size of the training set required to obtain an optimum classification accuracy is reduced, as well as the user's involvement in the labeling process. Compared to standard classification, where the goal is to minimize classification error, active learning has an additional goal: minimizing the amount of samples to be labeled.

In this paper, we evaluate the benefit of including active learning in the training phase of CAD systems in order to retrieve automatically representative training sets from large medical datasets. Particularly, we assess how the performance of a specific CAD application, namely the screening of diabetic retinopathy, is influenced by the active selection of training samples.

2 CAD System for Diabetic Retinopathy Screening

Diabetic Retinopathy (DR) is the most important cause of blindness in the working population of developed countries [3]. Early detection and diagnosis through screening programs is crucial for the prevention of visual loss and blindness in patients with diabetes [3]. A CAD system for the automatic large-scale screening of DR provides an effective way to obtain an early diagnosis and to prevent future complications. Figure 1(a) shows an example of a retinal image from a diabetic patient with DR.

The proposed CAD system for DR screening relies on four components: quality verification, normal anatomy detection, bright lesion detection and red lesion detection. These components are all based on previous work [4] and are therefore only briefly described here.

Quality verification: This component uses a statistical classifier to obtain the probability that the image quality is sufficient for diagnosis.

Normal anatomy detection: This component identifies blood vessels and the optic disk in retinal images using Gaussian derivative filters and k Nearest Neighbors (kNN) regressor.

Table 1. Set of features for retinal exam classification. PP: Posterior probability

Feature Description	
1	Quality of the exam Q . For exams with more than one image, the quality is given by $\frac{\max(Q_i) - \min(Q_i)}{2}$, where Q_k is the quality of image k in the exam.
2,3	The sum of all red/bright PPs as a measure for the total lesion load in the exam.
4,5	Highest red/bright PP in the exam.
6,7	Total red/bright lesion load weighted by the size of the detected lesions.
8,9	Average red/bright PP for those lesions with probability higher than 0.
10,11	Standard deviation of the red/bright PPs for those lesions with probability higher than 0.
12-19	A four bin histogram of the PPs of the red/bright lesion candidates.
20-27	A four bin histogram of the lesion area in pixels subdivided by PP.

Red lesion detection: Red lesions are pathological regions that usually appears in the earliest stages of DR. For their detection, a hybrid candidate extractor and a kNN classifier are applied to obtain a likelihood per candidate to be a red lesion.

Bright lesion detection: As well as red lesions, bright lesions are also important signs of DR. The algorithm relies on a candidate extraction step based on pixel classification and a kNN classifier.

The different findings from the aforementioned components are then fused to obtain a final outcome per patient: the likelihood of the patient to be referred to an ophthalmologist due to the presence of DR signs or because of insufficient quality. For the fusion procedure, we calculate a set of features for each exam based on the output from the different components (see Table 1). With these features, a kNN classifier is trained to obtain a probability per exam.

3 Methods

In order to train efficiently the proposed CAD system, an active learning approach is incorporated in the training phase of the fusion strategy. This approach is an iterative procedure where at each iteration the active learner is called to select an unlabeled sample from a pool of unlabeled data and an expert is asked for its label. The idea is to select efficiently a set of training samples from the unlabeled data in an active way to boost the performance of the classifier and reduce the number of samples that need to be labeled.

Assume that a small initial training set X_t , a classifier c , an active or query function F and unlabeled data X_u are given. The query function F assigns a value to each unlabeled sample in X_u depending on how informative the sample is. These values permit ranking the unlabeled objects and selecting the most informative sample x^* , which is expected to improve the classification performance the most [2].

The general framework of the active learning system can be described as follows [2]:

1. Train classifier c on the current training set X_t .
2. Select an object x^* from the unlabeled data X_u according to the active query function F .
3. Ask an expert for the label of x^* . Enlarge the training set X_t and reduce X_u : $X_t = X_t \cup \{x^*\}$, $X_u = X_u \setminus \{x^*\}$.
4. Repeat steps (1)-(3) until a stopping criterion has been reached.

The active function F determines the sampling selection, i.e., decides which sample in the unlabeled data X_u to query next. We investigate two different query functions in this study: uncertainty sampling and query-by-bagging (QBB) sampling.

Uncertainty sampling: This method queries unlabeled samples about which the current classifier is most uncertain and asks for their correct labels [2]. To measure uncertainty, the query function F_{US} can be defined as follows:

$$F_{US} \equiv x^* = \operatorname{argmax}_{x_i \in X_u} [- \sum_j P(w^j | x_i) \log(P(w^j | x_i))]. \quad (1)$$

with $j = 1, \dots, c$ and c the number of classes in the classification problem.

QBB sampling: In this sampling technique, a committee of classifiers are trained and the most informative sample is considered to be the sample over which the committee is in most disagreement about how to label [2]. In each round of active learning, X_t is sampling by replacement R times to create R modified training sets $X_t^1, \dots, X_t^r, \dots, X_t^R$. The classifier c is then trained with each modified set X_t^r to obtain a committee of R classifiers $\mathcal{C} = c^1, \dots, c^r, \dots, c^R$. To measure the disagreement among committee members, the query function F_{QBB} is defined as follows:

$$F_{QBB} \equiv x^* = \operatorname{argmax}_{x_i \in X_u} \frac{1}{R} \sum_{r=1}^R \sum_j P(w^j | x_i; c^r) \log \frac{P(w^j | x_i; c^r)}{P(w^j | x_i; \mathcal{C})} \quad (2)$$

with $P(w^j | x_i; \mathcal{C}) = \frac{1}{R} \sum_{r=1}^R P(w^j | x_i; c^r)$.

4 Materials and Experiments

Training set: For the creation of the training set, a group of 7500 unlabeled retinal exams (dataset A) were taken from an online retinal screening program [5]. Each exam consists of four images with resolution varying from 768x576 to 2048x1536 pixels. The exams were obtained using multiple types of fundus cameras while the field of view coverage varied between 35 and 45 degrees. A second publicly available dataset (dataset B) of 1200 exams was also used to train the CAD system and compare the performance with the results obtained using active learning [6]. In dataset B, only one image per exam is provided and they are acquired by 3 ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 non-mydratic retinography with a 45 degree field of view. The images were captured using 8 bits per color plane at 1440*960,

2240*1488 or 2304*1536 pixels. A human expert manually annotated each exam as normal (546 exams) or suspect (654 exams).

Test set: A totally different group of 7500 labeled retinal exams (dataset C) were also taken from the aforementioned screening program [5]. The exams were manually annotated by a human observer as normal (7080 exams) or suspect (420 exams). An exam is considered suspect if any DR signs are present or if the exam is ungradable.

To evaluate the influence of the training set in the CAD performance with and without the active learning approach, several experiments were performed:

Experiment 1: The CAD system was trained using dataset A and evaluated on dataset C. This was an ideal situation where a large labeled dataset from the same distribution was used for training. The area under the Receiver Operating Characteristic (ROC) curve (A_z) was used as performance metric.

Experiment 2: The CAD system was trained using dataset B and evaluated on dataset B. This experiment was done to evaluate the robustness of the system on a totally different data distribution. As the training and the test set was the same, a repeated ten-fold cross-validation was performed and the average A_z value was used as performance metric.

Experiment 3: The CAD system was trained using dataset B and evaluated on dataset C. This experiment assessed the CAD performance when a different distribution data was used for training. The A_z was also used as performance metric.

Experiment 4: The CAD system was trained using a subset retrieved from dataset A using active learning with uncertainty sampling and evaluated on dataset C. For iteration i of the active learner, the query strategy retrieved an exam from dataset A, a human observer annotated the selected exam and the updated training set $X_t^{(i)}$ was used to train the classifier $c^{(i)}$. The classification performance P was evaluated on dataset C based on A_z value when the training set reached sizes of $N = 10, 20, \dots, 100, 200, \dots, 1000, 2000, \dots, 6000$. With these performance values, a learning curve was calculated. The optimal performance was obtained at the point in the learning curve (and at the corresponding training size) for which the performance obtained with the retrieved training set was non-significantly different from the one obtained with the complete dataset A. Due to the number of comparisons (one per each point in the learning curve), the significance level was adjusted by the Bonferroni correction to $p < 0.0021$ ($0.05/24$). The training strategy was initialized with 10 samples chosen at random from dataset A. To reduce the influence of the random selection, the experiment was run 10 independent times and the results were averaged to obtain a mean A_z value as well as the standard deviation per each point in the learning curve.

Experiment 5: The same as experiment 4 but with QBB sampling as the query function.

Experiment 6: The same as experiment 4 but with random sampling as the query function, the approach typically adopted in CAD development. This method randomly selects the next sample from the pool of unlabeled data (dataset A). In this case the experiment was run 100 independent times.

In all the experiments, k was set at the square-root of the number of samples in the training set and, when QBB was used, the number of resampling was set to 3. Previous experiments in a different dataset showed that varying committee size (number of resampling) has little effect on the final performance.

5 Results

Table 2 summarizes the performance of the different experiments. For experiments 4-6, the optimal performance obtained using the learning curves was shown. Figure 1(b) shows the ROC curves for the experiment 1, 2 and 3.

Table 2. Characteristics and performance of the different experiments.* indicates value calculated performing cross-validation.

	<i>Training source</i>	<i>Training sampling</i>	<i>Training size</i>	<i>Test set</i>	<i>Az</i>
Experiment 1	Dataset A	-	7500	Dataset C	0.884
Experiment 2	Dataset B	-	1200	Dataset B	0.875*
Experiment 3	Dataset B	-	1200	Dataset C	0.689
Experiment 4	Dataset A	Uncertainty	200	Dataset C	0.856
Experiment 5	Dataset A	QBB	500	Dataset C	0.831
Experiment 6	Dataset A	Random	1000	Dataset C	0.837

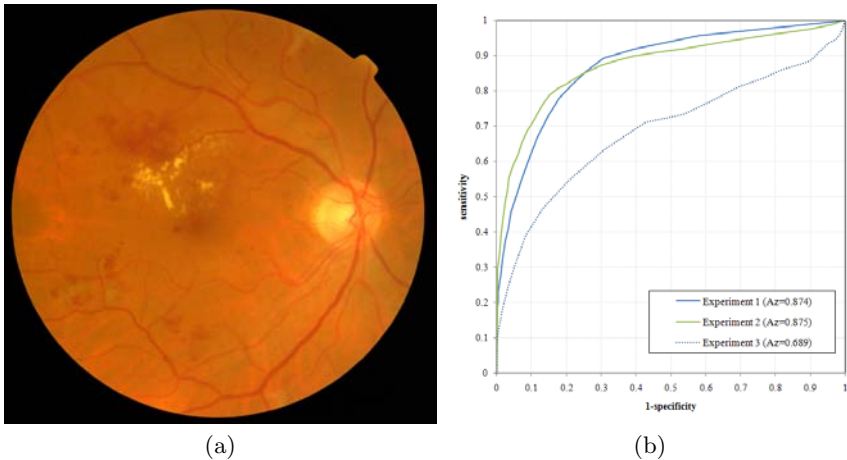


Fig. 1. (a) Example of an image from a diabetic patient with DR. Red and bright lesions appears as red and yellowish patches on the image, respectively. (b) ROC curves for the experiments 1, 2 and 3. In the case of experiment 2, a repeated ten-fold cross-validation was performed.

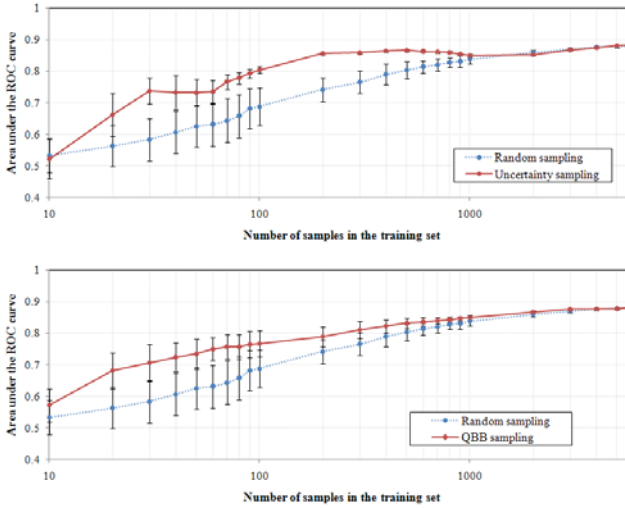


Fig. 2. Learning curves for the proposed active learning strategy using two different sampling functions, namely uncertainty sampling and QBB sampling. The learning curve for a training strategy using random sampling is also depicted. The experiments for each active query function were run 10 times whereas the experiment for the random sampling function was run 100 times. The learning curves show the average area under the ROC curves and the standard deviation of the values obtained for a fixed number of samples in the training set.

Figure 2 shows the learning curves for the experiments 4, 5 and 6. As we can see, at the beginning of the curves the CAD performance was similar to random guess but, after some iteration, CAD performance improved much faster with active learning than using random sampling. The three experiments converged finally to the same point where the training sets were similar to Dataset A.

6 Discussion

In this paper, an effective training strategy for CAD systems using active learning is proposed. Experiment 2 and 3 showed that defining an appropriate training set is a crucial step towards obtaining high quality results in clinical applications. A non-representative training set of the medical data distribution under study had a negative effect on the final CAD performance. In Figure 1, we can see that, although the CAD system worked well for different type of data when the system was trained using samples extracted from the specific distribution, the performance diminished significantly when the training set was from a different distribution. This can probably be attributed to the difference in the number of images, the disease prevalence or the quality between these datasets, which represent normal variations of data characteristics in medical applications. Therefore, this suggested the necessity of creating specific training sets per data type.

Retrieving representative training sets from large unlabeled medical sets is a difficult task, especially when the prevalence of the disease is low. This results in selecting large training set to achieve a good generalization of the distribution, increasing the time and effort spent on performing manual annotations and incurring in higher computational costs for the training phase. We have shown in experiment 4 and 5 that with active learning the training selection rapidly converged to an optimal small training set (see Figure 2 and Table 2). Compared to random sampling, the proposed approach reduced the number of training samples needed to obtain a similar performance when a larger dataset was used, reducing the labeling effort. Although a larger labeled dataset was available for training (dataset B), it was more beneficial to retrieve a small set from the unlabeled set in order to maximize the performance of the CAD system.

However, the selection of the query function can also influence on the final performance. As it is shown in Figure 2, a larger set was needed to obtain the same performance when the query strategy was changed from uncertainty to QBB sampling. Additionally, a stopping criterion needs to be defined for the active learner. Although this criterion can be set depending on a maximum number of manual annotations, this might not guarantee optimal accuracies. In future work, the influence of the stopping criterion will be evaluated.

In conclusion, an active learning approach incorporating in the CAD training phase was studied in order to reduce the number of training samples needed to obtain an optimum accuracy. The results show that the system accuracy can be maximized using small representative training sets, retrieved using an active learner. This approach allows an automatic efficient training stage of the CAD systems using the vast incompletely labeled databases that are now available in medical applications.

References

1. Doi, K.: Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics* 31(4), 198–211 (2007)
2. Juszczak, P.: Learning to recognise. PhD thesis, Delft University, the Netherlands, 2006.
3. Kinyoun, J., Barton, F., Fisher, M., Hubbard, L., Aiello, L., Ferris, F.: Detection of diabetic macular edema. Ophthalmoscopy versus photography—Early Treatment Diabetic Retinopathy Study Report Number 5. The ETDRS Research Group. *Ophthalmology* 96, 746–750 (1989)
4. Niemeijer, M., Abramoff, M.D., van Ginneken, B.: Information fusion for diabetic retinopathy CAD in digital color fundus photographs. *IEEE Transactions on Medical Imaging* 28(5), 775–785 (2009)
5. Abramoff, M.D., Suttorp-Schulten, M.: Web-based screening for diabetic retinopathy in a primary care population: the eyecheck project. *Telemedicine Journal and E-health* 11(6), 668–674 (2005)
6. Messidor database, <http://messidor.crihan.fr> (accessed March 11, 2010)

Sparse Bayesian Learning for Identifying Imaging Biomarkers in AD Prediction

Li Shen^{1,2,*}, Yuan Qi^{3,*}, Sungeun Kim^{1,2}, Kwangsik Nho^{1,2}, Jing Wan^{1,2}, Shannon L. Risacher¹, Andrew J. Saykin^{1,*}, and ADNI

¹ Center for Neuroimaging, Department of Radiology and Imaging Sciences,

² Center for Computational Biology and Bioinformatics,

Indiana University School of Medicine, 950 W Walnut St, Indianapolis, IN 46202

³ Departments of Computer Science, Statistics and Biology,

Purdue University, 305 N. University Street, West Lafayette, IN 47907

Abstract. We apply sparse Bayesian learning methods, automatic relevance determination (ARD) and predictive ARD (PARD), to Alzheimer’s disease (AD) classification to make accurate prediction and identify critical imaging markers relevant to AD at the same time. ARD is one of the most successful Bayesian feature selection methods. PARD is a powerful Bayesian feature selection method, and provides sparse models that is easy to interpret. PARD selects the model with the best estimate of the predictive performance instead of choosing the one with the largest marginal model likelihood. Comparative study with support vector machine (SVM) shows that ARD/PARD in general outperform SVM in terms of prediction accuracy. Additional comparison with surface-based general linear model (GLM) analysis shows that regions with strongest signals are identified by both GLM and ARD/PARD. While GLM P-map returns significant regions all over the cortex, ARD/PARD provide a small number of relevant and meaningful imaging markers with predictive power, including both cortical and subcortical measures.

1 Introduction

Neuroimaging is a powerful tool for characterizing neurodegenerative process in the progression of Alzheimer’s disease (AD) and can provide potential surrogate biomarkers for therapeutic trials. This paper is focused on identifying relevant imaging biomarkers from structural magnetic resonance imaging (MRI) data for AD classification. Machine learning methods have been applied to many problems in computational neuroscience, including computer-aided diagnosis for AD [13, 4, 6, 7, 9]. While popular methods like support vector machines (SVMs) [15]

* Correspondence to Li Shen (shenli@iupui.edu), Yuan Qi (alanqi@cs.purdue.edu), or Andrew J. Saykin (asaykin@iupui.edu). Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (U01 AG024904). This project was also supported in part by Indiana CTSI IUSM/CTR(RR025761), 1RC 2AG036535, NIA R01 AG19771, Foundation for the NIH, IEDC #87884, NIBIB R03 EB008674, U01 AG032984, and P30 AG10133.

can achieve decent prediction accuracy, most of them are not optimized for selecting sensitive features.

This paper presented the results of applying novel sparse Bayesian learning methods, automatic relevance determination (ARD) and predictive ARD (PARD) [13], to MRI-based AD classification for achieving two goals at the same time: (1) accurate prediction rate and (2) selection of relevant imaging biomarkers. Linear SVM and general linear model (GLM) based cortical thickness analyses were also performed on the same data for comparison to ARD/PARD. Our overarching goal is to learn from these data sparse Bayesian models so that they are easy to interpret while maintaining high predictive power.

2 Materials and Methods

MRI Data used in this study were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). ADNI is a landmark investigation sponsored by the NIH and industrial partners designed to collect longitudinal neuroimaging, biological and clinical information from 800 participants that will track the neural correlates of memory loss from an early stage. Further information can be found in [11] and at www.adni-info.org. Following a previous imaging genetics study [14], 378 non-Hispanic Caucasian participants (203 healthy control (HC) and 175 AD participants) were selected for this work. For one baseline scan of each participant, FreeSurfer V4 was employed to automatically label cortical and subcortical tissue classes [25] and to extract target region volume and cortical thickness, as well as to extract total intracranial volume (ICV), as previously described [14]. For each hemisphere, thickness measures of 34 cortical regions of interest (ROIs) (Fig. 1(a-f)) and volume measures of 15 cortical and subcortical ROIs (Fig. 1(c-f)) were included in this study. All these measures were adjusted for the baseline age, gender, education, handedness, and baseline ICV using the regression weights derived from the HC participants. Participant characteristics are summarized in Table 1.

ARD and Predictive ARD

We apply ARD and predictive ARD (PARD) [13] to classify the imaging features. ARD is one of the most successful Bayesian feature selection methods [8,12]. It is a hierarchical Bayesian approach where there are hyperparameters which explicitly represent the relevance of different input features. These relevance

Table 1. Participant characteristics

Category	HC	AD	<i>p</i> -value
Number of Subjects	203	175	-
Gender (M/F)	111/92	97/78	0.8840
Baseline Age (years; Mean±STD)	76.09±5.00	75.53±7.58	0.3884
Education (years; Mean±STD)	16.13±2.73	14.93±3.00	< 0.0001
Handedness (R/L)	188/15	163/12	0.8413

hyperparameters determine the range of variation for the parameters relating to a particular input, usually by modeling the width of a zero-mean Gaussian prior on those parameters. If the width of that Gaussian is zero, then those parameters are constrained to be zero, and the corresponding input cannot have any effect on the predictions, therefore making it irrelevant. ARD optimizes these hyperparameters to discover which inputs are relevant.

Predictive ARD improves upon ARD in the following aspects. First, the Laplace approximation used in ARD [8] is replaced by the more accurate expectation propagation (EP) [10]. Second, EP computes an estimate of leave-one-out predictive performance without requiring expensive cross-validation experiments. This estimate of predictive performance can be used as an important criterion for ARD to avoid the overfitting problem associated with evidence maximization. Last, predictive ARD uses a fast *sequential* optimization method such that we can efficiently prune and add new features without updating a full covariance matrix for the classifier.

Now we describe ARD for linear classification. A linear classifier classifies a point \mathbf{x} according to $t = \text{sign}(\mathbf{w}^T \mathbf{x})$ for some parameter vector \mathbf{w} (the two classes are $t = \pm 1$). Given a training set $D = \{(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_N, t_N)\}$, the likelihood for \mathbf{w} can be written as

$$p(\mathbf{t}|\mathbf{w}, X) = \prod_i p(t_i|\mathbf{x}_i, \mathbf{w}) = \prod_i \Psi(t_i \mathbf{w}^T \phi(\mathbf{x}_i)) \tag{1}$$

where $\mathbf{t} = \{t_i\}_{i=1}^N$, $X = \{\mathbf{x}_i\}_{i=1}^N$, $\Psi(\cdot)$ is the cumulative distribution function for a Gaussian. One can also use the step function or logistic function as $\Psi(\cdot)$. The basis function $\phi^T(\mathbf{x}_i)$ allows the classification boundary to be nonlinear in the original features. This is the same likelihood used in logistic regression and in Gaussian process classifiers. Given a new input \mathbf{x}_{N+1} , we approximate the predictive distribution:

$$p(t_{N+1}|\mathbf{x}_{N+1}, \mathbf{t}) = \int p(t_{N+1}|\mathbf{x}_{N+1}, \mathbf{w})p(\mathbf{w}|\mathbf{t})d\mathbf{w} \tag{2}$$

$$\approx P(t_{N+1}|\mathbf{x}_{N+1}, \langle \mathbf{w} \rangle) \tag{3}$$

where $\langle \mathbf{w} \rangle$ denotes the posterior mean of the weights, called the Bayes Point.

The basic idea in ARD is to give the feature weights independent Gaussian priors:

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_i \mathcal{N}(w_i|0, \alpha_i^{-1}),$$

where $\boldsymbol{\alpha} = \{\alpha_i\}$ is a hyperparameter vector that controls how far away from zero each weight is allowed to go. The hyperparameters $\boldsymbol{\alpha}$ are trained from the data by maximizing the Bayesian ‘evidence’ $p(\mathbf{t}|\boldsymbol{\alpha})$, which can be done using a fixed point algorithm or an expectation maximization (EM) algorithm treating \mathbf{w} as a hidden variable [8]. The outcome of this optimization is that many elements of $\boldsymbol{\alpha}$ go to infinity such that the classifier \mathbf{w} would have only a few nonzero weights w_j . This naturally prunes irrelevant features in the data.

Unlike previous approaches that use the EM algorithm and find a solution that maximize the evidence, the **predictive-ARD** (PARD) algorithm trains

the sparse classifier as follows: (1) Initialize the model so that it only contains a small fraction of features. (2) Sequentially update the classifiers via a fast sequential optimization method and calculate the required statistics by EP until the algorithm converges. The sparsity level of the classifiers increases along the optimization iterations. (3) From all the classifiers, choose the classifier with minimum predictive error probability estimate.

SVM and GLM

A linear support vector machine (SVM) was applied in our study to provide a comparison to ARD/PARD in terms of prediction accuracy. SVMs represent a new generation of learning systems based on recent advances in statistical learning theory [15]. The aim in training a linear SVM is to find the separating hyperplane with the largest margin; the expectation is that the larger the margin, the better the generalization of the classifier. We employed the OSU SVM Matlab Toolbox (sourceforge.net/projects/svm/) in this work.

We also performed surface based analysis for identifying thickness changes on the brain cortex and comparing these regions with the imaging markers detected by ARD/PARD. We consider the following general linear model (GLM): $y = X\Psi + Z\Phi + \epsilon$, where the dependent variable y is cortical thickness; $X = (x_1, \dots, x_p)$ are the variables of interest (*diagnosis* in our case); $Z = (z_1, \dots, z_k)$ are the variables whose effects we want to exclude (*age, gender, education, handedness* and *ICV* in our case); $\Psi = (\psi_1, \dots, \psi_p)^T$ and $\Phi = (\phi_1, \dots, \phi_k)^T$ are the coefficients; and ϵ is the error term. The goal is to test if X is significant (i.e., $\Psi \neq 0$) for some $y \in \partial\Omega$, where $\partial\Omega$ is the cortical surface manifold. To test GLMs, we used SurfStat [16], a Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory (RFT) [17].

3 Results

Classification was performed on each hemisphere separately, using two sets of imaging features: (1) 34 thickness measures (Fig. 1(a-b)), and (2) 34 thickness measures and 15 volume measures (Fig. 1(c-f)). 10-fold cross-validation was performed for accuracy estimation. Shown in Table 2 is the performance comparison among ARD, PARD and SVM. ARD and PARD outperformed SVM except for the case of using both thickness and volume measures from right hemisphere.

Table 2. Performance comparison. Training and testing error rates (mean±std) of 10-fold cross validation are shown for SVM, PARD (Predictive ARD) and ARD.

	SVM		PARD		ARD	
	training	testing	training	testing	training	testing
Left [†]	0.129 ± 0.011	0.156 ± 0.062	0.106 ± 0.009	0.147 ± 0.082	0.108 ± 0.010	0.154 ± 0.065
Right [†]	0.149 ± 0.009	0.185 ± 0.076	0.137 ± 0.009	0.168 ± 0.086	0.139 ± 0.006	0.175 ± 0.077
Left [‡]	0.112 ± 0.006	0.139 ± 0.051	0.078 ± 0.008	0.132 ± 0.056	0.078 ± 0.008	0.124 ± 0.056
Right [‡]	0.130 ± 0.006	0.142 ± 0.053	0.117 ± 0.005	0.160 ± 0.040	0.118 ± 0.006	0.162 ± 0.039

[†] Use cortical thickness measures only. [‡] Use both thickness and volume measures.

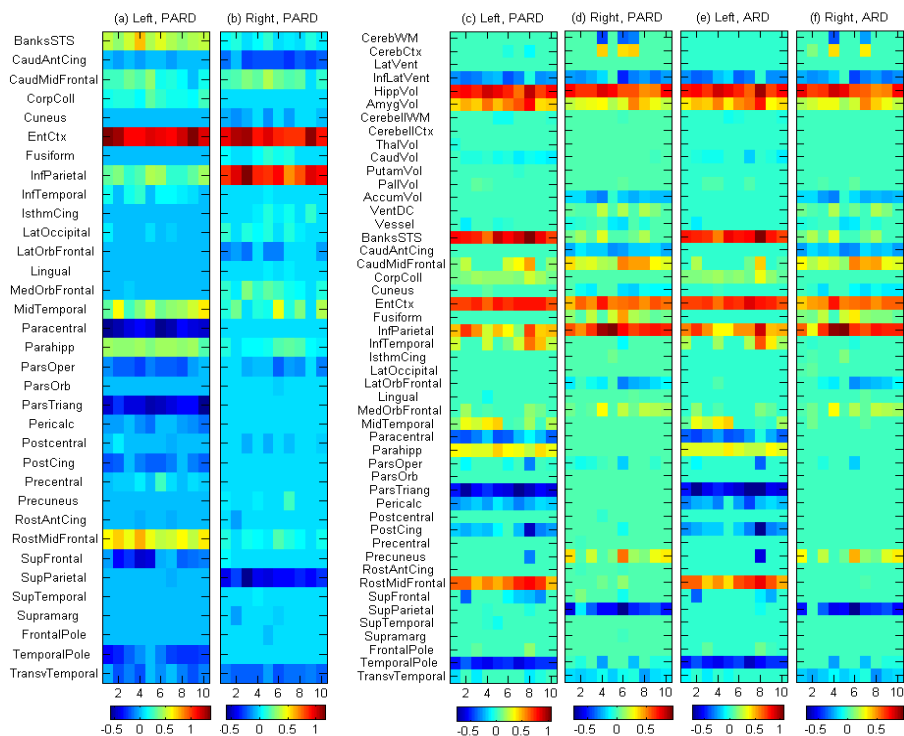


Fig. 1. (a-b) Heat maps of PARD weights $-w$ in cortical analyses using 34 thickness measures. (c-f) Heat maps of PARD (c,d) and ARD (e,f) weights $-w$ in analyses using 15 volume (top) and 34 thickness (bottom) measures. 10-fold cross-validation was performed for left (a,c,e) and right (b,d,f) hemisphere data. In each heat map, feature weights were plotted against 10 different trials in cross validation tests.

PARD outperformed ARD except for the case of using both thickness and volume measures from left hemisphere. PARD was designed for improving ARD predictive performance based on theoretical considerations, which empirically worked better for most cases but not all. Using thickness measures only, the best prediction rate was obtained at 85.3% by PARD for left hemisphere. Using both thickness and volume measures, the best prediction rate was improved to 87.6% by applying ARD to the left hemisphere data. In all cases, the prediction rates were improved after including 15 additional volume measures in the analyses, indicating both cortical and subcortical changes were related to AD.

A linear classifier is usually characterized by a weight vector w , which projects each individual data point (i.e., a feature vector) into a 1-D space for getting a discriminative value. Each weight measures the amount of the contribution of the corresponding feature to the final discriminative value. ARD and PARD aim to reduce the number of nonzero weights so that only relevant features are selected by examining these weights. For consistency, we always visualize

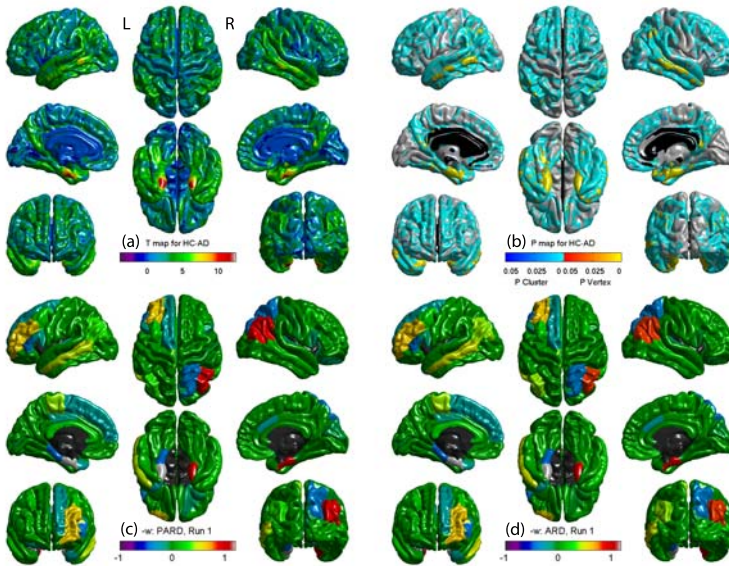


Fig. 2. (a-b) GLM results of diagnosis effect (HC-AD) on cortical thickness include (a) the map of the t statistics and (b) the map of corrected P values for peaks and clusters (only regions with corrected $p \leq 0.01$ are shown), where positive t values (red, yellow) indicate more grey matter in HC. (c-d) Back-projection of negative weights ($-w$) of the linear classifier for (c) PARD and (d) ARD, where positive values (gray, red, yellow) indicate more grey matter in HC.

negative weights $-w$ so that larger values (red) correspond to more grey matter in HC. Fig. 1(a-b) shows the heat maps of PARD weights $-w$ in cortical thickness analysis for one run of 10-fold cross validation for both hemispheres. The weight vectors (i.e., columns in the map) derived by different trials in cross validation are very similar. Most weights are close to zero, indicating a small number of relevant imaging markers. While entorhinal cortex (EntCtx) appears to be a strong predictor in both sides, rostral middle frontal gyri (RostMidFrontal) are strong only on the left and inferior temporal gyri (InfParietal) on the right.

These weights can be back-projected to the original image space for an intuitive visualization. Fig. 2(c-d) shows such a visualization for PARD and ARD results using thickness data. Since we only examine the mean thickness of each cortical subregion in our analysis, the entire region is painted with the same color defined by the corresponding weight. The patterns of imaging marker selection between PARD and ARD are very similar to each other. For comparison, surface-based GLM analysis using SurfStat is also performed to examine diagnosis effect (HC-AD) on cortical thickness and Fig. 2(a-b) shows the resulting T-map and P-map. Regions with strongest signals, such as entorhinal cortex on both sides and left middle temporal gyri are picked up by GLM and ARD/PARD. While

Table 3. Top imaging markers: “mean weight, rank” shown in each cell

ID	Description	Thick. & Vol., ARD		Thickness, PARD	
		Left	Right	Left	Right
HippVol	hippocampus	-0.804, 1	-0.641, 2	N/A	N/A
BanksSTS	banks of the superior temporal sulcus	-0.790, 2	-0.077, 10	-0.503, 3	-0.043, 9
EntCtx	entorhinal cortex	-0.745, 3	-0.523, 3	-1.199, 1	-0.954, 1
RostMidFrontal	rostral middle frontal gyri	-0.621, 4	-0.001, 17	-0.607, 2	-0.063, 7
InfParietal	inferior parietal gyri	-0.569, 5	-0.656, 1	-0.332, 6	-0.901, 2
AmygVol	amygdala	-0.514, 6	-0.271, 5	N/A	N/A
Parahipp	parahippocampal gyri	-0.316, 7	0.000, 19	-0.405, 5	-0.086, 6
InfTemporal	inferior temporal gyri	-0.189, 8	0.009, 38	-0.097, 9	-0.001, 15
MidTemporal	middle temporal gyri	-0.179, 9	-0.011, 13	-0.446, 4	-0.167, 4
CorpColl	corpus colosum	-0.168, 10	0.000, 20	-0.195, 8	0.000, 17

GLM P-map returns significant regions across the entire cortex, PARD/ARD maps provide a small number of selective regions with predictive power.

Heat maps of ARD/PARD weights $-w$ in combined thickness and volume analyses are shown in Fig. 1(c-f). Again, the patterns are very similar between ARD and PARD. Shown in Table 3 are top imaging markers selected by ARD using thickness and volume measures (PARD data not shown but extremely similar to ARD) and by PARD using thickness measures (ARD data not shown but extremely similar to PARD). While most top markers are thickness measures from cortical regions, two markers are volume measures from subcortical structures including hippocampus and amygdala.

4 Discussion

We presented a novel application of sparse Bayesian learning methods, ARD and PARD, to AD classification. Our strategy was to minimize the complexity of both data and methods for deriving a simple model easy to interpret. For methods, we focused on linear classifiers and showed that ARD/PARD in general outperformed SVM. For data, we focused on summary statistics (i.e., thickness and volume) of anatomically meaningful grey matter regions across the whole brain, and showed that promising prediction accuracy (87.6%) could be achieved with a small number of relevant imaging measures. Most prior studies (e.g., [13,4]) performed feature selection/extraction before classification. Our method integrated feature selection into the learning process to form a simple and principled procedure. Prior research [6] also integrated feature selection into classification and reported lower prediction rates (77-82%) for analyzing a subset of the same ADNI MRI data. Comparison to other feature selection schemes merits further investigation. While some prior studies [3,4,7,9] reported better prediction rates, they analyzed many more imaging variables in much smaller data sets. One interesting future topic is to apply our method to more detailed imaging features to determine if better prediction rates and refined imaging

marker maps can be achieved. It is unclear if disease duration of AD is comparable between ADNI cohort examined by us and [14,6] and others cohorts by [3,7,9], and this could have an effect on prediction rates. Incorporating disease duration in predictive models warrants further investigation. To sum up, contributions of this work include: (1) a simple and unified learning method that inherently does feature selection and enables biomarker discovery while maintaining high predictive power; (2) a much larger AD sample tested with much fewer variables, resulting in a better power; and (3) promising rates for predicting mild AD with identified biomarkers that are known to be related to AD.

References

1. Batmanghelich, N., Taskar, B., Davatzikos, C.: A general and unifying framework for feature construction, in image-based pattern classification. *Inf. Process Med. Imaging* 21, 423–434 (2009)
2. Dale, A., Fischl, B., Sereno, M.: Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9(2), 179–194 (1999)
3. Duchesne, S., Caroli, A., Geroldi, C., Barillot, C., Frisoni, G.B., Collins, D.L.: MRI-based automated computer classification of probable ad versus normal controls. *IEEE Trans. Med. Imaging* 27(4), 509–520 (2008)
4. Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C.: Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* 39(4), 1731–1743 (2008)
5. Fischl, B., Sereno, M., Dale, A.: Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9(2), 195–207 (1999)
6. Hinrichs, C., Singh, V., et al.: Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *Neuroimage* 48(1), 138–149 (2009)
7. Kloppel, S., Stonnington, C.M., et al.: Automatic classification of MR scans in Alzheimer’s disease. *Brain* 131(Pt. 3), 681–689 (2008)
8. MacKay, D.J.: Bayesian interpolation. *Neural Computation* 4(3), 415–447 (1992)
9. Magnin, B., Mesrob, L., Kinkingnehun, S., et al.: Support vector machine-based classification of Alzheimer’s disease from whole-brain anatomical MRI. *Neuroradiology* 51(2), 73–83 (2009)
10. Minka, T.P.: Expectation propagation for approximate Bayesian inference. In: 17th Conf. in Uncertainty in Artificial Intelligence, pp. 362–369 (2001)
11. Mueller, S.G., Weiner, M.W., et al.: The Alzheimer’s disease neuroimaging initiative. *Neuroimaging Clin. N. Am.* 15(4), 869–877 (2005)
12. Neal, R.M.: *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics, vol. 118. Springer, New York (1996)
13. Qi, Y., Minka, T., et al.: Predictive automatic relevance determination by expectation propagation. In: 21st Int. Conf. on Machine learning, pp. 671–678 (2004)
14. Shen, L., Kim, S., et al.: Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* (2010), doi:10.1016/j.neuroimage.2010.01.042
15. Vapnik, V.: *Statistical Learning Theory*. John Wiley and Sons, Chichester (1998)
16. Worsley, K.J.: SurfStat, <http://www.math.mcgill.ca/keith/surfstat>
17. Worsley, K.J., Andermann, M., Koulis, M., et al.: Detecting changes in non-isotropic images. *Human Brain Mapping* 8, 98–101 (1999)

Computer-Aided Detection of Pulmonary Pathology in Pediatric Chest Radiographs

André Mouton¹, Richard D. Pitcher², and Tania S. Douglas¹

¹ MRC/UCT Medical Imaging Research Unit, Department of Human Biology,
University of Cape Town, South Africa

² Department of Radiology, Tygerberg Hospital, Cape Town, South Africa
tania@ieee.org

Abstract. A scheme for triaging pulmonary abnormalities in pediatric chest radiographs for specialist interpretation would be useful in resource-poor settings, especially those with a high tuberculosis burden. We assess computer-aided detection of pulmonary pathology in pediatric digital chest X-ray images. The method comprises four phases suggested in the literature: lung field segmentation, lung field subdivision, feature extraction and classification. The output of the system is a probability map for each image, giving an indication of the degree of abnormality of every region in the lung fields; the maps may be used as a visual tool for identifying those cases that need further attention. The system is evaluated on a set of anterior-posterior chest images obtained using a linear slot-scanning digital X-ray machine. The classification results produced an area under the ROC of 0.782, averaged over all regions.

Keywords: classification, computer-aided diagnosis, tuberculosis, pulmonary abnormality, chest radiograph.

1 Introduction

Chest radiography remains the most common radiological examination, in both adults and children, and in both well-resourced and poorly resourced settings; poorly resourced settings, however have limited recourse to specialist radiological services. The chest X-ray is often the main diagnostic tool for tuberculosis (TB) in children in endemic settings [1]. The manner in which PTB manifests itself radiographically varies with the age of the patient as well as the stage of the TB infection (primary or post-primary). Children generally develop primary TB and display associated radiological features.

Chest X-rays are extremely difficult to interpret. Poor contrast and complex backgrounds comprised of superimposed anatomical structures overlapping with regions of interest make the detection of abnormalities difficult, even for experienced radiologists [2]. Thus a diagnostic tool which assists in triaging pediatric chest radiographs for specialist interpretation would be useful.

Although much research has been conducted into the use of computer-aided textural analysis for the detection of interstitial lung diseases in chest radiographs, research into

computer-aided methods for the detection and diagnosis of general pulmonary pathology in chest radiographs, especially in children, is limited. Explicit application to a pediatric population, has, to the authors' knowledge, not been reported. Pulmonary abnormality would include interstitial disease (of the supporting structures of the lung including the alveolar wall) and alveolar disease, which involves the alveolar airspace itself and manifests by diffuse, homogeneous opacification. We apply methods developed for the detection of adult TB and interstitial disease, to the detection of pulmonary abnormality in pediatric chest radiographs; we assess the utility of this framework with reference to the results obtained by others in adults.

2 Methods

Schemes for the detection of interstitial lung disease (ILD) have a common structure. Multiple regions of interest (ROIs) are manually or automatically selected within the lung fields. Features, in the form of textural measurements, are then extracted from each of these ROIs. The ROIs are classified using rule-based and/or pattern recognition techniques resulting in soft labels (posterior probabilities) for each of the regions in the lung fields. The results are combined to obtain an overall diagnosis for the image, indicating whether or not it contains any interstitial abnormalities. The approach of this study is guided by that presented in [3].

2.1 Study Data

The data set consisted of 119 anterior-posterior (AP) pediatric chest scans of children, obtained using a linear-slot-scanning digital X-ray machine [4]. The images were obtained as part of a study to evaluate their suitability for visualizing airway abnormalities in children having symptoms of TB. The children were aged between 0 and 5 years at the time the scans were taken. Although the pulmonary changes associated with TB are completely non-specific when viewed in isolation, chest x-rays are an important component of the diagnostic process for TB, especially in children.

To allow for the training of classifiers to distinguish between normal and abnormal regions, the abnormal regions in the images were outlined by the examining radiologist on a computer monitor using a mouse pointer. Of the 119 images, 6 contained no pathology while the remaining 113 images contained, in total, 263 abnormal regions, of which 168 occurred in the right lung field and 95 in the left lung field.

2.2 Lung Field Segmentation

Automatic segmentation of lung fields is considered a required procedure before computerized analysis of a chest radiograph can occur [5].

In this study the lung fields are segmented using the active shape model (ASM) technique [6], which consists of three elements: a global shape component, a multi-resolution grey-level appearance component and a multi-resolution search algorithm. The technique requires a set of annotated training images, in which a set of corresponding points (representing the lung borders) has been marked, from which the model can be built. Statistics about the model's global shape and appearance, in the vicinity of each model point, are captured. This statistical model is used in an

iterative search to locate the lung fields in unseen test images. A semi-automatic ASM initialization is used here, where the user is required to provide a rough starting point for the model search. This is done by clicking a mouse on three landmark points in the lung field to be segmented. The landmarks required are: the bottom lateral corner of the lung; the lung apex; and the bottom medial corner of the lung.

2.3 Lung Field Subdivision

Arguably the greatest challenge in textural analysis of the lungs is the presence of superimposed anatomical structures [5]. Dividing the lung fields into small ROIs allows for each ROI to be classified individually with a classifier trained with features extracted from the specific region only, thereby capturing common anatomical characteristics of particular lung regions in the feature vectors, reducing the effect of background structures [3]. The mean lung shapes are computed from ASM derived from the training data. The boundaries defining the regions are found using a simple search algorithm, with equal area as the objective. The region maps are illustrated in Figure 1; 42 regions of different sizes overlap. These regions are used to automatically position the ROIs in subsequent test images, with segmented lung fields, which are warped to these region maps to define the ROIs. This warping is done using radial basis functions with multiquadrics and interpolation [7].

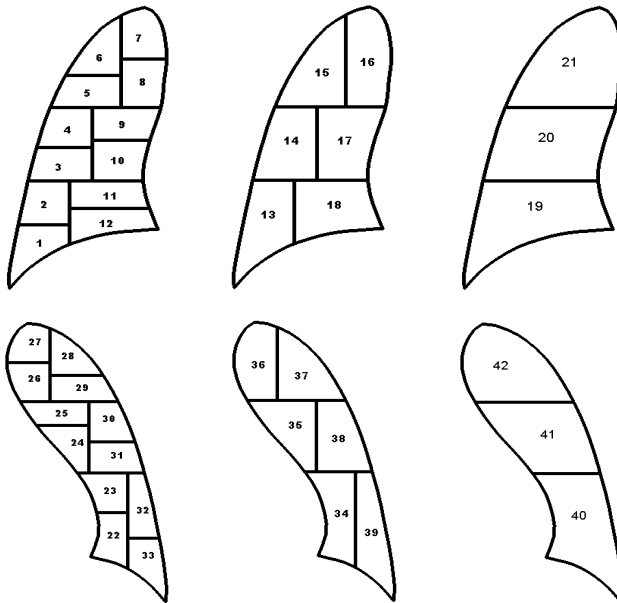


Fig.1. Subdivision of the lung fields into 42 overlapping regions of interest

2.4 Feature Extraction and Classification

Prior to the extraction of image features, several pre-processing steps are completed [8]: the left lung field is flipped about the centre line of the image so that the left lung

will resemble the right lung in orientation; pixel values outside the lungs are mirrored symmetrically with respect to the lung borders to prevent distortions in filter outputs at the lung borders.

A multiscale bank of Gaussian filters is used to extract the texture features [3,9]. The filter bank consists of Gaussian derivatives of orders, 0, 1 and 2 at five scales, $\sigma = 1, 2, 4, 8, 16$. By varying the degree of the filter (i.e. using multiple orders) more of the local image structure can be captured.

For each input image, 30 filtered versions of the image are computed, where the derivative of the image, $L(x)$, computed at an inner scale of σ is denoted by: $L_{n \times ny}$, with n_x and n_y denoting the order of the derivatives in the x and y directions. For each of the 5 scales the following 6 filtered images are computed: $L_{00}, L_{10}, L_{01}, L_{20}, L_{02}, L_{11}$.

The first four central moments (mean, standard deviation, skew and kurtosis) of the pixel intensity distribution from each ROI in the filtered and original images are computed as textural features [3,8].

Training separate classifiers for each of the 42 ROIs, thereby allowing them to be examined independently of one another, incorporates anatomical variation into the description of every region [3]. This is a particularly attractive approach in the present study, where the number of images containing no lung pathology is small; the number of normal regions is, however, much larger.

A kNN classifier is used. A leave-one-out method is used to construct the training and testing sets. The output of the classification process is a soft label for each feature vector in the range $[0, 1]$ where 0 = normal and 1 = abnormal. This allows receiver operating characteristic (ROC) curve analysis.

The per region classification results are used to create probability maps for the chest images where every pixel in the lung fields is replaced by the average of the soft labels of all the regions in which it falls - giving a posterior probability of that pixel being normal or abnormal. These maps are to be used as a visual tool by attracting the attention of a clinician reading the image to a region of abnormality and then allowing the clinician to make a decision regarding further interpretation of the image.

2.5 Performance Evaluation

The student's t -test is used to compare the classification results for the true (manual) and ASM segmentations; such a comparison gives an indication of the performance of the segmentation algorithm.

ROC analysis is performed independently on the 42 regions, resulting in a separate value for the area under the ROC curve (A_z) for each of the regions.

3 Results

The left and right lung fields were segmented independently of one another. The active shape models were trained using 25 images for the right lung field and 20 images for the left lung field.

Using an 11-NN classifier, the final classification results (in terms of area, A_z , under the ROC curve) of the system gave a per region score of $A_z = 0.782$ averaged over the 42 regions. No statistically significant difference was found between classifications on regions obtained by manual and ASM segmentation.

Table 1 shows a comparison between the results (in terms of per region Az scores) presented in this study and those presented in [3], where a similar scheme was used for the detection of TB in a screening database and interstitial disease in images obtained from daily clinical practice. Our database is similar to the TB database in [3], in that it contains a range of abnormalities encountered in chest screening, and we expect their results on the TB database to be a suitable basis for comparison with ours. The abnormalities identified in our images were primarily alveolar in origin, apart from a small group with nodular opacities, which could be either interstitial or alveolar.

Table 1. Comparison of Az values (across regions) with those of a similar study on adults [3]

	Mean	Min	Max
TB [3]	0.676	0.540	0.830
ILD [3]	0.835	0.660	0.930
Current study	0.782	0.655	0.937

The objective of a probability map is to indicate the degree of abnormality of every region in the lung field. The overlays on the lungs in Figures 2 and 3 define this degree of abnormality - a very white region indicates extreme abnormality, while a greyish-white region indicates that some subtle abnormalities could be present.

Figure 2 illustrates successful performance of the system, while Figure 3 illustrates common failures. The most frequent errors occurred in the basal and perihilar regions of the lungs, especially those of the left lung.

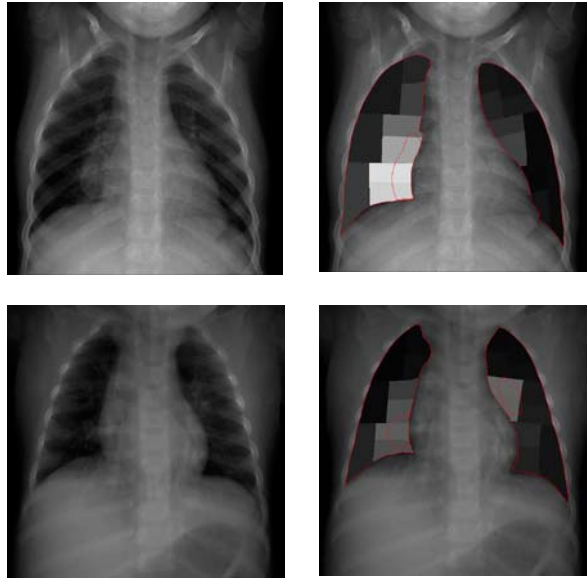


Fig 2. Successful probability maps: original image (left) and probability map (right). The healthy lung tissue is accurately depicted by darker grey values. Outlines of abnormal regions, made by a radiologist, are shown; the segmented lung borders are shown.

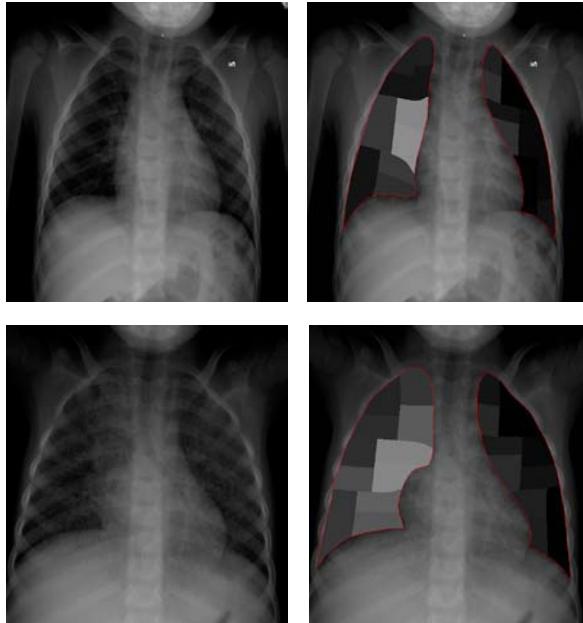


Fig 3. Common failures of the system. Top row: the probability map for a normal image indicates a greater degree of abnormality in the right perihilar region. Bottom: the image shows a diffuse nodular pattern in both lungs, which is only depicted in the right lung in the probability map.

4 Discussion

Frequent errors occurred in the basal and perihilar regions of the lungs, especially those of the left lung, and may be attributed largely to the sparseness of the data (in terms of number of abnormalities) in those regions of the lungs and the prominence of the vasculature in the lung fields.

Broncho-vascular markings (both normal and increased markings), proved problematic for the system. In many images, where the radiologist indicated increased broncho-vascular markings, the resulting probability maps displayed limited abnormality; while in many images where the radiologist indicated healthy vasculature the resulting probability maps showed abnormality in the regions containing vasculature. This is perhaps an unavoidable shortcoming of the classification system, which, in its current form, defines abnormality based solely on the textural characteristics of the lung tissue, while unhealthy broncho-vascular markings cannot always be distinguished from healthy counterparts using textural appearance alone. Experienced radiologists are able to take into consideration a large variety of factors when making a diagnosis, many of which are not quantifiable and therefore difficult to program. The probability map in Figure 3 (bottom) indicates diffuse abnormality in the right lung field but not in the left lung field for an input image containing a very obvious diffuse nodular pattern in both lung fields. This failure can be attributed to the fact that only three images containing this pattern in both lungs exist in the data set, while

several more exist displaying the pattern in the right lung only. The system is thus very poorly trained on such abnormalities for the left lung and since individual classifiers are trained for every region the results are poor.

The results presented in this study for pulmonary abnormality in children show an improvement over those presented in [3] for adult TB. Since similar methods are used in the two studies, the performance differences must be attributed to the data set. A few scenarios are possible. The abnormalities in this data set could be more obvious and therefore more easily detectable than those in the data set used in [3] – this is supported by the fact that the performance of the system presented in [3] on the interstitial lung disease data set is considerably higher. We used direct digital images, while the images in [3] were digitized from film. The digital X-ray machine we used may produce images which present textural pathologies in the lung fields more clearly. Testing on a larger and more diverse database and on image sets from different imaging systems may provide more clarity on system performance.

Acknowledgments. We thank Dr Rupesh Daya for assistance with data collection, and Dr Bram van Ginneken for guidance on algorithm implementation.

References

1. Marais, B.J., Pai, M.: New approaches and emerging technologies in the diagnosis of childhood tuberculosis. *Paediatric Respiratory Reviews* 8, 124–133 (2007)
2. Katsuragawa, S., Doi, K.: Computer-aided diagnosis in chest radiography. *Computerized Medical Imaging and Graphics* 31, 212–223 (2007)
3. van Ginneken, B., Katsuragawa, S., Romeny, B.M.T., Doi, K., Viergever, M.A.: Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Transactions on Medical Imaging* 21, 139–149 (2002)
4. Pitcher, R.D., Wilde, J.C.H., Douglas, T.S., van As, A.B.: The use of the Statscan digital X-ray unit in paediatric polytrauma. *Pediatric Radiology* 39, 433–437 (2009)
5. van Ginneken, B., Romeny, B.M.T.H., Viergever, M.A.: Computer-aided diagnosis in chest radiography: A survey. *IEEE Transactions on Medical Imaging* 20, 1228–1241 (2001)
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models - their training and application. *Computer Vision and Image Understanding* 61, 38–59 (1995)
7. Ruprecht, D., Muller, H.: Image warping with scattered data interpolation. *IEEE Computer Graphics and Applications* 15, 37–43 (1995)
8. Arzhaeva, Y., Tax, D.M.J., van Ginneken, B.: Dissimilarity-based classification in the absence of local ground truth: Application to the diagnostic interpretation of chest radiographs. *Pattern Recognition* 42, 1768–1776 (2009)
9. van Ginneken, B., Romeny, B.M.T.: Multi-scale texture classification from generalized locally orderless images. *Pattern Recognition* 36, 899–911 (2003)

Toward Precise Pulmonary Nodule Descriptors for Nodule Type Classification

Amal Farag, Shireen Elhabian, James Graham, Aly Farag, and Robert Falk*

Department of Electrical and Computer Engineering, University of Louisville
(*) Medical Imaging Division, Jewish Hospital, Louisville, KY, USA

Abstract. A framework for nodule feature-based extraction is presented to classify lung nodules in low-dose CT slices (LDCT) into four categories: juxta, well-circumscribed, vascularized and pleural-tail, based on the extracted information. The Scale Invariant Feature Transform (SIFT) and an adaptation to Daugman's Iris Recognition algorithm are used for analysis. The SIFT descriptor results are projected to lower-dimensional subspaces using PCA and LDA. Complex Gabor wavelet nodule response obtained from an adopted Daugman Iris Recognition algorithm revealed improvements from the original Daugman binary iris code. This showed that binarized nodule responses (codes) are inadequate for classification since nodules lack texture concentration as seen in the iris, while the SIFT algorithm projected using PCA showed robustness and precision in classification.

Keywords: texture descriptors, SIFT, Gabor wavelet, nodule type classification.

1 Introduction

Texture has been a subject of major investigation by researchers, especially dealing with imaging applications such as aerial, satellite, medical images, etc., since the texture can be defined as a function of the spatial variation in pixel gray levels [1]. In medical applications image analysis techniques have played a major role for such tasks like feature extraction, classification of normal and abnormal lung tissue, registration and segmentation. There are numerous methods within the scope of texture recognition such as parametric statistical model-based techniques, structural techniques and transform-based techniques [2]. In texture classification the main goal is to produce a map which enables classification of the input image(s) to the desired classes, object type classification in correlation to other images, etc.

There are various approaches for classification using texture, but all of the approaches fall-under linearly based or non-linearly-based approaches [2]. In this paper we investigate two approaches to extract texture information from lung nodules to automatically classify each nodule into one of four predefined categories identified in [3]. The first is an adoption of the linear-based algorithm known as the Daugman Iris Recognition Algorithm [4] and the second is the non-linear approach known as the Scale Invariant Feature Transform. These approaches and how they are used for lung nodule texture-based feature extraction will be described in details in the coming section.

The literature is rich with information on feature extraction whether it is domain specific; human face and fingerprinting, or general; color, texture and shape. The usage of texture approaches on lung nodules of low-dose CT (LDCT) slices, to the best of authors' knowledge, is not as common. The closest related works we found to our application are the following: Hara et al. [6] used 2nd order autocorrelation features to detect lung nodules in 3D chest images, there were no results pertaining to once the nodules were detected how to classify the nodules or if the 2nd order autocorrelation could provide information for further analysis. In [7] local texture analysis was used for identifying and classifying lung abnormalities such as tuberculosis. The k-nearest neighbor approach was implemented to extract the feature vector from the training set and leave out the feature vector that will be classified.

A nodule is defined as a small mass or lump of irregular or rounded shape, yet this definition is ambiguous when it comes to applying it in the fields of computer vision and machine learning, for example. Samala et al. [8] defined nine feature descriptors that describe the nodule characteristics that were used in assessments by radiologists. These descriptors are: 1. subtlety; 2. internal structure; 3. calcification; 4. sphericity; 5. margin; 6. lobulation; 7. speculation; 8. texture and 9. malignancy.

The assignment of the various nodule types can be formulated by allowing $I(x)$ to represent a CT slice, where $x = \{(x, y): 1 \leq x \leq N_x, 1 \leq y \leq N_y\}$ is a finite spatial grid supporting the slice and $x_0 = (x_0, y_0)$ be the centroid of a detected nodule region. The main objective of our framework is to assign a nodule type c to a given nodule region using texture-based descriptor $\mathcal{T}(x_0)$, where $c \in \{\mathcal{J}, \mathcal{W}, \mathcal{V}, \mathcal{P}\}$ which corresponds to juxta, well-circumscribed, vascular and pleural-trail respectively. This involves two main stages: first, detecting potential nodules for the given CT slice(s); second, building the nodule descriptor for each nodule type assignment/classification. In this paper we are concerned with the second stage. It is crucial that the local features extracted from the detected nodules are robust to various deformations due to scale, noise, acquisition artifacts, contrast and local geometric distortion[1].

This paper is organized as follows: section 2 describes the feature descriptor algorithms used in the classification analysis, section 3 discusses performance evaluation; and section 4 concludes the paper.

2 Feature Descriptors

Distinct object matching and description is an important goal for many medical imaging and computer vision applications. The success of the object description necessitate on two main conditions: invariance and distinction. The object description methodology must be robust to accommodate for various variations in imaging conditions and in the mean time producing a distinctive characterization of the desired object.

In this paper we use two algorithm designs for feature based description, Daugman Iris Recognition algorithm and the Scale Invariant Feature Transform (SIFT), on the nodule classification of Kostis et al. [3], which groups nodules into four main categories: (1) Well-circumscribed where the nodule is located centrally in the lung without being connected to vasculature; (2) Vascularized where the nodule has significant connection(s) to the neighboring vessels while located centrally in the lung; (3) Juxta-pleural where a significant portion of the nodule is connected to the pleural surface;

and (4) Pleural tail where the nodule is near the pleural surface, connected by a thin structure. Within these four categories we recognize variations in shape between the Juxta, Pleural Tail, Vascularized and Well-Circumscribed nodules and each other.

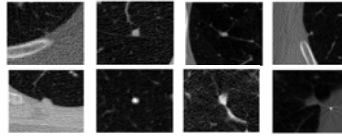


Fig. 1. Sample Nodules from Database. Column-wise: 1. Juxta; 2. Well-Circumscribed; 3. Vascularized and 4. Pleural-tail.

This difference in shape affects the gray level texture (pattern) distribution along the surface, where the nodule centroid is densely concentrated and becomes sparser as it dissipates to the outer boundaries of the nodule. Knowing this phenomenon from nodule regions we experiment with different feature descriptors to discover whether each nodule type has prominent feature descriptions that can be extracted and used for classification. Figure 1 depicts sample nodules from our database of the four types.

2.1 Daugman Iris Recognition Algorithm

John Daugman in 1994 [4] developed an algorithm for use in iris recognition which is also the basis of all current iris recognition systems. In this paper we adopt a similar framework to obtain nodule codes for each nodule type to classify them into one of four categories as described by Kostis et al.[3]. The overall Daugman Iris Recognition framework adopted for Nodule Classification is depicted in Figure 2. This paper concentrates on the steps after nodule detection through feature matching. The centroid of the nodule is located for computation as if it is the pupil of the eye. The radial distribution of the entire nodule is obtained, where each concentric circle area is then demodulated to extract its phase information using quadrature 2D Gabor wavelets.

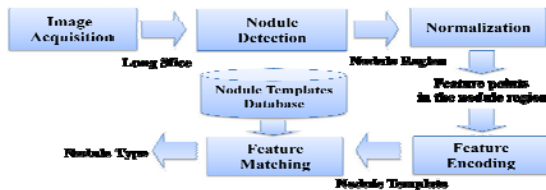


Fig. 2. Nodule Classification block-diagram adopted from Daugman's Iris Recognition framework

In Daugman's algorithm, binarization of the information was performed since the iris has concentrated texture information thus the phase binary information is enough for discrimination. In the case of the nodule we are not certain if the texture information is as concentrated as the iris, thus two techniques were implemented: (1) Daugman method with Gabor response binarization and (2) a modified Daugman method

in which the real complex values from the Gabor response are obtained, thus bypassing binarization in nodule feature encoding step.

$$\begin{aligned}
 &h_{\{Re,Im\}} \\
 &= sgn_{\{Re,Im\}} \int_{\rho} \int_{\emptyset} I(\rho, \emptyset) e^{-i\omega(\theta_o-\emptyset)} e^{-(r_o-\rho)^2/\alpha^2} e^{-(\theta_o-\emptyset)^2/\beta^2} \rho d\rho d\emptyset \tag{1}
 \end{aligned}$$

Where $h_{\{Re,Im\}}$ can be regarded as a complex-valued bit as either 1 or 0 depending on the sign of the 2D integral representing the real and imaginary parts; $I(\rho, \emptyset)$ is the raw nodule image in a dimensionless polar coordinate system that is translation and size invariant, α and β are the multi-scale 2D wavelet size parameters; ω is the wavelet frequency, spanning 3 octaves in inverse proportion to β ; $(r_o - \rho)$ represents the polar coordinated of each region of the nodule for which the phasor coordinates $h_{\{Re,Im\}}$ are computed. In the first technique the fractional Hamming Distance (HD) is computed, as conducted by Daugman. This distance is used as the dissimilarity measure between any 2 nodules, where 0 would represent a perfect match.

$$HD = \frac{\| (nodule A \otimes nodule B) \cap mask A \cap mask B \|}{\| mask A \cap mask B \|} \tag{2}$$

Where the XOR, \otimes , operator detects disagreement between any corresponding pair of bits while the AND, \cap , operator ensures that the compared bits are both deemed to have been uncorrupted by the lung nodule surroundings. Since the second technique is unbinarized, the Hamming Distance cannot be used thus the Euclidean Distance (ED) was computed, which examines the root of square difference between any 2 nodules. Given two descriptors $\mathbf{d}_i = [d_{i1}, d_{i2}, \dots, d_{iN}]^T$ and $\mathbf{d}_j = [d_{j1}, d_{j2}, \dots, d_{jN}]^T$, ED can be defined as follows;

$$ED_{ij} = \sqrt{\sum_{k=1}^n (d_{ik} - d_{jk})^2} \tag{3}$$

2.2 Scale Invariant Feature Transform (SIFT)

As detailed in [6], SIFT consists of four main steps: (1) Scale-space peak selection; (2) Key-point localization; (3) Orientation assignment and (4) Key-point descriptor. In the first step, potential interest points are detected using a scale-space continuous function $\mathbf{L}(\mathbf{x}, \sigma_s)$, it can be constructed by convolving the image $\mathbf{I}(\mathbf{x})$ with a cylindrical Gaussian kernel $\mathbf{G}(\mathbf{x}, \sigma_s)$ which can be viewed as a stack of 2D Gaussians one for each band. According to Lowe [5], the scale is discretized as $\sigma_s \in \{k^s\}$ where $k = 2^{1/3}$ and $s = \left\{ -1, 0, 1, 2, \dots, \frac{\log(s_{max})}{1/3 \log 2} \right\}$. Scale-space extrema detection searches over all scales σ_s and image locations $\mathbf{x} = \{(x, y)\}$ to identify potential interest points which are invariant to scale and orientation; this can be efficiently implemented using Difference-of-Gaussians $\mathbf{D}(\mathbf{x}, \sigma_s)$ which takes the difference between consecutive scales, i.e. $\mathbf{D}(\mathbf{x}, \sigma_s) = \mathbf{L}(\mathbf{x}, \sigma_s) - \mathbf{L}(\mathbf{x}, \sigma_{s-1})$, a point \mathbf{x} is selected to be a candidate interest point if it is larger or smaller than its 3×3 neighborhood system defined on

$\{D(\mathbf{x}, \sigma_{s-1}), D(\mathbf{x}, \sigma_s), D(\mathbf{x}, \sigma_{s+1})\}$, where σ_s is marked to be the scale of the point \mathbf{x} . This process leads to too many points some of which are unstable (sensitive to noise); hence removal of points with low contrast and points that are localized along edges is accomplished. In our framework, we assume that nodules have been already detected which correspond to interest/key points in Lowe's algorithm, hence this step can be bypassed. In order to obtain a nodule SIFT descriptor which is invariant to orientation, a consistent orientation should be assigned to the detected nodule which is represented by its centroid \mathbf{x}_0 . This orientation is based on the gradient of the nodule's local image patch. Considering a small window surrounding \mathbf{x}_0 , the gradient magnitude and orientation can be computed using finite differences. Local image patch orientation is then weighted by the corresponding magnitude and Gaussian window. Eventually the orientation is selected to be the peak of the weighted orientation histogram. Building a nodule SIFT descriptor is similar to orientation assignment, for example a 16×16 image window surrounding the nodule centroid point \mathbf{x}_0 is divided into sixteen 4×4 sub-windows, then an 8-bin weighted orientation histogram is computed for each sub-window, hence, we obtain $16 \times 8 = 128$ descriptors for each nodule. Thus, each detected nodule can now be defined at location (x_0, y_0) , specific scale σ , explicit orientation θ and descriptor vector $\mathbf{x}_0 = \{x_0, y_0, \sigma, \theta, \mathbf{d}\}$. Thus the SIFT operator $\mathcal{S}: I(x) \rightarrow X$ can be viewed as mapping a CT slice $I(x)$ to the nodule space with n -nodules, $X = \{\mathbf{x}_i\}_{i=1}^n$ detected from $I(x)$, where $\mathbf{x}_i = \{x_0^i, y_0^i, \sigma_i, \theta_i, \mathbf{d}_i\}$. Principle component analysis (PCA) [9] and linear discriminant analysis (LDA) [10] are used to project the extracted SIFT descriptors to a low-dimensional subspace where noise is filtered out.

3 Experimental Results

This work is based on the Early Lung Cancer Action Program (ELCAP) public database [11], which contains 50 sets of low-dose CT lung scans taken at a single breath-hold with slice thickness 1.25 mm. The locations of the 397 nodules are provided by the radiologists, where 39.12% are juxta-pleural nodules, 13.95% are vascularized nodules, 31.29% are well-circumscribed nodules and 15.65% are pleural-tail nodules. In this paper we created a subset database containing 294 nodules of the original 397. The ELCAP database is of resolution 0.5×0.5 mm [11]. Since we assume that the nodule region has been already detected, we use the groundtruth marked nodules by the radiologists to avoid sources of errors due to automated detection. Given a nodule's centroid, we extract texture descriptor information using two main techniques: Daugman coding and the SIFT descriptor. Tables 1 and 2 visualize the intermediate steps performed to generate a nodule code for the four nodule types while table 3 represents the SIFT descriptors of the nodule ensembles given in Tables 1 and 2.

Training was performed using two randomly drawn approaches; the bootstrapping resampling technique [12] and a one-time random sampling approach. The results using both methods were comparable, thus only the one time random sampling results are shown in this paper.

Classification was performed using the nearest-neighbor classifier with Euclidean distance as the similarity measure. To quantify nodule type classification performance, we measure true positives rates. A classification result is considered a true positive if a sample from class \mathbf{w}_i is classified as belonging to the same class.

Table 1. Visualization of Daugman Recognition process for the same Juxta and Well-Circumscribed Nodules in Fig. 1

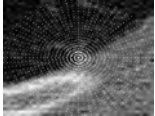
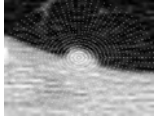
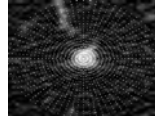













	Juxta		Well-Circumscribed	
Radial Distribution				
Rubber Sheet				
Binarized Code				
Unbinarized Code				

Table 2. Visualization of Daugman Recognition process for the same Pleural-Tail and Vascularized Nodules in Fig. 1

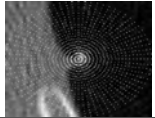
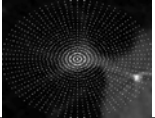
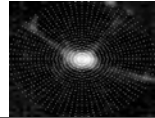













	Pleural-Tail		Vascularized	
Radial Distribution				
Rubber Sheet				
Binarized Code				
Unbinarized Code				

Table 4 shows the classification results for nodule codes for different percentages of training data ($x\%$ is the amount of ground-truth nodules taken into consideration in the training phase). It can be inferred that binarizing Gabor wavelet complex responses makes the nodule region lose discriminatory texture which assists in nodule type classification; this is not the case for iris recognition where it is known that the iris region is rich in texture. Gabor-based descriptor provides higher discrimination for juxta and well-circumscribed nodule types when compared to vascular and pleural tail. In general 50% training data can be used to provide overall excellent classification results.

Table 5 shows the classification results for nodule SIFT descriptor for different percentages of training data. We also use the projection of SIFT descriptor on a PCA-based and LDA-based subspaces trained by the descriptors of each nodule type. It can be observed that using the raw SIFT descriptor without statistical modeling provides the worst classification performance when compared to their PCA and LDA projection, yet, it provides similar performance to complex-valued nodule codes. LDA projection provides the greatest classification results when 100% nodule training is conducted, yet, in reality the model will not be re-trained each time new nodules are extracted to then classify them again (i.e. Input everything desired, output the same as input). As training percentage decreases, PCA projection results surpass those from the raw SIFT, LDA SIFT and Daugman nodule codes, this emphasizes the ability of

PCA as a statistical modeling approach to generalize the decision boundary between different classes (nodule types) with minimal amounts of training data. On the other hand, LDA requires larger amounts of training samples to provide a general subspace which has high discrimination power.

Table 3. Visualization of SIFT Recognition process for the same Juxta , Well-Circumscribed, Pleural-Tail and Vascularized Nodules in Fig. 1

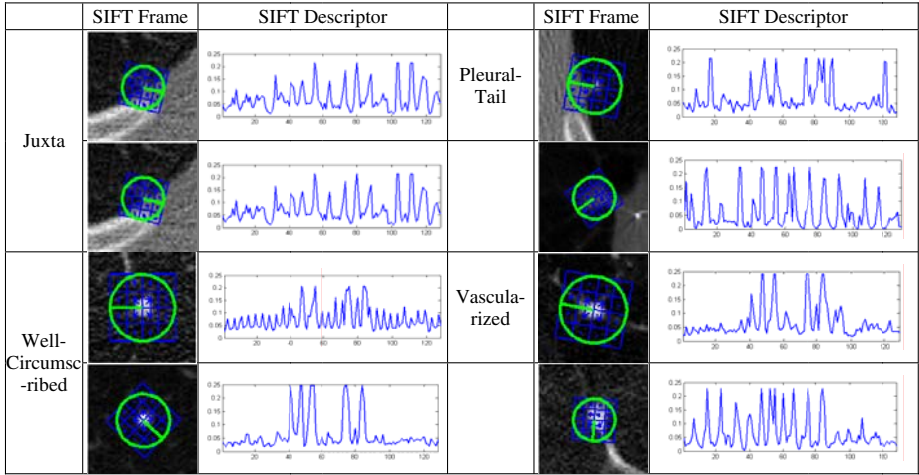


Table 4. Daugman Results for the Binarized and Unbinarized Nodules using the Hamming and Euclidean Distances (eqs. 2 and 3)

Training Percentage	Binary Nodule Codes					Complex-valued Nodule Codes				
	Overall	Juxta	Well	Vascular	Pleural Tail	Overall	Juxta	Well	Vascular	Pleural Tail
100%	41.84%	50.42%	43.48%	19.51%	36.96%	56.80%	74.78%	60.87%	21.95%	34.78%
75%	40.82%	49.57%	41.30%	19.51%	36.96%	61.56%	76.52%	67.39%	24.39%	45.65%
50%	40.48%	49.57%	43.48%	24.39%	26.09%	58.50%	73.04%	64.13%	21.95%	43.48%
25%	34.69%	41.74%	39.13%	19.51%	21.74%	54.42%	69.57%	58.70%	21.95%	36.96%

Table 5. SIFT Results obtained from the original SIFT, and results after applying the PCA and LDA methods to the SIFT output

Training Percentage	Raw SIFT					PCA SIFT					LDA SIFT				
	Overall	Juxta	Well	Vascular	Pleural Tail	Overall	Juxta	Well	Vascular	Pleural Tail	Overall	Juxta	Well	Vascular	Pleural Tail
100%	58.84%	70.43%	72.83%	26.83%	30.43%	67.69%	77.39%	83.70%	31.71%	43.48%	88.10%	91.30%	90.22%	85.37%	78.26%
75%	56.80%	66.96%	75%	19.51%	28.26%	74.49%	80.00%	88.04%	48.78%	56.52%	78.23%	83.48%	75%	82.93%	67.39%
50%	56.80%	68.70%	73.91%	14.63%	30.43%	76.53%	87.00%	81.52%	56.10%	58.70%	68.71%	73.91%	65.22%	65.85%	65.22%
25%	58.23%	64.35%	57.61%	12.20%	28.26%	78.57%	83.48%	82.61%	65.85%	69.57%	51.36%	60.00%	56.52%	39.02%	30.43%

4 Conclusion

In this paper we investigated the effects of texture analysis using two feature descriptor algorithms; Gabor wavelet nodule responses obtained from an adopted Daugman

Iris Recognition algorithm and the SIFT descriptor with lower-dimensional subspaces projections using PCA and LDA. The results from the descriptors were used to classify the nodules into their corresponding classes as defined by [3]. The results revealed the PCA SIFT method was more robust for lesser nodule training data. The Gabor wavelet unbinarized nodule responses provided better results than the binarized original framework of Daugman due to the lack of concentrated texture information in the lung nodules, yet these results were inferior to the overall SIFT performance. Future directions are geared toward generating a larger nodule database from other clinical data to expand our work. Further experimentations with this approach in terms of training and testing data will be conducted. We are also aiming to examine other feature descriptor approaches to compare with the results obtained in this paper.

References

1. Tuceryan, M., Jain, A.K.: Texture analysis. In: Chen, C., Pau, L., Wang, P. (eds.) *Handbook of Pattern Recognition and Computer Vision*, 2nd edn., pp. 207–248. World Scientific, River Edge (1998)
2. Tao, B., Dickinson, B.: Texture recognition and image retrieval using gradient indexing. *Journal of Visual Communication and Image Representation* 11(3), 327–342 (2000)
3. Kostis, W.J., et al.: Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up. *Radiology* 231, 446–452 (2004)
4. Daugman, J.: How iris recognition works. In: *Proceedings of 2002 International Conference on Image Processing*, vol. 1 (2002)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110.4 (2004)
6. Hara, T., Hirose, M., Zhou, X., Fujita, H., Kiryu, T.: Nodule detection in 3D chest CT images using 2nd order autocorrelation features. In: *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China (2005)
7. van Ginneken, B., Katsuragwa, S., Romney, B., Doi, K., Viergever, M.: Automatic Detection of Abnormalities in Chest Radiographs Using Local Texture Analysis. *IEEE Transactions on Medical Imaging* 21(2) (2002)
8. Samala, R., et al.: A Novel Approach to Nodule Feature Optimization on Thin Section Thoracic CT. *Acad. Radiology* 15, 1181–1197 (2009)
9. Shlens, J.: A Tutorial on Principal Component Analysis. Institute for Nonlinear Science, UCSD (2005)
10. Balakrishnama, S., Ganapathiraju, A.: Linear Discriminant Analysis- A brief tutorial, http://www.isip.msstate.edu/publications/reports/isip_interanal/1998/linearpublications/linear_discrim_analysis/
11. ELCAP public lung image database, <http://www.via.cornell.edu/databases/lungdb.html>
12. Polikar, R.: *Ensemble Based Systems in Decision Making* (2006)

Morphology-Guided Graph Search for Untangling Objects: *C. elegans* Analysis

Tammy Riklin Raviv¹, V. Ljosa², A.L. Conery³, F.M. Ausubel³, A.E. Carpenter², P. Golland¹, and C. Wählby²

¹ Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA

² Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA

³ Dept. of Molecular Biology and Center for Computational and Integrative Biology, Mass. General Hospital, Boston, MA

Abstract. We present a novel approach for extracting cluttered objects based on their morphological properties. Specifically, we address the problem of untangling *Caenorhabditis elegans* clusters in high-throughput screening experiments. We represent the skeleton of each worm cluster by a sparse directed graph whose vertices and edges correspond to worm segments and their adjacencies, respectively. We then search for paths in the graph that are most likely to represent worms while minimizing overlap. The worm likelihood measure is defined on a low-dimensional feature space that captures different worm poses, obtained from a training set of isolated worms. We test the algorithm on 236 microscopy images, each containing 15 *C. elegans* worms, and demonstrate successful cluster untangling and high worm detection accuracy.

1 Introduction

Recent progress in robotic sample preparation, combined with automated microscopy and image analysis, enables high-throughput screening experiments for testing biological processes such as immunity, behavior, and metabolism. For example, high-throughput screening of the roundworm *Caenorhabditis elegans* is used to test tens of thousands of chemical or genetic perturbations to identify promising drugs and regulators [1,2,3]. Automatic processing of the vast amount of data obtained from a screen is therefore necessary.

Existing methods compute image-based statistics, e.g., the ratio of fluorescing-stained area in the image to the area covered by worms [4]. However, many scientific questions require measurements on individual worms, such as shape and location of reporter signals. Worms in these images often overlap and cluster into clumps, making analysis based on individual worms challenging. A few solutions have been proposed for the analysis of individual worms in low-throughput video sequences where the worms are disambiguated by tracking them over time [5,6,7]. Methods for resolving high resolution 3-D images of worms have also been demonstrated [8,9]. These methods, however, are not suitable for the comparatively low-resolution, 2-D images produced in high-throughput experiments. Extracting overlapping worms in images has been recently demonstrated by thresholding the curvature of skeleton segments of a worm cluster [10]. Yet, this

method has a limited capability to detect curvy worms or clustered worms in complex configurations.

While isolated worms can be extracted easily based on the differences of image intensities, the image data alone is not sufficient to delineate clustered worms. Moreover, edges and intensity variations within the worms often mislead conventional segmentation algorithms. Nevertheless, while the different poses of the worms introduce significant extrinsic geometrical differences, all worms share similar intrinsic geometrical properties such as length and thickness profile. The shape characteristics should therefore play a key role in a segmentation algorithm.

There exist a few algorithms that incorporate shape priors into the segmentation process for extracting multiple, possibly overlapping, objects [11], but these shape-based algorithms assume that the expected number of objects to segment is known. In practice, due to computational constraints, only a few partially occluded objects can be segmented simultaneously.

We present a conceptually novel algorithm for untangling clusters of worms based exclusively on their poses. Our formulation leads to a computationally efficient, morphology-guided graph search that relies on a probabilistic model of the worm poses, learned from training data. We model the skeleton of a worm cluster by a graph whose vertices correspond to worm segments and search over paths in the graph that are more likely to represent complete individual worms while minimizing overlap. We formulate the problem as a minimization of a cost function. To reduce computational complexity, we use a greedy search strategy to approximate the optimal exhaustive search.

The elongated, thin structure of the worms, together with the similarity in their thickness profiles, motivate the use of their medial axis transform (skeleton) as their shape descriptor [12][13][14]. Since the deformations of the worms are nearly isometric (no stretching, shrinkage, etc.), the worms' variability in appearance can be captured by a low-dimensional feature space. We apply principal component analysis (PCA) to a comprehensive set of worm shapes from the training set. Only a limited number of eigen vectors are needed to reliably represent the entire population. We use the shape space to define a probability distribution that guides the detection of the most probable worm descriptors within the graph search.

While the PCA-based representation of the worms' shape has been demonstrated before [15], the contributions of this paper include the graph representation, the graph search algorithm and the greedy approximation. To the best of our knowledge, we demonstrate the first robust, automatic method for identifying a large number of objects (worms) in cluttered clusters.

We test the algorithm on 236 microscopy images, each containing a well with 15 *C. elegans* worms, and demonstrate successful cluster untangling and high worm-detection accuracy. The quality of the results obtained with a relatively low computational complexity shows the promise of our method to remove the computational bottleneck in large-scale biological experiments.

2 Probabilistic Shape Model

In this section, we briefly review the shape representation we use in our algorithm. The details are described elsewhere [15]. A worm descriptor consists of n equally spaced

control points $\mathbf{c} \in \mathcal{R}^2$ along the worm skeleton and the average worm thickness, obtained by using the medial axis transform [12][13]. This $(2n + 1)$ dimensional vector is used for reconstruction and retrieval. The worm descriptors are aligned by similarity transformation (translation and rotation) based on their first and second moments.

We use PCA to project the descriptors of the worms in the training set onto a low-dimensional feature space. The PCA model implies a probability measure on the space spanned by the most significant eigen vectors,

$$P(\mathbf{x}) \propto \exp(-\mathbf{w}^M \Sigma_M^{-1} \mathbf{w}), \quad (1)$$

where $\Sigma_M = \text{diag}(\lambda_1 \dots \lambda_M)$ are the highest eigen values and $\mathbf{w} = \{w_1, \dots, w_M\}$ are the weights obtained by projecting the worm descriptor \mathbf{x} onto the space spanned by the M most significant eigen vectors. The dimensionality M is chosen so that the selected eigen values explain 99% of the variance.

3 Graph Representation and Search Algorithm

We now focus on the main contribution of the paper, namely performing cluster un-tangling using morphology-guided graph search. We assume that the worm clusters are already segmented from the background and a worm likelihood measure $P(\mathbf{x})$ is defined based on the training set, according to Eq. (1).

3.1 Graph Representation of the Worm Cluster Skeleton

Let $G_s = \{V, E\}$ denote a sparse, directed graph that represents the skeleton S of a worm cluster, where $V = \{v_i\}$ and $E = \{e_{i,j}\}$ denote the graph vertices and edges, respectively. We denote the number of vertices by $\mathcal{V} = |V|$. Fig. 1b shows the skeleton that corresponds to the worm cluster shown in Fig. 1a with the intersection points marked in red. These cluster intersection points partition the skeleton into segments that are represented by the graph vertices, as demonstrated in Fig. 1c. The length of a skeleton segment is determined by the number of pixels it contains. An edge $e_{i,j}$

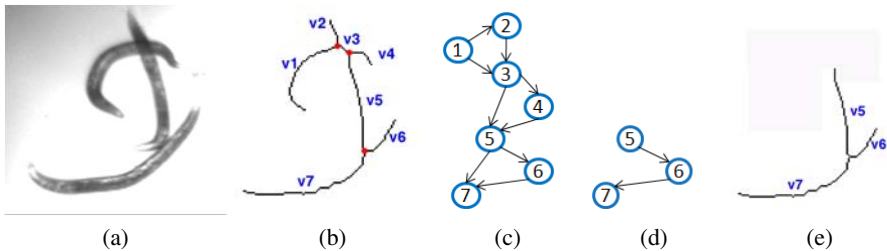


Fig. 1. Graph representation of a worm cluster. (a) Original image. (b) Cluster skeleton. Intersection points are marked in red. (c) Graph representation G_s of the skeleton shown in b. Each segment becomes a node and each intersection becomes a clique in the graph. (d) A simple path in G_s and (e) the corresponding skeleton segments. These skeleton segments form a branch which is unlikely to represent a worm skeleton.

connects two vertices $\{v_i, v_j\}$ if their corresponding segments share a common intersection point. To avoid path duplication we impose an order on the graph nodes, making the graph directed, i.e., $e_{i,j} = 1$ and $e_{j,i} = 0$ if $i < j$. A simple path in the graph is a chain of connected vertices without cycles. Note, however, that not every simple path in the graph represents a probable worm skeleton. A path in the graph that includes, for example, a 3-clique represents 3 worm segments with a common intersection point. These segments form a branch rather than an open curve and therefore are not likely to represent a worm skeleton. See Fig. 1d-e for an example of a path that forms a branch.

We construct an $\mathcal{V} \times \mathcal{V}$ adjacency matrix representing G_s . Neighborhoods of vertices are determined by detecting the intersections associated with each skeleton segment. We use the breadth-first search (BFS) algorithm to detect the paths in G_s . Let L denote the maximal length of a path, i.e., its total number of vertices minus one. We construct length- l paths, where $l = 0 \dots L$, by visiting the end vertices of each of the existing length- $(l-1)$ paths and attaching the neighboring vertices. That is, if the last node of a length- $(l-1)$ path has two neighbors, then two additional length- l paths are detected.

A general graph search of this form (as in the traveling-salesman problem) has a combinatorial complexity of $O(\mathcal{V}!)$ when $L = \mathcal{V}$. Yet, the nature of the problem reduces its complexity significantly. Since branches (cliques of more than two vertices) are excluded, the maximal length L of the likely paths corresponds to the number of intersection points. Let D denote the maximal size of the clique associated with an intersection point in the graph (i.e., the number of segments attached to it). Then the number of likely paths in the graph is in on the order of $O(L^2 D^2)$. In practice we limit the length of the paths even further so that the length in pixels of the corresponding skeleton segments does not exceed the length of the longest worm seen during the training.

In the example shown in Fig. 1 there are 3 intersection points, thus $L = 3$. For all intersection points in this example $D = 3$. In this particular graph the probable paths include 7 length-0 paths, 9 length-1 paths, 6 length-2 paths and 4 length-3 paths.

3.2 Search for the Most Probable Paths in the Graph

We obtain the approximated number of worms in a given cluster, denoted here by K , from the ratio of the cluster area and the median area of the training worms. Let ρ_k denote a path in G_s that represents the skeleton of the k -th worm. We can now formalize the cluster untangling problem as a search for K paths in G_s that minimize the cost function

$$E(\rho_1 \dots \rho_K) = - \sum_{k=1}^K \log P(\rho_k) + \alpha \sum_{k=2}^K \sum_{l=1}^{k-1} |\rho_k \cap \rho_l| + \beta |\tilde{V}_K|, \quad (2)$$

where $\tilde{V}_K = \{v | v \in G_s, v \notin [\rho_1 \cup \rho_2 \dots \cup \rho_K]\}$. Let us now take a closer look at this function. The first term (the *likelihood cost*) encourages selection of the paths that are most likely to represent complete, individual worms. The second term (the *overlap cost*) encourages minimization of the pairwise overlap between the selected paths:

$$|\rho_k \cap \rho_l| = \{v | v \in \rho_k \text{ and } v \in \rho_l\}. \quad (3)$$

The third term (the *leftover cost*) is the number of vertices $|\tilde{V}_K|$ in G_s that are not included in the union of the selected paths $\{\rho_k\}$. This term encourages the ‘coverage’

of the worm cluster skeleton S . The scalar weights α and β balance the terms in the function. In all of our experiments we set α and β to 2.

The global minimum of this optimization problem can be obtained by an exhaustive search for all subsets of K out of \mathcal{P} (all) paths in G_s . This is, however, a combinatorial problem of order $\binom{\mathcal{P}}{K}$. To reduce the computational time we apply a greedy strategy. At each stage we make a locally optimized choice of a path in the graph until we select K paths. Given $\{\rho_1 \dots \rho_{k-1}\}$ selected in the previous iterations of the search, the greedy algorithm seeks a path ρ_k that minimizes

$$E(\rho_k | \rho_1 \dots \rho_{k-1}) = -\log P(\rho_k) + \alpha \sum_{l=1}^{k-1} |\rho_k \cap \rho_l|. \tag{4}$$

We can represent the cost of the first selected path ρ_1 by reducing Eq. (4) for $k = 1$:

$$E(\rho_1) = -\log P(\rho_1) \tag{5}$$

Once we complete $K - 1$ iterations, we also test if K is the optimal number of paths to represent the cluster. We compare the incremental cost of having only $K - 1$ paths, i.e.,

$$E(\rho_K = \emptyset | \rho_1 \dots \rho_{K-1}) = \beta |\tilde{V}_{K-1}|, \tag{6}$$

to the cost of adding one additional path:

$$E(\rho_K | \rho_1 \dots \rho_{K-1}) = -\log P(\rho_K) + \alpha \sum_{l=1}^{K-1} |\rho_K \cap \rho_l| + \beta |\tilde{V}_K|. \tag{7}$$

In other words, we extract K rather than $K - 1$ paths from a cluster only if the cost of having an additional path ρ_K is less than the cost of not having its contribution to the 'coverage' of the skeleton represented by G_s . This condition can be viewed as a stopping criterion of the algorithm.

The greedy strategy still attempts to minimize the original cost function in Eq. (2) since the $K - 1$ cost functions in the form of Eq. (4) and the cost function in Eq. (7) sum up to the cost functional in Eq. (2) for selecting K paths in G_s . Replacing Eq. (7) with Eq. (6), we obtain the cost function in Eq. (2) for $K - 1$ paths.

3.3 Morphology-Guided Graph Search Algorithm

The proposed algorithm can be summarized as follows:

- **Initialize**
 - Calculate K .
 - Use a breadth-first search to detect \mathcal{P} paths in G_s as described in Section 3.1. This is the *path candidates* set.
 - Define an empty set of *selected paths*.
 - Calculate the *likelihood cost* $E(\rho) = -\log P(\rho)$ of each path.
- **For** $k = 1$ **to** $(K - 1)$:
 - Move the path ρ_k with the lowest cost $E(\rho_k | \rho_1 \dots \rho_{k-1})$ in the *path candidates* set to the *selected paths* set.
 - For each path ρ_l in the *path candidates* set compute the *overlap cost* (Eq. (3)) due to its intersection with the path ρ_k .

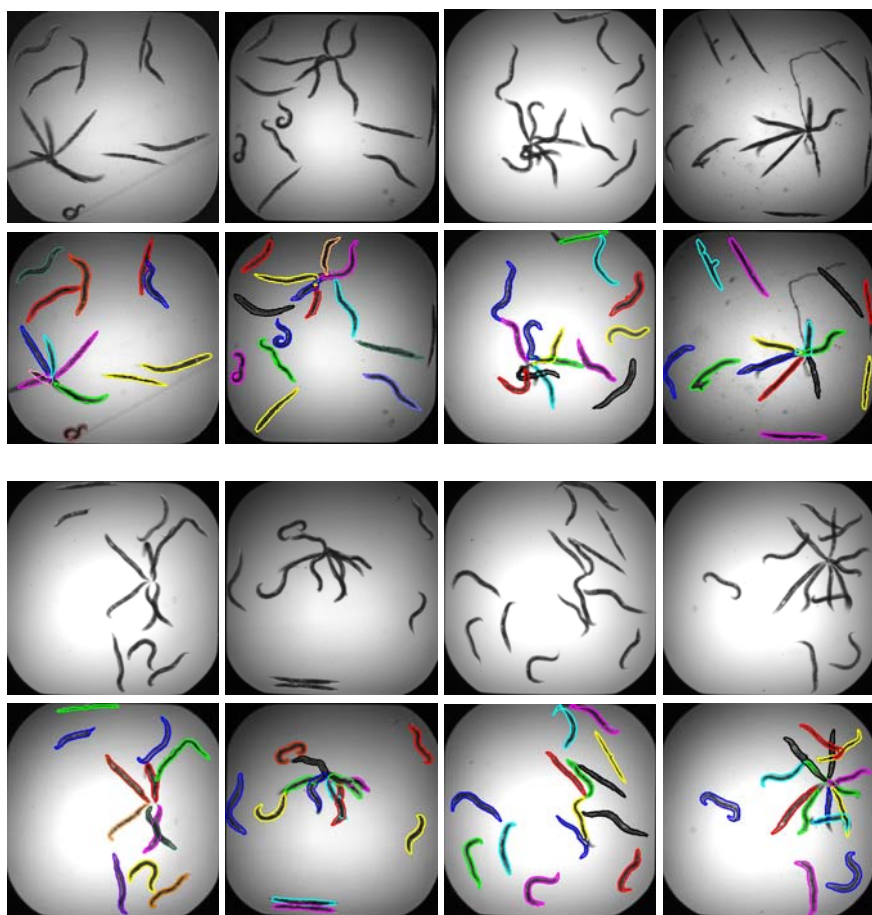


Fig. 2. Example images (odd rows) with their clusters resolved by the algorithm (even rows)

– **Determine the total number of selected paths**

- Update the costs of the remaining paths in the *path candidates* set by adding the *leftover cost*, (Eq. 7). Let C_K denote the lowest cost. This is the cost for adding the K -th path to the *selected paths* set.
- Calculate the *leftover cost* for not selecting an additional path (Eq. 6). Let C_{K-1} denote this cost.
- If $C_K < C_{K-1}$ output $\{\rho_1 \dots, \rho_K\}$, otherwise output $\{\rho_1 \dots \rho_{K-1}\}$.

4 Experiments

We evaluated the algorithm on an image set from a viability screen of *C. elegans* that were infected by a bacterial pathogen, washed, transferred to 384-well plates containing liquid medium with the compound to be tested and incubated until the infection killed untreated worms. The worm plates were imaged within 3, 24, 48, 72, 96, 120 and 144

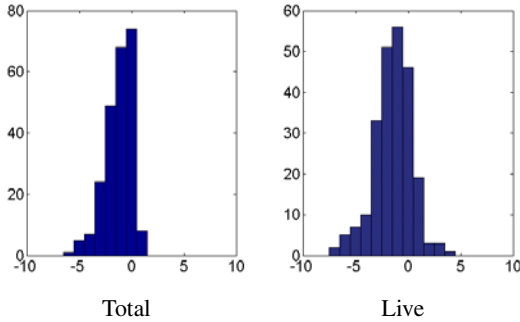


Fig. 3. Histograms of the differences in count between the automatic worm detection and the manual worm count for each of the 236 images, containing 3479 worms. Left: Total worm count differences. Right: Live worm count differences.

hours after treatment by an automated microscope with a $2\times$ magnification lens. Details of the sample preparation and image acquisition are available elsewhere [4]. The worms were segmented from the background using thresholding [15]. The training set contained 454 individual manually detected worms. Worm descriptors were generated using 21 equally-spaced control points along the worm skeletons, together with their average thickness. The feature space of the worm descriptors was spanned by the 10 most significant eigen vectors.

We applied the proposed algorithm to 236 images containing 3479 worms for which we had manual annotations by an expert. Fig. 2 presents successful untangling examples of images with complex worm clusters. To quantify the performance of the algorithm, we compared the number of worms detected by our algorithm with the manual count. In addition, we classified the detected worms as dead or alive, based on a simple classifier we trained on the training set used to construct the PCA space. Typically, dead worms are stiff and therefore straight. We then compared the number of worms classified as live to the manual live worm counts. Fig. 3 shows histograms of the count differences between the automatic worm detection and the manual worm count for each of the 236 images. Due to uneven illumination, worms near the well edges are not always detected by the preliminary segmentation. This leads to an underestimated worm counts. Overestimated total worm counts are caused by worm splitting or misclassifying debris as worms. In summary, 9% of the worms were not detected while the false detection percentage was 0.23%. Note, that underestimation in the total worm counts implies ignoring part of the data while overestimated counts (very few here) may introduce larger bias to the classification results.

5 Conclusions

This paper addresses a particularly challenging problem of extracting clustered objects. The objects are entangled and their poses vary, thus hierarchical models [16] or discriminative boosting algorithms [17] that have been proposed for image parcellation into multiple (brain) structures, are not applicable.

By exploiting the specific characteristics of our data, we developed a novel detection method based on concepts from machine learning and graph theory. Representing clustered worm segments as graph vertices, we search for paths in the graph that are more likely to represent complete, individual worms. The thin, elongated structure of

the worms motivates the use of their skeletons as a shape descriptor and allows a reduction in the computation complexity of the graph search. Yet, we believe that our morphology-guided graph search can be generalized to detect cluttered objects of different shapes with a suitable choice of descriptors.

Acknowledgments. This work was supported in part by the NIH grants R01-AI072508, P01-AI083214, R01-AI085581 and NAMIC U54-EB005149 and by the NSF CAREER Award 0642971.

References

1. Artal-Sanz, M., de Jong, L., Tavernarakis, N.: *Caenorhabditis elegans*: A versatile platform for drug discovery. *Biotechnol. J.* 1, 1405–1418 (2006)
2. Kamath, R.S., Ahringer, J.: Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* 30(4), 313–321 (2003)
3. Sifri, C.D., Begun, J., Ausubel, F.M.: The worm has turned – microbial virulence modeled in *Caenorhabditis elegans*. *Trends Microbiology* 13, 119–127 (2005)
4. Moy, T.I., Conery, A.L., Larkins-Ford, J., Wu, G., Mazitschek, R., Casadei, G., Lewis, K., Carpenter, A.E., Ausubel, F.M.: High-throughput screen for novel antimicrobials using a whole animal infection model. *ACS Chemical Biology* 4(7), 527–533 (2009)
5. Ramot, D., Johnson, B., Berry, T., Carnell, L., Goodman, M.: The parallel worm tracker: A platform for measuring average speed and drug-induced paralysis in nematodes. *PLoS One* 3(5) (2008)
6. Geng, W., Cosman, P., Berry, C.C., Feng, Z., Schafer, W.R.: Automatic tracking, feature extraction and classification of *C. elegans* phenotypes. *IEEE Transactions on Biomedical Engineering* 51(10), 1811–1820 (2004)
7. Restif, C., Metaxas, D.: Tracking the swimming motions of *C. elegans* worms with applications in aging studies. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 35–42. Springer, Heidelberg (2008)
8. Long, F., Peng, H., Liu, X., Kim, S.K., Myers, E.: A 3D digital atlas of *C. elegans* and its application to single-cell analyses. *Nature Methods* 6(9), 667–672 (2009)
9. Murray, J.I., Bao, Z., Boyle, T.J., Boeck, M.E., Mericle, B.L., Nicholas, T.J., Zhao, Z., Sandel, M.J., Waterston, R.H.: Automated analysis of embryonic gene expression with cellular resolution in *C. elegans*. *Nature Methods* (2008)
10. Rizvandi, N.B., Pizurica, A., Philips, W.: Machine vision detection of isolated and overlapped nematode worms using skeleton analysis. In: *Proc. of ICIP*, pp. 2972–2975 (2008)
11. Cremers, D., Sochen, N., Schnörr, C.: A multiphase dynamic labeling model for variational recognition-driven image segmentation. *IJCV* 66(1), 67–81 (2006)
12. Blum, H.: A transformation for extracting new descriptors of shape. *Models for the perception of speech and visual form*, 362–380 (1967)
13. Goutsias, J., Shonfeld, D.: Morphological representation of discrete and binary images. *IEEE Transactions on Signal Processing* 39(6), 1369–1379 (1991)
14. Stephens, G.J., Kerner, J.B., Bialek, W., Ryu, W.S.: Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput. Biol.* 4(4) (2008)
15. Wählby, C., Riklin-Raviv, T., Ljosa, V., Conery, A.L., Golland, P., Ausubel, F.M., Carpenter, A.E.: Resolving clustered worms via probabilistic shape models. In: *Proc. of ISBI* (2010)
16. Pohl, K., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M.: A hierarchical algorithm for MR brain image parcellation. *IEEE TMI* 26(9), 1201–1212 (2007)
17. Tu, Z., Narr, K., Dollár, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. *IEEE TMI* 27(4), 495–508 (2008)

Automatic Cephalometric Evaluation of Patients Suffering from Sleep-Disordered Breathing

Lior Wolf¹, Tamir Yedidya¹, Roy Ganor¹, Michael Chertok²,
Ariela Nachmani³, and Yehuda Finkelstein^{3,4}

¹ School of Computer Science, Tel-Aviv University

² School of Engineering, Bar-Ilan University

³ Center of Cleft Palate and Craniofacial Anomalies, Meir Medical Center

⁴ Palate Surgery Unit, the Department of Otolaryngology: Head and Neck Surgery, Meir Medical Center, affiliated with Sackler School of Medicine Tel-Aviv University

Abstract. We address the problem of automatically analyzing lateral cephalometric images as a diagnostic tool for patients suffering from Sleep Disordered Breathing (SDB). First, multiple landmarks and anatomical structures that were previously associated with SDB are localized. Then statistical regression is applied in order to estimate the Respiratory Disturbance Index (RDI), which is the standard measure for the severity of obstructive sleep apnea. The landmark localization employs a new registration method that is based on Local Affine Frames (LAF). Multiple LAFs are sampled per image based on random selection of triplets of keypoints, and are used to register the input image to the training images. The landmarks are then projected from the training images to the query image. Following a refinement step, the tongue, velum and pharyngeal wall are localized. We collected a dataset of 70 images and compare the accuracy of the anatomical landmarks with recent publications, showing preferable performance in localizing most of the anatomical points. Furthermore, we are able to show that the location of the anatomical landmarks and structures predicts the severity of the disorder, obtaining an error of less than 7.5 RDI units for 44% of the patients.

1 Introduction

Sleep-Disordered Breathing (SDB) consists of a continuous upper airway resistance with clinical indications ranging from snoring to Obstructive Sleep Apnea (OSA). In this disorder, the breathing temporarily stops during sleep as the throat muscles relax and block the patient's airway. The patient then wakes up in order to restart his breathing, and the quality of sleep is impaired. OSA symptoms include daytime sleepiness and fatigue, as well as an increased risk of cardiovascular disease, stroke, high blood pressure, arrhythmias, and diabetes.

The prevalence of sleep apnea is estimated to be 3.3% in the adult male population and increasing to almost 5% in the middle age group [1]. The objective diagnosis of OSA and its severity, and establishing a baseline for future changes require a referral to a sleep laboratory. The evaluation involves an overnight stay and monitoring of the breathing during sleep following an analysis by a specialist. While the resulting report is considered the gold standard in current diagnosis, its reliability is questionable since

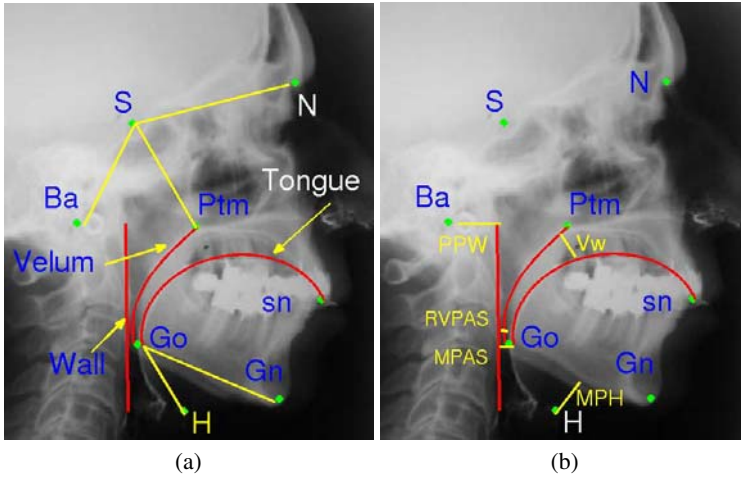


Fig. 1. Landmarks and anatomical structures. The 8 landmarks are marked with green points. (a) The tongue, velum and Pharyngeal Wall are marked in red (from right to left). The three angular measurements (in yellow) are: bony nasopharyngeal angle (**BaSPtm**), skull base angle (**BaSN**) and mandibular plane hyoid angle (**GnGoH**). (b) The five line measurements marked in yellow. **Vw** is the maximal width of the velum (velar width). **RVPAS** is the minimum distance between the velum and the posterior pharyngeal wall. **MPH** is the distance between landmark H (hyoid) to line GnGo (mandibular plain). **PPW** is the distance from the basion (Ba) to the wall (defined slightly differently from [4]). **MPAS** is the minimum distance between the tongue and the posterior pharyngeal wall. Notice that landmark sn (end of tongue) is different from Prn, which is detected in some previous work.

patients may fail to sleep naturally at the laboratory [23]. Moreover, the main drawback of the polysomnography is that it is only a descriptive tool which does not provide any information regarding the pathophysiology of SDB. Diagnosis of the anatomic factors that predispose the airway to collapse during sleep is required for tailoring the best treatment modality to the individual patient. Therefore, due to the high prevalence of OSA, its implications of the public health, the cost, length, and inherent limitations of the sleep-laboratory examination, an imaging based diagnosis might be beneficial.

Finkelstein et al. [4] lay the groundwork for 3D cephalometric analysis of SDB by the instillation of contrast media and adding a frontal view to the traditionally used lateral radiographs, demonstrating that worsening of SDB was generally associated with increased numbers of discovered *compromised cephalometric parameters*. The cephalometric analysis is performed by manually marking the anatomical landmarks and measuring various angles and distances, and interpreted by a skilled expert.

Here, we (i) automate the process described in [4]: The cephalometric landmarks are detected automatically using a new registration method, followed by an automatic detection of the anatomical structures; and (ii) the Respiratory Disturbance Index (RDI) that measures the severity of the SDB is predicted based on the image measurements. Therefore we switch from manual marking that requires a high level of expertise to an automatic one, and at the same time provide predictability and not just correlation.

Previous Work. To our knowledge, there is no existing research on automatic analysis of cephalometric images for SDB. Previous research focused on cephalometric landmark detection for orthodontics. The more recent methods use Active Shape Models (ASM) or Active Appearance Models (AAM) [5,6,7,8,9]. Typically, models with a large number (~ 250) of landmarks are learned, an initial localization is provided by heuristically detecting the easiest landmarks (e.g., Prn and Gn in [9]), and ASM/AAM is then used for fine-tuning.

2 Triplet-Based LAF for Image Registration

We represent each image by a set of local regions, each associated with its own affine coordinate frame. Such a representation is sometimes called a Local Affine Frames (LAF) representation, and was introduced in [10]. In the literature LAFs are derived based on Maximally Stable Extremal Regions (MSERs) [11], which do not provide desirable output on cephalometric images. Here, triplets of keypoints are used as the basis for our LAFs, each detected using the DoG interest point operator [12].

- 1. Keypoint detection.** The SIFT DoG operator [12] is applied to the image obtaining a large group of keypoints p_1, \dots, p_n (n varies between images).
- 2. Random triplet selection.** A large number of triplets ($n_T = 10,000$ in all of our experiments) is selected by random.

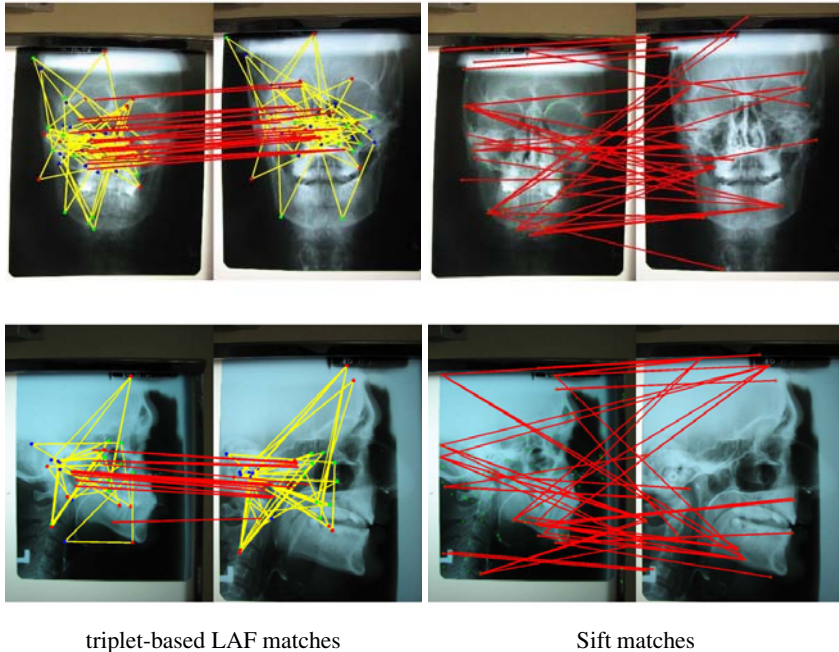


Fig. 2. Top triplets-based LAF matches (left) and SIFT matches (right). Coherent matches produce parallel lines. As can be seen the proposed LAF method provides matches that are much more coherent. The top row demonstrates results on frontal cephalometric images.

3. Triplet ordering. The points of each triplet are ordered to provide a 6 fold reduction in the number of potential triplets. The order is obtained by computing for each keypoint the local SIFT descriptor, and then projecting the descriptor vector to the first principle component obtained from a large group of SIFT features that were collected beforehand. Each triplet $T_i = \{p_{i,1}, p_{i,2}, p_{i,3}\}, i = 1..n_T$ is then ordered such that $p_{i,1} < p_{i,2} < p_{i,3}$, where the order is given by the projection of the SIFT vector.

4. Region of interest extraction. We compute an affine transformation that maps the three points of each triangle to the coordinates $(1, 1), (128, 1), (64, 128)$, and use it to warp the image. Then an enclosing square of corner coordinates $(1, 1)$ and $(128, 128)$ is used to obtain local appearance information for each triplet.

5. Appearance encoding. We record for each triangle the histogram of edge directions at the enclosing square, by using the SIFT descriptor [12].

After each image (I) is encoded as a collection of triplets and matching ROI descriptors, it can be compared to another image (J). The matching score between pairs of triplets, one from I and one from J , is provided by the distance measure of the SIFT descriptors. The highest scoring pairs provide information on the overall match quality. Figure 2 compares triplet-based matches to matches based on DoG (SIFT) keypoints.

3 Landmarks Detection

We are given m_t training images that are manually marked with the relevant anatomical landmarks. Two different sets of tracings and measurements were conducted by A.N. and Y.F. for each roentgenogram. The mean values of the two sets of measurements were used. Following [4], we employ a group of 8 cephalometric landmarks: $\mathbf{m} = \{Gn, Go, Ba, S, N, sn, Ptm, H\}$ (see Figure 1). Given an unannotated test image I_t , we employ the registration technique of Section 2 in order to find a multitude of matching LAFs between I_t and each of the training images $I_j, j = \{1, 2, \dots, m_t\}$. Each LAF matching provides: (i) a score computed from the distance of the SIFT descriptors; (ii) a training image index, and (iii) an affine transformation between that training image and the test image. We first select the 50 LAFs with the highest scores from among the pool of all LAFs. Then, the associated affine transformations are used to transform the locations of the landmarks from the training images to the test image.

This process results in a collection of 50 projected locations per anatomical point in the test image (see Fig. 3(a)). A single location per landmark is obtained by employing a mean shift algorithm [13] to locate the mode of each group of landmark projections.

Next, a fine-tuning step is applied in order to find the exact location of each landmark. For some landmarks, such as Go, localization can be improved by seeking the nearest edge points. For others, such as H and Ba that are not necessarily located on visible image edges or defined by a specific texture, a more sophisticated approach is needed. We choose to perform the fine-tuning step to all landmarks using the same local appearance based method.

To that end, we create a set of templates for each anatomical landmark, by projecting for each LAF the associated training image to the test image, and cropping a template around each anatomical point. For each landmark, an SVM classifier is trained to distinguish the relevant templates from nearby templates. The resulting detector is applied in

the vicinity of the initial mean-shift based estimation, and the maximum detector score is used to select the final location.

Once the landmarks are localized, we can evaluate the three angular measurements of [4]: the skull base angle (BaSN); the bony nasopharyngeal angle (BaSPtm); and the mandibular plane hyoid angle (GnGoH) (see Fig. 1(a)).

4 Detection of Cephalometric Structures

Next, we model the tongue, the velum and the posterior pharyngeal wall. These structures are critical for the evaluation of SDB since they are involved in the narrowing of airspace and other anatomical changes. For each structure, we fit a suitable geometric shape, whose parameters are searched in a range dictated by the training images.

In order to reduce the variability in the parameters of the structures, the images are first aligned with accordance to the locations of the eight anatomical feature points. For each landmark, we find the mean location in all training images $\bar{\mathbf{m}} = \sum_{i=1}^{m_t} \mathbf{m}_i$. For a given image I_j (training or testing), we compute the affine transformation that minimizes the least-squares error from \mathbf{m}_j to $\bar{\mathbf{m}}$ and apply it on I_j and its landmarks. All anatomical structures are detected using the aligned images, however, the actual measurements are performed after aligning the detections back to the original coordinate system.

Tongue. The relevant part of the tongue is from its intersection with the contour of the chin towards the mouth's end (sn). We found out that an ellipse $E = ax^2 + bxy + cy^2 + dx + ey + f = 0$ provides a good fitting in most cases, however, simpler models, such as a circle do not fit well. Since an ellipse is defined by five degrees of freedom, and since the two landmarks Go and sn are known to be on the tongue, three more degrees of freedom are to be evaluated.

One possible solution would be to find edge points in the vicinity of the tongue and use RANSAC to fit an ellipse, however, the velum is similar to the tongue and it is easy to confuse between the two. Instead, we perform a search in a parameter space that is defined by the anatomical points.

Observe Figure 3(b). We draw three lines: one from Gn to Ba, one from Gn to Ptm and a third line forming an angle twice as large as the angle BaGnPtm. The ellipse is defined by the points Go and sn and the intersection of its top part with the three lines. The distances of these three intersection points from Gn are used as the ellipse parameters. The suitable range for each parameter is estimated from the training images. This set of three parameters has a much more compact range than the range of the generic ellipse parameters, since the tongue might be tilted at various angles depending on the anatomy of each patient. The actual fitting is performed by an exhaustive search in these ranges, and the quality of each hypothesis ellipse is measured by the edge intensity along its path.

Velum and Pharyngeal Wall. In order to measure the Velar Width (VW) as in [4] (see Figure 3(c)), we detect the velum in the region to the left of landmark Ptm and find the maximum distance to the tongue. No simple geometric shape fits well to the velum, since it may curve strongly towards its top part. We model the velum using a cubic B-spline with four knots [14]. We set the coordinates of the first and last knots at Ptm and

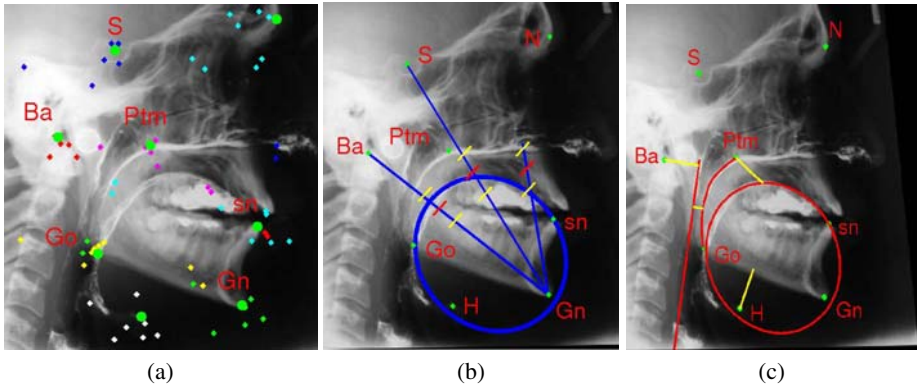


Fig. 3. The process of anatomical structures fitting. (a) Projected locations - the chosen points per each anatomical landmark are marked in the same color and the green dot is the correct location. One can see that at least one location per landmark is accurate. (b) Tongue Fitting. The tongue is modeled by using an ellipse. The green dots are the automatically detected landmarks from section 3. The lines that are used to define the three ellipse parameter points are plotted. The yellow marks on each line bound the valid searching distance from Gn and the red marks are the average distance from Gn. (c) Fitting of all structures. The yellow lines correspond to Vw, MPH, RVPAS and PPW. The tongue is detected correctly even though the teeth are not touching.

Go respectively. Using similar ideas to the tongue fitting, we draw two lines from the Gnathion (Gn) that intersect the velum. Then we measure the valid range from the tongue to the velum in the training images. These ranges define the searching area for the two additional knots and we iteratively fit B-splines to find the best fitting. The pharyngeal wall is modeled by a straight line, since it provides an accurate enough estimation in the region relevant to the line measurements. The model is parameterized by the intersection of the pharyngeal wall line with the two lines GnBa and GnGo, and the search is performed as above in a range dictated by the training images. Fig. 3(c) shows the detection of the three structures (in red) and the yellow lines refer to the 4 line measurements used.

5 Results

We have collected 70 cephalometric images of patients with varying degrees of OSA. All images were taken in accordance with recognized standards: the head was oriented in the Frankfurt horizontal plane and stabilized with a head-holder; the teeth were in habitual occlusion with lips together, and the tongue relaxed in the floor of mouth. Exposures were taken during mid-expiration following a moderately deep breath. Therefore, the variation in pose is minimal. The database is challenging since the quality varies, and in some cases the contrast is poor and the visibility of the anatomical structures is limited.

We randomly split the dataset to 25 testing images and 45 training images and repeat the experiments 10 times. The results discussed below are mean results that were computed automatically on the testing images. First, we compare the accuracy of the landmarks discussed in [4] with the manual detection and three other approaches (see Table 1). As far as we know, the datasets used in previous work are not available. Therefore,

Table 1. Automatic landmarks detection results. In parenthesis, when available: ^a Rueda et al [8], ^b Yue et al [9], and ^c Hutton et al [6]. Notice the high accuracy of S, Ptm, N and Gn. Landmark H has a relatively low detection rate. However, large variation exists also between clinicians marking this landmark. (In [9] the same landmarks are named differently, thus we compare Gn with Me, Ptm with Pns and Ba with Co).

Landmark	mean (mm)	≤ 2mm(%)	≤ 3mm(%)	≤ 4mm(%)	≤ 5mm(%)
S	0.48(2.29 ^a , 5.5 ^c)	100(39 ^a , 76 ^b)	100(70 ^a)	100(78 ^b)	100(98 ^a)
Ptm	0.53(2.67 ^a , 5.0 ^c)	100(37 ^a , 83 ^b)	100(69 ^a)	100(92 ^b)	100(92 ^a)
N	1.32(5.6 ^c)	91(86 ^b)	96	96(89 ^b)	96
Gn	1.21(1.58 ^a , 2.7 ^c)	83(73 ^a , 98 ^b)	83(86 ^a)	87(100 ^b)	91(99 ^a)
Go	2.32(3.88 ^a , 5.8 ^c)	52(26 ^a , 86 ^b)	74(44 ^a)	83(94 ^b)	95(67 ^a)
Ba	2.06(2.7 ^a)	65(38 ^a , 69 ^b)	74(68 ^a)	78(76 ^b)	87(92 ^a)
H	6.4	13	22	30	39

the comparison provides merely a general impression. We have achieved very accurate results for landmarks S, Ptm, Gn and N, even though S and Ptm are not located on clear edges. Comparing to [8] and [6], we have achieved better mean error for all reported landmarks. Comparing with [9], we have better results for S and Ptm and similar results for N, Ba and Gn. Regarding landmark Go, the authors of [9] assumed that the landmark can be found by tracing the edge that starts from Gn (Me in their paper), however, this assumption does not hold in many of our images, where the two sides of the mandible are visible (see the double line in Fig. 1). In such cases, the location of Go is not situated on this edge point. Results for H are not reported in previous work.

We also compare our automatic line measurements with the manual ones. The mean absolute errors and standard derivations ($x \pm y$) are: Vw 0.94 ± 0.78 , RVPAS 1.3 ± 1.3 , PPW 2.3 ± 2.2 and MPH 3.8 ± 3.4 . The errors in computing Vw and RVPAS are very low. The error in detecting MPH is due to the difficulty of finding landmark H. However, considering the inter-observer variance between clinicians in marking this landmark, our results are in the acceptable range.

Next, we predict the RDI and hence the severity of the SDB. In our testing images, the RDI varies from 5 (borderline healthy) to 77 (severe OSA). The prediction is performed via a linear Support Vector Regression model, that is based on either: (1) The three angles measured; (2) The 4 line measurements; (3) a concatenation of (1) and (2);

Table 2. RDI prediction results. See detailed description of the methods in section 5. Column 2: Mean squared RDI error; Columns 3 and 4: the percentage of cases that have a relatively low RDI error (≤ 7.5) and those with a more moderate error (≤ 15). The results could be compared to the night-to-night variability in RDI computations at sleep clinics which displays a standard deviation of 7.2 [2].

Method	Mean Squared Error (MSE)	error ≤ 7.5(%)	error ≤ 15(%)
Angles	324	26	61
Lines	393	30	53
Angles+Lines	361	44	57
BMI	511	17	43

and as a baseline, (4) The Body Mass Index (BMI). Table 2 summarizes the regression results. It is well known that BMI and RDI are well correlated. Nevertheless, the cephalometric measurements are much more predictive of the patients RDI. We therefore corroborate the claim of [4] that the anatomical structures are informative for OSA diagnosis and support the link between available airspace and the degree of the disorder.

6 Conclusions

Previously it was shown that manual measurements of cephalometric images are correlated with the Respiratory Disturbance Index. In this paper, we show, for the first time, that RDI measurements can be predicted from cephalometric images, moreover, our measurements are obtained automatically. Therefore, we make a significant advancement toward an imaging based OSA diagnosis tool. The detection of the underlying landmarks is based on a new registration technique that is potentially useful for several other structure detection applications.

References

1. Young, T., Peppard, P.E., Gottlieb, D.J.: Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective. *Am. J. Respir. Crit. Care Med.* 165(9), 1217–1239 (2002)
2. Mosko, S., Dickel, M., Ashurst, J.: Night-to-night variability in sleep apnea and sleep-related periodic leg movements in the elderly. *Sleep* 11(4), 340–348 (1988)
3. Ahmadi, N., Shapiro, G.K., Chung, S.A., Shapiro, C.M.: Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. *Sleep and Breathing* 13(3), 221–226 (2009)
4. Finkelstein, Y., Wexler, D., Horowitz, E., Berger, G., Nachmani, A., Shapiro-Feinberg, M., Ophir, D.: Frontal and lateral cephalometry in patients with sleep-disordered breathing. *Laryngoscope* 111, 623–641 (2001)
5. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 681–685 (2001)
6. Hutton, T.J., Cunningham, S., Hammond, P.: An evaluation of active shape models for the automatic identification of cephalometric landmarks. *Eur. J. Orthodont.* 22 (2000)
7. Kafieh, R., Mehri, A., Sadri, S.: Automatic landmark detection in cephalometry using a modified active shape model with sub image matching. In: *ICMV 2007*, pp. 73–78 (2007)
8. Rueda, S., Alcaniz, M.: An approach for the automatic cephalometric landmark detection using mathematical morphology and aam. In: Larsen, R., Nielsen, M., Sporning, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 159–166. Springer, Heidelberg (2006)
9. Yue, W., Yin, D., Li, C., Wang, G., Xu, T.: Automated 2-d cephalometric analysis on x-ray images by a model-based approach. *IEEE. Tran. Biomed. Eng.* 53(8) (2006)
10. Obdrzálek, S., Matas, J.: Object recognition using local affine frames on distinguished regions. In: *The British Machine Vision Conf.* (2002)
11. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: *The British Machine Vision Conf.*, pp. 384–393 (2002)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
13. Cheng, Y.: Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 790–799 (1995)
14. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical Recipes*, 3rd edn. *The Art of Scientific Computing*. Cambridge University Press, Cambridge (2007)

Fusion of Local and Global Detection Systems to Detect Tuberculosis in Chest Radiographs

Laurens Hogeweg^{1,3}, Christian Mol^{1,3}, Pim A. de Jong², Rodney Dawson⁴,
Helen Ayles⁵, and Bram van Ginneken^{1,3}

¹ Image Sciences Institute, University Medical Center Utrecht, The Netherlands

² Department of Radiology, University Medical Center Utrecht, The Netherlands

³ Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre,
The Netherlands

⁴ University of Cape Town Lung Institute, Cape Town, South Africa

⁵ Department of Infectious and Tropical Diseases,
London School of Hygiene & Tropical Medicine, London, United Kingdom
`l.hogeweg@rad.umcn.nl`

Abstract. Automatic detection of tuberculosis (TB) on chest radiographs is a difficult problem because of the diverse presentation of the disease. A combination of detection systems for abnormalities and normal anatomy is used to improve detection performance. A textural abnormality detection system operating at the pixel level is combined with a clavicle detection system to suppress false positive responses. The output of a shape abnormality detection system operating at the image level is combined in a next step to further improve performance by reducing false negatives. Strategies for combining systems based on serial and parallel configurations were evaluated using the minimum, maximum, product, and mean probability combination rules. The performance of TB detection increased, as measured using the area under the ROC curve, from 0.67 for the textural abnormality detection system alone to 0.86 when the three systems were combined. The best result was achieved using the sum and product rule in a parallel combination of outputs.

Keywords: chest radiography, tuberculosis, computer aided detection, system combination.

1 Introduction

Tuberculosis (TB) is a major cause of mortality and morbidity worldwide, with 9.3 million new cases and 1.8 million deaths reported in 2007 [10]. Chest radiography is becoming increasingly important in the fight against TB, because in populations with a high prevalence of AIDS existing screening diagnostics such as sputum staining are less reliable. With the increasing availability of digital radiography, computer-aided detection (CAD) systems can be developed that could facilitate mass population screening for TB. This work is part of the larger CAD4TB project aimed at developing such a system.

On chest radiographs (CXR) the presentation of TB is diverse: textural abnormalities are often present, but focal abnormalities (e.g. nodules) and changes of the lung shape may be found as well. So far, only the problem of textural abnormality detection in TB detection has been addressed [1]. If the lungs have different shapes due to disease, texture analysis might miss abnormalities, an example of this is shown in Figure 3d. Another problem is that chest radiographs have a complex appearance caused by the projection of overlapping structures. This increases the difficulty of characterizing normal structures in the lung, such as clavicles, and can result in false positive responses of a textural abnormality detection system (Figure 3b).

These issues provided the motivation for this work, in which multiple detection systems are combined to improve detection performance. It is expected that combining systems detecting different types of abnormalities as well as systems detecting normal structures (segmentation) will lead to an overall improvement of a CAD system. The idea of system or classifier fusion has been extensively studied theoretically (e.g. [6]) and has also been applied to medical imaging (e.g. [3]). To our knowledge, this work is the first to apply such a method to chest radiographs (CXRs). Different strategies for combining subsystems are discussed.

2 Method

2.1 Detection Systems

The detection systems considered in this study are derived from pixel classification of CXRs. Features are calculated for a number of samples (positions) in an image. A classifier is trained using labeled samples from a database of training images. Examples of labels are inside/outside lung field and normal/abnormal structure. Samples in a test image are classified and assigned a probability of belonging to a particular label. Systems for lung field, clavicle, textural abnormality and shape abnormality detection will be shortly described below.

Lung field detection and clavicle detection (CD). A segmentation system for CXRs based on pixel classification to detect the lung fields and clavicles described by van Ginneken et al. [9] was slightly modified. Features in that system are based on Gaussian derivative filtered images calculated at different scales and in different directions. For clavicle detection two extra position features derived from the lung segmentation were added: the distance of a pixel to the border of the lung field segmentation and the distance to the center of gravity of the two detected lung fields.

Textural abnormality detection (TAD). The detection of textural abnormalities is based on texture analysis of small circular image patches (radius = 32 pixels) sampled every 8 pixels from radiographs downsampled to 1024×1024 pixels. Features are based on the moments of intensity distributions of Gaussian derivative filtered images sampled in a patch. This method has recently successfully been used to detect textural abnormalities related to TB in chest

radiographs [1]. The two lung segmentation derived position features described above were added to the texture features. A total of 106 features per patch were extracted. Image patches were sampled inside the segmented lung fields and classified as normal or abnormal using linear discriminant analysis (LDA) preceded by feature reduction using principal component analysis (PCA) (retaining 95% of the variance).

Shape abnormality detection (SAD). Segmentation of the unobscured lung fields can be done accurately in normal images, or in abnormal images with only small abnormalities such as nodules [7]. When large abnormalities close to the lung walls are present, segmentation of the lung fields becomes much more difficult as can be seen in Figure 3d,f. This causes the boundaries of the detected lung fields to be displaced and abnormalities to fall outside the detected lung fields, rendering them undetectable for the TAD system. In such situations the changed shape of the detected lung fields can be used to detect abnormal images.

A simplified version of shape contexts [2] is used to describe shapes. Instead of using multiple points on the shape to construct angle-distance histograms the centroid of the shape is used. Rays are cast in multiple directions from the centroid and the distance to the intersection with the boundary is recorded. This creates a feature vector of length n (the number of directions used) for each shape. The procedure is performed separately for each lung and the feature vectors are concatenated to obtain one describing both lung shapes.

Using a training set of normal and abnormal lung shapes a classifier can be trained to detect abnormal shapes. It was observed from preliminary experiments that abnormal lungs in the training database were very varied in shape and few in number. In a situation where the number of positive examples is small, it is difficult for two-class classifiers to obtain good performance and one-class classifiers might perform better [8]. A one-class classifier based on a PCA model of normal shapes (retaining 90% of the variance) was used to detect abnormal shapes. A large Mahalanobis distance (generalisation of the standard score) of a test shape to the model indicates that the shape is more abnormal. By thresholding this value an image can be classified as normal or abnormal.

2.2 Combining Detections

The three detection systems (TAD, CD and SAD) are combined to improve the performance of the system in discriminating between normal and abnormal images. The texture and clavicle detection output is combined at the pixel level, fused to an image decision and combined with the shape decision to give a final decision (Figure 1). Each step is now described in detail.

Clavicle corrected TAD. Clavicle induced false positive responses in the TAD system are reduced by combining it with the probabilistic output of the clavicle detection system. This gives one clavicle corrected output $p_{tcc} = p_t \cdot (1 - p_c)$ per pixel, where p_t and p_c are the probabilistic outputs of the TAD and CD systems respectively. The combination reduces the probability of abnormality from the TAD system when the probability of the clavicle being located at that position is high.

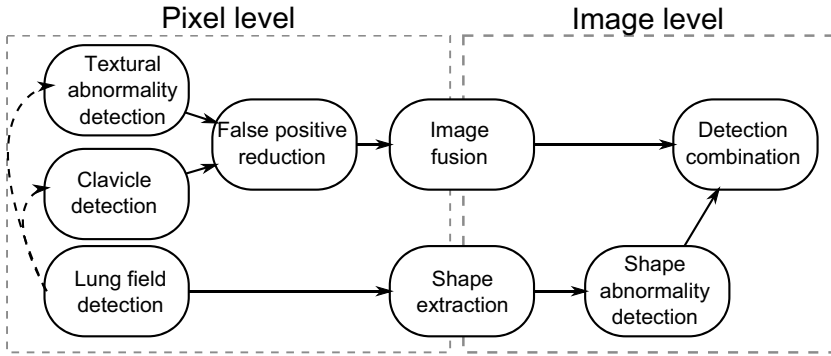


Fig. 1. Combination of detection systems. The detection combination is shown in detail in Figure 2

Image fusion. The texture detection system operates at the pixel level. Before combining it with the shape system the pixel decisions are fused to one image decision, which is performed using the quantile rule. This rule has been shown to have a good performance in texture classification for TB detection [1].

Detection combination. In general the output of a classification system - e.g. LDA, the quantile rule, Mahalanobis distance - is not a real probability estimate. Therefore, different outputs cannot be directly combined and first need to be converted to calibrated probability estimates in the range 0 – 1. In this work the probability of an output being normal is calibrated by taking the percentage of false positives (1 – specificity) produced by the classifier at a decision threshold equal to the output.

System combination (see Figure 2) can be performed in a serial fashion, where the output of one system is the input for the next one, or in a parallel fashion, in which outputs of multiple systems are combined. For serial combination the shape system serves as a filter for the texture system: if the output of the shape system is higher than a certain threshold the image is judged to be abnormal. The threshold is set to a calibrated probability of being abnormal of 0.95. Shape probabilities above this threshold are considered abnormal and set to 1, and 0 otherwise. In parallel combination the population probabilities were directly combined. Widely used combination rules - taking the minimum, maximum, product and mean value - were used to combine outputs.

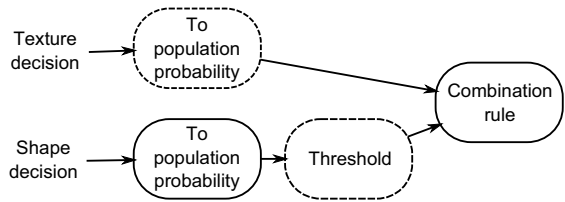


Fig. 2. Detail of detection combination of TAD and SAD system. Optional blocks are indicated by a dashed border.

2.3 Experiments

Data. A database containing 365 digital CXRs (2048×2048 resolution, pixel size 0.25 mm, Delft Imaging Systems, The Netherlands) was used. The training database consisted of 216 images (110 normal, 106 abnormal) for which full manual outlines of abnormalities were created by a chest radiologist in training. The training database contained 15 images with abnormal shapes as judged by the author of this work. The test database consisted of 149 consecutive images from a TB clinic in Africa of which the normal/abnormal decision was made by a local expert (69 normal, 80 abnormal). The database is part of a larger database of CXRs from areas in Africa with high TB incidence that is constructed within the CAD4TB project.

Analysis. Three types of systems were evaluated: (1) TAD, (2) TAD with clavicle correction, and (3) TAD with clavicle correction combined with SAD. The performance of the systems was evaluated using receiver operator characteristic (ROC) analysis and calculation of the area under the ROC curve (A_z). The different combination rules (minimum, maximum, product and mean) for system 3 were also evaluated.

3 Results

3.1 Detection Systems

Figure 3 shows examples of the effect of the clavicle correction and the combination of texture and shape system. Clavicle correction of the TAD system reduces false positive pixel responses (a,b,c) and the SAD system reduces false negatives when the texture inside the detected lung fields is normal (d,e,f).

3.2 Combination

Figure 4 shows a comparison of ROC curves for the different detection system combinations. The TAD system (1) achieved a performance, measured using A_z , of 0.69. When false positive responses were corrected using the CD system the performance of the TAD system increased to 0.75. The combination of outputs from the clavicle corrected TAD system and the SAD system increased the A_z to 0.85. Differences between A_z values were computed according to the method by Hanley et al. [4]: system 1 and 3, and 2 and 3 were significantly different ($p < 0.05$). The difference between 1 and 2 was not significant, but in the low false positive rate region of the ROC curve, which is important for screening, there was a clear improvement in sensitivity.

The effect of different combination strategies of the TAD system and the SAD system is shown in table 1. When outputs are combined in a parallel configuration (no threshold on shape probabilities) the sum and product rule give the best results. For a serial configuration (shape probabilities are thresholded) the sum and maximum rule perform best. In a serial configuration the maximum rule is



(a) Normal radiograph



(b) Clavicles give false re-
sponse in texture analysis

(c) Clavicle correction re-
duces the abnormality score



(d) Abnormal radiograph
with shape change



(e) Texture analysis gives a
low abnormality score



(f) Shape analysis gives a
high abnormality score

Fig. 3. Examples of results for the clavicle corrected TAD system (a,b,c) and the combination of the clavicle corrected TAD system and the SAD system (d,e,f). Figure b, c and e are maps indicating the probability of being abnormal for each pixel (white = high).

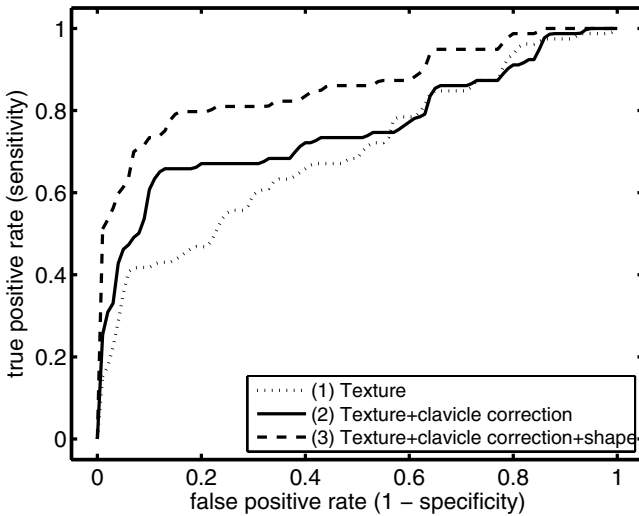


Fig. 4. ROC analysis of TAD system (1) and combined systems (2 and 3)

Table 1. Effect of different combination strategies (values indicate A_z)

Fusion method	Shape threshold		Practical effect of shape threshold
	No (parallel)	Yes (serial)	
Sum	0.86	0.85	Filtering of abnormal shapes
Maximum	0.84	0.85	Filtering of abnormal shapes
Minimum	0.81	0.76	Filtering of normal shapes
Product	0.86	0.76	Filtering of normal shapes

equivalent to preventing images with abnormal shapes to be further analyzed using texture analysis. The (unweighted) sum had the same practical effect. The minimum and product rule, which have the effect of analyzing only images which have abnormal shapes in the TAD system, perform worse than the other strategies.

4 Discussion and Conclusion

This paper described the combination of multiple subsystems to automatically detect tuberculosis in digital chest radiographs. The combination of a lung field, clavicle, textural abnormality and shape abnormality detection system improved detection performance.

The performance improvement of the addition of the shape abnormality detection system demonstrates the potential of combining systems at different levels (pixel and image). Alternatively the lung field segmentation algorithm could be made robust so that it can provide accurate lung field segmentations in the presence of pathology. Although an interesting research topic on its own the problem of segmenting abnormal images has not been solved yet [7]. Normal undiseased anatomy already shows substantial variation and this problem is even larger when pathology is present. Training a system which could solve that problem would require a far larger amount of training data than a system based on normal anatomy would use.

Parallel and serial configurations of the combination system were evaluated using a few well known probability combination rules. Generally, configurations and rules that retained the most information about the systems (sum rule and no thresholding) performed best. Depending on computational or other resource restraints (e.g. the choice of using expensive tests) other configurations might be useful in practical applications.

Instead of using multiple separate detection systems to detect disease in images it might be possible to design one large detection system to handle that task. However, there are a number of reasons to prefer multiple small systems over one large complex system. The generalization ability of a classification system drops when it becomes more complex [5], requiring larger amounts of training data to achieve good performance. From a perspective of system design, multiple highly specialized subsystems are also preferable since these can be more easily tested and evaluated than larger general ones. A practical reason to use multiple

subsystems, such as the clavicle and texture detection system, is that they can be developed parallelly by different research groups.

Within a framework for detection system combination such as presented in this paper other subsystems can be easily added. For example, in the application of TB detection in CXRs, detection of focal lesions and hilar abnormalities could further improve the performance of the whole system. It was demonstrated in this paper that the combination of multiple detection subsystems improves the detection of TB on CXRs.

References

1. Arzhaeva, Y., Hogeweg, L., de Jong, P.A., Viergever, M.A., van Ginneken, B.: Global and local multi-valued dissimilarity-based classification: Application to computer-aided detection of tuberculosis. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 724–731. Springer, Heidelberg (2009)
2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4), 509–522 (2002)
3. Das, S.K., Chen, S., Deasy, J.O., Zhou, S., Yin, F., Marks, L.B.: Combining multiple models to generate consensus: application to radiation-induced pneumonitis prediction. *Medical Physics* 35(11), 5098–5109 (2008)
4. Hanley, J.A., McNeil, B.J.: A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148(3), 839–843 (1983)
5. Ho, T.K., Basu, M.: Complexity measures of supervised classification problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(3), 289–300 (2002)
6. Kuncheva, L.I.: A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(2), 281–286 (2002)
7. Seghers, D., Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Minimal shape and intensity cost path segmentation. *IEEE Transactions on Medical Imaging* 26(8), 1115–1129 (2007)
8. Tax, D.: One-class classification. PhD thesis, Delft University of Technology (2001)
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. World Health Organization. WHO report 2009: Global tuberculosis control, epidemiology, strategy, financing (2009)

Novel Morphometric Based Classification via Diffeomorphic Based Shape Representation Using Manifold Learning

Rachel Sparks and Anant Madabhushi*

Department of Biomedical Engineering, Rutgers University, USA

Abstract. Morphology of anatomical structures can provide important diagnostic information regarding disease. Implicit features of morphology, such as contour smoothness or perimeter-to-area ratio, have been used in the context of computerized decision support classifiers to aid disease diagnosis. These features are usually specific to the domain and application (e.g. margin irregularity is a predictor of malignant breast lesions on DCE-MRI). In this paper we present a framework for extracting Diffeomorphic Based Similarity (DBS) features to capture subtle morphometric differences between shapes that may not be captured by implicit features. Object morphology is represented using the medial axis model and objects are compared by determining correspondences between medial axis models using a cluster-based diffeomorphic registration scheme. To visualize and classify morphometric differences, a manifold learning scheme (Graph Embedding) is employed to identify nonlinear dependencies between medial axis model similarity and calculate DBS. We evaluated our DBS on two clinical problems discriminating: (a) different Gleason grades of prostate cancer using gland morphology on a set of 102 images, and (b) benign and malignant lesions on 44 breast DCE-MRI studies. Precision-recall curves demonstrate DBS features are better able to classify shapes belonging to the same class compared to implicit features. A support vector machine (SVM) classifier is trained to distinguish between different classes utilizing DBS. SVM accuracy was $83 \pm 4.47\%$ for distinguishing benign from malignant lesions on breast DCE-MRI and over 80% in distinguishing between intermediate Gleason grades of prostate cancer on digitized histology.

1 Introduction

Morphological features of suspicious structures on imaging can provide useful diagnostic and prognostic cues. In BIRADSTM classification, benign breast masses

* This work was made possible via grants from the Wallace H. Coulter Foundation, National Cancer Institute Grant Nos. R01CA136535-01, R21CA12718601, and R03CA143991-01, the Cancer Institute of New Jersey, and Bioimagene, Inc. The authors also wish to thank Dr. Mark Rosen, Dr. Michael Feldman and Dr. John Tomaszewski from the University of Pennsylvania for providing the medical imagery and expert analysis.

often present as smooth round masses while cancerous masses tend to have spiculated edges [1]. Morphological cues are a critical component of Gleason grading, a scheme employed to assess the invasiveness of prostate cancer (CaP) [2]. Intermediate Gleason grade patterns 3 and 4 have poor interobserver reproducibility among trained pathologists [3]. Creating a reliable system to grade CaP on histology could have profound clinical impact on CaP survival rate [4]. Gland morphology is a particularly important feature for distinguishing between intermediate Gleason grades; higher Gleason grades of CaP are characterized by irregularly shaped glands while lower Gleason grades of CaP have smooth margins with a distinct lumen.

Computerized decision support classifiers have employed features implicitly related to object morphology for discriminating pathological structures [14]. Implicit features quantify particular traits typically relating to the contour or area of objects. For example the knowledge that a spiculated lesion boundary in breast DCE-MRI is associated with malignancy has led to inclusion of boundary irregularity within the BIRADSTM lexicon of descriptors for MR mammography [1]. However, it may not always be obvious which morphological traits are most beneficial in discriminating pathologies. Implicit features (such as area overlap ratio) may be unable to capture subtle shape differences between objects, differences which may be highly relevant to disease classification.

We present a new shape characterization framework to (1) accurately distinguish between the morphology of shapes (independent of domain or application) and (2) distinguish subtle differences between similar shapes (such as glands in Gleason grade 3 and grade 4 of CaP histology). The medial axis model represents shape as a connected skeleton with corresponding vectors to the object surface [5]. We employ a point-based diffeomorphic registration algorithm to find a mapping between two medial axis models. Diffeomorphic mapping allows for the alignment of medial axis models so that a pairwise shape dissimilarity can be formulated between corresponding points on two medial axis models to obtain a shape dissimilarity measure. A pairwise dissimilarity matrix is then used to define the nonlinear high dimensional shape space. Graph embedding [6], a nonlinear dimensionality reduction method that preserves local topology between points, is employed to reconstruct a low dimensional shape space within which shapes are arranged as a continuum on a smooth shape manifold; distances between shapes on this manifold being a function of morphological differences between shapes. Our integrated quantitative morphometric framework employing, (1) the medial axis model, (2) a diffeomorphic algorithm to evaluate shape dissimilarity, (3) a graph embedding scheme to obtain DBS features is illustrated in the flowchart in Figure 1.

In this work we evaluate our framework in the context of distinguishing (a) benign, Gleason grade 3, and Gleason grade 4 glands on digital histology images of needle core prostate biopsies (611 glands from 102 images) and (b) malignant and benign lesions on 44 DCE-MRI studies. We compared our framework against implicit features (e.g. contour smoothness, area overlap ratio) in terms of classification performance.

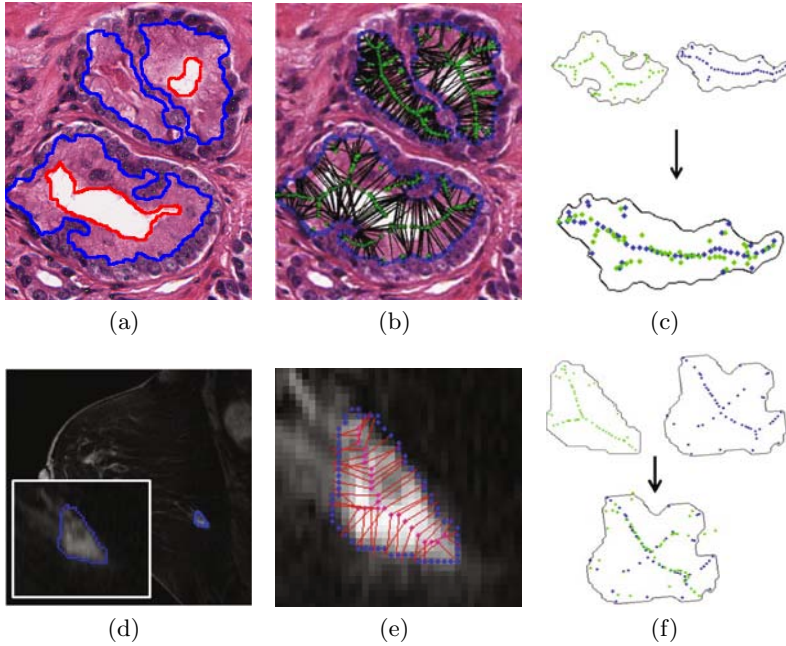


Fig. 1. Framework for extracting DBS features to (top row) distinguish Gleason grades of CaP and (bottom row) benign and malignant lesions on DCE-MRI. (a) For digital histology, both nuclear (blue) and lumen (red) boundaries of prostate glands are segmented [4]. (d) Lesion boundary (blue) segmentation is performed on breast DCE-MRI [1]. (b) Medial axis model (green, black) is fit to the nuclear or the luminal layer (not shown). (e) Lesion boundary on breast DCE-MRI is likewise fit with a model (red, magenta). (c), (f) A cluster-based diffeomorphic registration for mapping between template (blue) and registered (green) medial axis models helps guide a point-based dissimilarity measure. Object contours (black) are also displayed.

2 Shape Model Representation and Shape Dissimilarity

We define an image scene $\mathcal{C} = (C, f)$ where C is a 2D grid of image pixels $c \in C$ and an image intensity function $f(c) : c \in C$. The image pixels c are located in the X - Y Cartesian space. Objects of interest are segmented via application of a hybrid active contour model [7] such that $\mathcal{X} \subset C$ defines the set of pixels comprising the boundary of the object of interest.

2.1 Medial Axis Model

From an object contour \mathcal{X} a medial axis model M can be constructed, where $m \in M$ is a set of pixels $m \subset C$ pertaining to the skeletal backbone of the object [5]. To determine m backbone pixels, an image distance map for the object defined as $\mathcal{C}^e = (C, f^e)$ is constructed. The distance map function $f^e(c)$ is defined as:

$$f^e(c) = \begin{cases} 0 & c \in \mathcal{X}, \\ -\min_{p \in \mathcal{X}} \|c - p\| & c \in \mathcal{X}^h, \\ \min_{p \in \mathcal{X}} \|c - p\| & c \in \mathcal{X}^o, \end{cases} \tag{1}$$

where $\mathcal{X}^h \subset C$ is the set of pixels in the image contained within \mathcal{X} , and $\mathcal{X}^o \subset C$ refers to the pixels outside of \mathcal{X} . A gradient map of an object, $\widehat{C}^e = (C, \widehat{f}^e)$ is calculated for $c \in C$ as,

$$\widehat{f}^e(c) = \left(\frac{\partial f^e(c)}{\partial x} \right)^2 + \left(\frac{\partial f^e(c)}{\partial y} \right)^2. \tag{2}$$

where $\frac{\partial()}{\partial x}$, $\frac{\partial()}{\partial y}$ are the partial derivatives along the Cartesian X and Y axes respectively. Points on the medial axis are defined as, $M = \{m : m \in C, \widehat{f}^e(m) < \tau\}$. Empirically, $\tau = 0.05 \operatorname{argmin}_{c \in C} (\widehat{f}^e(c))$ was found to give a well defined medial axis with few spurious branches. Then for $m \in M$, the two closest points on the surface can be defined as

$$\widehat{p}_1 = \operatorname{argmin}_{p \in \mathcal{X}} \|m - p\|, \text{ and } \widehat{p}_2 = \operatorname{argmin}_{p \in \mathcal{X}, p \neq \widehat{p}_1} \|m - p\|,$$

and the corresponding surface vectors are defined $\mathbf{v}_1 = \widehat{p}_1 - m$, and $\mathbf{v}_2 = \widehat{p}_2 - m$.

2.2 Shape Model Dissimilarity Metric

To compare a set of N medial axis models $\{M_\alpha : \alpha \in \{1, \dots, N\}\}$, a dissimilarity metric based on model correspondence is calculated. M_α denotes sets of unlabeled points $m \in M_\alpha$. Hence the problem of determining the correspondence between models is reduced to the problem of determining point correspondences. We use a diffeomorphic registration guided by point clustering to find a mapping between point clouds [8].

Cluster Update. Point clusters are determined using simulated annealing [9]. For a set of medial axis points $m \in \{M_\alpha : \alpha \in \{1, \dots, N\}\}$, a set of K cluster centers at some iteration i may be represented as $\{\widehat{m}_\alpha^{k,i} : k \in \{1, \dots, K\}\}$. The number of clusters K is a user selected parameter. The probability of a point $m \in M_\alpha$ belong to the cluster centered at $\widehat{m}_\alpha^{k,i}$ is determined by,

$$P(m|\widehat{m}_\alpha^{k,i}) = \frac{e^{-\sigma_i \|m - \widehat{m}_\alpha^{k,i}\|^2}}{\sum_{k=1}^K e^{-\sigma_i \|m - \widehat{m}_\alpha^{k,i}\|^2}}. \tag{3}$$

Similarly, $P(m|\widehat{m}_\beta^{k,i})$ is defined for a different medial axis model $\{m \in M_\beta : \beta \in \{1, \dots, N\}\}$. The variable σ_i is defined as, $\sigma_i = (\epsilon)^i \sigma_0$, for some $\epsilon > 1$. So that σ_i become progressively larger as i increases, causing the probability function $P(m|\widehat{m}_\alpha^{k,i})$ to be more likely a member of closer cluster centers. The initialization term σ_0 is defined as $1/\sigma_0 = \max_{m \in M_\alpha} \|m - \mu_\alpha\| + \max_{m \in M_\beta} \|m - \mu_\beta\|$, where M_α, M_β represent any two shapes with corresponding centroids μ_α, μ_β .

Cluster centers at the next iteration $i + 1$ are updated as in [8]. As each point has partial membership in several clusters, cluster centers are updated to reflect this membership structure according to the following update equation,

$$\hat{m}_\alpha^{k,i+1} = \frac{\hat{m}_\beta^{k,i} + \sum_{m \in M_\alpha} m P(m|\hat{m}_\alpha^{k,i})}{1 + \sum_{m \in M_\alpha} P(m|\hat{m}_\alpha^{k,i})}. \tag{4}$$

The term $\hat{m}_\beta^{k,i}$ in Equation 4 ensures that $\hat{m}_\alpha^{k,i}$ does not move too far from its counterpoint in M_β . Similarly, $\hat{m}_\beta^{k,i+1}$ is defined for M_β . After each cluster center update a diffeomorphic mapping is used to align the corresponding cluster centers in $\hat{m}_\alpha^{k,i+1}$ and $\hat{m}_\beta^{k,i+1}$.

Diffeomorphic Mapping. A diffeomorphism over the image space C can be constructed from a set of paths $q^{k,i}(t) : \{t \in \{0, \dots, T\}, k \in \{1, \dots, K\}\}$. Each path $q^{k,i}(t)$ describes the movement of a point which starts at the location of the cluster centroid of one medial axis $q^{k,i}(0) = \hat{m}_\alpha^{k,i}$ and ends at the corresponding cluster centroid of the other medial axis $q^{k,i}(T) = \hat{m}_\beta^{k,i}$ transversing a subset of the image space C [10]. Additionally, we constrain $q^{k,i}(t)$ to minimize the deformation of the image space C using a kernel function G to smooth the image space. We choose to utilize Greene’s function defined as, $G(\Gamma, \Theta) = -(\Gamma - \Theta)^2 \log(\Gamma - \Theta)^2$, where $\Gamma \in C$ and $\Theta \in q^{k,i}(t)$.

An iterative update function is solved for where $q^{k,i}(t)$ is initialized as a line between the beginning and end points and iteratively updated to minimize the energy equation defined as,

$$\hat{q}^{k,i}(t) = \operatorname{argmin}_{\hat{q}^{k,i}(t)} \sum_{k=1}^K \sum_{t=0}^T \omega^i(t) \cdot \left(\sum_{\eta=1}^K \omega^\eta(t) G(q^{\eta,i}(t), \hat{q}^{k,i}(t)) \right), \tag{5}$$

where the variables $\omega^i(t)$ and $\omega^\eta(t)$ must also be solved for. We alternate minimizing over the variables $\omega^i(t)$, $\omega^\eta(t)$, and the path $q^{k,i}(t)$ [10].

Point Correspondence For the set of paths $\hat{q}^{k,i}(t)$ obtained, M_α can be mapped to \tilde{M}_α . The mapped medial axis model \tilde{M}_α shares the coordinate space of M_β . In this shared coordinate space point correspondences are determined by,

$$(\hat{u}, \hat{v}) = \operatorname{argmin}_{\hat{u}, \hat{v}} \|\tilde{m}_\alpha^u - m_\beta^v\|, \tag{6}$$

where $\tilde{m}_\alpha^{\hat{u}} \in \tilde{M}_\alpha : \hat{u} \in \{1, \dots, S\}$ is the set of S medial axis points contained in \tilde{M}_α ; similarly $m_\beta^{\hat{v}}$ is defined for M_β . The matched indices (\hat{u}, \hat{v}) are used to calculate dissimilarity between 2 shapes in the original space by,

$$A(\alpha, \beta) = \sum_{(\hat{u}, \hat{v}=1)}^N \kappa_1 \|\tilde{m}_\alpha^{\hat{u}} - m_\beta^{\hat{v}}\| + \kappa_2 \|\mathbf{v}_{\alpha,1}^{\hat{u}} - \mathbf{v}_{\beta,1}^{\hat{v}}\| + \kappa_3 \|\mathbf{v}_{\alpha,2}^{\hat{u}} - \mathbf{v}_{\beta,2}^{\hat{v}}\|. \tag{7}$$

For N objects $A \in \mathbb{R}^{N \times N}$ is a high dimensional dissimilarity matrix. A is positive definite when the user selected weights κ_1, κ_2 , and $\kappa_3 > 0$.

2.3 Manifold Learning

Graph embedding [6] reduces A to a low dimensional space $\mathbf{y} \in \mathbb{R}^d$, where $N \gg d$. Optimal low dimensional projections $\mathbf{y} = [y_1, y_2, \dots, y_N]$ and can be found by,

$$\mathbf{y} = \arg \min \left(\sum_{a,b=1}^N \|y_a - y_b\|^2 w_{ab} \right), \tag{8}$$

where $w_{ab} = e^{-A(a,b)/\gamma}$, and $\gamma > 0$ is used to normalize A . If W is positive definite, Equation 8 reduces to a minimum eigenvalue decomposition problem,

$$W\mathbf{y} = \lambda D\mathbf{y}, \tag{9}$$

where D is a diagonal matrix $D_{aa} = \sum_b W_{ab}$. The top d eigenvalues in λ correspond to the d -dimensional embeddings in \mathbf{y} , and the top d DBS.

3 Quantitative Evaluation

3.1 Datasets

Prostate needle core biopsy tissue samples stained for hematoxylin and eosin were digitally imaged at 40x optical magnification. An expert pathologist manually graded the tissue samples. A total of 24 samples containing 94 glands were identified as benign, 67 samples containing 470 glands were identified as grade 3, and 11 samples containing 47 glands were identified as grade 4. The 44 patient breast lesion DCE-MRI comprised of 16 benign and 28 malignant masses.

We compared DBS to implicit features: contour smoothness, contour standard deviation, compactness, area overlap ratio, average radial ratio, radial ratio standard deviation [1].

3.2 Precision Recall Curves

Precision-recall (PR) curves are generated as follows: the closest objects within differently sized neighborhoods of a query object are identified. Relevant objects are defined as belonging to the same class as the query object. Precision is relevant objects in the neighborhood over the neighborhood size. Recall is all relevant objects in the neighborhood over all relevant objects in the dataset. PR curves are generated by measuring average precision and recall over different neighborhood sizes for several query objects in the database with the same class label. Table 1 reports area under the PR curve. DBS consistently outperforms implicit features.

3.3 Support Vector Machine Accuracy

The ability of DBS or implicit features to determine disease state was evaluated using a SVM classifier [11]. The SVM used 5-fold cross validation where 2/3 of the dataset was used to train the SVM and 1/3 of the dataset used to test. For the multiclass digital histology problem a one-against-all SVM scheme was implemented. DBS outperformed implicit features for all tasks (see Table 2).

Table 1. Area under the PR curve (AUC) for implicit and DBS features. A perfect AUC score is dependent on the distribution of the query object in the database and feature performance, higher numbers represent better distinction between classes. The best AUC values for each task is bolded.

Morphological Feature	Query Object				
	DCE-MRI		Histology		
	Benign	Malignant	Benign	Grade 3	Grade 4
Implicit	0.39 ± 0.08	0.63 ± 0.70	0.12 ± 0.05	0.82 ± 0.03	0.10 ± 0.02
DBS	0.40 ± 0.07	0.62 ± 0.05	0.21 ± 0.07	0.85 ± 0.08	0.20 ± 0.09

Table 2. SVM classification accuracy evaluated on two feature sets with 5-fold cross validation. For DCE-MRI the dataset was evaluated on lesions classified as benign (n=16) or malignant (n=28). For the digital histology glands were classified as belonging to one of three classes: benign (n=94), Gleason grade 3 (n=470), and Gleason grade 4 (n=47). The best feature set for all classification tasks is bolded.

Morphological Feature	DCE-MRI	Digital Histology		
		Benign	Grade 3	Grade 4
Implicit	77.5 ± 3.73%	79.47 ± 4.71%	69.47 ± 7.58%	75.26 ± 5.77%
DBS	83.00 ± 4.47%	82.89 ± 3.97%	86.58 ± 7.39%	84.74 ± 3.23%

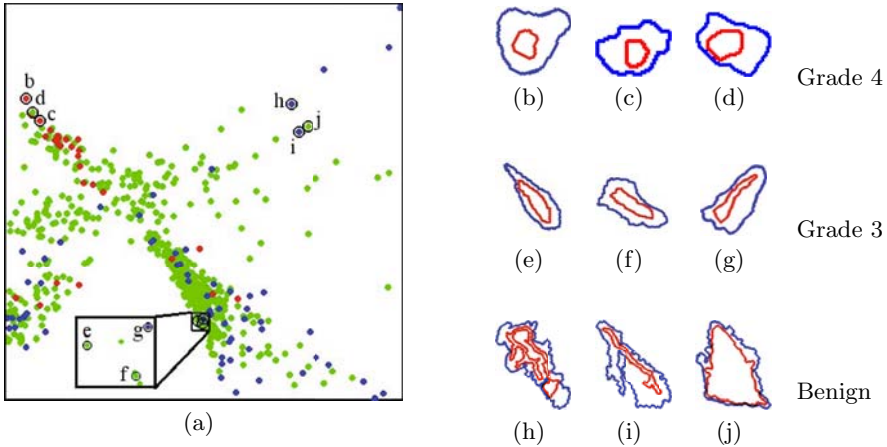


Fig. 2. (a) DBS features for prostate digital histology with benign (blue), Gleason grade 3 (green), and Gleason grade 4 (red) glands. The first and second DBS features are plotted on the X and Y axes respectively. Lumen (red) and nuclear (blue) layers are shown, for glands labeled (b),(d) grade 4, (e),(g) grade 3, and (h),(j) benign. Ground truth for mislabeled glands, displayed in the far right row, are (d) grade 3, (g) benign, (j) grade 3. Similar shapes are nearby on the manifold while dissimilar shapes are far.

4 Concluding Remarks

We presented a framework to calculate a set of new morphological features DBS. Pairwise comparisons of medial axis models describes a high dimensional shape space. Graph Embedding of the shape space allows for extraction of DBS features which were evaluated on two medical classification problems: (a) Gleason grading of CaP digital histology (611 glands on 102 images), and (b) distinguishing benign from malignant lesions on DCE-MRI (44 studies). DBS outperform implicit features for both classification tasks considered in this paper, both in terms of area under precision-recall curves and SVM classifier accuracy. We believe this improved performance was due to DBS capturing subtle shape differences that could not be appreciated by the implicit features. Future work will involve evaluating DBS for more applications.

References

1. Agner, S., Soman, S., Libfeld, E., McDonald, M., Thomas, K., Englander, S., Rosen, M., Chin, D., Noshier, J., Madabhushi, A.: Textural kinetics: A novel dynamic contrast enhanced (DCE)- MRI feature for breast lesion classification. *Journal of Digital Imaging* (2010)
2. Epstein, J., Allsbrook, W., Amin, M., Egevad, L.: The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *American Journal of Surgical Pathology* 29, 1228–1242 (2005)
3. Madabhushi, A., Doyle, S., Lee, G., Basavanthally, A., Monaco, J., Masters, S., Tomaszewski, J., Feldman, M.: Review: Integrated diagnostics: a conceptual framework with examples. *CCLM* (2010) (in Press)
4. Monaco, J., Tomaszewski, J., Fledmand, M., Hagemann, I., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A.: High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *MeDIA* 14, 617–629 (2010)
5. Blum, H.: A transformation for extracting new descriptors of shape. In: *Models for the Perception of Speech and Visual Form*, pp. 367–380. MIT Press, Cambridge (1967)
6. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. PAMI* 22, 888–905 (2000)
7. de Luis-Garcia, R., Deriche, R., Alberola-Lopez, C.: Texture and color segmentation based on the combined use of the structure tensor and the image components. *Signal Process.* 88, 776–795 (2008)
8. Guo, H., Rangarajan, A., Joshi, S.: Diffeomorphic point matching. In: *Handbook of Mathematical Models in Computer Vision*, pp. 205–219. Springer, US (2005)
9. Rose, K., Gurewitz, G., Fox, G.: A deterministic annealing approach to clustering. *Pattern Recogn. Lett.* 11, 589–594 (1990)
10. Twining, C., Marland, S., Taylor, C.: Measuring geodesic distances on the space of bounded diffeomorphisms. *BMVC* 2, 847–856 (2002)
11. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* 20, 273–297 (1995)

Semi Supervised Multi Kernel (SeSMiK) Graph Embedding: Identifying Aggressive Prostate Cancer via Magnetic Resonance Imaging and Spectroscopy

Pallavi Tiwari^{1,*}, John Kurhanewicz², Mark Rosen³, and Anant Madabhushi¹

¹ Department of Biomedical Engineering, Rutgers University, USA
pallavit@eden.rutgers.edu, anantm@rci.rutgers.edu

² Department of Radiology, University of California, San Francisco, USA

³ Department of Radiology, University of Pennsylvania, Philadelphia, USA

Abstract. With the wide array of multi scale, multi-modal data now available for disease characterization, the major challenge in integrated disease diagnostics is to be able to represent the different data streams in a common framework while overcoming differences in scale and dimensionality. This common knowledge representation framework is an important pre-requisite to develop integrated meta-classifiers for disease classification. In this paper, we present a unified data fusion framework, Semi Supervised Multi Kernel Graph Embedding (SeSMiK-GE). Our method allows for representation of individual data modalities via a combined multi-kernel framework followed by semi-supervised dimensionality reduction, where partial label information is incorporated to embed high dimensional data in a reduced space. In this work we evaluate SeSMiK-GE for distinguishing (a) benign from cancerous (CaP) areas, and (b) aggressive high-grade prostate cancer from indolent low-grade by integrating information from 1.5 Tesla *in vivo* Magnetic Resonance Imaging (anatomic) and Spectroscopy (metabolic). Comparing SeSMiK-GE with unimodal T2w, MRS classifiers and a previous published non-linear dimensionality reduction driven combination scheme (ScEPTre) yielded classification accuracies of (a) 91.3% (SeSMiK), 66.1% (MRI), 82.6% (MRS) and 86.8% (ScEPTre) for distinguishing benign from CaP regions, and (b) 87.5% (SeSMiK), 79.8% (MRI), 83.7% (MRS) and 83.9% (ScEPTre) for distinguishing high and low grade CaP over a total of 19 multi-modal MRI patient studies.

1 Introduction

With the rapid growth of new imaging modalities and availability of multi-scale, multi-modal information, data fusion has become extremely important for improved disease diagnostics. However one of the major challenges in integrating independent channels of heterogeneous information is representing them in a unified framework prior to data integration [1]. Typically, information fusing algorithms may be categorized as being either combination of data (COD) or interpretation (COI) methodologies [2]. In

* This work was supported by the Wallace H. Coulter Foundation, the National Cancer Institute under Grants R01CA136535, R21CA127186, R03CA128081, and R03CA143991, the Cancer Institute of New Jersey, Department of Defense (W81XWH-09), and Bioimage Inc.

COD, features \mathbf{F}_{m_1} and \mathbf{F}_{m_2} from two disparate modalities m_1 and m_2 may be combined as $\mathbf{F}_{m_1 m_2} = [\mathbf{F}_{m_1}, \mathbf{F}_{m_2}]$. However directly aggregating data from very different sources without accounting for differences in the number of features and relative scaling can lead to classifier bias towards the modality with more attributes. In [3], Lanckriet et al transformed data from amino acid sequences, protein complex data, gene expression data, and protein interactions into a common kernel space. Kernels are positive definite functions which capture the similarities of the input data into a dot product space such that $K(\mathbf{F}(c_i), \mathbf{F}(c_j)) = \langle \Phi_K(\mathbf{F}(c_i)), \Phi_K(\mathbf{F}(c_j)) \rangle$, where Φ is the implicit pairwise embedding between points $\mathbf{F}(c_i)$ and $\mathbf{F}(c_j)$. This multi-kernel learning (MKL) (Figure 1(a)) involves similarity matrices for kernels from individual modalities being combined and used to train classifiers (within the fused kernel space) in order to make meta-predictions. However, due to the large amount of information present in each input source, all COD methods, including MKL, suffer from the curse of dimensionality.

In [1], we introduced ScEPTre (Figure 1(b)) which employed graph embedding (GE) [4] to combine low dimensional data representations obtained from individual modalities. GE accounts for the non-linearities in the data by constructing a similarity graph $G = (V, W)$, where V corresponds to the vertex between pairwise points and W is a $n \times n$ weight matrix of n data points. However GE, like most other dimensionality reduction (DR) schemes, is unsupervised and does not include any domain knowledge while transforming the data to lower dimensions which often leads to overlapping embeddings. A few supervised DR schemes such as linear discriminant analysis (LDA) employ class label information to obtain low dimensional embeddings. However obtaining labels for biomedical data is extremely expensive and time consuming. Recently semi-supervised DR (SSDR) schemes based on GE have been proposed [5], which construct a weight matrix leveraging the known labels such that higher weights are given to within-class points and lower weights to points from different classes. The proximity of labeled and unlabeled data is then used to construct the low dimensional manifold.

In this work, we present a unified data fusion DR framework called Semi Supervised Multi Kernel Graph Embedding (SeSMiK-GE), a novel data fusion and dimensionality reduction scheme that leverages the strengths of GE, semi-supervised learning, and MKL into a single integrated framework for simultaneous data reduction, fusion, and classification. Only the work of Lin et al [6], that we are aware of, has used MKL in conjunction with GE. However their approach does not leverage learning in constructing the embeddings. SeSMiK-GE involves first transforming each individual modality in a common kernel framework, followed by weighted combination of individual kernels as $\hat{K} = \sum_{m=1}^M \beta_m K_m$, where $K_m, m \in \{1, 2, \dots, M\}$ is the kernel obtained from each modality, β_m is the weight assigned to each kernel, and M is the total number of kernels employed. DR is then performed on \hat{K} using semi-supervised GE (SSGE) which incorporates partial labels to provide a better low dimensional representation of the data allowing for better class separation and hence improved classification with limited training samples.

In this paper we show an application of SeSMiK-GE to combine structural information obtained from T2-weighted Magnetic Resonance (MR) Imaging (T2w MRI) and metabolic information obtained from MR Spectroscopy (MRS) for detection of high-grade prostate cancer (CaP) *in vivo*. The Gleason grading system is the most commonly used system world-wide for identifying aggressivity of CaP, and hence patient outcome.

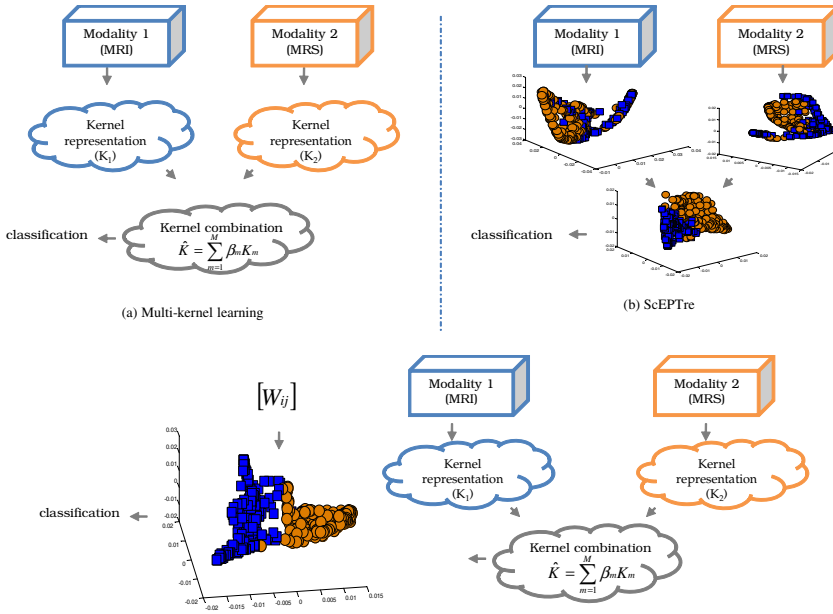


Fig. 1. (a) MKL employs a kernel combination strategy, (b) ScEPTre-based data fusion where low dimensional embedding representations are combined, and (c) SeSMiK-GE method where MKL is performed to first combine the data in a common kernel space followed by semi-supervised GE. The two colors in each 3D embedding plot represent two different classes.

High Gleason scores are associated with poor outcome, while lower scores are typically associated with better patient outcome. Recently, researchers have been attempting to identify MR imaging signatures for high- and low-grade CaP *in vivo* [7][8].

2 Graph Embedding Framework

The aim of GE [4] is to reduce the data matrix $\mathcal{F} \in \mathbb{R}^D$ into a low-dimensional space $\mathbf{y} \in \mathbb{R}^d$ ($D \gg d$), such that object adjacencies are preserved from \mathbb{R}^D to \mathbb{R}^d . Let $\mathcal{F} = [\mathbf{F}(c_1), \mathbf{F}(c_2), \dots, \mathbf{F}(c_n)] \in \mathbb{R}^D$ be a data matrix of n objects, $i \in \{1, \dots, n\}$, with dimensionality D , and $\mathbf{y} = [y_1, y_2, \dots, y_n]$ be the corresponding optimal low dimensional projection matrix. \mathbf{y} can be obtained by solving,

$$\mathbf{y} = \arg \min_y \left(\sum_{i,j=1}^n ||y_i - y_j||^2 w_{ij} \right), \tag{1}$$

where $W = [w_{ij}]$ is a similarity matrix which assigns edge weights to characterize similarities between pairwise points c_i and c_j , $i, j \in \{1, \dots, n\}$. The minimization of Equation 1 reduces it to an eigenvalue decomposition problem,

$$W\mathbf{y} = \lambda \mathcal{D}\mathbf{y}, \tag{2}$$

where \mathcal{D} is a diagonal matrix, $\mathcal{D}_{ii} = \sum_i W_{ij}$. According to the Representer Theorem [9], to calculate the kernel representation $K(\mathbf{F}(c_i), \mathbf{F}(c_j))$ of input data, it is assumed that the optimal embedding \mathbf{y} lies in the input space such that $\mathbf{y} = \sum_{j=1}^n \alpha_j K(\mathbf{F}(c_i), \mathbf{F}(c_j))$. Thus, the kernel formulation of Equation 2 can be re-written as,

$$KWK^T \alpha = \lambda K \mathcal{D} K^T \alpha, \tag{3}$$

where K is a valid positive semi-definite kernel and α is the d dimensional eigenvector of the kernel representation in Equation 3.

3 Semi-Supervised Multi-kernel Graph Embedding (SeSMiK-GE)

1. *Constructing Kernels for each modality:* Kernel functions embed input data in the implicit dot product space, evaluating which yields a symmetric, positive definite matrix (gram matrix). A kernel gram matrix K_m defining the similarities between n data points in each modality m may be obtained as $K_m = [K(\mathbf{F}(c_i), \mathbf{F}(c_j))] \forall i, j \in \{1, \dots, n\}$.

2. *Combining Multiple kernels:* A linear combination of different kernels has the advantage of also yielding a symmetric, positive definite matrix. Assuming we have M base kernel functions for M modalities, $\{K_m\}_{m=1}^M$, corresponding individual kernel weights β_m , the combined kernel function may be expressed as,

$$\begin{aligned} \hat{K}(\mathbf{F}(c_i), \mathbf{F}(c_j)) &= \sum_{m=1}^M \beta_m K_m(\mathbf{F}(c_i), \mathbf{F}(c_j)), \beta_m \geq 0, \forall (i, j) \in \{1, \dots, n\} \\ &= \sum_{m=1}^M \beta_m K_m, \beta_m \geq 0. \end{aligned} \tag{4}$$

3. *Constructing the adjacency graph using partial labels:* Assuming the first l of n samples are labeled $\omega_i \in \{+1, -1\}$, we can incorporate the partial known labels into the similarity matrix $W = [w_{ij}]$. A \mathcal{N} nearest neighbor graph, $\mathcal{N} > 0$, is created to obtain W such that pairwise points in \mathcal{N} neighborhood with same labels are given high weights and points with different class labels are given low weights [5]. If the points are not in \mathcal{N} , the corresponding edges are not connected. Thus the weight matrix is,

$$w_{ij} = \begin{cases} \gamma(1 + \gamma), & \text{if } F(c_i) \in \mathcal{N}_j \text{ or } F(c_j) \in \mathcal{N}_i \text{ and } \omega_i = \omega_j, \\ \gamma(1 - \gamma), & \text{if } F(c_i) \in \mathcal{N}_j \text{ or } F(c_j) \in \mathcal{N}_i \text{ and } \omega_i \neq \omega_j, \\ \gamma, & \text{if } F(c_i) \in \mathcal{N}_j \text{ or } F(c_j) \in \mathcal{N}_i, i > l \text{ or } j > l, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

where $\gamma = e^{-\frac{\|F(c_i) - F(c_j)\|^2}{\sigma}}$ and σ is the scaling parameter. The weight matrix W_m obtained from each modality may be averaged to obtain $\hat{W} = \frac{1}{m} \sum_{m=1}^M W_m$.

4. *Obtaining the low dimensional embedding:* The combined kernel \hat{K} and associated weight matrix \hat{W} obtained from Equations (4) and (5) can be used to reduce Equation (3) to the eigenvalue decomposition problem,

$$\hat{K} \hat{W} \hat{K}^T \alpha = \lambda \hat{K} \hat{\mathcal{D}} \hat{K}^T \alpha, \tag{6}$$

where $\hat{\mathcal{D}} = \sum_j \hat{w}_{ji}$. The optimal d dimensional eigenvectors $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ are obtained using standard kernel ridge regression optimization as described in [10].

4 SeSMiK-GE for Prostate Cancer Detection and Grading

4.1 Data Description

A total of 19 1.5 Tesla (T) T2w MRI and corresponding MRS pre-operative endorectal *in vivo* prostate studies were obtained from the University of California, San Francisco. The 3D prostate T2w MRI scene is represented by $\mathcal{C} = (\hat{C}, \hat{f})$, where \hat{C} is a 3D grid of voxels $\hat{c} \in \hat{C}$ and $\hat{f}(\hat{c})$ is a function that assigns an intensity value to every $\hat{c} \in \hat{C}$. We also define a spectral scene $\mathcal{C} = (C, \mathbf{F})$ where C is a 3D grid of MRS metavoxels, $c \in C$, and \mathbf{F} is a spectral vector associated with each $c \in C$. Note that multiple MRI resolution voxels are present within the region R_{cd} between any two adjacent MRS resolution metavoxels $c, d \in C$. An expert spectroscopist (JK) manually annotated individual MRS metavoxels across all 19 patient studies as firstly (a) CaP/benign, and secondly, (b) as low/high grade CaP. The 19 1.5 T studies comprised a total of (a) 573 CaP and 696 benign metavoxels, and (b) 175 low and 96 high grade CaP metavoxels.

4.2 Feature Extraction from MRI and MRS

(a) *Feature extraction from MRS*: For each $c \in C$, $\mathbf{F}(c) = [f_a(c)|a \in \{1, \dots, U\}]$, represents the MR spectral vector, reflecting the frequency component of each of U metabolites. The corresponding spectral data matrix is given as $\mathcal{F} = [\mathbf{F}_1(c); \mathbf{F}_2(c), \dots; \mathbf{F}_n(c)] \in \mathbb{R}^{n \times U}$ where $n = |C|$, $|C|$ is the cardinality of C .

(b) *Feature extraction from MRI*: 38 texture features were extracted to define CaP appearance on *in vivo* T2w MRI [11]. We calculated the feature scenes $\mathcal{G}_u = (\hat{C}, \hat{f}_u)$ for each \mathcal{C} by applying the feature operators $\Phi_u, u \in \{1, \dots, 38\}$ within a local neighborhood associated with every $\hat{c} \in \hat{C}$. 13 gradient, 12 first order statistical and 13 Haralick features were extracted at each $\hat{c} \in \hat{C}$. We define a T2w MRI texture feature vector for each metavoxel $c \in C$ by taking the average of the feature values within the corresponding metavoxel as $g_u(c) = \frac{1}{|R_{cd}|} \sum_{\hat{c} \in R_{cd}} [\hat{f}_u(\hat{c})]$. The corresponding feature vector is then given as $\mathbf{G}(c) = [g_u(c)|u \in \{1, \dots, 38\}]$, $\forall c \in C$, and the MRI data matrix is given as $\mathcal{G} = [\mathbf{G}_1; \mathbf{G}_2; \dots; \mathbf{G}_n] \in \mathbb{R}^{n \times 38}$.

4.3 SeSMiK-GE for Integration of MRI and MRS

A Gaussian kernel $K(\mathbf{F}(c_i), \mathbf{F}(c_j)) = e^{-\frac{\|\mathbf{F}(c_i) - \mathbf{F}(c_j)\|^2}{\sigma}}$ was employed within SeSMiK-GE to obtain K_{MRS} and K_{MRI} from input MRS data \mathcal{F} and MRI data \mathcal{G} . 40% of the total samples were randomly selected to train the algorithm over 25 iterations of cross validation. W_{MRI} and W_{MRS} were obtained using Equation (5) and averaged to obtain the fused gram matrix \hat{W} . The algorithm was evaluated over different values of $\beta \in [0, 1]$ over intervals of 0.1 to obtain 11 embedding outputs $\alpha_q, q \in \{1, \dots, 11\}$, where α_1 represents the embedding obtained purely from K_{MRS} ($\beta = 0$) and α_{11} represents the embedding obtained purely from K_{MRI} ($\beta = 1$). A probabilistic boosting tree (PBT) classifier [12], was then trained using the same set of samples exposed for SeSMiK over each iteration of 25 cross validation runs. During each iteration, the optimal α_q which results in the maximum classification accuracy is selected as the final embedding result. The algorithm for SeSMiK-GE is presented below.

Algorithm *SeSMiK-GE***Input:** $\mathcal{F}, \mathcal{G}, \mathcal{N}, d$ **Output:** α *begin*

0. Obtain $K_{MRS} \leftarrow \mathcal{F}$ and $K_{MRI} \leftarrow \mathcal{G}$
1. Obtain W_{MRI} and W_{MRS} using \mathcal{N} ; obtain \hat{W} using Equation (6)
2. Initialize $\beta = 0, q = 1$
3. **while** $\beta = 1$, compute $\hat{K}_q = \beta \times K_{MRI} + (1 - \beta) \times K_{MRS}$
4. Substitute \hat{K}_q and \hat{W} in Equation (3)
5. Obtain d -dimensional α_q by solving Equation (7)
6. **return** α_q
7. $\beta = \beta + 0.1, q = q + 1$
8. **endwhile**

end

The algorithm above was applied to the problems of (a) discriminating CaP vs. benign regions, and (b) identifying high-grade CaP using multi-protocol MRI.

5 Results and Discussion

5.1 Qualitative

A PBT classifier [12] was trained on the low dimensional embedding representations obtained from application of SeSMiK-GE to the 19 MRI, MRS studies. Figure 2(a) shows a T2w MRI slice with high grade CaP, while Figure 2(b) shows signature spectra corresponding to low and high grade CaP, in turn illustrating the difficulty in visually

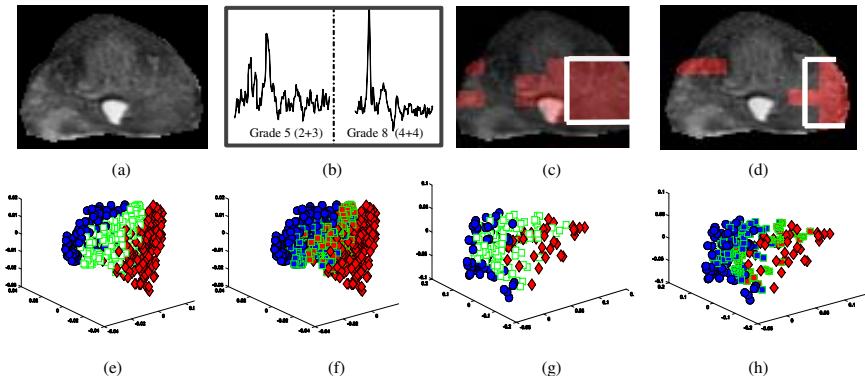


Fig. 2. (a) Original T2w MRI, (b) MRS signatures for high/low grade CaP. Red regions on (c)-(d) show the classification results obtained using SeSMiK-GE for identifying CaP and high grade CaP on *in vivo* MRI, MRS. White ROI delineates the ground truth for CaP extent in (c) and high-grade CaP in (d). (e), (g) show 3D embedding plots obtained from SeSMiK-GE (with partial training labels for each class) for cancer metavoxels (red) and benign metavoxels (blue). The spectra in the evaluation (test set) are shown via green squares. (f), (h) illustrate the classification results via PBTs on the same embedding for detection of CaP ((f)), and high-grade CaP ((h)).

identifying high grade CaP on *in vivo* MRI and MRS. Figures 2(c)-(d) show the PBT classifier prediction results (as red regions) on the same T2w MRI slice using SeSMiK-GE for identifying CaP (Figure 2(c)) and high-grade CaP (Figure 2(d)). Note the high detection accuracy obtained using SeSMiK-GE for both CaP (Figure 2(c)) and high-grade CaP (Figure 2(d)) identification. Ground truth for CaP and high-grade CaP extent is shown via a white ROI on Figures 2(c) and 2(d).

Embedding plots obtained from SeSMiK-GE for cancer (red)/benign (blue) and high (red)/low (blue) grade CaP are shown in Figures 2(e)-(h). Figures 2(e), (g) show the partial labels provided to SeSMiK-GE, allowing for better separation between the classes (green squares represent the unlabeled samples). Figures 2(f), (h) show corresponding PBT classification labels for identifying CaP (2(f)) and high-grade CaP (2(h)). Note that in Figures 2(f) and (h), a majority of the unlabeled samples are accurately identified by our scheme, despite using limited partial labels for training.

5.2 Quantitative

Table 1(a) shows mean area under the ROC curve (AUC) and accuracy results averaged over 19 studies for identifying cancer vs. benign using SeSMiK-GE on (a) only MRI, (b) only MRS, and (c) in combining MRS and MRI (results shown are for optimal α_q obtained at $\beta = 0.8$). We compared our results with ScEPTre [11], where GE [4] was first performed on each of \mathcal{F} and \mathcal{G} followed by concatenation of the resulting low

Table 1. (a) Average AUC and accuracy for CaP detection, compared to MRI-MRS alone, and ScEPTre [11] based data fusion, averaged over a total of 19 MRI-MRS studies using the 30 top-ranked eigen values, (b) Average CaP detection accuracy and AUC results of SeSMiK-GE and ScEPTre for different dimensions $\mathbf{d} \in \{10, 20, 30\}$.

(a)			(b)				
Method	AUC	Accuracy	\mathbf{d}	AUC		Accuracy	
T2w MRI	66.1 ± 1.5	61.9 ± 1.3		SeSMiK	ScEPTre	SeSMiK	ScEPTre
MRS	82.6 ± 1.3	76.8 ± 1.3	10	89.8 ± 0.8	86.8 ± 0.9	84.2 ± 1.1	80.6 ± 1.3
ScEPTre	86.8 ± 1.26	78.2 ± 1.2	20	90.7 ± 0.9	87.5 ± 0.8	84.6 ± 0.1	79.1 ± 1.2
SeSMiK-GE	91.3 ± 0.2	83.0 ± 0.1	30	91.3 ± 0.2	86.8 ± 1.26	83.0 ± 0.1	78.2 ± 1.2

Table 2. (a) Average AUC and accuracy for high-grade CaP detection, compared to MRI or MRS alone, and ScEPTre [11] data fusion averaged over a total of 19 MRI-MRS studies using the 10 top-ranked eigen values, (b) high-grade CaP detection accuracy and AUC results of SeSMiK-GE and ScEPTre for different dimensions $\mathbf{d} \in \{10, 20, 30\}$.

(a)			(b)				
Method	AUC	Accuracy	\mathbf{d}	AUC		Accuracy	
T2w MRI	79.8 ± 3.3	74.1 ± 4.0		SeSMiK	ScEPTre	SeSMiK	ScEPTre
MRS	83.7 ± 3.5	78.5 ± 3.0	10	86.9 ± 2.2	84.4 ± 2.7	80.5 ± 2.6	79.1 ± 3.6
ScEPTre	83.9 ± 3.5	76.8 ± 3.1	20	87.5 ± 2.5	83.9 ± 3.5	82.5 ± 2.6	76.8 ± 3.1
SeSMiK-GE	87.5 ± 2.5	82.5 ± 2.6	30	86.5 ± 2.8	83.8 ± 3.5	79.5 ± 3.3	77.2 ± 3.5

dimensional eigenvectors. Note the high detection accuracy obtained using SeSMiK-GE for CaP (Table 1(a)) and high-grade CaP (Table 2(a)) detection. Table 1(b) shows the low variability in AUC and accuracy results over different values of reduced dimensions (**d**). Similar results for discriminating high and low grade CaP are shown in Table 2(b).

6 Conclusions and Future Work

We presented a novel semi-supervised multi-kernel (SeSMiK) scheme which is well integrated in a graph embedding framework for simultaneous data fusion and dimensionality reduction. Multi-kernel learning is first used to combine heterogeneous information from various data sources in a common kernel framework. The method leverages partial domain knowledge to create an optimal embedding from the combined data such that object classes are optimally separable. We demonstrated the application of our scheme in discriminating cancer/benign and high/low grade prostate cancer regions using metabolic information obtained from MRS and anatomic information obtained from T2w MRI. Quantitative results demonstrate a high detection accuracy in identifying cancer and high-grade prostate cancer regions, suggesting that SeSMiK can serve as a powerful tool for both computer aided diagnosis and prognosis applications. In future work we intend to explore the application of SeSMiK in other domains and problems.

References

1. Tiwari, P., et al.: Spectral Embedding Based Probabilistic Boosting Tree (ScEPTre): Classifying High Dimensional Heterogeneous Biomedical Data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 844–851. Springer, Heidelberg (2009)
2. Rohlfing, T., et al.: Information fusion in biomedical image analysis: Combination of data vs. combination of interpretations. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 150–161. Springer, Heidelberg (2005)
3. Lanckriet, G., et al.: Kernel-Based Data Fusion and Its Application to Protein Function Prediction in Yeast. In: Pacific Symposium on Biocomputing, pp. 300–311 (2004)
4. Shi, J., Malik, J.: Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 888–905 (2000)
5. Zhao, H.: Combining labeled and unlabeled data with graph embedding. *Neurocomputing* 69(16–18), 2385–2389 (2006)
6. Lin, Y., et al.: Dimensionality Reduction for Data in Multiple Feature Representations. In: Proc. NIPS, vol. 21, pp. 961–968 (2008)
7. Zakian, K., et al.: Correlation of Proton MRSI with Gleason Score Based on Step-Section Pathologic Analysis after Radical Prostatectomy. *Radiology* 234(3), 804–814 (2005)
8. Wang, L., et al.: Assessment of Biologic Aggressiveness of Prostate Cancer: Correlation of MRI with Gleason Grade after Radical Prostatectomy. *Radiology* 246(1), 168–176 (2008)
9. Scholkopf, B., et al.: A Generalized Representer Theorem. In: Proc. Computational Learning Theory, pp. 416–426 (2001)
10. Cai, D., et al.: Semi-Supervised Discriminant Analysis. In: Proc. ICCV, pp. 1–7 (2007)
11. Madabhushi, A., et al.: Automated detection of prostatic adenocarcinoma from high-resolution ex vivo mri. *IEEE Transactions on Medical Imaging* 24(12), 1611–1625 (2005)
12. Tu, Z.: Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering. In: Proc. ICCV, vol. 2, pp. 1589–1596 (2005)

Author Index

- Abolmaesumi, Purang II-68, III-311
Abràmoff, Michael D. III-33, III-603
Abu El-Ghar, Mohammed I-10
Abugharbieh, Rafeef II-331
Achanta, Radhakrishna II-463
Adebar, Troy K. II-626
Adeshina, Steve A. II-635
Aganj, Iman II-84
Ahmidi, Narges III-295
Aja-Fernandez, S. I-518
Aksoy, Murat I-259
Al-Sanawi, Hisham III-359
Alexander, Daniel C. I-183, I-534,
I-623, I-640, II-404
Alhonnoro, Tuomas I-45
Aljabar, P. III-1
Allain, Baptiste II-514
Altinay, Murat III-105
An, Hongyu II-274
Anderson, Jeffrey S. II-363
André, Barbara II-480
Andrews, Shawn III-9
Angelini, Elsa D. II-34
Arbel, Tal I-127, II-290, II-643, III-41
Armspach, Jean-Paul II-117
Arnold, Douglas L. II-290, III-41
Asano, Takaharu II-50
Ashraf, Haseem I-37
Atasoy, Selen II-437
Atkins, M. Stella I-582
Audière, Stéphane II-34
Ausubel, F.M. III-634
Avants, Brian I-324, III-105
Ayache, Nicholas I-111, II-151,
II-420, II-480, II-652
Ayles, Helen III-650
Azernikov, Sergei III-555

Bab-Hadiashar, Alireza II-193
Baka, Nora I-452
Bakker, N.H. I-526
Balachandran, Ramya III-587
Balda, Michael III-547
Balicki, Marcin III-303

Balocco, Simone II-59
Baloch, Sajjad III-113, III-555
Bammer, Roland I-259
Barber, David II-380
Barbu, Adrian I-28
Barkovich, A. James II-339
Barmpoutis, Angelos I-582
Barnes, Josephine II-125
Barnes, Nick II-266
Baronnet, Flore I-200
Barron, John III-205
Batchelor, Philip G. I-418
Baust, Maximilian II-586
Bax, Jeff II-17
Bellani, M. II-177
Ben Ayed, Ismail I-409
Ben-Bashat, Dafna I-103
Ben-Sira, Liat I-103
Benali, Habib I-316
Berezney, Ronald II-530
Bernardis, Elena I-119
Beuthien, Björn II-546
Biesdorf, Andreas I-444
Biller, Lisa I-151
Blum, Christian I-291
Blum, Tobias III-400
Bó, Antônio P.L. I-267
Boardman, J.P. III-1
Bocan, Thomas I-308, III-57
Bock, Jelena III-416
Boctor, Emad M. II-9
Boese, Jan I-375, I-476
Boisvert, Jonathan II-68
Boomen, R.v.d. I-526
Boone, Darren III-497
Borschneck, Dan II-68
Bourgeat, Pierrick II-125, II-185
Bove, Susan III-57
Brambilla, P. II-177
Breteler, Monique M.B. II-101
Brockmann, Gernot I-476
Brown, Tracy I-308
Brugada, Josep II-1
Brunenberg, Ellen I-175

- Buchner, Anna M. II-480
 Buhmann, Joachim M. II-209
 Buonaccorsi, G.A. III-121
 Burdette, Clif III-311
- Caan, Matthan W.A. I-167, II-249
 Cabeen, Ryan P. I-357
 Cagniard, Cedric III-237
 Camara, Oscar II-1
 Carpenter, A.E. III-634
 Carvajal-Gonzalez, Santos III-57
 Castellani, U. II-177
 Cavallaro, Alexander I-95
 Chakrapani, Shruthi I-357
 Chan, Kap Luk II-522
 Chan, Tony II-323
 Chaney, Ed III-335
 Chang, Yu-Bing III-278
 Changizi, Neda III-17, III-563
 Charbit, Maurice II-34
 Chatelin, Simon I-235
 Chelikani, Sudhakar I-53
 Chen, Elvis III-205
 Chen, Hanbo II-412
 Chen, Hua-mei I-340
 Chen, Mei I-1, I-209
 Chen, Sean Jy-Shyang II-92
 Chen, Taoyi III-473
 Chen, Terrence III-269
 Chen, Ting III-65
 Chen, Yasheng II-274
 Chen, Zhe I-53
 Cheng, Jack C.Y. III-538
 Cheng, Jian I-590, I-648
 Chertok, Michael III-642
 Cheung, Carling L. III-408
 Chiao, Ping-Chun III-57
 Chinchapatnam, P. II-420
 Chitphakdithai, Nicha I-367
 Cho, Daniel S. III-205
 Chowdhury, Ananda S. III-89
 Chu, Winnie C.W. III-538
 Chu, Xinqi II-522
 Chung, Adrian I-69
 Chung, Moo K. III-505
 Ciompi, Francesco II-59
 Clarkson, Matthew J. I-534, II-125
 Clatz, Olivier I-111
 Coe, Christopher I-690
 Cointepas, Yann I-550
- Collins, D. Louis II-92, II-290, II-643,
 III-41, III-129, III-181
 Colliot, Olivier I-316
 Comaniciu, Dorin I-28, I-95, I-218,
 I-383, I-460, I-476, III-269
 Comas, Olivier II-371
 Combès, Benoît II-594
 Commowick, Olivier III-25, III-155
 Conery, A.L. III-634
 Constantini, Shlomi I-103
 Cook, Philip A. I-324
 Cook, Richard J. II-514
 Cootes, Timothy F. II-635
 Corani, G. II-488
 Cotin, Stéphane II-371
 Coulthard, A. II-185
 Counsell, S. III-1
 Coupé, Pierrick III-129
 Cowan, Brett I-383
 Craige, Caryne III-105
 Criminisi, Antonio I-111
 Crozier, Stuart II-185, II-388
 Cuingnet, Rémi I-316
- D'hooge, Jan II-1
 Dabbah, M.A. I-300
 Dabool, Elad III-457
 Daghli, M. II-185
 Dalal, Pahal I-349
 Danilchenko, Andrei III-587
 Darzi, Ara W. III-245, III-319
 Das, Sandhitsu III-105
 Dassopoulos, T. II-454
 Davatzikos, Christos II-160,
 II-257, II-676
 Davydov, Oleg I-666
 Dawson, Rodney III-650
 de Boer, Renske II-101
 de Bruijne, Marleen I-37, I-452,
 II-193, III-595
 De Camilli, Pietro II-315
 De Craene, Mathieu II-1
 de Groot, Marius II-101
 de Jong, Pim A. II-396, III-650
 de La Gorce, Martin II-668
 De Nigris, Dante II-643
 De Silva, Tharindu III-213
 Dehghan, Ehsan I-283
 Del Maestro, Rolando II-643
 Delingette, Hervé I-235, II-420, II-652

- Delmaire, Christine II-217
 Delmas, Patrice III-481
 Demiralp, Çağatay I-542
 den Dekker, Arjan I-167
 Deriche, Rachid I-590, I-631, I-648
 Descoteaux, Maxime I-550
 Díaz, Alejandro III-163
 Dietemann, Jean-Louis I-574, II-355
 Dinov, Ivo I-357, III-49, III-81
 Dirksen, Asger I-37, II-193
 Doltra, Adelina II-1
 Dong, Bin I-143
 Doria, V. III-1
 Dormont, Didier I-316
 dos Santos, Thiago R. I-251, II-660
 Douglas, Tania S. III-619
 Dowson, N. II-185
 Duchateau, Nicolas II-1
 Duda, Jeffrey T. II-282
 Duin, Robert P.W. I-37
 Duits, Remco I-175
 Duncan, James S. I-53, I-367,
 I-502, II-538, II-315
 Duriez, Christian II-371
 Durr, Alexandra II-217
- Ecabert, O. I-400
 Edwards, A.D. III-1
 Eisenschenk, Stephan J. III-65
 El-Baz, Ayman I-10
 Elhabian, Shireen III-626
 Elliott, Colm II-290
 Ellis, Randy E. III-359
 Elson, Daniel III-245
 Engin, Kayihan III-137
 Eskildsen, Simon F. III-181
 Essafi, Salma III-189
 Esses, S.J. III-73
 Estroff, Judy A. II-109
- Falk, Robert III-626
 Fallavollita, Pascal III-311
 Fang, Tong III-113, III-555
 Farag, Aly III-626
 Farag, Amal III-626
 Fay, M. II-185
 Fenster, Aaron II-17, III-213
 Feußner, Hubertus III-400
 Feulner, Johannes I-95
- Fichtinger, Gabor I-283, II-68,
 III-295, III-311, III-383, III-424
 Fillard, Pierre I-200, I-550
 Finkelstein, Yehuda III-642
 Fischer, Bernd II-546
 Fitzpatrick, J. Michael I-251, III-587
 Flanagan, Ronan I-45
 Fleming, Ioana N. II-9
 Fletcher, P. Thomas II-363, III-529
 Flöry, Simon III-555
 Folkerts, Michael III-449
 Fonov, Vladimir III-129
 Forman, Christoph I-259
 Foroughi, Pezhman II-9
 Foskey, Mark III-335
 Fox, Nick C. I-534, II-125
 Francis, Simon J. II-290, III-41
 Frangi, Alejandro F. I-468, I-518, II-1
 Franz, A.M. I-251
 Freiman, M. III-73
 Frigstad, Sigmund I-510
 Friman, Ola III-416
 Fripp, Jurgen II-125
 Fritz, Andrew II-530
 Fua, Pascal I-291, II-463
 Fuchs, Siegfried III-113
 Fuchs, Thomas J. II-209
- Gallia, Gary L. III-295
 Gamage, Pavan III-481
 Gambardella, L. II-488
 Gammon, Braden III-359
 Gangeh, Mehrdad J. III-595
 Ganor, Roy III-642
 Gao, Dashan II-446
 Gao, Fei III-571
 Gao, Hang II-1
 Gao, Wei II-298
 Gardi, Lori II-17
 Garvin, Gregory J. I-340
 Garvin, Mona K. III-33
 Gaser, Christian II-169
 Gatenó, Jaime III-278
 Gatta, Carlo II-59
 Gauvrit, Jean-Yves II-92
 Gee, James C. I-324, II-282
 Gehlbach, Peter III-303, III-465
 Geng, Xiujuan I-598
 Georgescu, Bogdan I-218, I-383,
 I-460, I-476

- Geremia, Ezequiel I-111
 Gerig, Guido II-602
 Gholipour, Ali II-109
 Ghosh, Aurobrata I-590, I-648
 Ghotbi, Reza III-237
 Gianaroli, L. II-488
 Gill, Jaswinder I-391
 Gill, Sean II-68
 Gilmore, John H. I-690, II-602
 Gimel'farb, Georgy I-10
 Giovanello, Kelly II-298
 Giusti, A. II-488
 Glenn, Orit A. II-339
 Glocker, Ben II-676
 Goela, Aashish I-340
 Gogin, Nicolas I-391
 Golby, Alexandra J. II-225
 Goldstein, Jill I-657
 Golland, Polina I-191, II-151, III-634
 Gong, Ren Hui II-68
 González, Germán I-291
 Gorbunova, Vladlena II-193
 Graham, J. I-300,
 Graham, James III-626
 Grbić, Saša I-218
 Grigis, Antoine II-117
 Grimbergen, Kees I-167
 Grossman, Murray I-324, II-282
 Grzeda, Victor III-424
 Gu, Hong I-598
 Gu, Xianfeng II-323, III-538
 Gu, Xuejun III-449
 Guehring, Jens I-383
 Guevara, Pamela I-550, II-217, II-347
 Guion, Peter III-383
 Guo, Lei II-143, II-412
 Gur, Ruben C. I-558, I-631
 Gutman, Boris I-357
 Gyacskov, Igor II-17

 Ha, Linh II-602, III-529
 Habas, Piotr A. II-339
 Hager, Gregory D. II-9, II-454,
 III-295, III-465
 Hajnal, J.V. III-1
 Hall, Matt G. II-404
 Halligan, Steve III-497
 Hamamci, Andac III-137
 Hamarneh, Ghassan II-331, III-9,
 III-17, III-563

 Hameeteman, Reinhard III-97
 Handa, James III-303, III-465
 Harloff, Andreas III-416
 Hassan Zahraee, Ali III-432
 Hawkes, David J. I-435, II-514,
 III-497
 Hawkins, Maria A. I-69
 He, Huiguang III-489
 He, Tiancheng III-392
 Hege, H.-C. I-227
 Heibel, Hauke III-343
 Heimann, Tobias I-235
 Heismann, Björn III-547
 Heitz, Fabrice II-117
 Hellier, Pierre II-92
 Hemetsberger, Rayyan II-59
 Hennemuth, Anja III-416
 Hicks, Carol I-308
 Ho, Jeffrey I-682
 Ho, Khek Yu II-522
 Hoffman, Eric A. II-578
 Hogeweg, Laurens II-396, III-650
 Højgaard, Liselotte III-253
 Hojjatoleslami, Ali I-666
 Holm, Darryl D. II-610
 Honnorat, Nicolas III-440
 Hornegger, Joachim I-95, I-151, I-259,
 I-460, III-547
 Hose, Rod II-380
 Htwe, That Mon II-522
 Hu, Jiayi I-332
 Hu, Mingxing II-514, III-497
 Hu, Xintao II-143, II-412
 Hu, Zhihong III-33
 Hua, Jing I-332, III-489, III-538
 Huang, Junzhou I-135
 Huang, Xiaolei I-86
 Huber, Martin I-95, I-460

 Iglesias, Juan Eugenio III-81, III-147
 Ikram, M. Arfan II-101
 Ingallhalikar, Madhura I-558
 Ionasec, Razvan Ioan I-218, I-460
 Iordachita, Iulian III-303
 Ishaq, Omer III-17
 Ishii, Lisa III-295
 Ishii, Masaru III-295
 Ishikawa, Hiroshi I-1
 Islam, Ali I-409

- Jackson, A. III-121
 Jacobs, Sander S.A.M. II-193
 Jacques, Robert III-521
 Jähne, B. I-251
 James, David R.C. III-319
 Janoos, F. II-201
 Jayson, G.C. III-121
 Ji, Songbai II-274
 Jia, Hongjun II-570, II-618, II-684
 Jia, Xun I-143, III-449
 Jiang, Steve B. I-143, III-449
 Jiang, Tianzi I-590, I-648
 John, Matthias I-375, I-476
 Jolly, Marie-Pierre I-383
 Joshi, Anand A. I-357
 Joshi, Sarang II-602, III-529
 Joshi, Shantanu H. I-357
 Joskowicz, Leo I-103, III-73, III-457
 Jürgens, Philipp I-61

 Kadoury, Samuel III-579
 Kadowaki, Takashi II-50
 Kainmueller, D. I-227
 Kainz, Bernhard I-45
 Kamel, Mohamed S. III-595
 Kamen, Ali II-546, II-586
 Kanade, Takeo I-209
 Kanterakis, Efstathios I-558, I-631
 Karamalis, Athanasios I-243
 Karimaghloo, Zahra III-41
 Karnik, Vaishali V. II-17
 Karpikov, Alexander II-538
 Kaushal, Aradhana III-383
 Kaynig, Verena II-209
 Kazhdan, Michael III-521
 Kempfert, Jörg I-375, I-476
 Khalifa, Fahmi I-10
 Khallaghi, Siavash II-68
 Khedoe, Ganesh I-167
 Kim, Hyungmin I-61
 Kim, Jinman II-562
 Kim, Kio II-339, II-355
 Kim, Minjeong II-306
 Kim, Sungeun III-611
 Kindlmann, Gordon I-674, III-163
 King, Alex III-521
 King, Andy P. I-391
 King, Richard III-529
 Kirisli, Hortense I-452

 Kirschstein, Uwe I-375, I-476
 Klein, Stefan I-452
 Kleinschmidt, Andreas I-200
 Klinder, Tobias III-327
 Kneser, R. I-400, I-526
 Knopp, M.V. II-201
 Koizumi, Norihiro II-50
 Kong, Linglong I-690
 Konukoglu, Ender I-111
 Koob, Meriam I-574, II-355
 Korosoglou, G. I-526
 Kozerke, Sebastian I-418
 Krieger, Axel III-383
 Krishnan, Arun I-19
 Kronman, A. III-73
 Kroon, Dirk-Jan III-221
 Kruecker, Jochen II-42
 Kubicki, Marek I-191
 Kubota, Naoto II-50
 Kucuk, Nadir III-137
 Kumar, R. II-454
 Kurhanewicz, John III-666
 Kwok, Ka-Wai III-229, III-319
 Kyujin Paik, Jamie III-432

 Ladikos, Alexander III-237
 Lai, Rongjie III-49
 Lai, Zhaoqiang I-332
 Laidlaw, David H. I-542
 Lallemand, Joe II-437
 Lamata, Pablo II-380
 Lamecker, H. I-227
 Laporte, Catherine I-127
 Larsen, Rasmus III-253
 Lashkari, Danial II-151
 Lasso, Andras III-383
 Lauritsch, Günter I-151
 Lee, Huai-Ping III-335
 Lee, Jack II-380
 Lee, Junghoon I-283
 Lee, Kyungmoo III-33
 Lee, Su-Lin I-69
 Lee, Tzumin II-472
 Leff, Daniel R. III-319
 Lehéricy, Stéphane I-316, II-217
 Lelieveldt, Boudewijn I-452
 Lenglet, Christophe II-84
 Lepetit, Vincent II-463
 Lerotic, Mirna I-69
 Lessick, J. I-400

- Leung, Kelvin K. I-435, I-534, II-125
 Levy, Josh III-335
 Lewis, John H. III-449
 Li, Hongsheng I-86
 Li, Kaiming II-143
 Li, Kang I-209
 Li, Liyuan II-522
 Li, Ruijiang III-449
 Li, Shuo I-340, I-409
 Li, Yang II-133
 Li, Yimei I-690
 Liachenko, Serguei III-57
 Liang, Liang II-315
 Liao, Hongen II-50
 Liao, Rui I-375, I-476
 Liao, Wei-shing III-269
 Lilja, Mikko I-45
 Lim, Joo Hwee II-522
 Lin, Ching-Long II-578
 Lin, Weili I-690, II-133, II-274, II-298
 Linguraru, Marius George III-89
 Linte, Cristian A. III-205
 Litjens, G.J.S. II-396
 Littmann, Arne I-383
 Liu, Chang I-332
 Liu, Cheng-Yi III-147
 Liu, David I-28
 Liu, Huafeng III-571
 Liu, Jianfei II-505
 Liu, Jiang II-522
 Liu, Qingshan I-484
 Liu, Tianming II-143, II-412
 Liu, Wei II-363
 Liu, Yinxiao III-172
 Liu, Yunlong III-172
 Liu, Yu-Ying I-1
 Ljosa, V. III-634
 Lo, Pechin I-37, II-193
 Loog, Marco I-37, III-595
 Lorenz, Cristian III-327
 Lou, Yifei I-143
 Lovat, Laurence B. II-514
 Lu, Chao I-53
 Lu, Le I-19
 Lu, Xiaoguang I-383
 Lucchi, Aurélien II-463
 Lui, Lok Ming II-323, III-538
 Lv, Jinglei II-143
 Lythgoe, Mark F. II-404
 Ma, Jun I-19
 Ma, YingLiang I-391
 Maal, Thomas J.J. III-221
 MacFarlane, D. II-185
 Machiraju, R. II-201
 Madabhushi, Anant III-197,
 III-658, III-666
 Magli, C. II-488
 Magnenat-Thalmann, Nadia II-562
 Mahapatra, Dwarikanath I-493
 Mahdavi, S. Sara II-76
 Maier-Hein, Lena I-251, II-660
 Maitrejean, Serge II-668
 Majumdar, Angshul III-513
 Malandain, Grégoire III-155
 Malcolm, James G. I-657, II-233
 Malik, R.A. I-300
 Malone, Ian B. I-534
 Mangin, Jean-François I-550,
 II-217, II-347
 Manjón, José V. III-129
 Mansi, T. II-652
 Marchal, Maud II-92
 Marchesseau, Stéphanie I-235
 Markl, Michael III-416
 Marrakchi-Kacem, Linda I-550, II-217
 Marron, J. Stephen III-529
 Martin-Fernandez, M. I-518
 Marvasty, Idean I-484
 Mateus, Diana II-437
 Mayrhauser, Ursula I-45
 McCarley, Robert W. I-657
 McClelland, Jamie III-497
 McKay, Richard II-446
 McMillan, Corey I-324, II-282
 McNutt, Todd III-521
 Meining, Alexander II-437
 Meinzer, Hans-Peter I-251, II-660
 Melkisetoglu, Rupen III-555
 Men, Chunhua III-449
 Menze, Bjoern H. I-111, II-151
 Mercier, Laurence II-643
 Meriaudeau, Fabrice II-125
 Merrifield, Robert III-261
 Metaxas, Dimitris N. I-135, I-484
 Metz, Coert I-452
 Michailovich, Oleg I-607, I-657
 Michel, Fabrice III-189
 Mochizuki, Takashi II-50
 Modat, Marc I-534, III-497

- Mol, Christian III-650
 Mollet, Nico I-452
 MomayyezSiahkal, Parya I-566
 Monaco, James P. III-197
 Montreuil, Jacques II-17
 Moore, John III-205, III-408
 Moradi, Mehdi I-283, II-76
 Morandi, Xavier II-92
 Morel, Guillaume III-432
 Mórocz, I.Á. II-201
 Morris, William J. II-76
 Morton, Daniel I-308
 Mountney, Peter II-496
 Mousavi, Parvin II-68
 Mouton, André III-619
 Muehl, Judith I-45
 Mueller, Susanne III-105
 Mullin, G. II-454
 Murgasova, M. III-1
 Murino, V. II-177
 Myers, Eugene II-472
 Mylonas, George P. III-319
- Nachmani, Ariela III-642
 Narasimhamurthy, Anand I-510
 Navab, Nassir I-218, I-243, II-26, II-437,
 II-586, III-237, III-343, III-400
 Neeffjes, Lisan I-452
 Nenadic, Igor II-169
 Newton, Richard III-245
 Ng, Bernard II-331
 Nguan, Christopher Y. II-626
 Nho, Kwangsik III-611
 Nie, Jingxin II-412
 Niederer, Steven II-380
 Nielsen, Mads II-193
 Niemeijer, Meindert III-33, III-603
 Niessen, Wiro J. I-452, II-101, III-97
 Noblet, Vincent II-117
 Nolte, Lutz-Peter I-61
 Noonan, David P. III-245
 Norsletten, David II-380
 Norton, Isaiah II-225
 Nöttling, Alois I-375
- O'Connor, J.P.B. III-121
 O'Donnell, Lauren J. II-225
 Oentoro, Anton III-359
 Olabarriaga, Silvia I-167
 Olesen, Oline Vinter III-253
- Ong, Eng Hui II-522
 Orihuela-Espina, Felipe III-319
 Ostermann, Jörn III-327
 Ou, Yangming II-676
 Oubel, Estanislao I-574
 Oudry, Jennifer II-34
 Ourselin, Sébastien I-435, I-534, II-125,
 II-388, II-514, III-497
- Padfield, Dirk I-510, II-446
 Pai, Darshan I-332
 Palencia, C. I-518
 Panagiotaki, Eleftheria II-404
 Papademetris, Xenophon I-53
 Paragios, Nikos II-668, II-676,
 III-189, III-440, III-579
 Parker, G.J.M. III-121
 Patel, Rajni III-205
 Paulsen, Rasmus R. III-253
 Pauly, Olivier III-343
 Pautler, Stephen E. III-408
 Pavlidis, I. III-351
 Payne, Christopher J. III-245
 Pearlman, Paul C. I-502
 Peitgen, Heinz-Otto III-416
 Peng, Hanchuan II-472
 Pennec, X. II-652
 Perina, A. II-177
 Peters, J. I-400, I-526
 Peters, Terry M. III-205, III-408
 Petropoulos, I. I-300
 Pichora, David R. II-68, III-359
 Piella, Gemma II-1
 Pieper, Steve II-554
 Pike, G. Bruce III-181
 Pitcher, Richard D. III-619
 Platel, Bram I-175
 Pluta, John III-105
 Poh, Chee Khun II-522
 Poignet, Philippe I-267
 Poline, Jean-Baptiste II-241
 Pollari, Mika I-45
 Poot, Dirk I-167, I-615
 Portugaller, Horst I-45
 Poupon, Cyril I-550, II-217, II-347
 Poupon, Fabrice II-217
 Prabhu, Sanjay P. II-109
 Prastawa, Marcel II-602
 Pratt, Philip I-77, I-275
 Precel, Ronit I-103

- Preston, J. Samuel III-529
 Prima, Sylvain II-594
 Propper, Ruth II-225
 Pruessner, Jens III-129
 Punithakumar, Kumaradevan I-409
 Pura, John A. III-89

 Qi, Yuan III-611
 Qian, Zhen I-484
 Qu, Zhenshen III-473

 Radeva, Petia II-59
 Rajagopalan, Vidya II-339
 Rambaldelli, G. II-177
 Ramus, Liliane III-155
 Rangarajan, Anand III-65
 Rathi, Yogesh I-191, I-607, I-657, II-233
 Ratnarajah, Nagulan I-666
 Razavi, Reza I-391, I-435, II-420
 Rehg, James M. I-1
 Reiber, Johan H.C. III-367
 Reiser, Maximilian III-237
 Relan, J. II-420
 Renard, Félix II-117
 Resnick, Susan M. II-160
 Reyes, Mauricio I-61
 Rhode, Kawal S. I-391, I-435, II-420
 Richa, Rogério I-267
 Richardson, John B. III-181
 Ridgway, Gerard R. II-125
 Rigolo, Laura II-225
 Riklin Raviv, T. III-634
 Rinaldi, C. Aldo I-391
 Rinehart, Sarah I-484
 Risacher, Shannon L. III-611
 Risholm, Petter II-554
 Risser, Laurent II-610
 Rittscher, Jens II-446
 Rivaz, Hassan II-9
 Rivière, Denis I-550, II-347
 Roberts, C. III-121
 Roberts, Timothy P.L. I-558
 Robles, Montserrat III-129
 Roca, Pauline II-217, II-347
 Rodriguez Leor, Oriol II-59
 Roed, Bjarne III-253
 Rohkohl, Christopher I-151
 Rohling, Robert N. II-626
 Rohr, Karl I-444
 Romagnoli, Cesare II-17

 Rose, C.J. III-121
 Rose, S. II-185
 Rosen, Mark III-666
 Ross, Ian G. I-409
 Ross, James C. III-163
 Ross, Thomas J. I-598
 Rosso, Charlotte I-316
 Roth, Holger III-497
 Rother, Diego III-465
 Rousseau, Francois I-574, II-339, II-355
 Rowe, Matthew C. I-183
 Rueckert, Daniel II-610, III-1
 Rumbach, Lucien II-117
 Rutherford, M.A. III-1
 Ryan, Natalie I-534

 Saad, Ahmed III-9
 Saboo, Rohit III-335
 Saha, Punam K. III-172
 Sahin, Mustafa II-109
 Sakuma, Ichiro II-50
 Salcudean, Septimiu E. I-283, II-76,
 II-429, II-626
 Salganicoff, Marcos I-19
 Salvado, Olivier II-125, II-185
 Samarabandu, Jagath III-213
 Sammet, S. II-201
 Samset, Eigil II-554
 Samson, Yves I-316
 Sánchez, C.I. III-603
 San José Estépar, Raúl III-163
 Sandrin, Laurent II-34
 Sapiro, Guillermo II-84
 Sauer, Frank III-287
 Sauvage, Vincent III-245
 Savadjiev, Peter II-233
 Savinaud, Mickael II-668
 Saykin, Andrew J. III-611
 Schaap, Michiel I-452, II-101, III-97
 Schievano, Silvia I-460
 Schilham, A.M.R. II-396
 Schmid, Jérôme II-562
 Schmidt, M. I-251
 Schneider, Torben I-623
 Schultz, Thomas I-674
 Schuman, Joel S. I-1
 Scott, Julia II-339
 Seidman, Larry I-657
 Seim, H. I-227
 Seitel, Alexander I-251, II-660

- Seo, Seongho III-505
 Sermesant, Maxime I-418, II-420, II-652
 Seshamani, S. II-454
 Shah, Mohak III-41
 Shahid, Muhammad Waseem II-480
 Shaker, Saher B. III-595
 Shamir, Reuben R. III-457
 Shamonin, Denis III-367
 Shang, Jianzhong III-245
 Shen, Chunhua II-266
 Shen, Dinggang I-349, II-133, II-274,
 II-306, II-570, II-618, II-684
 Shen, Hongying II-315
 Shen, Kai-Kai II-125
 Shen, Li III-611
 Shen, Tian I-86
 Shen, Weijia II-522
 Shenton, Martha E. I-657, II-233
 Sherbondy, Anthony J. I-183
 Shi, Feng I-349, II-133
 Shi, Lin III-538
 Shi, Pengcheng I-159, III-571
 Shi, Yonggang III-49
 Shi, Yundi I-690
 Shin, Wanyong I-598
 Shoshan, Yigal III-457
 Siddiqi, Kaleem I-566
 Sijbers, J. I-615
 Silva, Cláudio T. II-602
 Silva, Etel II-1
 Silverman, Edwin K. III-163
 Simari, Patricio III-521
 Simmons, Andrew I-666
 Singh, Jaskaran III-81
 Singh, Nikhil III-529
 Sinusas, Albert J. I-502
 Siow, Bernard II-404
 Sitges, Marta II-1
 Slabaugh, Greg III-497, III-555
 Slump, Cornelis H. III-221
 Smith, Erin Janine III-359
 Smith, J. II-185
 Smith, Kevin II-463
 Smith, Nic II-380
 Song, Danny III-311
 Song, Qi III-172
 Sonka, Milan III-172
 Sørensen, Lauge I-37, III-595
 Sosna, J. III-73
 Sotiras, Aristeidis II-676
 Sparks, Rachel III-658
 Srinivasan, L. III-1
 St. John, Paul III-359
 Staib, Lawrence H. I-53
 Staring, Marius III-367
 Stephenson, Diane I-308
 Stiegler, Philipp I-45
 Stoeck, Christian T. I-418
 Stoel, Berend C. III-367
 Stojkovic, Branislav II-530
 Stolk, Jan III-367
 Stoyanov, Danail I-77, I-275, III-261
 Studholme, Colin I-574, II-339, II-355
 Styner, Martin I-690
 Subramanian, Kalpathi R. II-505
 Subramanian, Navneeth I-510
 Suehling, Michael I-28
 Sukno, Federico M. I-468
 Summers, Ronald M. III-89
 Sun, Bo I-357, III-49
 Sun, Hui I-468
 Sun, Ying I-493
 Sundar, Hari III-287
 Szewczyk, Jérôme III-432
 Sznitman, Raphael III-465
 Tagare, Hemant D. I-502
 Taimouri, Vahid I-332, III-489
 Tait, Diana I-69
 Tam, Roger III-17
 Tamir, Idit III-457
 Tang, Hui III-97
 Tansella, M. II-177
 Tardif, Christine L. III-181
 Tavakoli, M. I-300
 Taylor, Andrew I-460
 Taylor, Charles A. I-426, III-375
 Taylor, Russell III-303, III-465, III-521
 Taylor, Zeike A. II-388
 Teboul, Olivier III-189
 ter Haar Romeny, Bart I-175
 Thirion, Bertrand I-200, I-550, II-241
 Thiruvenkadam, Sheshadri I-510
 Thomas, P. II-185
 Thompson, Paul M. II-169,
 II-323, III-147
 Tian, Zhen I-143
 Tiwari, Pallavi III-666
 Tobon-Gomez, Catalina I-468
 Toga, Arthur W. I-357, III-49

- Tong, Gregory III-81
 Toomre, Derek II-538
 Tori, Fufa. L. III-97
 Toussaint, Nicolas I-418
 Tscheliessnigg, Karlheinz I-45
 Tsekos, N. III-351
 Tsiamyrtzis, P. III-351
 Tsin, Yanghai III-113, III-287
 Tsymbal, Alexey I-460
 Tu, Zhuowen III-81, III-147
 Tucholka, Alan II-217, II-347
 Türetken, Engin I-291
 Tustison, Nicholas J. I-324

 Unal, Gozde III-137
 Uneri, Ali III-303

 Vaillant, Régis III-440
 van den Bouwhuijsen, Quirijn J.A.
 III-97
 van der Graaff, Maaïke II-249
 van der Lijn, Fedde II-101
 van der Lugt, Aad II-101, III-97
 van Ginneken, Bram II-396,
 III-603, III-650
 Van Leemput, Koen II-151
 Van Meir, V. I-615
 van Noorden, Sander II-249
 van Onkelen, Robbert S. III-97
 van Velsen, Evert F.S. II-101
 van Vliet, Lucas I-167, II-249, III-97
 van Walsum, Theo I-452, III-97
 Varoquaux, Gaël I-200, II-241
 Vavylonis, Dimitrios I-86
 Vecerova, Jaromira II-530
 Vegas-Sanchez-Ferrero, G. I-518
 Vembar, M. I-400
 Vemuri, Baba C. I-682, III-65
 Venkataraman, Archana I-191
 Vercauteren, Tom II-480, II-514
 Verma, Ragini I-558, I-631
 Vernooij, Meike W. II-101
 Vialard, François-Xavier II-610
 Viergever, M.A. II-396
 Vikal, Siddharth III-383
 Visentini-Scarzanella, Marco I-77,
 I-275, III-261
 Vitanovski, Dime I-218, I-460
 Vitiello, Valentina III-229
 Voigt, Ingmar I-218

 von Tengg-Koblighk, Hendrik I-444
 Voros, Szilard I-484
 Vorperian, Hourì K. III-505
 Vos, Frans I-167, II-249
 Vrooman, Henri A. II-101

 Wachinger, Christian II-26
 Wächter, I. I-400, I-526
 Wählby, C. III-634
 Wallace, Michael B. II-480
 Walther, Thomas I-375, I-476
 Wan, Jing III-611
 Wang, Changhong III-473
 Wang, Chaohui III-189
 Wang, Defeng III-538
 Wang, Hongzhi I-468, III-105
 Wang, Lei II-266
 Wang, Linwei I-159
 Wang, Peng III-269
 Wang, Qian II-570, II-618, II-684
 Wang, Song I-349
 Wang, Yang I-218
 Wang, Yaping II-133
 Wang, Ying II-160
 Ward, Aaron D. II-17, III-17, III-213
 Ward, Rabab K. III-513
 Warfield, Simon K. II-109, III-25
 Washko, George R. III-163
 Wassermann, Demian I-631
 Watson, Y. III-121
 Weber, Marc-André II-151
 Wedlake, Chris III-205, III-408
 Weese, J. I-400, I-526
 Wein, Wolfgang I-243
 Weiner, Michael III-105
 Weizman, Lior I-103
 Weldelesassie, Yonas T. I-582
 Wells III, William M. II-554
 Wen, Xu I-283
 Westin, Carl-Fredrik I-191, I-657, I-674,
 II-225, II-233, III-163
 Whalen, Stephen II-225
 Wheeler-Kingshott, Claudia A.M. I-623
 Whitcomb, Louis L. III-383
 Wiener, Michael III-529
 Wiles, Andrew D. III-587
 Willinger, Rémy II-235
 Winter, C. II-185
 Witteman, Jacqueline C.M. III-97
 Wolf, Lior III-642

- Wollstein, Gadi I-1
 Wolz, Robin II-610, III-1
 Wong, Kelvin K. III-473
 Wong, Ken C.L. I-159
 Wong, Stephen T.C. III-392, III-473
 Wong, Tsz Wai II-323
 Woods, Roger P. I-357
 Wörz, Stefan I-444
 Wu, Binbin III-521
 Wu, Guorong II-306, II-570,
 II-618, II-684
 Wu, Xiaodong III-172
 Wu, Xunlei II-274

 Xia, James J. III-278
 Xiao, Changyan III-367
 Xie, Jun II-472
 Xie, Sheng Quan III-481
 Xie, Weixin III-392
 Xie, Yuchen I-682
 Xie, Zhiyong I-308, III-57
 Xiong, Guanglei I-426, III-375
 Xiong, Zixiang III-278
 Xu, Helen III-383
 Xu, Jianrong I-332
 Xu, Jinhui II-530
 Xu, Jun III-197
 Xu, Wei Liang III-481
 Xu, Xun I-28
 Xue, Zhong II-133, III-392, III-473

 Yan, Pingkun II-42
 Yan, Shuicheng II-522
 Yang, Dewen I-308
 Yang, Guang-Zhong I-69, I-77, I-275,
 II-437, II-496, III-229, III-245, III-261,
 III-319
 Yang, Jianfei II-143
 Yang, Qian II-538
 Yang, Yihong I-598
 Yap, Pew-Thian II-306
 Yau, Shing-Tung II-323, III-538
 Yedidya, Tamir III-642
 Yelnik, Jérôme II-217

 Yeniaras, E. III-351
 Yigitsoy, Mehmet II-26
 Yin, Youbing II-578
 Yin, Zhaozheng I-209
 Yip, Michael C. II-626
 Yoo, Terry S. II-505
 Yotter, Rachel Aine II-169
 Young, Alistair I-383
 Yu, Stella X. I-119
 Yuhashi, Kazuhito II-50
 Yurgelun-Todd, Deborah II-363
 Yushkevich, Paul A. I-468, III-105

 Zachow, S. I-227
 Zamanyan, Alen I-357
 Zeitz, Michael J. II-530
 Zeng, Wei III-538
 Zhan, Yiqiang I-19
 Zhang, Degang II-143
 Zhang, Heye I-159
 Zhang, Hui I-640, II-404
 Zhang, Pei II-635
 Zhang, Shaoting I-135
 Zhang, Tianhao II-257
 Zhang, Tuo II-143, II-412
 Zhao, Ting II-472
 Zheng, Yefeng I-375, I-476
 Zheng, Yuanjie I-324
 Zhou, Luping II-266
 Zhou, S. Kevin I-28, I-95, I-460,
 I-476, III-269
 Zhou, Xiang I-19
 Zhou, Xiaobo III-278
 Zhou, Y. III-351
 Zhou, You II-50
 Zhu, Hongtu I-690, II-274, II-298
 Zhu, Jiong I-332
 Zhu, Mengchen II-429
 Zhu, Peihong II-363
 Zhu, Ying III-287
 Zhu, Yongding II-530
 Zhuang, X. I-435
 Zikic, Darko II-586
 Zouhar, Alexander III-113, III-555