

Kensaku Mori Ichiro Sakuma
Yoshinobu Sato Christian Barillot
Nassir Navab (Eds.)

LNCS 8151

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013

16th International Conference
Nagoya, Japan, September 2013
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

Kensaku Mori Ichiro Sakuma
Yoshinobu Sato Christian Barillot
Nassir Navab (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013

16th International Conference
Nagoya, Japan, September 22-26, 2013
Proceedings, Part III

Volume Editors

Kensaku Mori
Nagoya University, Japan
kensaku@is.nagoya-u.ac.jp

Ichiro Sakuma
University of Tokyo, Japan
sakuma@bmpe.t.u-tokyo.ac.jp

Yoshinobu Sato
Osaka University, Japan
yoshi@image.med.osaka-u.ac.jp

Christian Barillot
IRISA, Rennes, France
E-mail: christian.barillot@irisa.fr

Nassir Navab
Technical University of Munich
Germany
nassir.navab@tum.de

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-40759-8

e-ISBN 978-3-642-40760-4

DOI 10.1007/978-3-642-40760-4

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013946897

CR Subject Classification (1998): I.4, I.5, I.3.5-8, I.2.9-10, J.3, I.6

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

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 16th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2013, was held in Nagoya, Japan during September 22–26, 2013 at Toyoda Auditorium, Nagoya University. The conference was held on a university campus, unlike the past three conferences. Toyoda Auditorium is memorable for all Nagoya University students, because entrance and graduation ceremonies are held in it during cherry-blossom season. Since MICCAI is the premier conference in the field of medical image computing and computer assisted surgery, it was our great honor to host it. Nagoya University has more than 50 years of history in medical image processing, which was initiated by Prof. Jun-ichiro Toriwaki. Nagoya also is famous for transportation and aerospace industries that utilize many robotics technologies. These robots are also manufactured in the Nagoya area and have become indispensable in current medical interventions.

This is the second time that the MICCAI conference has been held in Japan; the 5th MICCAI was held in Tokyo in 2002, which was the first MICCAI in Asia or Oceania. In MICCAI 2002, 184 papers were accepted among 321 submissions, and the conference included five satellite half-day tutorials. Since then, MICCAI has become a much larger event and typically includes 250 accepted papers from 800 submissions and 30 satellite events. At MICCAI 2013, 262 papers were accepted from 798 submissions; 34 satellite events (workshops, challenges, tutorials) were accepted.

The Program Committee (PC) of MICCAI 2013 was comprised of 101 members coordinated by a program chair and two program co-chairs from three countries. Each of the 798 papers was assigned to one primary and two secondary PC members. The primary member knew the identity of the authors, but the secondary ones did not. Each PC member had five to ten papers as the primary member and another ten to twenty as the secondary member, according to their expertise and the subject matter of the paper. The primary PC member assigned three or more external reviewers to each paper. 835 external reviewers provided 2794 reviews (359 words on average per review): 3.5 reviews per paper. At this stage, 76 papers, which failed to receive sufficient support from the external reviews, were rejected without further consideration. The authors of the remaining 722 papers were given the opportunity to rebut the anonymous reviews, based on which discussions among the reviewers took place. Finally, two secondary members independently provided meta-reviews by taking all input (the reviews, rebuttal, discussion, and the paper itself) into account to make an acceptance or rejection recommendation. For a few papers that had only two external reviews, the secondary members provided detailed reviews in addition to the meta-reviews.

A two-day PC meeting was held in Tokyo with 32 of its members. Prior to the meeting, the initial acceptance of 198 papers was decided, because they were ranked high by the external reviewers as well as two secondary PC members. 362 papers were rejected because they did not receive enough support from the reviewers or the two secondary members. Each of the remaining 162 borderline papers was considered in the following three-phase decision process.

- First stage: Six groups of five or six PC members ranked the 162 papers to select the best 36 papers for acceptance and rejected the lowest 72 papers.
- Second stage: A different set of groups selected the best 18 papers for acceptance from the remaining 54 papers and rejected 18 papers.
- Third stage: The program chair and the co-chairs selected an additional ten papers from the remaining 18 papers by considering the topics, the institutional variety, and the quality.

262 papers were finally accepted, for a 32.8% acceptance rate. The PC members also selected a set of papers suitable for oral presentation, from which the program chair and co-chairs finally decided a list of 37 oral papers by taking the variety of topics as well as the suitability for oral presentation into account. During all the review processes, possible conflicts of interests were carefully monitored and avoided as far as possible. The geographic and keyword distributions of the accepted papers are summarized in the figures.

All accepted papers were presented during three poster sessions. Oral papers were further presented during six single-track plenary oral sessions. We are greatly indebted to the reviewers and the PC members for their extraordinary efforts of careful evaluations of the submissions within a very short time frame.

In addition to the three days of the MICCAI main conference, the annual MICCAI event hosted satellite workshops, tutorials, and challenges that were organized on the day before and after the main conference. This year's call for submissions for workshops and tutorials recorded 30 workshop / challenge proposals (including four half-day proposals) and seven tutorial proposals (also including four half-day proposals). These proposals were independently reviewed by the workshop, tutorial and challenge chair teams, headed by Hongen Liao (Tsinghua University), Pierre Jannin (University of Rennes 1), Simon Warfield (Harvard Medical School), and Akinobu Shimizu (Tokyo University of Agriculture and Technology).

In the review process for the proposals for these events, we emphasized the following points. The workshop proposals were reviewed under criteria that addressed whether the workshop emphasized an open problem addressed in the MICCAI community. Tutorial proposals were reviewed based on whether they provided educational material for training new professionals in the field, including students, clinicians, and new researchers. Also, we emphasized tutorials that focused on existing sub-disciplines of MICCAI with known material, approaches, and open problems. Challenge proposals were reviewed based on whether they were interactive and encouraged problem solving. Although all of the workshop proposals were very strong, the workshop chairs selected 22 workshops (including three half-day workshops), six tutorials (including four half-day tutorials), and

six challenges (including one half-day challenge and one challenge included in the workshop). We thank the workshop, tutorial, and challenge chairs for their hard work organizing such a comprehensive and unique program.

The highlights of the MICCAI 2013 events were the keynote lectures by Dr. Atsushi Miyawaki (Riken) and Prof. Toshio Fukuda (Meijo University). Dr. Miyawaki's talk focused on new imaging technology that enables us to cruise inside a cell. Prof. Fukuda discussed simulation-based medicine for intravascular surgery. We believe these two talks provided deep insights into new technologies and highlighted the future and emerging trends in these areas.

A public lecture, which was held on the day before MICCAI's main conference, widely introduced MICCAI to the public. Three distinctive guest speakers show the state-of-the-art technologies in the MICCAI field. Prof. Koji Ikuta presented exciting nano-robotics technologies. Prof. Yoshihiro Muragaki presented technologies for advanced intelligent operating theaters. Prof. Hidefumi Kobatake demonstrated the technologies and medical applications of computational anatomy. This wonderful public lecture was managed by Prof. Ken Masamune (The University of Tokyo.)

The First International Workshop on Medical Imaging and Computer-assisted Intervention (MICI Workshop) was independently organized just after the PC meeting at The University of Tokyo under the support. This workshop shared knowledge among the public audience and PC members who are experts in the MICCAI field.

MICCAI 2013 would not have been possible without the efforts of many people behind the scenes. We thank the Organizing, Executive, and Local Executive Committee members. The Scientific Council of Japan provided great assistance organizing this conference in Japan. The Japan Society of Computer Aided Surgery (JSCAS), headed by Prof. Masaki Kitajima (International University of Wealth and Health), also helped organize it. Prof. Takeyoshi Dohi (Tokyo Denki University) supervised a successful MICCAI meeting as a founders of the MICCAI Society and the general chair of MICCAI 2002. We also thank Prof. Etsuko Kobayashi (The University of Tokyo) and Prof. Takayuki Kitasaka (Aichi Institute of Technology) for handling the financial issues. Dr. Toshiyuki Okada (Osaka University) efficiently organized the review process and compiled the proceedings. Prof. Masahiro Oda (Nagoya University) solved facility management problems. Dr. Takehiro Ando and Dr. Junchen Wang made local arrangements for the PC meeting. Prof. Daniel Rueckert (Imperial College) helped us from the preparation of MICCAI 2013 proposal to actual conference management.

We also thank the MICCAI Secretaries, Janette Wallace, Jackie Williams, and Johanne Langford of the team from Canada. We communicated with them by e-mail around midnight every day (the time difference between Nagoya and Toronto is 11 hours) for advice regarding the conference organization. Without their help, the MICCAI 2013 conference would not have been successful. We thank the MICCAI Board headed by Prof. James Duncan (Yale University) and Prof. Alison Noble (University of Oxford) for trusting us with the organization of the MICCAI 2013 conference. They gave us a lot of freedom and advice.

We also thank our secretaries, Mizuru Suzuki, Kengo Suzuki, and Emi Tanahashi (Inter Group Corp.) for their hard work handling so many requests from attendees. We say a special thanks to Rie Ohashi (Nagoya University), Ai Okano (The University of Tokyo), and Naho Obata (The University of Tokyo). The original MICCAI 2013 logos and banners were sketched by the following four students of the Aichi Institute of Technology: Miki Takahashi, Kaori Suzuki, Hikaru Sekiguchi, and Yuiko Kori.

We appreciate the financial support from the Nagoya Convention and Visitors Bureau, The Murata Science Foundation, and the Daiko Foundation. We are deeply grateful to Nagoya University for allowing us to use the Toyoda Auditorium for MICCAI 2013.

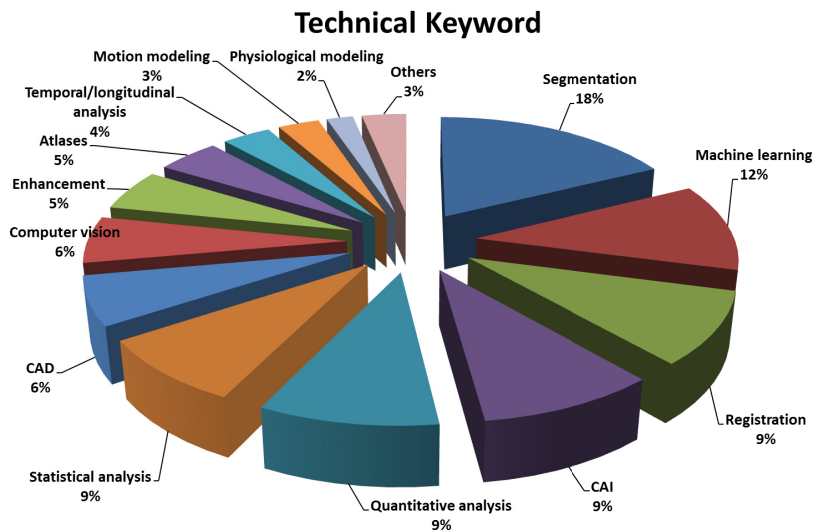
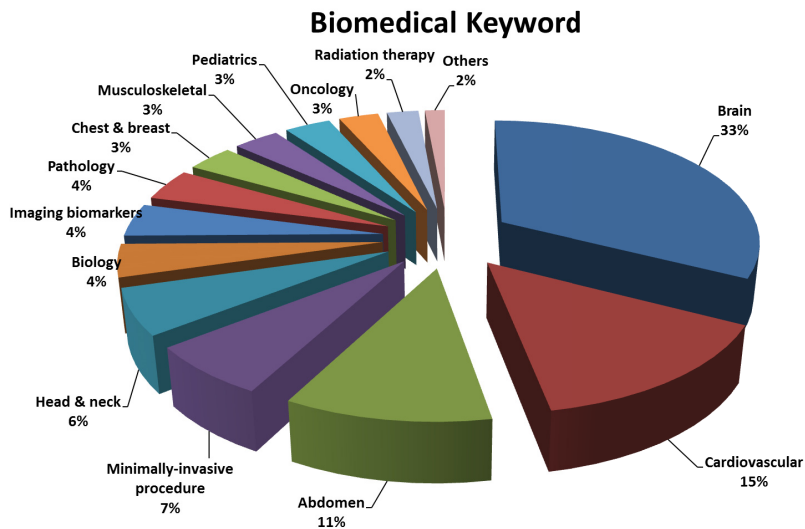
We also deeply thank our sponsors and exhibitors for their financial support.

Our initial proposal for MICCAI 2013 was accepted during MICCAI 2010 in Beijing. Six months later, a huge earthquake devastated North East Japan. Thousands of people lost their lives. We encountered many difficult situations, including the threat of radiation from the Fukushima Nuclear Power Plant. Many people from countries all over the world helped Japan and offered assistance. We are deeply grateful.

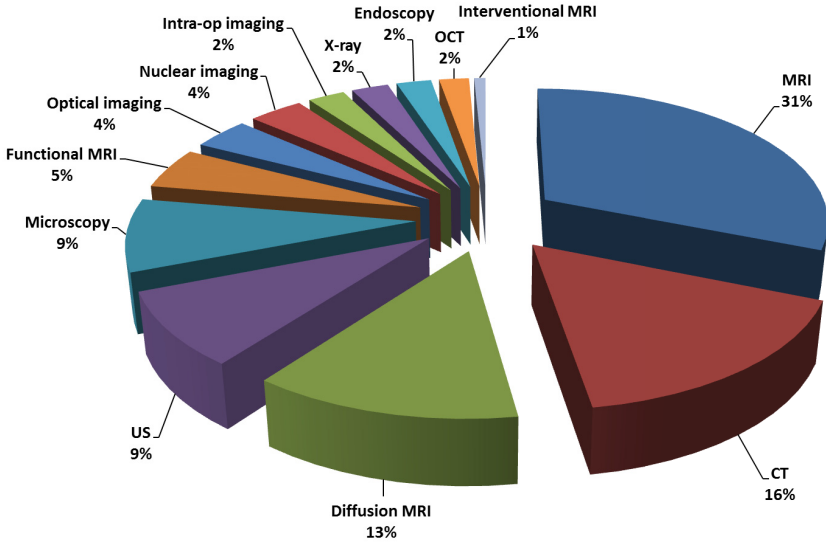
The next MICCAI conference will be held during September 14–18, 2014 in Boston, which is the one of the most beautiful cities in the world. It hosted the 1st MICCAI conference in 1998. We are looking forward to seeing all of you in Boston in 2014!

September 2013

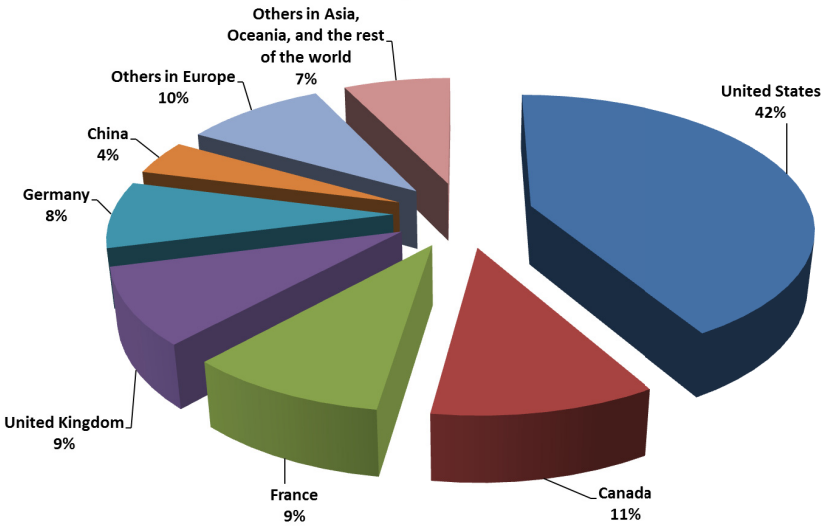
Kensaku Mori
Ichiro Sakuma
Yoshinobu Sato
Christian Barillot
Nassir Navab



Imaging Modality Keyword



Country of First Author



Organization

General Chair and Co-chair

Kensaku Mori	Nagoya University, Japan
Ichiro Sakuma	The University of Tokyo, Japan

Program Chair and Co-chairs

Yoshinobu Sato	Osaka University, Japan
Christian Barillot	INSERM, France
Nassir Navab	TU Munich, Germany

Workshop Chair and Co-chairs

Hongen Liao	Tsinghua University, China
Simon Warfield	Harvard Medical School, USA
Pierre Jannin	University of Rennes 1, France
Akinobu Shimizu	Tokyo University of Agriculture and Technology, Japan

Organizers

MICCAI 2013 Organizing Committee
Japan Society of Computer Aided Surgery
Scientific Council of Japan

Co-organizers

Information and Communications Headquarters, Nagoya University
Graduate School of Information Science, Nagoya University

MICCAI Society, Board of Directors

Alison Noble (President)	University of Oxford, United Kingdom
Sebastien Ourselin (Treasurer)	University College London, United Kingdom
Wiro Niessen (Exec. Director)	Erasmus MC - University Medical Centre, The Netherlands
Gabor Fichtinger (Secretary)	Queen's University, Canada
Stephen Aylward	Kitware, Inc., USA

Nicholas Ayache	INRIA, France
Polina Golland	MIT, USA
David Hawkes	University College London, United Kingdom
Kensaku Mori	Nagoya University, Japan
Xavier Pennec	INRIA, France
Josien Pluim	University Medical Center Utrecht, The Netherlands
Daniel Rueckert	Imperial College London, United Kingdom
Dinggang Shen	UNC, USA

Consultants to Board

Alan Colchester	University of Kent, United Kingdom
Terry Peters	University of Western Ontario, London, Canada
Richard Robb	Mayo Clinic College of Medicine, USA

Executive Officers

President:	Alison Noble, United Kingdom
Executive Director	Wiro Niessen, The Netherlands
Secretary	Gabor Fichtinger, Canada
Treasurer	Sebastien Ourselin, United Kingdom
Elections Officer	Richard Robb, USA
Awards Coordinator	Gabor Fichtinger, Canada

Non-executive Officers

Society Secretariat	Janette Wallace, Canada
Recording Secretary	Jackie Williams, Canada
Fellow Nomination Coordinator	Terry Peters, Canada

Student Board Members

President	Hakim Achterberg, The Netherlands
Website Officer	Katherine Gray, United Kingdom
Treasurer	Sinara Vijayan, Norway
Profession Events Officer	Maxime Taquet, Belgium
Student Communication and Social Events Officer	Kristin McLeod, France

MICCAI 2013 Program Committee

Purang Abolmaesumi	University of British Columbia, Canada
Burak Acar	Boğaziçi University, Turkey
Daniel Alexander	University College London, UK
Stephen Aylward	Kitware, USA
Wolfgang Birkfellner	Medical University of Vienna, Austria
Albert C. S. Chung	HKUST, Hong Kong
Ela Claridge	University of Birmingham, UK
D. Louis Collins	McGill University, Canada
Dorin Comaniciu	Siemens, USA
Tim Cootes	University of Manchester, UK
Stephane Cotin	Inria, France
Antonio Criminisi	Microsoft Research, UK
Christos Davatzikos	University of Pennsylvania, USA
Benoit Dawant	Vanderbilt University, USA
Marleen de Bruijne	Erasmus MC & University of Copenhagen, The Netherlands & Denmark
Hervé Delingette	Inria, France
Rachid Deriche	Inria, France
James S Duncan	Yale University, USA
Philip Edwards	Imperial College London, UK
Randy Ellis	Queen's University, Canada
Gabor Fichtinger	Queen's University, Canada
P. Thomas Fletcher	University of Utah, USA
Alejandro Frangi	University of Sheffield, UK
James Gee	University of Pennsylvania, USA
Guido Gerig	University of Utah, USA
Ben Glocker	Microsoft Research, UK
Leo Grady	HeartFlow, USA
Hayit Greenspan	Tel Aviv University, Israel
Alexander Hammers	Neurodis Foundation, France
Nobuhiko Hata	Harvard Medical School, USA
David Hawkes	University College London, UK
Tobias Heimann	Siemens, Germany
Joachim Hornegger	University of Erlangen-Nuremberg, Germany
Ameet Jain	Philips, USA
Pierre Jannin	Inserm, France
Tianzi Jiang	Chinese Academy of Sciences, China
Marie-Pierre Jolly	Siemens, USA
Leo Joskowicz	Hebrew University of Jerusalem, Israel
Ioannis Kakadiaris	University of Houston, USA
Nico Karssemeijer	Radboud University, The Netherlands
Ron Kikinis	Harvard Medical School, USA
Rasmus Larsen	Technical University of Denmark, Denmark
Shuo Li	GE Healthcare, Canada

Hongen Liao	Tsinghua University, China
Marius George Linguraru	Children's National Medical Center, USA
Tianming Liu	University of Georgia, USA
Cristian Lorenz	Philips, Germany
Anant Madabhushi	Case Western Reserve University, USA
Frederik Maes	KU Leuven, Belgium
Jean-Francois Mangin	CEA, France
Anne Martel	University of Toronto, Canada
Ken Masamune	University of Tokyo, Japan
Yoshitaka Masutani	University of Tokyo, Japan
Dimitris Metaxas	Rutgers University, USA
Mehdi Moradi	University of British Columbia, Canada
Mads Nielsen	University of Copenhagen, Denmark
Poul Nielsen	University of Auckland, New Zealand
Wiro Niessen	Erasmus MC & TU Delft, The Netherlands
Alison Noble	Oxford University, UK
Sebastien Ourselin	University College London, UK
Nicolas Padoy	University of Strasbourg, France
Nikos Paragios	Centale & Pontois-Paris Tech, France
Xavier Pennec	Inria, France
Terry M Peters	Roberts Research Institute, Canada
Josien Pluim	UMC Utrecht, The Netherlands
Kilian Pohl	University of Pennsylvania, USA
Torsten Rohlfing	SRI International, USA
François Rousseau	CNRS, France
Daniel Rueckert	Imperial College London, UK
Mert Rory Sabuncu	Harvard Medical School, USA
Tim Salcudean	University of British Columbia, Canada
Julia A. Schnabel	Oxford University, UK
Dinggang Shen	University of North Carolina, USA
Akinobu Shimizu	Tokyo University of A & T, Japan
Kaleem Siddiqi	McGill University, Canada
Lawrence Staib	Yale University, USA
Danail Stoyanov	University College London, UK
Colin Studholme	Washington, USA
Martin Styner	University of North Carolina, USA
Chris Taylor	University of Manchester, UK
Russell Taylor	Johns Hopkins University, USA
Bertrand Thirion	Inria, France
Paul Thompson	UCLA, USA
Jocelyne Troccaz	CNRS, France
Regis Vaillant	GE Healthcare, France
Bram van Ginneken	Radboud University, The Netherlands
Koen Van Leemput	Harvard Medical School, USA
Baba Vemuri	University of Florida, USA
Ragini Verma	University of Pennsylvania, USA

Rene Vidal	Johns Hopkins University, USA
Christian Wachinger	MIT, USA
Simon Warfield	Harvard Medical School, USA
Jürgen Weese	Philips, Germany
Wolfgang Wein	TU Munich, Germany
William Wells	Harvard Medical School, USA
Carl-Fredrik Westin	Harvard Medical School, USA
Guang Zhong Yang	Imperial College London, UK
Ziv Yaniv	Children's National Medical Center, USA
Alistair Young	University of Auckland, New Zealand
Guoyan Zheng	University of Bern, Switzerland
Darko Zikic	Microsoft Research, UK

MICCAI 2013 Organizing Committee

Kensaku Mori	Yasuhiro Kodera
Ichiro Sakuma	Hongen Liao
Yoshinobu Sato	Ken Masamune
Yen-Wei Chen	Yoshitaka Masutani
Kiyoyuki Chinzei	Yoshito Mekada
Takeyoshi Dohi	Mamoru Mitsuishi
Masakatsu G. Fujie	Ken'ichi Morooka
Hiroshi Fujita	Yoshihiro Muragaki
Hidemi Goto	Shinji Naganawa
Hideaki Haneishi	Masato Nagino
Yoshinori Hasegawa	Toshiya Nakaguchi
Makoto Hashizume	Yoshikazu Naka-jima
Hidekata Hontani	Ryoichi Nakamura
Koji Ikuta	Shigeru Nawano
Atsushi Imiya	Noboru Niki
Hiroshi Iseki	Atsushi Nishikawa
Shoji Kido	Makoto Nokata
Masaki Kitajima	Akinobu Shimizu
Takayuki Kitasaka	Toru Tamaki
Hidefumi Kobatake	Morimasa Tomikawa
Etsuko Kobayashi	Toshihiko Wakabayashi

MICCAI 2013 Executive Committee

Kensaku Mori	Yuichiro Hayashi
Ichiro Sakuma	Shingo Iwano
Yoshinobu Sato	Yasukazu Kajita
Takehiro Ando	Takayuki Kitasaka
Jumpei Arata	Etsuko Kobayashi

Hongen Liao
Ken Masamune
Yoshito Mekada
Shinji Mizuno
Ryoichi Nakamura
Yukitaka Nimura

Masahiro Oda
Toshiyuki Okada
Jun Okamoto
Shinya Onogi
Takashi Suzuki
Junchen Wang

MICCAI 2013 Local Executive Committee

Kensaku Mori
Yuichiro Hayashi
Takayuki Kitasaka
Xiongbiao Luo
Yoshito Mekada
Shinji Mizuno
Yoshihiko Nakamura

Yukitaka Nimura
Masahiro Oda
Daniel Rueckert
Kengo Suzuki
Mizuru Suzuki
Emi Tanahashi

MICCAI 2013 Reviewers

Abugharbieh, Rafeef
Achterberg, Hakim
Acosta-Tamayo, Oscar
Adluru, Nagesh
Afacan, Omur
Afsari, Bijan
Aganj, Iman
Ahmadi, Seyed-Ahmad
Aja-Fernández, Santiago
Akhondi-Asl, Alireza
Alam, Kaisar
Alander, Jarmo
Alexander, Andrew
Ali, Sahirzeeshan
Alic, Lejla
Aljabar, Paul
Allan, Maximilian
An, Jung-ha
Andres, Bjoern
Angelini, Elsa
Angelopoulou, Elli
Antony, Bhavna
Anwander, Alfred
Arbel, Tal

Arimura, Hidetaka
Ashburner, John
Assemlal, Haz-Edine
Atasoy, Selen
Atkins, Stella
Aubert-Broche, Berengere
Audette, Michel
Auzias, Guillaume
Avants, Brian
Awate, Suyash
Axel, Leon
Ayad, Maria
Bach Cuadra, Meritxell
Baka, Nora
Baldock, Richard
Baloch, Sajjad
Barbu, Adrian
Barmpoutis, Angelos
Barratt, Dean
Bartoli, Adrien
Basavanhally, Ajay
Batmanghelich, Nematollah
Batmanghelich, Kayhan
Bauer, Stefan

Baumann, Michael
 Becker, Tim
 Beichel, Reinhard
 Bekkers, Erik
 Ben Ayed, Ismail
 Bergeles, Christos
 Berger, Marie-Odile
 Bergmeir, Christoph
 Bernal, Jorge Luis
 Bernardis, Elena
 Betrouni, Nacim
 Bhatia, Kanwal
 Bhotika, Rahul
 Biesdorf, Andreas
 Bilgic, Berkin
 Bismuth, Vincent
 Blaschko, Matthew
 Bloy, Luke
 Blum, Tobias
 Boctor, Emad
 Bodenstedt, Sebastian
 Bogunovic, Hrvoje
 Boisvert, Jonathan
 Boroczky, Lilla
 Bosch, Johan
 Bouarfa, Loubna
 Bouix, Sylvain
 Bourgeat, Pierrick
 Brady, Michael
 Bria, Alessandro
 Brost, Alexander
 Buelow, Thomas
 Butakoff, Constantine
 Caan, Matthan
 Cahill, Nathan
 Cai, Weidong
 Camara, Oscar
 Cao, Kunlin
 Cardenes, Ruben
 Cardoso, Manuel Jorge
 Carmichael, Owen
 Caruyer, Emmanuel
 Castañeda, Victor
 Castro-Gonzalez, Carlos
 Cater, John
 Cattin, Philippe C.
 Cebal, Juan
 Celebi, M. Emre
 Cetingul, Hasan Ertan
 Chakravarty, M. Mallar
 Chan, Raymond
 Chefd'hotel, Christophe
 Chen, Ting
 Chen, Chao
 Chen, George
 Chen, Xinjian
 Chen, Elvis C. S.
 Chen, Thomas Kuiran
 Chen, Terrence
 Cheng, Jian
 Cheriet, Farida
 Chinzei, Kiyoyuki
 Chitphakdithai, Nicha
 Chou, Yiyu
 Chowdhury, Ananda
 Christensen, Gary
 Chung, Moo
 Cifor, Amalia
 Cimen, Serkan
 Cinquin, Philippe
 Ciuciu, Philippe
 Clarkson, Matthew
 Clarysse, Patrick
 Clouchoux, Cédric
 Cobzas, Dana
 Colliot, Olivier
 Commowick, Olivier
 Cook, Philip
 Corso, Jason
 Costa, Maria
 Coulon, Olivier
 Counsell, Serena J.
 Coupe, Pierrick
 Cowan, Brett
 Crimi, Alessandro
 Crum, William
 Cui, Xinyi
 Cuingnet, Remi
 Daducci, Alessandro
 Daga, Pankaj

Dahl, Anders L.
Darkner, Sune
Dauguet, Julien
David, Liu
De Craene, Mathieu
De Raedt, Sepp
Dehghan, Ehsan
Deligianni, Fani
DeLong, Andrew
Demiralp, Cagatay
Demirci, Stefanie
Deng, Xiang
Dequidt, Jeremie
Descoteaux, Maxime
Desvignes, Michel
Dibella, Edward
Diciotti, Stefano
Dijkstra, Jouke
Dimaio, Simon
Ding, Kai
Donner, René
Douiri, Abdel
Dowling, Jason
Doyle, Scott
Drechsler, Klaus
Du, Yuhui
Duan, Qi
Duchateau, Nicolas
Duchesnay, Edouard
Duchesne, Simon
Dufour, Pascal
Duriez, Christian
Durreleman, Stanley
Dzyubachyk, Oleh
Ecabert, Olivier
Egger, Jan
Ehrhardt, Jan
El-Baz, Ayman
Elen, An
Elliott, Colm
Elson, Daniel
Ennis, Daniel
Enquobahrie, Andinet
Erdt, Marius
Eskildsen, Simon
Eslami, Abouzar
Essert, Caroline
Fahmi, Rachid
Fallavollita, Pascal
Fan, Yong
Farag, Aly
Fedorov, Andriy
Fei, Baowei
Fenster, Aaron
Figl, Michael
Figuroa, C. Alberto
Fishbaugh, James
Fitzpatrick, J Michael
Florack, Luc
Fogtmann, Mads
Fonov, Vladimir
Forestier, Germain
Foroughi, Pezhman
Fouard, Celine
Freiman, Moti
Freysinger, Wolfgang
Friman, Ola
Fripp, Jurgen
Frouin, Vincent
Fua, Pascal
Funka-Lea, Gareth
Fuster, Andrea
Gangeh, Mehrdad
Ganz, Melanie
Gao, Mingchen
Gao, Wei
Gao, Yaozong
Garcia-Lorenzo, Daniel
Garyfallidis, Eleftherios
Gaser, Christian
Georgescu, Bogdan
Ghanbari, Yasser
Gholipour, Ali
Ghosh, Aurobrata
Giannarou, Stamatia
Gibson, Eli
Giger, Maryellen
Gilles, Benjamin
Gilson, Wesley
Ginsburg, Shoshana

Gobbi, David
 Goh, Alvina
 Goksel, Orcun
 Gonzalez Ballester, Miguel Angel
 Gooya, Ali
 Gorospe, Giann
 Graham, Jim
 Gramfort, Alexandre
 Gray, Katherine
 Grbic, Sasa
 Guerrero, Julian
 Guetter, Christoph
 Gulsun, Mehmet Akif
 Gupta, Aditya
 Gur, Yaniv
 Gutman, Boris
 Guye, Maxime
 Hacihaliloglu, Ilker
 Haeck, Tom
 Haeffele, Ben
 Hager, Gregory D
 Hahn, Horst
 Hajnal, Joseph
 Haldar, Justin
 Hamamci, Andac
 Hamarneh, Ghassan
 Hamm, Jihun
 Hanaoka, Shouhei
 Haneishi, Hideaki
 Hanson, Dennis
 Hao, Xiang
 Harders, Matthias
 Hatt, Chuck
 Haynor, David
 He, Huiguang
 Heckemann, Rolf
 Heese, Harald
 Heinrich, Mattias Paul
 Heldmann, Stefan
 Hernandez, Monica
 Hinkle, Jacob
 Hipwell, John
 Hirano, Yasushi
 Holmes, David
 Hong, Jaesung
 Hong, Byung-Woo
 Honnorat, Nicolas
 Hontani, Hidekata
 Howe, Robert
 Hu, Mingxing
 Hu, Zhihong
 Hu, Yipeng
 Huang, Heng
 Huang, Xiaolei
 Huang, Junzhou
 Huisman, Henkjan
 Hyde, Damon
 Iglesias, Juan Eugenio
 Ingallhalikar, Madhura
 Ionasec, Razvan
 Isgum, Ivana
 Jagadeesan, Jayender
 Jain, Aastha
 Jain, Saurabh
 Janoos, Firdaus
 Janowczyk, Andrew
 Jbabdi, Saad
 Jian, Bing
 Jiang, Yifeng
 Johnson, Hans
 Jomier, Julien
 Jordan, Petr
 Joshi, Anand
 Joshi, Sarang
 Joung, Sanghyun
 Kabus, Sven
 Kachelrieß, Marc
 Kaden, Enrico
 Kadoury, Samuel
 Kahl, Fredrik
 Kainmueller, Dagmar
 Kang, Xin
 Kapoor, Ankur
 Kapur, Tina
 Karamalis, Athanasios
 Karimaghhaloo, Zahra
 Kataoka, Hiroyuki
 Katouzian, Amin
 Kazanzides, Peter
 Keeve, Erwin

Kerckhoffs, Roy
Kerrien, Erwan
Khalvati, Farzad
Khan, Ali R.
Khurd, Parmeshwar
Kim, Minjeong
Kim, Boklye
Kim, Kio
Kindlmann, Gordon
Kirchberg, Klaus
Kirisli, Hortense
Kitasaka, Takayuki
Klein, Martina
Klein, Tassilo
Klein, Stefan
Klinder, Tobias
Koay, Cheng
Kobayashi, Yo
Kohlberger, Timo
Komodakis, Nikos
Konukoglu, Ender
Krieger, Axel
Krissian, Karl
Kruggel, Frithjof
Kumar, Rajesh
Kumar, Ankur
Kumar, Ritwik
Kunz, Manuela
Kurkure, Uday
Kwok, Ka-Wai
Kwon, Dongjin
Ladikos, Alexander
Lalys, Florent
Landman, Bennett
Langs, Georg
Lapeer, Rudy
Laporte, Catherine
Lartizien, Carole
Lasser, Tobias
Lasso, Andras
Lauze, Francois
Law, Max W.K.
Lecoeur, Jeremy
Ledesma-Carbayo, Maria-J
Ledig, Christian
Lee, George
Lee, Tim
Lee, Su-Lin
Lee, Junghoon
Lefèvre, Julien
Lekadir, Karim
Lelieveldt, Boudewijn
Lenglet, Christophe
Lensu, Lasse
Lepore, Natasha
Leung, Kelvin
Li, Chunming
Li, Ying
Li, Hongsheng
Li, Ming
Li, Yang
Li, Kaiming
Li, Fuhai
Li, Bo
Li, Gang
Liao, Shu
Liao, Rui
Liao, Jun
Lin, Ming
Linte, Cristian
Litjens, Geert
Liu, Huafeng
Liu, Sidong
Liu, Xiaoxiao
Liu, Jianfei
Liu, Xiaofeng
Liu, Manhua
Liu, Meizhu
Lo, Pechin
Loew, Murray
Lombaert, Herve
Loog, Marco
Lorenzi, Marco
Lu, Le
Lu, Xiaoguang
Lu, Chao
Luboz, Vincent
Lucas, Blake
Lueth, Tim
Lui, Lok Ming

Luo, Xiongbiao
Lézoray, Olivier
Ma, Burton
Machiraju, Raghu
Mackay, Alex
Maddah, Mahnaz
Maduskar, Pragnya
Magee, Derek
Mahdavi, Seyedeh Sara
Maier-Hein (né Fritzsche), Klaus H.
Maier-Hein, Lena
Major, David
Majumdar, Angshul
Makram-Ebeid, Sherif
Malandain, Gregoire
Manduca, Armando
Manjon, Jose V.
Manniesing, Rashindra
Mansi, Tommaso
Marchal, Maud
Mariottini, Gian Luca
Marrakchi-Kacem, Linda
Marsland, Stephen
Martin-Fernandez, Marcos
Martinez-Perez, Elena
Martí, Robert
Mateus, Diana
Matsumiya, Kiyoshi
Mattes, Julian
Maurel, Pierre
Mcclelland, Jamie
McCormick, Matthew
Medrano-Gracia, Pau
Mehrabian, Hatef
Meier, Dominik
Meinzer, Hans-Peter
Melbourne, Andrew
Menze, Bjoern
Merlet, Sylvain
Mertzaniidou, Thomy
Metz, Coert
Meyer, Chuck
Meyer, Francois
Michailovich, Oleg
Michel, Fabrice
Miga, Michael
Miller, James
Miller, Karol
Mirota, Daniel
Modat, Marc
Modersitzki, Jan
Mohamed, Ashraf
Momayyez, Parya
Montiel, J.M. Martiínez
Montillo, Albert
Morooka, Ken'ichi
Mory, Benoit
Mountney, Peter
Mousavi, Zahra
Mousavi, Parvin
Mozer, Pierre
Mueller, Susanne
Murgasova, Maria
Murphy, Keelin
Mylonas, George
Müller, Henning
Nageotte, Florent
Najman, Laurent
Napel, Sandy
Nappi, Janne
Narayana, Ponnada
Natarajan, Shyam
Negahdar, Mohammadreza
Neumuth, Thomas
Ng, Bernard
Niaf, Emilie
Nichols, Thomas
Nickisch, Hannes
Nicolau, Stephane
Nie, Jingxin
Niederer, Steven
Niethammer, Marc
Nikou, Christophoros
Nir, Guy
Noble, Jack
Noblet, Vincent
Nolte, Lutz
Nordsletten, David
Novak, Carol
O'Donnell, Thomas

O'Donnell, Lauren
 Oda, Masahiro
 Oguz, Ipek
 Okada, Toshiyuki
 Okada, Kazunori
 Okur, Asli
 Olabbarriaga, Silvia
 Oliver, Arnau
 Onogi, Shinya
 Oost, Elco
 Oshinski, John
 Otake, Yoshito
 Ou, Yangming
 Ozarslan, Evren
 Padfield, Dirk
 Palaniappan, Kannappan
 Pallavaram, Srivatsan
 Panagiotaki, Eleftheria
 Paniagua, Beatriz
 Papademetris, Xenios
 Papadopoulo, Theo
 Parisot, Sarah
 Park, Jinhyeong
 Park, Mi-Ae
 Passat, Nicolas
 Patriciu, Alexandru
 Paul, Perrine
 Paulsen, Rasmus
 Pauly, Olivier
 Payne, Christopher
 Pearlman, Paul
 Pedemonte, Stefano
 Penney, Graeme
 Pernus, Franjo
 Peter, Loic
 Peterlik, Igor
 Peters, Jochen
 Petersen, Jens
 Petitjean, Caroline
 Peyrat, Jean-Marc
 Pham, Dzung
 Pike, Bruce
 Pitiot, Alain
 Piuze, Emmanuel
 Pizer, Stephen
 Platel, Bram
 Poignet, Philippe
 Poline, Jean-Baptiste
 Polzehl, Joerg
 Poot, Dirk
 Pop, Mihaela
 Poynton, Clare
 Pozo, Jose Maria
 Prasad, Gautam
 Prastawa, Marcel
 Pratt, Philip
 Prevost, Raphael
 Prevrhal, Sven
 Prince, Jerry
 Punithakumar, Kumaradevan
 Qazi, Arish A.
 Qian, Zhen
 Qiu, Anqi
 Quellec, Gwenole
 Qureshi, Hammad
 Radeva, Petia
 Radulescu, Emil
 Rahmatullah, Babbibi
 Rajagopalan, Vidya
 Rajpoot, Nasir
 Ramezani, Mahdi
 Rangarajan, Anand
 Raniga, Parnesh
 Rao, Anil
 Rasouljian, Abtin
 Rathi, Yogesh
 Ray, Nilanjan
 Redouté, Jérôme
 Reichl, Tobias
 Reinertsen, Ingerid
 Reisert, Marco
 Reiter, Austin
 Rettmann, Maryam
 Reuter, Martin
 Reyes-Aldasoro, Constantino
 Reyes, Mauricio
 Rhode, Kawal
 Ribbens, Annemie
 Richa, Rogério
 Riddell, Cyril

Riklin Raviv, Tammy
 Risser, Laurent
 Rit, Simon
 Rittscher, Jens
 Rivaz, Hassan
 Riviere, Denis
 Riviere, Cameron
 Robinson, Emma
 Roche, Alexis
 Roehl, Sebastian
 Rohling, Robert
 Rohr, Karl
 Ropinski, Timo
 Roth, Holger
 Rothgang, Eva
 Roux, Ludovic
 Roysam, Badrinath
 Rueda, Sylvia
 Russakoff, Daniel
 Rusu, Mirabela
 Saalbach, Axel
 Sadeghi-Naini, Ali
 Salvado, Olivier
 San Jose Estepar, Raul
 Sanchez, Clarisa
 Sarrut, David
 Savadjiev, Peter
 Schaap, Michiel
 Scherrer, Benoit
 Schneider, Caitlin
 Schultz, Thomas
 Schweikard, Achim
 Seiler, Christof
 Sermesant, Maxime
 Seshamani, Sharmishta
 Shah, Shishir
 Shamir, Reuben R
 Shekhovtsov, Alexander
 Shen, Tian
 Shen, Li
 Shi, Yundi
 Shi, Feng
 Shi, Kuangyu
 Shi, Wenzhe
 Shi, Yonggang
 Shi, Pengcheng
 Shi, Yonghong
 Simpson, Amber
 Simpson, Ivor
 Singanamalli, Asha
 Singh, Nikhil
 Singh, Vikas
 Sinkus, Ralph
 Slabaugh, Greg
 Smal, Ihor
 Smeets, Dirk
 Sofka, Michal
 Soler, Luc
 Sommer, Stefan
 Song, Xubo
 Song, Gang
 Sotiras, Aristeidis
 Sparks, Rachel
 Sporning, Jon
 Staring, Marius
 Staroswiecki, Ernesto
 Stauder, Ralf
 Stehle, Thomas
 Stewart, James
 Stolka, Philipp
 Styles, Iain
 Subramanian, Navneeth
 Suetens, Paul
 Suinesiaputra, Avan
 Suk, Heung-II
 Summers, Ronald
 Sundar, Hari
 Suzuki, Kenji
 Swanson, Kristin
 Syeda-Mahmood, Tanveer
 Sznitman, Raphael
 Sørensen, Lauge
 Tahmasebi, Amir
 Taimouri, Vahid
 Talbot, Hugues
 Tan, Tao
 Tanner, Christine
 Tao, Xiaodong
 Taquet, Maxime
 Taron, Maxime

Tasdizen, Tolga
Taylor, Zeike
Thielemans, Kris
Thienphrapa, Paul
Thiriet, Marc
Thompson, Chris
Tiwari, Pallavi
Toews, Matthew
Tohka, Jussi
Tokuda, Junichi
Tomas Fernandez, Xavier
Tosun, Duygu
Toth, Robert
Totz, Johannes
Toussaint, Nicolas
Tristán-Vega, Antonio
Tsoumpas, Charalampos
Tu, Zhuowen
Tunc, Birkan
Turkheimer, Federico
Tustison, Nicholas
Twining, Carole
Türetken, Engin
Ukwatta, Eranga
Ullrich, Sebastian
Unal, Gozde
Unay, Devrim
Ungi, Tamas
Uzunbas, Mustafa
Van Assen, Hans
Van Der Laak, Jeroen
Van Rikxoort, Eva
Van Stralen, Marijn
Van Vliet, Lucas J.
Van Walsum, Theo
Vannier, Michael
Varoquaux, Gael
Veenland, Jifke
Venkataraman, Archana
Vercauteren, Tom
Veta, Mtiko
Vialard, Francois-Xavier
Vidal, Camille
Vignon, Francois
Villard, Pierre-Frederic
Visentini-Scarzanella, Marco
Visvikis, Dimitris
Viswanath, Satish
Vitanovski, Dime
Vogel, Jakob
Vogelstein, Joshua
Voigt, Ingmar
Von Berg, Jens
Voros, Sandrine
Vos, Frans
Vos, Pieter
Vosburgh, Kirby
Vrooman, Henri
Vrtovec, Tomaz
Waechter-Stehle, Irina
Waelkens, Paulo
Wahle, Andreas
Wan, Tao
Wang, Haibo
Wang, Zhijie
Wang, Li
Wang, Qian
Wang, Song
Wang, Lichao
Wang, Liansheng
Wang, Yalin
Wang, Chaohui
Wang, Lejing
Wang, Peng
Wang, Zhimin
Wang, Hongzhi
Ward, Aaron
Wassermann, Demian
Weber, Frank Michael
Wee, Chong-Yaw
Wei, Liu
Weller, Daniel
Wels, Michael
Werner, Rene
Wesarg, Stefan
Whitaker, Ross
Whittingstall, Kevin
Wiemker, Rafael
Wiles, Andrew
Witz, Jean-François

Wolf, Ivo
Wolz, Robin
Wright, Graham
Wu, Xiaodong
Wu, Guorong
Wuenschel, Burkhard
Wörz, Stefan
Xie, Yuchen
Xie, Hua
Xie, Jun
Xiong, Guanglei
Xu, Lei
Xu, Sheng
Xu, Rui
Xu, Jun
Xue, Zhong
Yamashita, Hiromasa
Yan, Pingkun
Yang, Lin
Yankam Njiwa, Josiane A.
Yao, Jianhua
Yap, Pew-Thian
Yaqub, Mohammad
Ye, Dong Hye
Yendiki, Anastasia
Yeniaras, Erol
Yeo, B.T. Thomas
Yigitsoy, Mehmet
Yin, Zhaozheng
Yoo, Terry
Yoshida, Hiro
Young, Jonathan
Yushkevich, Paul
Zagorchev, Lyubomir
Zaidi, Habib
Zappella, Luca
Zawadzki, Robert
Zeng, Wei
Zerubia, Josiane
Zhan, Liang
Zhan, Yiqiang
Zhang, Jingya
Zhang, Shaoting
Zhang, Li
Zhang, Daoqiang
Zhang, Weidong
Zhang, Pei
Zhang, Hui
Zhao, Tao
Zhao, Qian
Zheng, Yefeng
Zheng, Yuanjie
Zhong, Hua
Zhou, X. Sean
Zhou, S. Kevin
Zhou, Yan
Zhou, Kevin
Zhou, Luping
Zhou, Jinghao
Zhu, Hongtu
Zhu, Ning
Zhu, Dajiang
Zhuang, Xiahai
Zollei, Lilla
Zosso, Dominique
Zuluaga, Maria A.
Zwiggelaar, Reyer

Awards Presented at MICCAI 2012, Nice, France

MICCAI Society Enduring Impact Award: The Enduring Impact Award is the highest award of the MICCAI Society. It is a career award for continued excellence in the MICCAI research field. The 2012 Enduring Impact Award was presented to *Jerry Prince*, Johns Hopkins University, 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. In 2012, fellowships were awarded to:

- *Alison Noble* (Oxford University, UK)
- *Wiro Niessen* (Erasmus Medical Centre, The Netherlands)
- *Nassir Navab* (Technical University of Munich, Germany)

Medical Image Analysis Journal Award Sponsored by Elsevier: *Benoit Scherrer*, for his paper entitled “Super-Resolution Reconstruction to Increase the Spatial Resolution of Diffusion Weighted Images from Orthogonal Anisotropic Acquisitions”, authored by Benoit Scherrer, Ali Gholipour and Simon K. Warfield.

Best Paper in Computer-Assisted Intervention Systems and Medical Robotics: *Benjamin Bejar* for his paper entitled “Surgical Gesture Classification from Video Data”, authored by Benjamin Bejar, Luca Zappella, Rene Vidal.

Young Scientist Publication Impact Award Sponsored by Kitware Inc.: MICCAI papers by a young scientist from the past 5 years were eligible for this award. It is made to a researcher whose work had an impact on the MICCAI field in terms of citations, secondary citations, subsequent publications, h-index. The 2012 Young Scientist Publication Impact Award was given to *Caroline Brun*: “A Tensor-Based Morphometry Study of Genetic Influences on Brain Structure using a New Fluid Registration Method” authored by C. Brun, N. Lepore, X. Pennec, Y.-Y. Chou, K. McMahon, G.I. de Zubicaray, M. Meredith, M.J. Wright, A.D. Lee, M. Barysheva, A.W. Toga, P.M. Thompson.

MICCAI Young Scientist Awards: The Young Scientist Awards are stimulation prizes awarded for the best first authors of MICCAI contributions in distinct subject areas. The nominees had to be full-time students at a recognized university at, or within, two years prior to submission. The 2012 MICCAI Young Scientist Awards were given to:

- *Hang Su* for his paper entitled: “Phase Contrast Image Restoration Via Dictionary Representation of Diffraction Patterns”, authored by Hang Su, Zhaozheng Yin, Takeo Kanade, and Seungil Huh

- *Eli Gibson*, for his paper entitled: “Registration Accuracy: How Good is Good Enough? A Statistical Power Calculation Incorporating Image Registration Uncertainty”, authored by Eli Gibson, Aaron Fenster and Aaron D. Ward
- *Stephanie Marchesseau* for her paper entitled: “Cardiac Mechanical Parameter Calibration Based on the Unscented Transform”, authored by Stephanie Marchesseau, Herve Delingette, Maxime Sermesant, Kawal Rhode, Simon G. Duckett, C. Aldo Rinaldi, Reza Razavi, and Nicholas Ayache
- *Roland Kwitt* for his paper entitled: “Recognition in Ultrasound Videos: Where am I?”, authored by Roland Kwitt, Nuno Vasconcelos, Sharif Razaque, and Stephen Aylward
- *Robin Wolz*, for his paper entitled: “Multi-Organ Abdominal CT Segmentation Using Hierarchically Weighted Subject-Specific Atlases”, authored by Robin Wolz, Chengwen Chu, Kazunari Misawa, Kensaku Mori, Daniel Rueckert

Table of Contents – Part III

Image Reconstruction and Motion Modeling

Improved Multi B-Value Diffusion-Weighted MRI of the Body by Simultaneous Model Estimation and Image Reconstruction (SMEIR)	1
<i>Moti Freiman, Onur Afacan, Robert V. Mulkern, and Simon K. Warfield</i>	

Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch	9
<i>Wenzhe Shi, Jose Caballero, Christian Ledig, Xiahai Zhuang, Wenjia Bai, Kanwal Bhatia, Antonio M Simoes Monteiro de Marvao, Tim Dawes, Declan O'Regan, and Daniel Rueckert</i>	

Self-gated Radial MRI for Respiratory Motion Compensation on Hybrid PET/MR Systems	17
<i>Robert Grimm, Sebastian Fürst, Isabel Dregely, Christoph Forman, Jana Maria Hutter, Sibylle I. Ziegler, Stephan Nekolla, Berthold Kiefer, Markus Schwaiger, Joachim Hornegger, and Tobias Block</i>	

Complex Lung Motion Estimation via Adaptive Bilateral Filtering of the Deformation Field	25
<i>Bartłomiej W. Papież, Mattias Paul Heinrich, Laurent Risser, and Julia A. Schnabel</i>	

Helical Mode Lung 4D-CT Reconstruction Using Bayesian Model	33
<i>Tiancheng He, Zhong Xue, Paige L. Nitsch, Bin S. Teh, and Stephen T. Wong</i>	

3D Tongue Motion from Tagged and Cine MR Images	41
<i>Fangxu Xing, Jonghye Woo, Emi Z. Murano, Junghoon Lee, Maureen Stone, and Jerry L. Prince</i>	

Machine Learning in Medical Image Computing

Learning without Labeling: Domain Adaptation for Ultrasound Transducer Localization	49
<i>Tobias Heimann, Peter Mounthey, Matthias John, and Razvan Ionasec</i>	

Segmentation of 4D Echocardiography Using Stochastic Online Dictionary Learning	57
<i>Xiaojie Huang, Donald P. Dione, Ben A. Lin, Alda Bregasi, Albert J. Sinusas, and James S. Duncan</i>	
Atlas Encoding by Randomized Forests for Efficient Label Propagation	66
<i>Darko Zikic, Ben Glocker, and Antonio Criminisi</i>	
Robust and Accurate Coronary Artery Centerline Extraction in CTA by Combining Model-Driven and Data-Driven Approaches	74
<i>Yefeng Zheng, Huseyin Tek, and Gareth Funka-Lea</i>	
Incorporating Shape Variability in Image Segmentation via Implicit Template Deformation	82
<i>Raphael Prevost, Remi Cuingnet, Benoit Mory, Laurent D. Cohen, and Roberto Ardon</i>	
Automatic 3D Motion Estimation of Left Ventricle from C-arm Rotational Angiocardiology Using a Prior Motion Model and Learning Based Boundary Detector	90
<i>Mingqing Chen, Yefeng Zheng, Yang Wang, Kerstin Mueller, and Guenter Lauritsch</i>	

Imaging, Reconstruction, and Enhancement II

Interventional Digital Tomosynthesis from a Standard Fluoroscopy System Using 2D-3D Registration	98
<i>Mazen Alhrishy, Andreas Varnavas, Tom Carrell, Andrew King, and Graeme Penney</i>	
Calibrationless Parallel MRI with Joint Total Variation Regularization	106
<i>Chen Chen, Yeqing Li, and Junzhou Huang</i>	
Denosing PET Images Using Singular Value Thresholding and Stein’s Unbiased Risk Estimate	115
<i>Ulas Bagci and Daniel J. Mollura</i>	
Super-Resolution Reconstruction Using Cross-Scale Self-similarity in Multi-slice MRI	123
<i>Esben Plenge, Dirk H.J. Poot, Wiro J. Niessen, and Erik Meijering</i>	
Dynamic CT Reconstruction by Smoothed Rank Minimization	131
<i>Angshul Majumdar and Rabab K. Ward</i>	
Harnessing Group-Sparsity Regularization for Resolution Enhancement of Lung 4D-CT	139
<i>Arnav Bhavsar, Guorong Wu, and Dinggang Shen</i>	

Improved Myocardial Scar Characterization by Super-Resolution Reconstruction in Late Gadolinium Enhanced MRI	147
<i>Oleh Dzyubachyk, Qian Tao, Dirk H.J. Poot, Hildo Lamb, Katja Zeppenfeld, Boudewijn P.F. Lelieveldt, and Rob J. van der Geest</i>	
Direct Parametric Image Reconstruction of Rapid Multi-tracer PET	155
<i>Xiaoyin Cheng, Nassir Navab, Sibylle I. Ziegler, and Kuangyu Shi</i>	

Registration II

Non-rigid 2D-3D Medical Image Registration Using Markov Random Fields	163
<i>Enzo Ferrante and Nikos Paragios</i>	
Learning Nonrigid Deformations for Constrained Multi-modal Image Registration	171
<i>John A. Onofrey, Lawrence H. Staib, and Xenophon Papademetris</i>	
Non-Rigid 2D-3D Registration Using Anisotropic Error Ellipsoids to Account for Projection Uncertainties during Aortic Surgery	179
<i>Alexis Guyot, Andreas Varnavas, Tom Carrell, and Graeme Penney</i>	
A Variational Formulation for Discrete Registration	187
<i>Karteek Popuri, Dana Cobzas, and Martin Jügersand</i>	
FLOOR: Fusing Locally Optimal Registrations	195
<i>Dong Hye Ye, Jihun Hamm, Benoit Desjardins, and Kilian M. Pohl</i>	
Particle-Guided Image Registration	203
<i>Joohee Lee, Ilwoo Lyu, İpek Oğuz, and Martin A. Styner</i>	

Segmentation II

Contour-Driven Regression for Label Inference in Atlas-Based Segmentation	211
<i>Christian Wachinger, Gregory C. Sharp, and Polina Golland</i>	
Discriminative Parameter Estimation for Random Walks Segmentation	219
<i>Pierre-Yves Baudin, Danny Goodman, Puneet Kumar, Noura Azzabou, Pierre G. Carlier, Nikos Paragios, and M. Pawan Kumar</i>	
Fully Automatic X-Ray Image Segmentation via Joint Estimation of Image Displacements	227
<i>Cheng Chen, Weiguo Xie, Jochen Franke, Paul A. Grützner, Lutz-P. Nolte, and Guoyan Zheng</i>	

IntellEditS: Intelligent Learning-Based Editor of Segmentations	235
<i>Adam P. Harrison, Neil Birkbeck, and Michal Sofka</i>	
Automatic Nuchal Translucency Measurement from Ultrasonography . . .	243
<i>JinHyeong Park, Michal Sofka, SunMi Lee, DaeYoung Kim, and S. Kevin Zhou</i>	
Automated Segmentation of CBCT Image Using Spiral CT Atlases and Convex Optimization	251
<i>Li Wang, Ken Chung Chen, Feng Shi, Shu Liao, Gang Li, Yaozong Gao, Steve GF Shen, Jin Yan, Philip K.M. Lee, Ben Chow, Nancy X. Liu, James J. Xia, and Dinggang Shen</i>	
Automatic Analysis of Pediatric Renal Ultrasound Using Shape, Anatomical and Image Acquisition Priors	259
<i>Carlos S. Mendoza, Xin Kang, Nabile Safdar, Emmarie Myers, Aaron D. Martin, Enrico Grisan, Craig A. Peters, and Marius George Linguraru</i>	
Joint Model-Pixel Segmentation with Pose-Invariant Deformable Graph-Priors	267
<i>Bo Xiang, Jean-Francois Deux, Alain Rahmouni, and Nikos Paragios</i>	
Abdominal Multi-organ CT Segmentation Using Organ Correlation Graph and Prediction-Based Shape and Location Priors	275
<i>Toshiyuki Okada, Marius George Linguraru, Masatoshi Hori, Ronald M. Summers, Noriyuki Tomiyama, and Yoshinobu Sato</i>	
Physiological Modeling, Simulation, and Planning II	
Multimodal Image Driven Patient Specific Tumor Growth Modeling . . .	283
<i>Yixun Liu, Samira M. Sadowski, Allison B. Weisbrod, Electron Kebebew, Ronald M. Summers, and Jianhua Yao</i>	
Patient-Specific Biomechanical Modeling of Ventricular Enlargement in Hydrocephalus from Longitudinal Magnetic Resonance Imaging	291
<i>Yasheng Chen, Zheng Fan, Songbai Ji, Joseph Muenzer, Hongyu An, and Weili Lin</i>	
Simulation of Lipofilling Reconstructive Surgery Using Coupled Eulerian Fluid and Deformable Solid Models	299
<i>Vincent Majorczyk, Stéphane Cotin, Christian Duriez, and Jeremie Allard</i>	

Towards a Better Understanding of Pelvic System Disorders Using Numerical Simulation	307
<i>Pauline Lecomte-Grosbras, Mouhamadou Nassirou Diallo, Jean-Francois Witz, Damien Marchal, Jeremie Dequidt, Stéphane Cotin, Michel Cosson, Christian Duriez, and Matthias Brieu</i>	
Constructive Real Time Feedback for a Temporal Bone Simulator	315
<i>Yun Zhou, James Bailey, Ioanna Ioannou, Sudanthi Wijewickrema, Gregor Kennedy, and Stephen O’Leary</i>	
Lattice Boltzmann Method for Fast Patient-Specific Simulation of Liver Tumor Ablation from CT Images	323
<i>Chloé Audigier, Tommaso Mansi, Hervé Delingette, Saikiran Rapaka, Viorel Mihalef, Puneet Sharma, Daniel Carnegie, Emad Boctor, Michael Choti, Ali Kamen, Dorin Comaniciu, and Nicholas Ayache</i>	
Registration of a Validated Mechanical Atlas of Middle Ear for Surgical Simulation	331
<i>Guillaume Kazmitcheff, Christian Duriez, Mathieu Miroir, Yann Nguyen, Olivier Sterkers, Alexis Bozorg Grayeli, and Stéphane Cotin</i>	
Surgical Gesture Segmentation and Recognition	339
<i>Lingling Tao, Luca Zappella, Gregory D. Hager, and René Vidal</i>	
 Intraoperative Guidance and Robotics II 	
Towards Intra-operative OCT Guidance for Automatic Head Surgery: First Experimental Results	347
<i>Jesús Díaz Díaz, Dennis Kundrat, Kim-Fat Goh, Omid Majdani, and Tobias Ortmaier</i>	
Configurable Automatic Detection and Registration of Fiducial Frames for Device-to-Image Registration in MRI-Guided Prostate Interventions	355
<i>Junichi Tokuda, Sang-Eun Song, Kemal Tuncali, Clare Tempany, and Nobuhiko Hata</i>	
Robust Intraoperative US Probe Tracking Using a Monocular Endoscopic Camera	363
<i>Uditha L. Jayarathne, A. Jonathan McLeod, Terry M. Peters, and Elvis C.S. Chen</i>	
Automatic Detection of Multiple and Overlapping EP Catheters in Fluoroscopic Sequences	371
<i>Fausto Milletari, Nassir Navab, and Pascal Fallavollita</i>	

Validation of Catheter Segmentation for MR-Guided Gynecologic Cancer Brachytherapy 380
Guillaume Pernelle, Alireza Mehrtash, Lauren Barber, Antonio Damato, Wei Wang, Ravi Teja Seethamraju, Ehud Schmidt, Robert A. Cormack, Williams Wells, Akila Viswanathan, and Tina Kapur

A Novel High Intensity Focused Ultrasound Robotic System for Breast Cancer Treatment 388
Taizan Yonetsuji, Takehiro Ando, Junchen Wang, Keisuke Fujiwara, Kazunori Itani, Takashi Azuma, Kiyoshi Yoshinaka, Akira Sasaki, Shu Takagi, Etsuko Kobayashi, Hongen Liao, Yoichiro Matsumoto, and Ichiro Sakuma

Microscope, Optical Imaging, and Histology III

Cell Orientation Entropy (CORe): Predicting Biochemical Recurrence from Prostate Cancer Tissue Microarrays 396
George Lee, Sahirzeeshan Ali, Robert Veltri, Jonathan I. Epstein, Christhunesa Christudass, and Anant Madabhushi

Robust Selection-Based Sparse Shape Model for Lung Cancer Image Segmentation 404
Fuyong Xing and Lin Yang

Flash Scanning Electron Microscopy 413
Raphael Sznitman, Aurelien Lucchi, Marco Cantoni, Graham Knott, and Pascal Fua

Superpixel Classification Based Optic Cup Segmentation 421
Jun Cheng, Jiang Liu, Dacheng Tao, Fengshou Yin, Damon Wing Kee Wong, Yanwu Xu, and Tien Yin Wong

Learning from Partially Annotated OPT Images by Contextual Relevance Ranking 429
Wenqi Li, Jianguo Zhang, Wei-Shi Zheng, Maria Coats, Frank A. Carey, and Stephen J. McKenna

Phenotype Detection in Morphological Mutant Mice Using Deformation Features 437
Sharmili Roy, Xi Liang, Asanobu Kitamoto, Masaru Tamura, Toshihiko Shiroishi, and Michael S. Brown

Efficient Reconstruction-Based Optic Cup Localization for Glaucoma Screening 445
Yanwu Xu, Stephen Lin, Damon Wing Kee Wong, Jiang Liu, and Dong Xu

Automatic Detection of Blue-White Veil by Discrete Colour Matching in Dermoscopy Images	453
<i>Ali Madooei, Mark S. Drew, Maryam Sadeghi, and M. Stella Atkins</i>	

Separation of Benign and Malignant Glands in Prostatic Adenocarcinoma	461
<i>Sabrina Rashid, Ladan Fazli, Alexander Boag, Robert Siemens, Purang Abolmaesumi, and Septimiu E. Salcudean</i>	

Diffusion MRI II

A Cross-Sectional Piecewise Constant Model for Segmenting Highly Curved Fiber Tracts in Diffusion MR Images	469
<i>Brian G. Booth and Ghassan Hamarneh</i>	

Improving DTI Resolution from a Single Clinical Acquisition: A Statistical Approach Using Spatial Prior	477
<i>Vikash Gupta, Nicholas Ayache, and Xavier Pennec</i>	

Adaptively Constrained Convex Optimization for Accurate Fiber Orientation Estimation with High Order Spherical Harmonics	485
<i>Giang Tran and Yonggang Shi</i>	

Fiber Continuity Based Spherical Deconvolution in Spherical Harmonic Domain	493
<i>Marco Reisert and Henrik Skibbe</i>	

A 4D Hyperspherical Interpretation of q -space	501
<i>Ameer Pasha Hosseinbor, Moo K. Chung, Yu-Chien Wu, Andrew L. Alexander, and Barbara B. Bendlin</i>	

Diffusion Propagator Estimation from Sparse Measurements in a Tractography Framework	510
<i>Yogesh Rathi, Borjan Gagoski, Kawin Setsompop, Oleg Michailovich, P. Ellen Grant, and Carl-Fredrik Westin</i>	

Characterizing the DIstribution of Anisotropic MicrO-structural eNvironments with Diffusion-Weighted Imaging (DIAMOND)	518
<i>Benoit Scherrer, Armin Schwartzman, Maxime Taquet, Sanjay P. Prabhu, Mustafa Sahin, Alireza Akhondi-Asl, and Simon K. Warfield</i>	

A Generative Model for Resolution Enhancement of Diffusion MRI Data	527
<i>Pew-Thian Yap, Hongyu An, Yasheng Chen, and Dinggang Shen</i>	

Brain Segmentation and Atlases III

Multi-atlas Segmentation without Registration: A Supervoxel-Based Approach	535
<i>Hongzhi Wang and Paul A. Yushkevich</i>	
Adaptive Voxel, Texture and Temporal Conditional Random Fields for Detection of Gad-Enhancing Multiple Sclerosis Lesions in Brain MRI	543
<i>Zahra Karimaghloo, Hassan Rivaz, Douglas L. Arnold, D. Louis Collins, and Tal Arbel</i>	
Minimizing Joint Risk of Mislabeling for Iterative Patch-Based Label Fusion	551
<i>Guorong Wu, Qian Wang, Shu Liao, Daoqiang Zhang, Feiping Nie, and Dinggang Shen</i>	
Lateral Ventricle Segmentation of 3D Pre-term Neonates US Using Convex Optimization	559
<i>Wu Qiu, Jing Yuan, Jessica Kishimoto, Eranga Ukwatta, and Aaron Fenster</i>	
Semi-automatic Brain Tumor Segmentation by Constrained MRFs Using Structural Trajectories	567
<i>Liang Zhao, Wei Wu, and Jason J. Corso</i>	
A Probabilistic, Non-parametric Framework for Inter-modality Label Fusion	576
<i>Juan Eugenio Iglesias, Mert Rory Sabuncu, and Koen Van Leemput</i>	
Weighted Functional Boxplot with Application to Statistical Atlas Construction	584
<i>Yi Hong, Brad Davis, J.S. Marron, Roland Kwitt, and Marc Niethammer</i>	
Bayesian Estimation of Probabilistic Atlas for Anatomically-Informed Functional MRI Group Analyses	592
<i>Hao Xu, Bertrand Thirion, and Stéphanie Allassonnière</i>	

Functional MRI and Neuroscience Applications II

Exhaustive Search of the SNP-SNP Interactome Identifies Epistatic Effects on Brain Volume in Two Cohorts	600
<i>Derrek P. Hibar, Jason L. Stein, Neda Jahanshad, Omid Kohannim, Arthur W. Toga, Katie L. McMahon, Greig I. de Zubicaray, Grant W. Montgomery, Nicholas G. Martin, Margaret J. Wright, Michael W. Weiner, and Paul M. Thompson</i>	

Sparse Representation of Group-Wise fMRI Signals	608
<i>Jinglei Lv, Xiang Li, Dajiang Zhu, Xi Jiang, Xin Zhang, Xintao Hu, Tuo Zhang, Lei Guo, and Tianming Liu</i>	
Anatomy-Guided Discovery of Large-Scale Consistent Connectivity-Based Cortical Landmarks	617
<i>Xi Jiang, Tuo Zhang, Dajiang Zhu, Kaiming Li, Jinglei Lv, Lei Guo, and Tianming Liu</i>	
Sparse Representation of Higher-Order Functional Interaction Patterns in Task-Based fMRI Data	626
<i>Shu Zhang, Xiang Li, Jinglei Lv, Xi Jiang, Dajiang Zhu, Hanbo Chen, Tuo Zhang, Lei Guo, and Tianming Liu</i>	
Fusing Functional Signals by Sparse Canonical Correlation Analysis Improves Network Reproducibility	635
<i>Jeffrey T. Duda, John A. Detre, Junghoon Kim, James C. Gee, and Brian B. Avants</i>	
Multi-resolutional Brain Network Filtering and Analysis via Wavelets on Non-Euclidean Space	643
<i>Won Hwa Kim, Nagesh Adluru, Moo K. Chung, Sylvia Charchut, Johnson J. GadElkarim, Lori Altshuler, Teena Moody, Anand Kumar, Vikas Singh, and Alex D. Leow</i>	
Implications of Inconsistencies between fMRI and dMRI on Multimodal Connectivity Estimation	652
<i>Bernard Ng, Gael Varoquaux, Jean Baptiste Poline, and Bertrand Thirion</i>	
Erratum	
Dynamic CT Reconstruction by Smoothed Rank Minimization	E1
<i>Angshul Majumdar and Rabab K. Ward</i>	
Author Index	661

Improved Multi B-Value Diffusion-Weighted MRI of the Body by Simultaneous Model Estimation and Image Reconstruction (SMEIR)*

Moti Freiman¹, Onur Afacan¹, Robert V. Mulkern², and Simon K. Warfield¹

¹ Computational Radiology Laboratory, Boston Children's Hospital, Harvard Medical School, MA, USA

² Department of Radiology, Boston Children's Hospital, Harvard Medical School, MA, USA

Abstract. Diffusion-weighted MRI images acquired with multiple b-values have the potential to improve diagnostic accuracy by increasing the conspicuity of lesions and inflammatory activity with background suppression. Unfortunately, the inherently low signal-to-noise ratio (SNR) of DW-MRI reduces enthusiasm for using these images for diagnostic purposes. Moreover, lengthy acquisition times limit our ability to improve the quality of multi b-value DW-MRI images by multiple excitations acquisition and signal averaging at each b-value. To offset these limitations, we propose the Simultaneous Model Estimation and Image Reconstruction (SMEIR) for DW-MRI, which substantially improves the quality of multi b-value DW-MRI images *without* increasing acquisition times. Our model introduces the physiological signal decay model of DW-MRI as a constraint in the reconstruction of the DW-MRI images. An in-vivo experiment using 6 low-quality DW-MRI datasets of a healthy subject showed that SMEIR reconstruction of low-quality data improved SNR by 55% in the liver and by 41% in the kidney without increasing acquisition times. We also demonstrated the clinical impact of our SMEIR reconstruction by increasing the conspicuity of inflamed bowel regions in DW-MRI of 12 patients with Crohn's disease. The contrast-to-noise ratio (CNR) of the inflamed regions in the SMEIR images was higher by 12.6% relative to CNR in the original DW-MRI images.

1 Introduction

Diffusion-weighted MRI (DW-MRI) of the body is a non-invasive imaging technique sensitive to the incoherent motion of water molecules inside the area of interest. This motion is characterized by a combination of a slow diffusion component associated primarily with the Brownian motion of water molecules, and a

* This investigation is supported in part by NIH grants R01 LM010033 R42 MH086984 R01 EB008015 UL1 TR000170 and by research grants from the Translational Research Program at Boston Children's Hospital. M.F is also supported in part by a Research Fellow Award from the Crohn's and Colitis Foundation of America.

fast diffusion component associated primarily with the bulk motion of intravascular molecules in the micro-capillaries [8].

The signal in DW-MRI images decays as a function of the amount of incoherent motion present in the tissue and a diffusion-weighting parameter known as “b-value.” The rapid DW-MRI signal decay in tissue with normal diffusion characteristics, combined with the reduced signal decay in regions of restricted diffusion, increases the conspicuity of abnormal regions in DW-MRI images acquired with sufficiently high b-value and aids in detecting abnormal regions.

Recent studies demonstrate the potential of sufficiently high b-value DW-MRI images to improve the detection rate of different types of carcinoma [4,5,11], focal hepatic lesions [6], and inflammatory activity in the bowel [9] without ionized radiation and/or exogenous contrast media.

Unfortunately, DW-MRI images have an inherently low signal-to-noise ratio (SNR), which reduces enthusiasm for using these images for diagnostic purposes. Koh and Collins [7] recommend acquiring the DW-MRI data with multiple excitations (i.e., 5 to 6 excitations) and using the averaged signal to improve DW-MRI image quality. However, acquiring *multi b-value* DW-MRI data with multiple excitations to achieve both sufficient image quality and sufficient information for quantitative assessment of fast and slow diffusion will substantially increase the overall acquisition time, thus making this method less suitable for routine clinical use.

A unique feature in DW-MRI images in particular, and in parametric imaging techniques in general, is the addition of a 4th dimension to the control parameters in the acquired data. This 4th parametric dimension, which is the diffusion-weighting factor (b-value) at play in DW-MRI, can be exploited as an additional source of prior information that can be utilized in reconstructing images.

Several groups have suggested incorporating the 4th parametric dimension as a constraint to increase the quality of reconstructed images in quantitative T1 and T2 images [1,2,12]. However, these models are difficult to optimize and have not been successfully applied to body DW-MRI reconstruction.

In this work, we reduce the number of excitations required to obtain multi b-value DW-MRI images of the body with sufficient SNR by introducing a Bayesian model of the expected signal with the signal decay model utilized as the prior information. With this model, we are able to simultaneously obtain high-quality DW-MRI images for *multiple* b-values at once and estimate the signal decay model parameter values instead of generating estimates of each b-value image independently and without estimating the signal decay model parameter values.

We also introduce an efficient iterative solution based on the Expectation-Maximization framework through which we obtain high SNR DW-MRI images and parameter estimates with the “Simultaneous Model Estimation and Image Reconstruction” (SMEIR) solver, a novel approach that simultaneously estimates the intra-voxel incoherent motion signal decay model parameters and reconstructs high-quality DW-MRI images. In our experiments, we have shown substantial improvements in DW-MRI image quality. Specifically, our SMEIR reconstruction approach improves the SNR of 6 DW-MRI datasets of a healthy

volunteer by 55% in the liver and by 41% in the kidney. We have also assessed clinical impact, namely, by demonstrating increased conspicuity of inflammatory bowel regions in a study cohort of 12 pediatric Crohn’s disease patients. Our results show that SMEIR-reconstructed images have a contrast-to-noise ratio that is 12.6% higher than the CNR produced by the original DW-MRI data.

2 Method

2.1 DW-MRI Reconstruction Model

DW-MRI images have an inherently low SNR, which reduces enthusiasm for using these images for diagnostic purposes. Specifically, the SNR of DW-MRI images decreases as the b-values used to acquire the images increases. As a rule, the SNR of DW-MRI images obtained with high b-values is increased by acquiring the images with multiple excitations at each b-value. An SNR-optimized DW-MRI image is then reconstructed by solving a Maximum-Likelihood estimation problem for each b-value independently:

$$\hat{S}_i = \underset{S_i}{\operatorname{argmin}} \sum_{j=1}^M (S_i - S'_{i,j})^2 \quad (1)$$

where $S'_{i,j}$ is the observed signal at excitation j with b-value b_i , S_i is the unknown signal, and M is the number of excitations. The solution is simply obtained by averaging the signal at the different excitations:

$$\hat{S}_i = \frac{1}{M} \sum_{j=1}^M S'_{i,j} \quad (2)$$

Unfortunately, by obtaining images with M as the number of excitations (NEX), we are only able to achieve a SNR increase of \sqrt{M} , the low rate of which requires acquisitions of long duration.

As an alternative, we propose to incorporate a prior knowledge on signal evolution in the b-value dimension to reconstruct the DW-MRI images by adding a regularization term to Eq. 1:

$$\hat{S}_i = \underset{S_i}{\operatorname{argmin}} \sum_{j=1}^M (S_i - S'_{i,j})^2 + \alpha (S_i - f(\Theta, i))^2 \quad (3)$$

Where α is the regularization weighting parameter, $f(\Theta, i)$ is the expected signal at b-value b_i given the signal decay model parameters Θ . In this work we used the intra-voxel incoherent motion (IVIM) model of DW-MRI signal decay proposed by Le Bihan et al. [8]. This model assumes a signal decay function of the form:

$$f(\Theta, i) = S_0 (f \exp(-b_i(D + D^*)) + (1 - f)(\exp(-b_i D))) \quad (4)$$

Where $f(\Theta, i)$ is the expected signal at b-value b_i , $\Theta = \{s_0, f, D^*, D\}$ are the intra-voxel incoherent motion model parameters describing the baseline (i.e., without any diffusion effect) signal (s_0); the fast-diffusion fraction (f); the fast-diffusion coefficient (D^*) characterizing primarily the bulk motion of intravascular molecules in the micro-capillaries; and the slow-diffusion coefficient (D) characterizing primarily the Brownian motion of water molecules in the extra-cellular space.

Unfortunately, the expected signal is dependent on the parameters of the signal decay model (i.e., Θ) which are unknown. Therefore, we cannot optimize Eq. 3 directly. Instead, we formulate the reconstruction problem as a simultaneous reconstruction of the DW-MRI images acquired with different b-values and an estimation of the signal decay model:

$$[\hat{S}, \hat{\Theta}] = \underset{S, \Theta}{\operatorname{argmin}} \sum_{i=1}^N \sum_{j=1}^M (S_i - S'_{i,j})^2 + \alpha \sum_{i=1}^N (S_i - f(\Theta, i))^2 \quad (5)$$

where $S = \{S_1, \dots, S_N\}$.

2.2 Optimization Scheme

We used an Expectation-Maximization-like approach to solve Eq. 5 by iteratively estimating the signal decay model parameters Θ given the current estimate of signal S and then estimating signal S given the current estimate of the model parameters Θ . We describe these steps in detail next.

E-Step: Signal decay model (Θ) estimation:

The expected DW-MRI signal decay at each voxel is described by the intra-voxel incoherent motion model (Eq. 4). We used the spatially constrained incoherent motion signal decay model described by Freiman et al. [3] to robustly estimate the model parameters Θ . Given the current estimate of the DW-MRI signal S^t , the estimate of the model parameters Θ^t was obtained by minimizing:

$$\hat{\Theta}^t = \underset{\Theta^t}{\operatorname{argmin}} \sum_{i=1}^N (S_i^t - f(\Theta^t, i))^2 + \sum_{v_p \sim v_q} g(\Theta_{v_p}^t, \Theta_{v_q}^t) \quad (6)$$

where $g(\cdot, \cdot)$ is the spatial constraint given by:

$$g(\Theta_{v_p}^t, \Theta_{v_q}^t) = \alpha W |\Theta_{v_p}^t - \Theta_{v_q}^t| \quad (7)$$

and $\alpha \geq 0$ is the spatial coupling factor; W is a diagonal weighting matrix that accounts for the different scales of the parameters in Θ ; and v_p, v_q are the neighboring voxels established by the employed neighborhood system utilized in this model. We then estimated the model parameters Θ by minimizing Eq. 6 using the ‘‘fusion bootstrap moves’’ combinatorial solver introduced by Freiman et al. [3].

M-Step: DW-MRI Signal (S) reconstruction:

Given the current estimate of the DW-MRI signal decay model Θ^t , we minimized Eq. 5 using the BOBYQA non-linear optimization algorithm [10] to get the next estimate of the signal S^{t+1} .

First, we initialized the algorithm with the acquired DW-MRI data as the current estimate of the signal; and second, using the acquired DW-MRI data and the current estimate of the model parameters, we iteratively alternated between estimating the model parameters from the current signal (Eq. 6) and estimating the true DW-MRI signal (Eq. 5) until the change in the estimated signal became negligible.

3 Experimental Results

3.1 *In-vivo* Evaluation Using Healthy Control

We conducted an *in-vivo* study with healthy control volunteer data to analyze the improvement in SNR achieved by using our SMEIR approach.

We acquired DW-MRI data of a healthy volunteer using a 1.5-T unit (Magnetom Avanto, Siemens Medical Solutions, Erlangen, Germany). We performed free-breathing single-shot echo-planar imaging using the following parameters: repetition/echo time (TR/TE) = 6800/59 ms; matrix size = 192×156; field of view = 300×260 mm; slice thickness/gap = 5 mm/0.5 mm; 40 axial slices; 7 b-values = 0,50,100,200,400,600,800 s/mm² with 6 excitations (i.e. NEX=6). The acquisition time for each excitation was 3:30 min. with an overall acquisition time of 21 min.

We generated high-quality images by averaging the data from the 6 excitations and 6 low-quality datasets - each one consisting of data acquired with 1 NEX. For each low-quality dataset, we reconstructed the images using our SMEIR reconstruction approach. We experimentally set the value of α in Eq. 5 to 0.01. The average (std) running time required to reconstruct DW-MRI images of 256 × 256 × 40 voxels on an 8 processors machine Intel[®] Xeon[®] at 2.40 GHz with cache size of 12 MB and overall memory of 48 GB using the SMEIR reconstruction was 8:04 (2:35) min.

For purposes of evaluation, we defined 2 spherical regions of interest (ROI) in the liver and in the kidney, respectively. We defined SNR at each voxel as the average signal over the 6 low-quality datasets divided by the standard deviation of the signal over these datasets. We likewise calculated SNR for the b-value=800 s/mm² DW-MRI image for the raw low-quality datasets (RAW) and for the SMEIR-reconstructed datasets (SMEIR). Next, we averaged SNR for the RAW and SMEIR data, respectively, over the liver and kidney ROIs.

Fig. 1 presents a high b-value (i.e. 800 s/mm²) image acquired in high-quality (NEX=6); in low-quality (NEX=1); in low-quality (NEX=1) combined with SMEIR reconstruction; and a bar-plot representation of SNR of low-quality data *with and without* SMEIR reconstruction. The average±std SNR of SMEIR data (12.2±2.5 in the liver and 11±2.5 in the kidney) was higher than the SNR of the low-quality data (7.9±4.2 in the liver and 7.8±2.7 in the kidney) - a difference

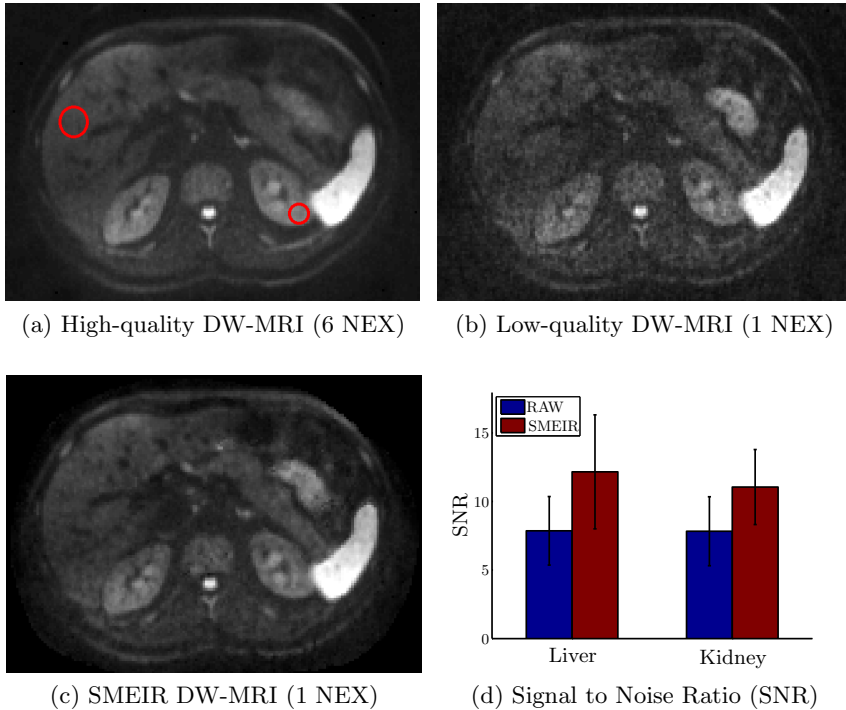


Fig. 1. *In-vivo* example. This figure presents b -value= 800 s/mm^2 images acquired in high-quality (NEX=6); in low-quality (NEX=1); and in low-quality (NEX=1) combined with SMEIR reconstruction, and a bar-plot representation of SNR of low-quality data *with and without* SMEIR reconstruction.

that was statistically significant (Paired Student’s t -test, $p < 0.0001$). Notably, the SMEIR reconstruction of low-quality data improved SNR by 55% in the liver and by 41% in the kidney *without* additional acquisition time.

3.2 Clinical Impact

To demonstrate the actual clinical impact of using our SMEIR reconstruction approach instead of the raw low-quality DW-MRI data, we assessed the conspicuity of inflamed bowel regions in b -value= 800 s/mm^2 images by means of contrast-to-noise ratio (CNR) between regions with active inflammation and surrounding normal tissues in DW-MRI data of Crohn’s disease patients.

We retrospectively reviewed DW-MRI data of 30 patients who underwent clinical MRI exams including a MR enterography (MRE) protocol that included polyethylene glycol administration for bowel distention; gadolinium-enhanced, dynamic 3D VIBE (volume-interpolated breath hold exam); and DW-MRI with the same protocol described in section 3.1 acquired with 1 NEX.

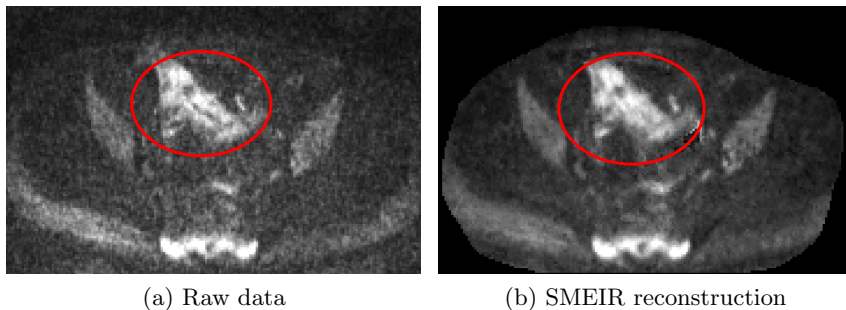


Fig. 2. Representative Crohn’s disease patient with active inflammation in the ileum. (a) The acquired raw DW-MRI data; and (b) SMEIR-reconstructed data. The region with active inflammation is more conspicuous in the SMEIR-reconstructed image than in the raw DW-MRI data.

Two board-certified radiologists reviewed the MRE data independently. Disease activity was defined as abnormal bowel wall thickening and enhancement in the gadolinium-enhanced images by each of the readers. In case of disagreement between the two reviewers, consensus was reached by joint reading of the data. The consensus decision identified 12 patients with active inflammation in the ileum. Another board-certified radiologist, blinded to the MRE data and to the review, identified the ileum on the DW-MRI data for each patient. We manually annotated the ileum wall on the DW-MRI images with $b\text{-value}=200\text{ s/mm}^2$. Next, we calculated the CNR of the inflamed ileum in $b\text{-value}=800\text{ s/mm}^2$ images by subtracting the background signal from the signal of the inflamed ileum and dividing by the standard deviation of the signal in the ileum.

Fig. 2 depicts the acquired raw DW-MRI data and SMEIR-reconstructed data of a representative Crohn’s disease patient with active inflammation in the ileum. Visually, the region with active inflammation is more conspicuous in the SMEIR-reconstructed image than in the raw DW-MRI data. Quantitatively, the average (std) CNR between the inflamed regions and the surrounding neighborhood in the SMEIR-reconstructed images was higher (2.52 ± 0.69) than in the raw DW-MRI data (2.23 ± 0.47) - a difference that was statistically significant (Paired Student’s t-test, $p<0.05$). The SMEIR-reconstructed images improved CNR by 12.6%.

4 Conclusions

We have presented a new model and method for reconstructing high-quality multi b-value DW-MRI images of the body *without* increasing overall acquisition times. This novel approach features the signal decay model as a prior knowledge in the image reconstruction, effectively enabling us to simultaneously reconstruct DW-MRI images and estimate the signal decay model parameters using the Expectation-Maximization framework.

As demonstrated in our experiments, our method improves overall image quality by increasing the signal-to-noise ratio (SNR) by up to 55% and by increasing the conspicuity of inflamed bowel regions of pediatric Crohn's disease patients by 12.6% *without* increasing overall acquisition times. The proposed method permits the acquisition of high-quality DW-MRI images for diagnostic purposes within a clinically acceptable acquisition timeframe.

References

1. Block, K.T., Uecker, M., Frahm, J.: Model-based iterative reconstruction for radial fast spin-echo MRI. *IEEE Trans. Med. Imaging* 28(11), 1759–1769 (2009)
2. Doneva, M., Börner, P., Eggers, H., Stehning, C., S enegas, J., Mertins, A.: Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magn. Reson. Med.* 64(4), 1114–1120 (2010)
3. Freiman, M., Voss, S.D., Mulkern, R.V., Perez-Rossello, J.M., Callahan, M.J., Warfield, S.K.: Reliable Assessment of Perfusivity and Diffusivity from Diffusion Imaging of the Body. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 1–9. Springer, Heidelberg (2012)
4. Ichikawa, T., Erturk, S.M., Motosugi, U., Sou, H., Iino, H., Araki, T., Fujii, H.: High-B-value diffusion-weighted MRI in colorectal cancer. *AJR Am. J. Roentgenol.* 187(1), 181–184 (2006)
5. Ichikawa, T., Erturk, S.M., Motosugi, U., Sou, H., Iino, H., Araki, T., Fujii, H.: High-b value diffusion-weighted MRI for detecting pancreatic adenocarcinoma: preliminary results. *AJR Am. J. Roentgenol.* 188(2), 409–414 (2007)
6. Kanematsu, M., Goshima, S., Watanabe, H., Kondo, H., Kawada, H., Noda, Y., Aomatsu, A., Moriyama, N.: Detection and characterization of focal hepatic lesions with diffusion-weighted MR imaging: a pictorial review. *Abdom. Imaging* (July 2012)
7. Koh, D.M., Collins, D.J.: Diffusion-weighted MRI in the body: applications and challenges in oncology. *AJR Am. J. Roentgenol.* 188(6), 1622–1635 (2007)
8. LeBihan, D., Breton, E., Lallemand, D., Aubin, M.L., Vignaud, J., Laval-Jeantet, M.: Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology* 168(2), 497–505 (1988)
9. Oto, A., Kayhan, A., Williams, J.T.B., Fan, X., Yun, L., Arkani, S., Rubin, D.T.: Active Crohn's disease in the small bowel: evaluation by diffusion weighted imaging and quantitative dynamic contrast enhanced MR imaging. *J. Magn. Reson. Imaging* 33(3), 615–624 (2011)
10. Powell, M.: The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report NA2009/06, Dep. App. Math. and Th. Physics, Cambridge, England (2009)
11. Sugita, R., Yamazaki, T., Furuta, A., Itoh, K., Fujita, N., Takahashi, S.: High b-value diffusion-weighted MRI for detecting gallbladder carcinoma: preliminary study and results. *Eur. Radiol.* 19(7), 1794–1798 (2009)
12. Velikina, J.V., Alexander, A.L., Samsonov, A.: Accelerating MR parameter mapping using sparsity-promoting regularization in parametric dimension. *Magn. Reson. Med.* (December 2012) (in press), <http://dx.doi.org/10.1002/mrm.24577>

Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch

Wenzhe Shi¹, Jose Caballero¹, Christian Ledig¹, Xiahai Zhuang², Wenjia Bai¹,
Kanwal Bhatia¹, Antonio M Simoes Monteiro de Marvao³,
Tim Dawes³, Declan O'Regan³, and Daniel Rueckert¹

¹ Biomedical Image Analysis Group, Imperial College London, UK

² Shanghai Advanced Research Institute, Chinese Academy of Sciences, China

³ Institute of Clinical Science, Imperial College London, UK

Abstract. The accurate measurement of 3D cardiac function is an important task in the analysis of cardiac magnetic resonance (MR) images. However, short-axis image acquisitions with thick slices are commonly used in clinical practice due to constraints of acquisition time, signal-to-noise ratio and patient compliance. In this situation, the estimation of a high-resolution image can provide an approximation of the underlying 3D measurements. In this paper, we develop a novel algorithm for the estimation of high-resolution cardiac MR images from single short-axis cardiac MR image stacks. First, we propose to use a novel approximate global search approach to find patch correspondence between the short-axis MR image and a set of atlases. Then, we propose an innovative super-resolution model which does not require explicit motion estimation. Finally, we build an expectation-maximization framework to optimize the model. We validate the proposed approach using images from 19 subjects with 200 atlases and show that the proposed algorithm significantly outperforms conventional interpolation such as linear or B-spline interpolation. In addition, we show that the super-resolved images can be used for the reproducible estimation of 3D cardiac functional indices.

1 Introduction

3D cardiac magnetic resonance (MR) imaging has developed rapidly during the past few years, particularly in the acquisition of 3D cine MR images [1,2]. Near isotropic 3D cardiac MR images allow reliable assessment of complex cardiac morphology. Using 3D images also allows for a more accurate and reproducible estimation of cardiac functional indices [3]. However, 3D cardiac MR imaging is not always available due to several limitations: First, 3D cardiac MR imaging often involves breath-holding for periods that are too long for many patients. In addition, it often has a low signal-to-noise ratio (SNR). Finally, advanced 3D cardiac MR imaging is not yet widely available in clinical practice and still requires substantial specialist expertise.

Image super-resolution is an active field of research in computer vision. Most super-resolution algorithms use an observation model which establishes a relationship between the high-resolution image and the observed low-resolution images. The observed low-resolution images are considered to be warped, blurred,

down-sampled and noisy versions of the original high-resolution image. One of the most common approaches to the super-resolution problem is to use the maximum likelihood (ML) or maximum a posteriori (MAP) estimation [4]. In these approaches, a distance measure between the reconstructed image and the observed images is iteratively reduced. Example-based image super-resolution [5] is another popular approach where correspondences between low- and high-resolution image patches are learned from a database and then applied to a new low-resolution image to recover its most likely high-resolution version. In both approaches, a distance measure between the current estimation and the low-resolution images must be computed. Takeda [6] proposed a distance estimation approach which requires no explicit motion estimation by inverting the position of the patch selection operator and the resampling operator.

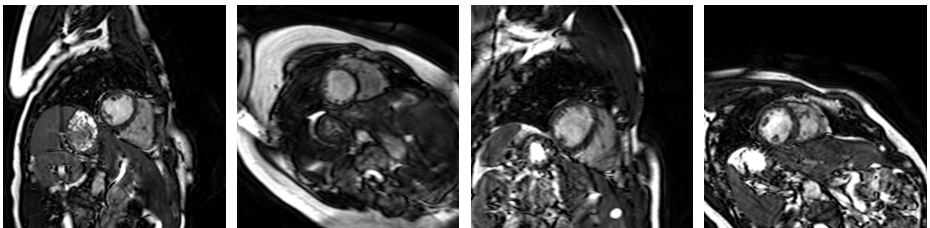


Fig. 1. Variability of heart orientation, position and shape across subjects

The idea of super-resolution has been applied in medical imaging too: Gholipo [7] reconstructed a high-resolution volume from multiple low resolution (LR) images using image priors based on total variation constraint with MAP estimation. However, this method cannot be directly applied to our problem because it requires multiple instances of low resolution (LR) images from different views. To take advantage of the information redundancy in similar patches across different subjects, patch-based methods have been shown to be highly efficient in applications such as segmentation [8,9]. Rousseau [10,11] proposed to combine registration with a patch-based approach to create super-resolution brain MR images from atlases of multiple LR images of different subjects. In this approach the high-resolution image is constructed via non-local fusion of those patches. However, the method requires rough correspondence between images either via explicit motion estimation or other means. This is difficult to guarantee in cardiac MR images due to the large variation in the orientation, position and shape of the heart across subjects (see Fig.1). Moreover, the complexity of these non-local patch-based methods increases with the number of atlases.

In this paper we aim to reconstruct a super-resolution (SR) cardiac MR image from a single short axis (SA) cardiac MR image with a set of 3D atlases available as the training database. Three different aspects are challenging: First, the slice thickness of the SA image is much larger than the slice thickness of the 3D image (approximately 5 times, e.g. 2mm vs. 10mm) while the up-sampling factor of classic super-resolution algorithms is usually around two [12]. Second, the search

for the best match of patches in 3D with multiple atlases using conventional approaches is very expensive. Finally, cardiac images exhibit significantly more variability in terms of orientation and anatomy compared to brain images. Local search methods used in brain imaging [10] are thus not suitable. In addition an exhaustive global search for patches is impossible given the computational cost.

To solve this problem, we propose a framework to combine classic and example-based super-resolution approaches using an approximation graph based search based on the recently proposed PatchMatch algorithm [13]. Inspired by [10], we assume that information redundancy in similar patches across different subjects can be exploited. Thus, we reformulate the PatchMatch approach to find patch correspondence between a single image and an atlas database. We then use the principles in [6,7] to estimate the super-resolved image using the expectation-maximization (EM) framework.

The novelty and contributions of this paper are the introduction of a global search strategy as well as an observation model with non-explicit motion estimation that avoids any spatial alignment or registration of the images. Furthermore, the computational cost is kept low by using PatchMatch and a closed-form solution in the observation model [6,9]. The number of atlases does not influence the computational cost and thus allows full exploitation of a large atlas database. Our results demonstrate that the algorithm can robustly estimate a SR image in the presence of thick slice data and performs both extrapolation and interpolation by recovering missing apical and basal slices.

2 Methods

2.1 Multi-Atlas PatchMatch

The PatchMatch algorithm proposed by Barnes [13] finds corresponding patches across two images or regions. In contrast with the original PatchMatch algorithm, our multi-atlas PatchMatch (MAPM) finds patch correspondences \mathbf{N} between an image and a database of atlases. Given an image \mathbf{I} and an atlas database \mathbf{A} (individual atlases are denoted as \mathbf{A}_i), we would like to find for each point $\mathbf{x} = (x, y, z)$ in image \mathbf{I} a match in the atlas database \mathbf{A} , $\mathbf{N}(\mathbf{x}) = (\mathbf{p}, i)$ where $\mathbf{p} = (x', y', z')$ is the closest match in atlas \mathbf{A}_i for a given distance function D between patches. The distance function to be used during the search is independent from MAPM and can be customized to different applications.

The MAPM algorithm consists of four different steps which will be described in the following. The reader can find additional figures showing a graphical illustration of the four different steps in the supplementary material¹. The mapping \mathbf{N} can be initialized either by random assignment or by using prior information (Fig.1 in supplementary material). In our case, we assign $\mathbf{N}(\mathbf{x}) = (\mathbf{x}, R(n))$ where $R(n)$ generates a random selection uniformly between \mathbf{A}_1 and \mathbf{A}_n . After initialization, we perform an iterative process of improving the mapping \mathbf{N} using propagation and random search. During the propagation of \mathbf{N} , from point

¹ <https://www.dropbox.com/s/eoeqbviq5kqcdix/MAPdiagram.pdf>

\mathbf{p} neighbouring to point \mathbf{x} (Fig.2 in supplementary material), we attempt to improve $\mathbf{N}(\mathbf{x})$ using the known mapping of $\mathbf{N}(\mathbf{p})$ as in [13]. During the random search step, we attempt to improve $\mathbf{N}(\mathbf{x})$ by testing a sequence of candidate points at an exponentially decreasing distance from $\mathbf{N}(\mathbf{x})$. Different from [13], in our case, the atlas index i can be fixed (Fig.3 in supplementary material) or relaxed (Fig.4 in supplementary material). Each iteration of the algorithm proceeds as follows:

- for each \mathbf{x} propagation from $(x - 1, y, z)$, $(x, y - 1, z)$ and $(x, y, z - 1)$;
- for each \mathbf{x} random search with \mathbf{A}_i fixed then relaxed
- for each \mathbf{x} propagation from $(x + 1, y, z)$, $(x, y + 1, z)$ and $(x, y, z + 1)$;
- for each \mathbf{x} random search with \mathbf{A}_i fixed then relaxed

This process is performed until the sum of all distances in image \mathbf{I} converges.

2.2 Super-Resolution Model with No Explicit Motion Estimation

In the classical observation model, the SR image is reconstructed from a LR training database. The LR images are considered to be degraded versions of the SR image undergoing blurring, downsampling and the addition of noise [6,7,4]. In our case, we aim to reconstruct the SR image from a SR atlas database constrained by a single LR image.

Takeda [6] suggested that the patch selection should be applied before rather than after the downsampling in order to avoid an explicit motion estimation. Gholipour [7] proposed the following formulation designed for MR images:

$$\mathbf{I}_k^L = \mathbf{R}\mathbf{B}_k\mathbf{S}_k\mathbf{M}_k\mathbf{I}^H, \quad (1)$$

Here k denotes a slice, \mathbf{M} denotes motion operator which is no longer needed in our case, \mathbf{S}_k denotes the slice selection operator which can be replaced by patch selection operator \mathbf{P} , \mathbf{B}_k is a blurring kernel representing the point spread function (PSF) of the MR imaging signal acquisition process and \mathbf{R} is the down-sampling operator.

By combining patch redundancy [10] and the formulation proposed in [7], we propose a novel model with two terms $\Phi_{SR} = \Phi_{SR}^1 + \Phi_{SR}^2$ to reconstruct the SR image \mathbf{I} where Ω is the image domain. In this model, the first term constrains \mathbf{N} using the observed LR image so that the selected patches after downsampling operations should be as similar as possible to the LR image:

$$\Phi_{SR}^1 := \sum_{\mathbf{x} \in \Omega} w[\mathbf{x}, \mathbf{N}(\mathbf{x})] \|\mathbf{P}_x \mathbf{I}^L - \mathbf{R}\mathbf{P}_{\mathbf{N}(\mathbf{x})} \mathbf{B}\mathbf{A}\|_2. \quad (2)$$

The second term constrains \mathbf{I} using \mathbf{N} and \mathbf{A} based on the fact that the reconstructed images should be as similar as possible to the selected patches:

$$\Phi_{SR}^2 := \sum_{\mathbf{x} \in \Omega} w[\mathbf{x}, \mathbf{N}(\mathbf{x})] \|\mathbf{P}_x \mathbf{I} - \mathbf{P}_{\mathbf{N}(\mathbf{x})} \mathbf{A}\|_2, \quad (3)$$

Here $\mathbf{P}_{\mathbf{x}}$ selects a patch from an image with radius in mm around \mathbf{x} and $\mathbf{P}_{\mathbf{N}(\mathbf{x})}$ with $\mathbf{N}(\mathbf{x}) = (\mathbf{p}, i)$ selects a patch from \mathbf{A}_i with radius in mm around \mathbf{p} . $w[\mathbf{x}, \mathbf{N}(\mathbf{x})]$ is chosen as $\exp\left\{-\frac{D(\mathbf{N}(\mathbf{x}))^2}{2\sigma^2}\right\}$ according to [6] and controls the contribution of the selected patch to the final reconstruction. Finally, we define $D(\mathbf{N}(\mathbf{x})) = \|\mathbf{P}_{\mathbf{x}}\mathbf{I}^L - \mathbf{R}\mathbf{P}_{\mathbf{N}(\mathbf{x})}\mathbf{B}\mathbf{A}\|_2 + \|\mathbf{P}_{\mathbf{x}}\mathbf{I} - \mathbf{P}_{\mathbf{N}(\mathbf{x})}\mathbf{A}\|_2$. We blur the atlases before patch selection to save computation time.

2.3 Expectation Maximization Framework

In this subsection, we construct the whole super-resolution approach within an EM framework: In this context the atlases \mathbf{A} and the LR image \mathbf{I}^L correspond to the observed data, \mathbf{I} is the unobserved data and \mathbf{N} are the parameters. The EM algorithm is initialized by assuming \mathbf{I} to be empty and $\mathbf{N}(\mathbf{x}) = (\mathbf{x}, R(n))$. The distance between an empty patch and any patch is defined as $+\infty$.

In the M-step we optimize \mathbf{N} using the MAPM described in Sec. 2.1. Then, the weighting matrix W is updated according to the distance computed. In the E-step we estimate the SR image \mathbf{I} by optimizing the observation model Φ_{SR} . We can calculate the penalty at each patch $\mathbf{P}_{\mathbf{x}}\mathbf{I}$ independently if \mathbf{N} is fixed similar to the multi-point estimation in [9]:

$$\arg \min_{\mathbf{P}_{\mathbf{x}}\mathbf{I}} \Phi_{SR}(\mathbf{P}_{\mathbf{x}}\mathbf{I}) := \sum_{\mathbf{p} \in \Omega^P} w[\mathbf{x}, \mathbf{N}(\mathbf{p})] \|\mathbf{P}_{\mathbf{x}}\mathbf{I} - \mathbf{P}_{\mathbf{N}(\mathbf{p})}\mathbf{A}\|_2, \quad (4)$$

Here Ω^P is a neighborhood with all patches which contain point \mathbf{x} and centered at point \mathbf{p} and the distance is calculated on overlapping areas as in [9]. This leads to a closed-form solution:

$$\mathbf{P}_{\mathbf{x}}\mathbf{I} = \frac{\sum_{\mathbf{p} \in \Omega^P} w[\mathbf{x}, \mathbf{N}(\mathbf{p})] \mathbf{P}_{\mathbf{N}(\mathbf{p})}\mathbf{A}}{\sum_{\mathbf{p} \in \Omega^P} w[\mathbf{x}, \mathbf{N}(\mathbf{p})]}, \quad (5)$$

3 Application to Cardiac MR Images

The proposed framework was applied to cardiac MR images and evaluated its performance in two scenarios using both simulated and real cardiac MR images. Two hundred healthy volunteers were scanned using a 1.5T Philips Achieva system with a 32-channel cardiac coil. A single breath-hold 3D balanced steady-state free precession (b-SSFP) sequence is acquired. The final voxel size is 1.25 x 1.25 x 2 mm. The typical breath-hold time is 20 seconds. 11 good quality images were selected and used to build a synthetic data set and the remaining 189 images were used as the atlases. The LR images (1.25 x 1.25 x 10 mm) were generated from the 3D images using the operator defined by Eq.1. In addition, 19 normal volunteers were scanned twice on the same day. A standard acquisition was performed including an axial stack of cine b-SSFP MR images in the left ventricular short axis plane. The voxel sizes for these image is 1.25 x 1.25 x 10 mm. The images were then super-resolved using the previous 200 3D images.

Table 1. The median and interquartile range of PSNR for different methods from 11 synthetic cases. There is significant difference between the PSNRs of interpolation methods and the proposed method (p -value < 0.05 indicated by *).

	linear	B-spline	cubic B-spline	MAPM
PSNR (dB)	19.05 (1.17)*	19.62 (1.28)*	19.9 (1.22)*	20.96 (1.1)

There are three pre-processing steps which occur before applying the EM algorithm. First, the SA slices are spatially aligned to remove the inter-slice shifting caused by respiratory motion. The inter-slice shifts between SA slices are corrected by registering SA slices to long-axis (LA) slices [14]. Second, a region of interest (ROI) is detected using a Haar feature classifier [15]. Finally, all atlases are intensity normalized [16] to the spatially corrected image. During the experiments we have set our patch size to 14 x 14 x 14 mm.

3.1 Quantitative Evaluation

In this evaluation, we compare the PSNR between the image reconstructed from the synthetic LR image and the original image. We reconstruct the SR image using linear interpolation, spline interpolation [17] and the proposed approach. The result is shown in Tab. 1. During the down-sampling process, part of the apex and base might be missing due to the reduced field of view. This is also a common problem in SA images. It can be seen from Fig.2 that the missing parts of the apical and basal slices can be recovered. This is due to the fact that a patch is copied from the atlases instead of a single voxel. Thus, during the iterative process, the missing topology can be gradually repaired.

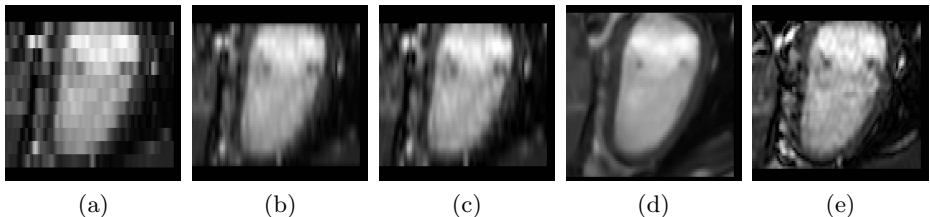


Fig. 2. This figure shows the results of the synthetic evaluation from long-axis view. (a) shows the down-sampled images; (b) shows the linear interpolation; (c) shows the cubic B-spline interpolation; (d) shows the proposed method and (e) shows the original 3D image.

3.2 Reproducibility Analysis

In the second experiment, we attempt to super-resolve the SA cardiac MR images using the proposed algorithm (Fig 3). The super-resolved image has better contrast and less noise compared to 3D image of the same subject. In addition, we segment both the SA images and super resolved images using the patch-based

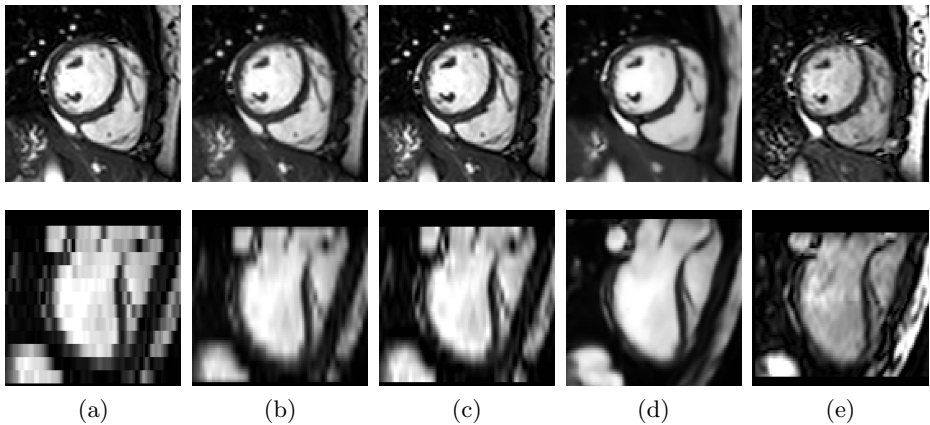


Fig. 3. This figure shows the results from the super-resolution of the SA MR images. (a) original SA image; (b) linear interpolation; (c) cubic B-spline interpolation; (d) proposed and (e) corresponding 3D image of the same subject rigidly align to the SA MR image.

segmentation [8]. We calculate the mean and the standard deviation of absolute differences \mathbf{d} between the left ventricle (LV) volume obtained from two scans of the 19 subjects. The results from SR images ($\mathbf{d}_{SR} : 4.94 \pm 4.36 \text{ ml}$) are more reproducible compare to results from SA images ($\mathbf{d}_{SA} : 6.58 \pm 6.76 \text{ ml}$).

4 Conclusion

In this paper, we developed a MAPM based framework for medical images. We have shown that our framework works well in cases where hundreds of atlases are used as the training database to super-resolve one LR image. In addition, there is no need for any spatial alignment with atlases. The computational time is 2 hours on average per case and does not change with an increasing number of atlases. Finally, the algorithm performs extrapolation as well as interpolation of the images. This is desirable in cardiac images where apical and basal slices may be missing due to limited field of view and thick slices. In the SR image, the original SA slice is a little blurred due to the fusion of multiple patches [6]. This is a trade-off for improved through-plane resolution. Future work will include exploring the possibility to extend MAPM to patch-based segmentation and to exploit neighboring correspondence [18] to preserve the original SA slice and image self similarity [19] to increase the robustness.

References

1. Uribe, S., Muthurangu, V., Boubertakh, R., Schaeffter, T., Razavi, R., Hill, D., Hansen, M.: Whole-heart cine MRI using real-time respiratory self-gating. *Magnetic Resonance in Medicine* 57(3), 606–613 (2007)

2. Davarpanah, A., Chen, Y., Kino, A., Farrelly, C., Keeling, A., Sheehan, J., Ragin, A., Weale, P., Zuehlsdorff, S., Carr, J.: Accelerated two- and three-dimensional cine MR imaging of the heart by using a 32-channel coil. *Radiology* 254(1), 98–108 (2010)
3. Sørensen, T., Körperich, H., Greil, G., Eichhorn, J., Barth, P., Meyer, H., Pedersen, E., Beerbaum, P.: Operator-independent isotropic three-dimensional magnetic resonance imaging for morphology in congenital heart disease. *Circulation* 110(2), 163–169 (2004)
4. Tian, J., Ma, K.: A survey on super-resolution imaging. *Signal, Image and Video Processing* 5(3), 329–342 (2011)
5. Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. *IEEE Computer Graphics and Applications* 22(2), 56–65 (2002)
6. Takeda, H., Milanfar, P., Protter, M., Elad, M.: Super-resolution without explicit subpixel motion estimation. *IEEE Transactions on Image Processing* 18(9), 1958–1975 (2009)
7. Gholipour, A., Estroff, J.A., Warfield, S.K.: Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. *IEEE Transactions on Medical Imaging* 29(10), 1739–1758 (2010)
8. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, D.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *Neuroimage* 54(2), 940–954 (2011)
9. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. *IEEE Transactions on Medical Imaging* 30(10), 1852–1862 (2011)
10. Rousseau, F., Kim, K., Studholme, C.: A groupwise super-resolution approach: application to brain MRI. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 860–863 (2010)
11. Rousseau, F.: A non-local approach for image super-resolution using intermodality priors. *Medical Image Analysis* 14(4), 594 (2010)
12. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(9), 1167–1183 (2002)
13. Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized patch-match correspondence algorithm. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part III*. LNCS, vol. 6313, pp. 29–43. Springer, Heidelberg (2010)
14. Lötjönen, J., Pollari, M., Kivistö, S., Lauerma, K.: Correction of movement artifacts from 4-D cardiac short- and long-axis MR data. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3217, pp. 405–412. Springer, Heidelberg (2004)
15. Viola, P., Jones, M.: Robust real-time object detection. *International Journal of Computer Vision* 57(2), 137–154 (2002)
16. Nyúl, L., Udupa, J., et al.: On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine* 42(6), 1072 (1999)
17. Unser, M., Aldroubi, A., Eden, M.: B-spline signal processing. i. Theory. *IEEE Transactions on Signal Processing* 41(2), 821–833 (1993)
18. Ólafsdóttir, H., Pedersen, H., Hansen, M.S., Larsson, H., Larsen, R.: Improving image registration by correspondence interpolation. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1524–1527. IEEE (2011)
19. Manjón, J.V., Coupé, P., Buades, A., Fonov, V., Louis Collins, D., Robles, M.: Non-local MRI upsampling. *Medical Image Analysis* 14(6), 784–792 (2010)

Self-gated Radial MRI for Respiratory Motion Compensation on Hybrid PET/MR Systems

Robert Grimm¹, Sebastian Fürst², Isabel Dregely², Christoph Forman¹,
Jana Maria Hutter¹, Sibylle I. Ziegler², Stephan Nekolla², Berthold Kiefer³,
Markus Schwaiger², Joachim Hornegger¹, and Tobias Block⁴

¹ Pattern Recognition Lab, FAU Erlangen, Erlangen, Germany

² Department of Nuclear Medicine, TU Munich, Munich, Germany

³ Siemens Healthcare MR, Erlangen, Germany

⁴ Department of Radiology, NYU Langone Medical Center, New York City, NY, USA

Abstract. Accurate localization and uptake quantification of lesions in the chest and abdomen using PET imaging is challenging due to the respiratory motion during the exam. The advent of hybrid PET/MR systems offers new ways to compensate for respiratory motion without exposing the patient to additional radiation. The use of self-gated reconstructions of a 3D radial stack-of-stars GRE acquisition is proposed to derive a high-resolution MRI motion model. The self-gating signal is used to perform respiratory binning of the simultaneously acquired PET raw data. Matching μ -maps are generated for every bin, and post-reconstruction registration is performed in order to obtain a motion-compensated PET volume from the individual gates. The proposed method is demonstrated in-vivo for three clinical patients. Motion-corrected reconstructions are compared against ungated and gated PET reconstructions. In all cases, motion-induced blurring of lesions in the liver and lung was substantially reduced, without compromising SNR as it is the case for gated reconstructions.

1 Introduction

With scan times of between two and ten minutes per bed position, positron emission tomography (PET) imaging of the lung and abdomen is affected by respiratory motion. Breath-hold techniques cannot be applied for such long time, and respiratory gating approaches lack clinical acceptance because they reduce the scan efficiency and, thus, lead to increased noise or excessive scan time. Clinical PET scans are affected by two types of artifacts due to respiratory motion: First, the mismatch between the MR-based attenuation correction map (μ -map) used for PET reconstruction, which is typically acquired in an end-expiratory breathhold, and the PET image itself, which is acquired during free breathing, causes an under- or overestimation of tracer activity especially in the vicinity of the diaphragm [1]. Second, respiration leads to local image blurring (smearing) along the direction of motion, i.e., primarily in the cranio-caudal direction. This can result in an incorrectly estimated volume, shape, and uptake of lesions [2–4] as well as in reduced conspicuity especially of small lesions.

Respiration is a predominantly periodic type of motion that can be compensated for if a model of the motion is available. Typically, this model consists of a displacement vector field describing the nonrigid deformation that maps voxels between different respiratory states. A comprehensive review of respiratory motion models has recently been published by McClelland et al. [5].

In hybrid PET/MR scanners, respiratory motion models can be formed either by near-realtime 3D MR sequences [6–8] or by retrospective gating and averaging over multiple respiratory cycles [4, 9]. Retrospectively gated MR motion models can further be subdivided into three categories according to the acquisition method: 1) The displacement fields can be measured directly using tagged MRI [9–11], 2) the volume can be sampled using a 2D multi-slice technique [4, 7] or 3) using a 3D acquisition [6, 7].

According to these characteristic features, the proposed approach for respiratory motion compensation on integrated PET/MR scanners can be classified as follows. It utilizes a 3D radial stack-of-stars MRI sequence. This sequence allows for self-gating, i.e. a respiratory signal can be derived without the need for additional MR navigator echoes or sensors attached to the patient. Moreover, the trajectory allows for retrospective gating, without having to face discontinuities between slices that affect 2D multi-slice methods. The model is acquired for the whole PET acquisition and thereby assures high spatial fidelity and can adapt to changes in the respiratory pattern.

After derivation of the MR-based motion model, it is utilized for motion-corrected PET image reconstruction. The PET gates are reconstructed using conventional algorithms, then warped to a common respiratory phase, and finally averaged [4, 6]. We followed this approach rather than incorporating the motion model directly into the PET reconstruction, as recently suggested in literature [7, 12–14], because we focused on the formation of the motion model in this work. The utilized post-reconstruction registration approach is well-studied, reproducible, and provides comparable results given that the respiratory bins have similar and sufficient count statistics [13–15].

2 Methods

Our approach is based on the following scan protocol for a combined whole-body PET/MR scanner. First, the μ -map is acquired conventionally (20 s breathhold scan at end-expiration). Afterwards, MR and PET data are acquired simultaneously. Detailed sequence parameters can be found in Section 4.

In the subsequent data processing chain, described in more detail in the following sections, a motion model is computed from the MR data. It is subsequently applied to deform the μ -map to different respiratory states. A self-gating signal extracted from the MR scan is used to perform respiratory gated PET reconstructions at different levels of inspiration, each utilizing the matching μ -map. Finally, the inverse deformations are used to co-register and combine the individual PET volumes.

2.1 Self-gated Radial MRI

The respiratory motion model is generated with the help of a T1-weighted 3D radial stack-of-stars spoiled gradient echo sequence with fat suppression. A golden-angle increment of 111.25° between subsequently acquired radial angles distributes the sampling incoherently but approximately uniformly [16] over the readout plane, facilitating retrospectively gated reconstruction.

This particular k-space trajectory, shown in Fig. 1, allows to derive a self-gating signal (SGS) from the central k-space partition $k_z = 0$ that is sampled every $N_z \cdot \text{TR}$ [17]. Each sample of the self-gating signal is computed as the mean absolute value of the central three k-space samples in a readout. A suitable coil element to derive the signal is determined automatically by computing the Fourier transform of the SGS for every coil and selecting the one with the highest ratio of the peak in the range of expected respiratory frequencies to the energy in the remaining high-frequency coefficients. It should be noted that the SGS is of qualitative nature and, unlike a 1D navigator echo or projection-based self-gating [6], does not indicate an absolute displacement e.g. of the liver.

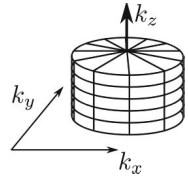


Fig. 1. Radial stack-of-stars

Variable amplitude-based binning is applied to partition the radial readouts into N_{Bins} bins containing equal amounts of data, according to the respective self-gating signal amplitude. This scheme ensures comparable statistics also for the PET listmode data that were gated in the same manner (see Section 2.3).

2.2 Motion Modeling

A state-of-the-art nonrigid registration algorithm that was recently proposed for lung registration [18] is employed to compute the deformation between the respiratory-gated MR volumes. Its output is a 3D deformation field $\mathbf{T}_{i,j}$ that maps a volume at respiratory phase i to a volume at respiratory phase j . The deformations are estimated from each respiratory phase to the reference phase 1 (end-exhale), and vice versa.

2.3 Motion-Compensated PET Reconstruction

To generate matching μ -maps \hat{M}_i for the respiratory phases $i \in \{2 \dots N_{\text{Bins}}\}$, the original end-exhale μ -map M_1 is warped to each respiratory state using the corresponding deformation field $\mathbf{T}_{1,i}$.

The self-gating signal from the MRI acquisition is used to reconstruct gated images from the PET listmode data. For this purpose, each listmode event is assigned to the same bin as the closest SGS sample according to the respective timestamps. Static (ungated) images are reconstructed from all available PET data as well. The vendor-provided clinical reconstruction algorithm (ordered-subset expectation maximization, OSEM) is applied. The respective μ -maps (warped / native) are utilized for attenuation correction. After reconstruction, the gated PET volumes P_i are co-registered to the end-expiratory state by applying $\mathbf{T}_{i,1}$. Finally, the co-registered images are averaged.

3 Experiments

3.1 Self-gating

To study the accuracy of the respiratory self-gating, images were reconstructed using $N_{\text{Bins}} \in \{2 \dots 15\}$ respiratory bins. Partitioning the data into more bins reduces the SNR in every bin but allows more accurate separation of different respiratory states. A virtual 1D navigator column (averaged over 3×3 voxels in the transversal plane) was extracted at the apex of the liver dome in the reconstructed volumes. The position of the liver edge at each respiratory state was detected by Canny filtering ($\sigma = 1.5, \theta_{\text{low}} = 0.8, \theta_{\text{high}} = 0.95$) along the column.

3.2 PET Reconstruction

The three datasets were analyzed visually as well as quantitatively, in terms of apparent lesion volume that was computed by placing ellipsoidal VOIs using *syngo TrueD* (Siemens Healthcare, Erlangen, Germany) and applying a fixed threshold for segmentation. The following reconstructions were compared:

1. R_{All} : Ungated reconstruction using all PET data and original μ -map.
2. R_{40} : *Gold standard*: Gated reconstruction using the 40% of the data with the least variation in the SGS amplitude.

To examine the effects of post-reconstruction registration, the gated PET images were computed with the adapted μ -map and warped to the reference state before fusion. This gives the corrected reconstruction:

3. G_5 : Gating to 5 bins, warped μ -maps, post-reconstruction registration.

4 Clinical Study

The protocol described in Section 2 was used for three oncological patient studies P_1, P_2, P_3 , conducted on a 3 Tesla hybrid PET/MR system (Biograph mMR; Siemens Healthcare, Erlangen, Germany). Written consent from the subjects and approval from the local ethics committee was obtained prior to the examinations. Between 401 and 455 MBq of ^{18}F -FDG were administered as radionuclide agent, 96 - 148 min before the study. The μ -maps were acquired with a 3D Dixon GRE sequence while the patients were instructed to hold their breath at end-expiration. No particular respiratory instruction was given for the following PET listmode acquisition, during which the motion modeling scan was carried out:

P_1 (f/58y/82kg): coronal slab orientation (FOV $450 \times 450 \times 245 \text{ mm}^3$), spatial resolution $1.6 \times 1.6 \times 1.7 \text{ mm}^3$, 288 pixel matrix, 145 slices (50% slice resolution, 5/8 partial Fourier, 10% oversampling), 3296 radial angles, 12 min scan time.

P_2 (m/79y/96kg) and P_3 (m/67y/115kg): sagittal slab orientation (FOV $400 \times$

$400 \times 360 \text{ mm}^3$), spatial resolution $1.65 \times 1.65 \times 5 \text{ mm}^3$, 256 pixel matrix, 72 slices (61% slice res., 5/8 partial Fourier), 4416 radial angles, 10 min scan time.

To demonstrate the applicability of the derived motion model, $N_{\text{Bins}} = 5$ respiratory states were used. MR data were reconstructed using regridding. The empirically selected parameters for the deformable registration were smoothing $\alpha = 5$, 75 % randomized sampling, and 3 levels with a grid spacing of 8, 4, and 2. The PET reconstruction (OSEM3D 3i21s) used 3 iterations on 21 subsets, with a matrix size of 172×172 and 4 mm Gaussian post-reconstruction filtering.

5 Results and Discussion

5.1 Self-gating

In all cases, the self-gating signal was successfully extracted. Fig. 2a) shows the detected maximal edge displacement depending on the number of bins chosen. An asymptotic behavior, caused by successive reduction of intra-bin motion with a higher number of bins [19], was observed. The maximal displacement in the three patients was 10-13 mm. Due to the uniform bin size, a relatively high number of bins is required to capture the full respiratory range, eventually compromising image quality in the exhale bins. A cumulative respiratory histogram was computed from the detected liver edge displacements using $N_{\text{Bins}} = 15$. As can be seen in Fig. 2b), 60-75% of the detected respiration was in a range of 5 mm from maximal end-expiration, while $\sim 20\%$ are within 5 mm from maximal inspiration. Thus, when using only 5 bins, the intra-bin motion blur causes the edge of the last bin to be detected at about 2-3 mm below maximal inspiration (cf. also Fig. 2a)). The choice of $N_{\text{Bins}} = 5$ can be considered a compromise between accuracy and sufficient statistics for all bins.

Upon comparing the motion model with the Dixon images acquired for the μ -map it was noticed that P_2 and P_3 had held their breath at *end-inspiration* rather than at end-expiration as instructed. Therefore, in both cases the deformed μ -maps \hat{M}_i were generated by applying the model $T_{5,i}$. Moreover, the end-exhale μ -map for reconstructions R_{All}, R_{40} was estimated by applying $T_{5,1}$.

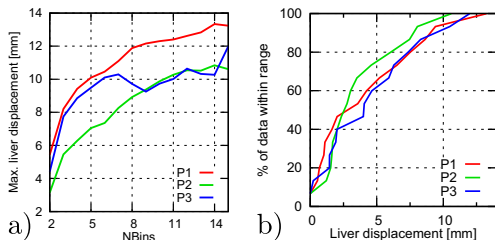


Fig. 2. a) Maximal displacement of liver edge depending on N_{Bins} . b) Cumulative respiratory histogram for $P_1 - P_3$.

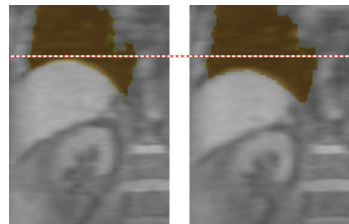


Fig. 3. Self-gated MRI of P_3 overlaid with warped μ -map (brown: lung compartment)

5.2 PET Reconstruction

The tissue interfaces in the deformed μ -maps were visually consistent with the acquired motion model, as shown in Fig. 3 depicting a fusion of the motion model at end-expiration and end-inspiration with the corresponding μ -maps in P_3 .

In P_1 , a lesion (\varnothing 20 mm) was found to move by approx. 6 mm in head-feet direction in the individual gated PET reconstructions of G_5 . Fig. 4 shows a line profile through the lesion, averaged over 3×3 pixels. In the ungated reconstruction, the lesion appears elongated by ~ 5 mm. As shown in Fig. 4b), the true shape was restored after post-reconstruction registration. Corresponding PET images are given in Fig. 5a).

In P_2 , a larger lesion (\varnothing 45 mm) was found in the liver. Significant respiratory movement in the anterior-posterior direction was visible in the model and confirmed by the direction of motion blur in the ungated PET reconstruction. The SUV=6 iso-intensity contour of the lesion for all reconstructions is shown in Fig. 5b), as an overlay in the end-exhale motion model. Gated and corrected reconstructions match the outlines of the slightly hypointense region in the MR image well, while the ungated contour is displaced by ~ 6 mm.

In P_3 , multiple foci (\varnothing up to 10 mm) in both kidney cortices showed increased uptake of FDG. The largest spot was considered pathological. Due to respiratory motion, the foci were less conspicuous in the ungated images, cf. Fig. 5c).

The lesions were segmented with fixed thresholds of SUV=2.5 (P_1) and SUV=6 (P_2, P_3) and the apparent lesion volume was compared against the gated reconstruction R_{40} . For (P_1, P_2, P_3), the difference was reduced from (+27.5%, +18.2%, -58.9%) in the ungated reconstructions, respectively, to (+1.1%, +6.4%, +15.3%) by applying the proposed method. Loss of contrast in the motion-compensated reconstruction, as also reported by [4], was noticed for few very small details. A possible reason is residual intra-bin motion especially in end-inspiration, as discussed in Section 5.1. Moreover, R_{40} utilizes 40% of the data for reconstruction, while the bins of G_5 are reconstructed from only 20%. The non-linear reconstruction may recover such small structures in the former case, but fail to do so in the latter. Here, applying the motion compensation already *during* rather than after PET reconstruction [12] promises improved sensitivity.

6 Summary and Conclusion

We presented first in-vivo results of self-gated MRI motion modeling applied to respiratory motion compensation in PET/MR scanners. It requires no additional physiological signal sensors and captures an averaged respiratory motion cycle throughout the measurement. This information is used to correct for respiratory

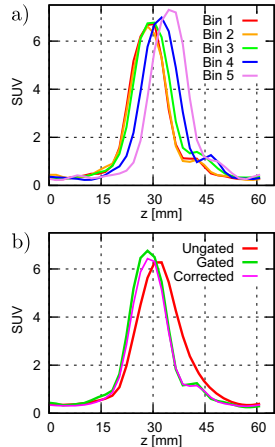


Fig. 4. Line profiles through pulmonary lesion in P_1

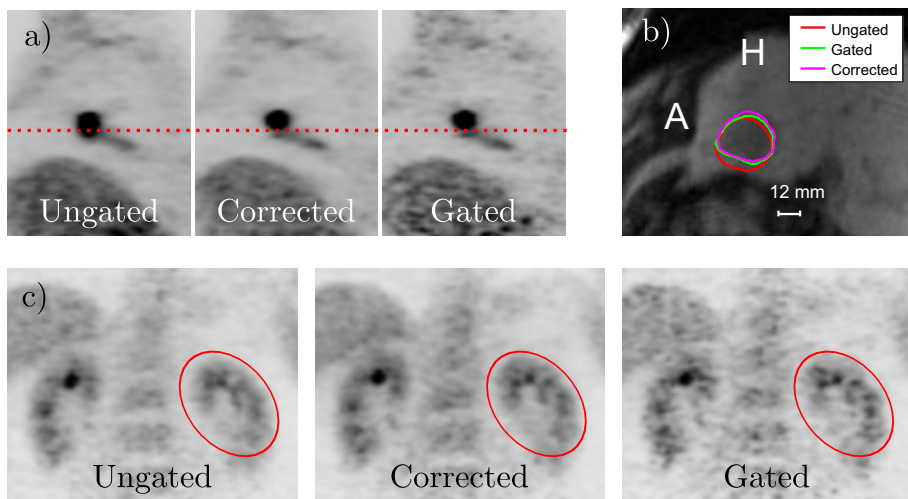


Fig. 5. a) Sagittal slice through pulmonary lesion in P_1 . b) SUV=6 iso-contours as overlay over MR image of liver lesion in P_2 . c) Spots in left kidney cortex in P_3 are more clearly enhanced in the motion-compensated and gated reconstructions.

motion in the PET reconstruction, resulting in a clear reduction of motion blur but improved SNR compared to conventional gated reconstructions. The approach was validated on clinical patients with lesions in three different organs.

Further investigations will consider applying the motion model in motion-compensation incorporated PET reconstructions [12], especially utilizing a higher number of respiratory bins. Moreover, despite the advantage of capturing the global course of respiration during the examination, the scan time for motion modeling is relatively long. A reduction is easily possible by acquiring fewer radial spokes at the cost of increased radial streaking artifacts. Robust registration methods on the one hand and Compressed Sensing reconstruction enforcing a temporal regularization on the MRI images on the other hand can help to overcome the undersampling artifacts. Finally, the proposed method will be extended to automatically detect the actual respiratory position of the acquired μ -map.

References

1. Keller, S.H., Holm, S., Hansen, A.E., Sattler, B., Andersen, F., Klausen, T.L., Højgaard, L., Kjær, A., Beyers, T.: Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI. *MAGMA* 26, 173–181 (2013)
2. Liu, C., Pierce, L.A., Alessio, A.M., Kinahan, P.E.: The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging. *Phys. Med. Biol.* 54(24), 7345–7362 (2009)
3. Bundschuh, R.A., Martínez-Möller, A., Essler, M., Nekolla, S.G., Ziegler, S.I., Schwaiger, M.: Local motion correction for lung tumours in PET/CT—first results. *Eur. J. Nucl. Med. Mol. Imaging* 35(11), 1981–1988 (2008)

4. Würslin, C., Schmidt, H., Martirosian, P., Brendle, C., Boss, A., Schwenzer, N.F., Stegger, L.: Respiratory motion correction in oncologic PET using T1-weighted MR imaging on a simultaneous whole-body PET/MR system. *J. Nucl. Med.* (2013)
5. McClelland, J.R., Hawkes, D.J., Schaeffter, T., King, A.P.: Respiratory motion models: A review. *Med. Image Anal.* 17(1), 19–42 (2013)
6. Buerger, C., Tsoumpas, C., Aitken, A., King, A.P., Schleyer, P., Schulz, V., Marsden, P.K., Schaeffter, T.: Investigation of MR-based attenuation correction and motion compensation for hybrid PET/MR. *IEEE Trans. Nucl. Sci.* 59(5), 1967–1976 (2012)
7. Dikaios, N., Izquierdo-Garcia, D., Graves, M.J., Mani, V., Fayad, Z.A., Fryer, T.D.: MRI-based motion correction of thoracic PET: initial comparison of acquisition protocols and correction strategies suitable for simultaneous PET/MRI systems. *Eur. Radiol.* 22(2), 439–446 (2012)
8. King, A.P., Buerger, C., Tsoumpas, C., Marsden, P.K., Schaeffter, T.: Thoracic respiratory motion estimation from MRI using a statistical model and a 2-D image navigator. *Med. Image Anal.* 16(1), 252–264 (2012)
9. Ouyang, J., Li, Q., El Fakhri, G.: Magnetic resonance-based motion correction for positron emission tomography imaging. *Semin. Nucl. Med.* 43(1), 60–67 (2013)
10. Guérin, B., Cho, S., Chun, S.Y., Zhu, X., Alpert, N.M., El Fakhri, G., Reese, T., Catana, C.: Nonrigid PET motion compensation in the lower abdomen using simultaneous tagged-MRI and PET imaging. *Med. Phys.* 38(6), 3025–3038 (2011)
11. Chun, S.Y., Reese, T.G., Ouyang, J., Guérin, B., Catana, C., Zhu, X., Alpert, N.M., El Fakhri, G.: MRI-based nonrigid motion correction in simultaneous PET/MRI. *J. Nucl. Med.* 53(8), 1284–1291 (2012)
12. Lamare, F., Ledesma Carbayo, M.J., Cresson, T., Kontaxakis, G., Santos, A., Le Rest, C.C., Reader, A.J., Visvikis, D.: List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations. *Phys. Med. Biol.* 52(17), 5187–5204 (2007)
13. Polycarpou, I., Tsoumpas, C., Marsden, P.K.: Analysis and comparison of two methods for motion correction in PET imaging. *Med. Phys.* 39(10), 6474–6483 (2012)
14. Chun, S.Y., Fessler, J.A.: Spatial resolution properties of motion-compensated tomographic image reconstruction methods. *IEEE Trans. Med. Imaging* 31(7), 1413–1425 (2012)
15. Rahmim, A., Tang, J., Zaidi, H.: Four-dimensional image reconstruction strategies in cardiac-gated and respiratory-gated PET imaging. *PET Clinics* 8(1), 51–67 (2013)
16. Winkelmann, S., Schaeffter, T., Koehler, T., Eggers, H., Doessel, O.: An optimal radial profile order based on the golden ratio for time-resolved MRI. *IEEE Trans. Med. Imaging* 26(1), 68–76 (2007)
17. Grimm, R., Bauer, S., Kiefer, B., Hornegger, J., Block, T.: Optimal channel selection for respiratory self-gating signals. In: *Proc. 21st Annual Meeting ISMRM*, Salt Lake City, Utah, USA, p. 3749 (2013)
18. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012*, Part III. LNCS, vol. 7512, pp. 115–122. Springer, Heidelberg (2012)
19. Dawood, M., Büther, F., Stegger, L., Jiang, X., Schober, O., Schäfers, M., Schäfers, K.P.: Optimal number of respiratory gates in positron emission tomography: a cardiac patient study. *Med. Phys.* 36(5), 1775–1784 (2009)

Complex Lung Motion Estimation via Adaptive Bilateral Filtering of the Deformation Field

Bartłomiej W. Papież¹, Mattias Paul Heinrich¹,
Laurent Risser², and Julia A. Schnabel¹

¹ Institute of Biomedical Engineering,

Department of Engineering Science, University of Oxford, UK

² CNRS, Institut de Mathématiques de Toulouse (UMR5219), France

Abstract. Estimation of physiologically plausible deformations is critical for several medical applications. For example, lung cancer diagnosis and treatment requires accurate image registration which preserves sliding motion in the pleural cavity, and the rigidity of chest bones. This paper addresses these challenges by introducing a novel approach for regularisation of non-linear transformations derived from a bilateral filter. For this purpose, the classic Gaussian kernel is replaced by a new kernel that smoothes the estimated deformation field with respect to the spatial position, intensity and deformation dissimilarity. The proposed regularisation is a spatially adaptive filter that is able to preserve discontinuity between the lungs and the pleura and reduces any rigid structures deformations in volumes. Moreover, the presented framework is fully automatic and no prior knowledge of the underlying anatomy is required. The performance of our novel regularisation technique is demonstrated on phantom data for a proof of concept as well as 3D inhale and exhale pairs of clinical CT lung volumes. The results of the quantitative evaluation exhibit a significant improvement when compared to the corresponding state-of-the-art method using classic Gaussian smoothing.

Keywords: nonrigid registration, respiratory motion, sliding motion modeling, adaptive bilateral filtering.

1 Introduction

Image registration is a key processing step in medical image analysis. However, common deformation models such as diffusion, elasticity or fluid methods usually do not reflect the underlying mechanisms (true tissue properties) of the tissue changes between the consecutive volumes. Therefore, additional constraints need to be introduced such as displacement field discontinuities (sliding motion) [8,7,9,1], rigidity [11] or incompressibility [5]. Applications such as diagnosis and image guided radiotherapy (IGRT) have attracted active research on accurate lung motion estimation over the last years [3,9,1].

The sliding motion pattern that naturally occurs at e.g. lungs and liver boundaries has been addressed by various image registration approaches. Direction dependent regularisation [9] decouples diffusion regularisation into normal and

tangential direction around lung boundaries, while the registration of the remaining part of the volume is based on a classic diffusion model. Similarly, anisotropic diffusion regularisation was utilised for a lung phantom data study in [6]. A different concept was proposed in [8] where the deformation is decomposed into basic components that are then regularised separately. Recently, Large Deformation Diffeomorphic Metric Mapping (LDDMM) was extended towards the piecewise-diffeomorphic registration that enables explicit sliding motion modelling [7]. These approaches have some limitations, as they require preprocessing steps in form of segmentation of some structures where sliding motion may occur [6,9], very accurate domain splitting strategy to ensure diffeomorphism [7], or a sliding motion detection system [9]. For approaches where no segmentations are required or available, the estimated deformations remain smooth at lung boundaries, or discontinuity preserving regularisation is applied to the whole volume domain [8].

Several methodologies that enforce the rigidity on some volume objects have been described in literature. The most related to the method presented in this paper was proposed in [11]. It is based on an iterative procedure of adaptive filtering of the deformation field that is employed for the area of rigid objects. The simplicity is however deceptive as it requires a *stiffness* coefficient which has to be derived either from segmentation or from Hounsfields unit if CT data are available. One such example is the recent work [1], where a fluid registration framework with preservation of topology and rigidity is proposed, which however also relies on the segmentation of the lung surface and bony structures.

This work aims to develop automated method for deformable registration to address both, the problem of sliding motion estimation and the local rigidity preservation. The contributions of this paper are as follows. First, we adapt the bilateral filter technique previously introduced for image filtering and occlusion detection in computer vision applications [13,15] to regularise the estimated deformation field in a Demons formulation [12]. The classic Gaussian kernel is then replaced by a new kernel that is dependent on the anisotropic diffusion, the intensity and deformation dissimilarity. The presented framework does not require any prior knowledge about the organs' properties and therefore it forms a fully automated technique.

2 Classic Diffusion Based Regularisation

In a classic non-linear image registration formulation [4,16], the optimal displacement field $\hat{\mathbf{u}}$ describing the geometrical transformation between a reference image I_R and a source image I_S is estimated by minimising a global energy:

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}} (Sim(I_R, I_S(\mathbf{u})) + \alpha Reg(\mathbf{u})) \quad (1)$$

where *Sim* is a similarity measure, *Reg* is a regularisation term, and α is a positive weighting factor. A common choice of similarity criterion is the sum of the squared differences (SSD) and a diffusion regularisation yielding:

$$\int_{\Omega} (I_R(\mathbf{x}) - I_S(\mathbf{x} + \mathbf{u}(\mathbf{x})))^2 + \alpha \text{tr}(\nabla \mathbf{u}(\mathbf{x})^T \nabla \mathbf{u}(\mathbf{x}))^2 dx \quad (2)$$

The diffusion regularisation is performed as Gaussian smoothing of the deformation field, therefore solving the Euler-Lagrange equations of Eq. (1) can be divided into two steps [15]: finding an update $\mathbf{d}\mathbf{u}$ that is related to the similarity measure Sim , and performing smoothing of the estimated deformation field using an isotropic Gaussian kernel G_{iso} instead of explicit regularisation $Reg(\mathbf{u})$:

$$\mathbf{u}_{new}(\mathbf{x}) = G_{iso} * (\mathbf{u}_{old}(\mathbf{x}) \circ \mathbf{d}\mathbf{u}(\mathbf{x})) \quad (3)$$

where \mathbf{u}_{new} is a new estimate of the deformation field, \mathbf{u}_{old} is a deformation field calculated in the previous iteration, and \circ is a composition operation. Filtering the deformation field using an isotropic Gaussian kernel leads to smooth deformations. In medical image registration, this framework can be related to the commonly used Demons algorithm [12,14].

3 Adaptive Bilateral Smoothing

In order to prevent the deformation field to be smoothed across object boundaries which would not be physically realistic, we propose to replace the standard Gaussian filtering of the deformation field by a more powerful non-linear filtering technique originally proposed for image denoising [13]. The bilateral filter smoothes an input image I_i by two Gaussian kernels in the following way:

$$I_o(\mathbf{x}) = \frac{1}{W} \sum_{\mathbf{y} \in \mathcal{N}} \underbrace{\exp\left(-\frac{(\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y})}{2\sigma_x^2}\right)}_{G_{iso}(\mathbf{x}, \mathbf{y})} \cdot \underbrace{\exp\left(-\frac{\|I_i(\mathbf{x}) - I_i(\mathbf{y})\|^2}{2\sigma_r^2}\right)}_{G_r(I(\mathbf{x}), I(\mathbf{y}))} \cdot I_i(\mathbf{y}) \quad (4)$$

where G_{iso} is Gaussian kernel on the spatial domain (with variance σ_x^2) and G_r is another Gaussian kernel but defined on the intensity domain I_i (with variance σ_r^2), \mathbf{y} is a spatial position within the image neighbourhood \mathcal{N} , and W is a normalisation factor for this image neighbourhood \mathcal{N} . Even though the additional kernel G_r does not allow for smoothing when the neighbourhood intensity values are different, direct substitution of the Gaussian kernel by the bilateral kernels in our registration framework can lead to several deformation field discontinuities (*motion over-segmentation*) in the whole volume domain. This discontinuity can occur at each intensity change and is typical for image-driven regularisation [16]. In addition to this, some organs have very similar intensity values, however they can slide along each other. Therefore, a supplementary kernel is employed [15] and the bilateral filtering of deformation field is extended in the following way:

$$\mathbf{u}_{new}(\mathbf{x}) = \frac{1}{W} \sum_{\mathbf{y} \in \mathcal{N}} G_{iso}(\mathbf{x}, \mathbf{y}) \cdot G_r(I(\mathbf{x}), I(\mathbf{y})) \cdot \underbrace{\exp\left(-\frac{(\mathbf{u}_{cur}(\mathbf{x}) - \mathbf{u}_{cur}(\mathbf{y}))^T(\mathbf{u}_{cur}(\mathbf{x}) - \mathbf{u}_{cur}(\mathbf{y}))}{2\sigma_u^2}\right)}_{G_u(\mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{y}))} \cdot \mathbf{u}_{cur}(\mathbf{y}) \quad (5)$$

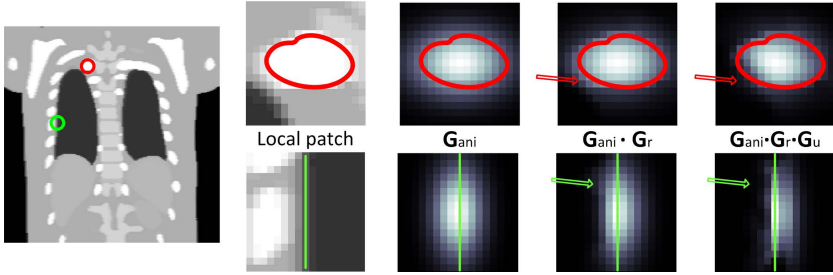


Fig. 1. Comparison between different kernels used for deformation filtering. Two local patches from distinctive areas of the chest were taken from the NCAT data. The patch presenting the chest bone (red circle) and its corresponding kernels are shown in the top row, while the patch including the lung boundary (green circle) and its corresponding kernels are shown in the bottom row.

where G_u describes a Gaussian kernel based on the local deformation field dissimilarity and $\mathbf{u}_{cur}(\mathbf{x}) = \mathbf{u}_{old}(\mathbf{x}) \circ d\mathbf{u}(\mathbf{x})$. Thus, the combination of three kernels G_{iso} , G_r , and G_u leads to a joint image- and deformation-driven regularisation method, which compromises mutual benefits such as discontinuous deformation without motion over-segmentation. Finally, we can also replace the isotropic Gaussian kernel G_{iso} by an oriented Gaussian kernel G_{ani} . After that substitution, the kernel G_{ani} varies at different image position \mathbf{x} with respect to the diffusion tensor \mathbf{D} . The diffusion tensor \mathbf{D} for n -dimensional volumes is defined as [4]:

$$\mathbf{D} = \frac{(\lambda + \|\nabla I\|^2)\mathbf{Id} - \nabla I \nabla I^T}{(n-1)\|\nabla I\|^2 + n\lambda} \quad (6)$$

where ∇I is the gradient of image I , λ is an (an)isotropy parameter, \mathbf{Id} is a $n \times n$ identity matrix. As can be expected, if the intensity values around point \mathbf{x} are (close to) constant ($\|\nabla I\| \approx 0$), the eigenvalues of \mathbf{D} are equal and the kernel G_{ani} is equivalent to the isotropic kernel G_{iso} .

Examples of different kernels are presented in Fig. 1. Filtering the deformation field with the anisotropic Gaussian kernel leads to smooth flow at lung boundaries (Fig. 1 bottom row) as the kernel averages the deformation field between neighbourhood areas. Whereas the bilateral filter working on both the spatial G_{ani} and intensity domain G_r improves the shape of the kernel, still some deformation field averaging is done (see area marked by arrows). While better separation between different structures can be only captured by limiting the size of the bilateral filter using only an intensity kernel (lower value of variance σ_r^2), it will lead to the aforementioned motion over-segmentation problem. Such example of motion over-segmentation in case of lung data registration is illustrated in Fig. 2c. The proposed composition of three kernels G_{ani} , G_r and G_u produces a kernel which visually has better overlap with the underlying anatomical structure than the two others. Similarly, the kernel comprising all of G_{ani} , G_r and G_u acts mostly inside the rigid structure (Fig. 1 top row).

Table 1. Average ratio of overlap obtained for the NCAT data set using the Demon framework with four different smoothing kernels. The last column shows statistical significance of improvement between **iso-dem** compared to others methods.

	lungs	liver	ribs	p-value
iso-dem	0.86±0.08	0.89±0.10	0.76±0.05	N/A
ani-dem	0.90±0.05	0.91±0.07	0.79±0.05	0.475
iso-bil	0.92±0.02	0.93±0.05	0.80±0.05	0.061
ani-bil	0.92±0.02	0.93±0.05	0.80±0.05	0.067

4 Results

Materials. The proposed approach is assessed using two publicly available data sets. The first data set is a set of synthetically generated 4D CT volumes modelling consecutive respiratory cycle phases from the NCAT phantom [10]. The spatial resolution of the data is $2.0 \times 2.0 \times 2.0 \text{ mm}^3$. The second data set consists of ten 4D CT data from the *Dir-Lab* data set [2]. The spatial resolution of that data varies between $0.97 \times 0.97 \times 2.5 \text{ mm}^3$ and $1.16 \times 1.16 \times 2.5 \text{ mm}^3$. For all data sets, the end-of-inspiration volume was chosen as a reference, and the end-of-expiration volume as a source image.

Experiments. For quantitative evaluation of the proposed regularisation filters, a Demon approach with an update composition scheme was implemented (see [14] for details). Four different kernels for smoothing the deformation field were used for comparison: isotropic Gaussian G_{iso} (**iso-dem**), anisotropic Gaussian G_{ani} (**ani-dem**), bilateral kernel with isotropic Gaussian $G_{iso} \cdot G_r \cdot G_u$ (**iso-bil**) and with anisotropic Gaussian $G_{ani} \cdot G_r \cdot G_u$ (**ani-bil**). In most cases, the registration with a bilateral filter without deformation similarity kernel $G_{iso} \cdot G_r$ produces unrealistic deformation fields (see example in Fig. 2c), therefore the quantitative results obtained are not included. Additionally, a two-sample Wilcoxon rank sum test was performed between **iso-dem** and the other evaluated methods. Filtering of the deformation field was performed once after each update of deformation field, but in principle it could be done several times at each iteration following the approach presented in [11]. Although the best design parameters ($\sigma_x, \sigma_r, \sigma_u, \lambda$) were determined empirically by an extensive search over parameter space for each method, experiments on both data sets showed that the proposed algorithm is very robust to their choice, hence these parameters do not need to be tuned for each volume separately (for this particular application i.e. lung CT).

Results on NCAT Data. The ratio of overlap (RO) for the organ of interest (lungs, liver, ribs) was calculated and the registration outcomes for the NCAT data are presented in Tab. 1. The RO exhibits an improvement for methods based on the bilateral filtering when compared with the Gaussian smoothing. Although this does not necessarily ensure deformation field plausibility, it can highlight differences between methods in terms of the anatomical correspondence.

Results on CT Dir-Lab Data. The TRE was calculated for the landmarks which are included in this data set (300 per case) and the results of quantitative evaluation can be found in Tab. 2. The initial average TRE is 8.46 ± 5.48 mm. As can be seen, the deformation fields obtained using both frameworks based on the bilateral filtering produce significantly lower TRE when compared to the classic Demon algorithm. Contrary to expectation, the bilateral filtering with anisotropic kernel G_{ani} performs slightly worse than the method with isotropic kernel G_{iso} . This indicates that the bilateral filtering with isotropic Gaussian kernel can effectively adapt smoothing across different structures. Moreover, the proposed methods based on bilateral filtering yields a lower TRE (2.34 mm) whereas the classic Demon has a TRE=2.88 mm. The results reported in the literature [9] were 3.02 mm for diffusion regularisation and 2.13 mm for direction dependent regularisation. It must be noted that further improvement might be expected when a more advanced similarity Sim would be applied to capture local intensity variations apparent in the Dir-Lab data due to lung compression [2].

Fig. 2 is an illustrative example of the deformation field magnitudes when registering case *c5* from the Dir-Lab data set using different smoothing kernels. The results from the quantitative evaluation (shown in fifth row of Tab. 2) exhibit a statistically significant improvement in terms of the TRE between different methods, and consequently some noticeable differences between the estimated deformation fields can be identified especially close to the lung boundaries (compare Fig. 2b and Fig. 2d, and its corresponding zoomed images in Fig. 2f and Fig. 2h). Employing bilateral filtering derived both from intensity and deformation field similarity preserves discontinuity between the lungs and the pleura, while satisfying smoothness of the deformation field inside the lungs. Contrary, applying the bilateral filter based only on intensity difference generates discontinuities inside and outside the pleural cavity (depicted by black arrows in Fig. 2c).

Table 2. Target Registration Error and its standard deviations obtained for Dir-Lab data set using the Demon framework with four different smoothing kernels. The last column shows statistical significance of improvement between **iso-dem** compared to others methods with p -value level below 0.05 (marked as (+)). The proposed **iso-bil** achieves the lowest average TRE among all methods.

No. data	before	iso-dem	ani-dem	iso-bil	ani-bil	sign.
<i>c1</i>	3.89 ± 2.78	1.08 ± 0.57	1.09 ± 0.58	1.05 ± 0.54	1.07 ± 0.57	- - -
<i>c2</i>	4.34 ± 3.90	1.11 ± 0.64	1.10 ± 0.63	1.08 ± 0.58	1.09 ± 0.60	- - -
<i>c3</i>	6.94 ± 4.05	1.54 ± 0.98	1.52 ± 0.91	1.47 ± 0.86	1.49 ± 0.89	- ++
<i>c4</i>	9.83 ± 4.86	2.38 ± 2.04	2.38 ± 2.03	2.28 ± 1.82	2.40 ± 2.04	- + -
<i>c5</i>	7.48 ± 5.51	2.26 ± 1.93	2.22 ± 1.90	2.04 ± 1.71	2.13 ± 1.78	- ++
<i>c6</i>	10.9 ± 6.97	3.26 ± 2.84	3.14 ± 2.67	2.72 ± 2.01	3.05 ± 2.49	+++
<i>c7</i>	11.0 ± 7.43	3.81 ± 3.69	3.62 ± 3.47	3.14 ± 2.76	3.53 ± 3.33	+++
<i>c8</i>	15.0 ± 9.01	8.22 ± 8.27	7.73 ± 8.26	4.99 ± 5.14	7.19 ± 7.67	+++
<i>c9</i>	7.92 ± 3.98	2.56 ± 2.02	2.46 ± 1.89	2.08 ± 1.45	2.36 ± 1.79	+++
<i>c10</i>	7.30 ± 6.35	2.65 ± 3.27	2.61 ± 3.21	2.45 ± 2.78	2.57 ± 3.11	- ++
<i>mean</i>	8.46 ± 5.48	2.88 ± 2.07	2.79 ± 1.92	2.34 ± 1.16	2.67 ± 1.77	

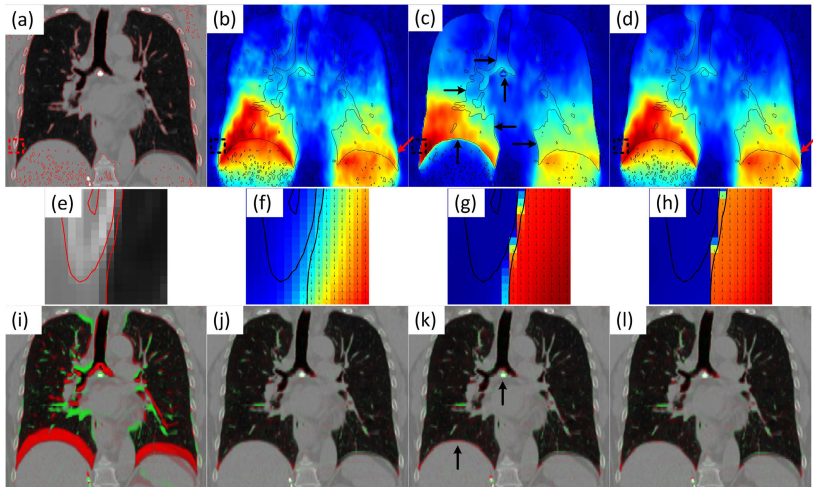


Fig. 2. Results for case *c5* of the Dir-Lab data set. (a) coronal view of the reference image with the corresponding contours. The magnitude of the deformation fields estimated using: (b) isotropic Gaussian kernel G_{iso} (Demon), (c) original bilateral kernel $G_{iso} \cdot G_r$, (d) the proposed bilateral kernel $G_{iso} \cdot G_r \cdot G_u$, (e)-(h) zoomed images of the region of interest (labelled by squared box in the top row), the intensity differences between input images (i) before registration, and after using (j) **iso-dem**, (k) original bilateral kernel $G_{iso} \cdot G_r$, and (l) **iso-bil**. Registration using **iso-bil** yields smooth deformation inside the pleura cavity whilst preserving sliding motion at the lung boundary.

5 Discussion and Conclusions

This paper presents an image registration framework which is able to estimate deformation fields preserving both the sliding motion in the cavity of the pleura whilst preserving the rigidity of the chest bones and yielding desirable smooth deformation field inside the lungs. The overall deformation field is regularised within a one step procedure that is performed via adaptive deformation field filtering. The kernel which is used for the purpose of deformation field filtering, is based on three components: spatial smoothness, local image intensity and deformation field similarity. Evaluation of the proposed regularisation scheme was done both on the NCAT phantom data and clinical lung CT data. In cases where noticeable sliding motion occurs in the data, the presented results exhibit significant improvements when the new filtering procedure is applied compared to the classic Gaussian smoothing. Moreover, for *non-sliding* cases (where statistical significance of the improvements for the proposed filtering procedure was not achieved) the slightly lower average TRE was obtained, and in addition, the visual inspection of the estimated deformation fields still exhibited physiologically more plausible results. Future work will perform a sensitivity analysis of design parameters of the proposed filtering procedure such as σ_x , σ_r , σ_u to this

application, and comparison to methods that require segmentation or explicit sliding motion detection to achieve desired properties [7,9].

Acknowledgements. We would like to acknowledge funding from the CRUK/EPSRC Oxford Cancer Imaging Centre. JAS and LR also wish to acknowledge the INSMI-CNRS/John Fell Oxford University Press (OUP) Research Fund.

References

1. Baluwala, H., Risser, L., Schnabel, J.A., Saddi, K.: Toward physiologically motivated registration of diagnostic CT and PET/CT of lung volumes. *Med. Phys.* 40, 021903 (2013)
2. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys. Med. Biol.* 54, 1849–1870 (2009)
3. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., Handels, H.: Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. *IEEE Trans. Med. Imag.* 30, 251–265 (2011)
4. Hermosillo, G., Chef d’Hotel, C., Faugeras, O.: Variational Methods for Multimodal Image Matching. *Int. J. Comput. Vision* 50, 329–343 (2002)
5. Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N.: iLogDemons: A Demons-Based Registration Algorithm for Tracking Incompressible Elastic Biological Tissues. *Int. J. Comput. Vision* 92, 92–111 (2011)
6. Pace, D.F., Enquobahrie, A., Yang, H., Aylward, S.R., Niethammer, M.: Deformable image registration of sliding organs using anisotropic diffusive regularization. In: *IEEE ISBI*, pp. 407–413 (2011)
7. Risser, L., Vialard, F.X., Baluwala, H., Schnabel, J.A.: Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions. *Med. Image Anal.* 17, 182–193 (2013)
8. Ruan, D., Esedoglu, S., Fessler, J.A.: Discriminative Sliding Preserving Regularization in Medical Image Registration. In: *IEEE ISBI*, pp. 430–433 (2009)
9. Schmidt-Richberg, A., Werner, R., Handels, H., Ehrhardt, J.: Estimation of slipping organ motion by registration with direction-dependent regularization. *Med. Image Anal.* 16, 150–159 (2012)
10. Segars, W.P.: Development and application of the new dynamic NURBS-based cardiac-torso (NCAT) phantom. PhD thesis, University of North Carolina (2001)
11. Staring, M., Klein, S., Pluim, J.P.W.: Nonrigid registration with tissue-dependent filtering of the deformation field. *Phys. Med. Biol.* 52, 6879–6892 (2007)
12. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell’s demons. *Med. Image Anal.* 2, 243–260 (1998)
13. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: *IEEE ICCV*, pp. 839–846 (1998)
14. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic Demons: Efficient non-parametric image registration. *NeuroImage* 45, 61–72 (2009)
15. Xiao, J., Cheng, H., Sawhney, H., Rao, C., Isnardi, M.: Bilateral filtering-based optical flow estimation with occlusion detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006, Part I. LNCS*, vol. 3951, pp. 211–224. Springer, Heidelberg (2006)
16. Zimmer, H., Bruhn, A., Weickert, J.: Optic Flow in Harmony. *Int. J. Comput. Vision* 93, 368–388 (2011)

Helical Mode Lung 4D-CT Reconstruction Using Bayesian Model

Tiancheng He, Zhong Xue*, Paige L. Nitsch, Bin S. Teh, and Stephen T. Wong

The Methodist Hospital Research Institute, The Methodist Hospital,
Weill Cornell Medical College, Houston, TX, US
zxue@tmhs.org

Abstract. 4D computed tomography (CT) has been widely used for treatment planning of thoracic and abdominal cancer radiotherapy. Current 4D-CT lung image reconstruction methods rely on respiratory gating to rearrange the large number of axial images into different phases, which may be subject to external surrogate errors due to poor reproducibility of breathing cycles. New image-matching-based reconstruction works better for the cine mode of 4D-CT acquisition than the helical mode because the table position of each axial image is different in helical mode and image matching might suffer from bigger errors. In helical mode, not only the phases but also the un-uniform table positions of images need to be considered. We propose a Bayesian method for automated 4D-CT lung image reconstruction in helical mode 4D scans. Each axial image is assigned to a respiratory phase based on the Bayesian framework that ensures spatial and temporal smoothness of surfaces of anatomical structures. Iterative optimization is used to reconstruct a series of 3D-CT images for subjects undergoing 4D scans. In experiments, we compared visually and quantitatively the results of the proposed Bayesian 4D-CT reconstruction algorithm with the respiratory surrogate and the image matching-based method. The results showed that the proposed algorithm yielded better 4D-CT for helical scans.

Keywords: Bayesian estimation, respiratory motion, 4D-CT reconstruction.

1 Introduction

4D-CT has been widely used for radiation therapy planning of lung cancer for defining the clinical target volume (CTV) and planning target volume (PTV) to ensure that the radiation dose covers CTV, is within PTV, and does not damage neighboring critical tissues during respiratory cycles [1, 2]. 4D-CT scanning captures a large number of axial images during multiple breathing cycles using cine or helical modes, and reconstructs them to a series of 3D-CT images [3]. The cine mode captures multiple axial images in respiratory cycles at each table position; while the helical mode performs the scans when the table is slowly and continuously moving.

Efforts have been made to either using respiratory sensors such as gating and optical tracking to capture detailed respiratory motion patterns or using image computing

* Corresponding author.

methods to retrospectively improve the image sorting. Traditionally, using surrogate respiratory signals from a chest height marker, a strain gauge or a spirometer [4], respiratory cycles are detected and divided into a number of respiratory phases. Then, the synchronized axial images are rearranged to reconstruct the serial 3D images. However, because of the lack of reproducibility of breathe cycles, such gating signals appear to be not exactly periodical and may miss group some axial images, resulting discontinuity of anatomical structures in the images [5-7]. Recent studies have attempted to reconstruct 4D-CT through image computing in the cine mode [8-11]. For helical mode, since each axial image has a different table position, such additional variable should be considered in the reconstruction, and few works have been reported in the literature. As more helical 4D scans are being used in radiotherapy planning, it is highly desirable to study the methods for its image reconstruction.

This paper proposes an automated 4D-CT reconstruction algorithm for helical scanning based on the Bayesian framework, referred to as Bayesian 4D-CT reconstruction. The objective for lung 4D-CT image reconstruction is to preserve the anatomical structures at each time-point, while the image sequence reflects underlying respiratory motion. Spatial and temporal smoothness of certain surfaces of anatomical structures can be used as constraints in the reconstruction. In the Bayesian 4D-CT reconstruction algorithm, image sorting is jointly estimated with an underlying ideal image sequence whose surface's spatial-temporal properties are subject to such smoothness constraints. A novel energy function is designed and formulated in the Bayesian framework, and the optimization is achieved by iteratively assigning axial images to their best phase, and at the same time, enforcing spatial-temporal surface smoothness. Finally, due to the nature of helical 4D scanning, the image and surface matching also takes into account the unequal inter-slice distances of axial images in each respiratory phase, and the final reconstructed images are generated using a cubic B-Spline-based interpolation.

In experiments, we used the images from thirty nine patients undergoing radiotherapy planning to validate the algorithm. The final reconstructed images were compared visually and quantitatively with the external surrogate-based reconstruction and the image matching-based method [10] that are currently used in radiotherapy planning. For quantitative comparison, we compared the spatial and temporal smoothness of the surfaces extracted from all the results. The results indicated that our method outperformed both methods: visualization of the CT images showed less artifacts, particularly in the regions close to the diaphragm; and quantitative results showed that the surfaces extracted from the resultant images were smoother, so there was less sudden bumps along the image boundaries.

2 Method

2.1 Algorithm Formulation

During the helical mode scanning, axial images are captured while the table is slowly and continuously moving. Depending on the slice thickness and the number of simultaneous slices the scanner can capture (e.g., multiple row detector CT), the table speed can be determined so that the axial images captured within a small position range cover an entire respiratory cycle. Using the synchronized surrogate respiratory signal

the axial images can be initially resorted to different respiratory phases, and each 3D image is formed by the axial images according to their table positions. The goal for the proposed Bayesian 4D-CT reconstruction algorithm is to assess such assignment and correct the miss grouped ones so that the 3D images at each phase preserve anatomical structures, and the 4D data reflect respiratory motion well. Thus, spatial and temporal smoothness of surfaces act as the key for enforcing this requirement. We used the smoothness constraint of chest surface to ensure the quality of the reconstructed images. The reason is that it is smoother compared to internal organs, and it is also possible to extend our algorithm to use computer vision-based method to track the chest surface. Suppose the set of all axial images is D , the chest surfaces of an underlying ideal image sequences are $R = \{R_1, R_2, \dots, R_K\}$, and R and D are independent, the goal of our reconstruction algorithm is to jointly estimate a new image series $S = \{s_1, s_2, \dots, s_K\}$ and the ideal surfaces R by maximizing the following joint posterior distribution, where K is number of respiratory phases in one breathing cycle:

$$P(S, R|D) = \frac{P(D|S)P(S,R)}{P(D)}. \quad (1)$$

The joint probability of S and R can be expressed as,

$$P(S, R) = P(S|R)P(R). \quad (2)$$

Combining Eq. (1) and Eq. (2), and assuming that the probability of the known axial image set D is 1 ($P(D) = 1$), S and R can be estimated by

$$(S^*, R^*) = \operatorname{argmax}\{P(S, R|D)\} = \operatorname{argmax}\{P(D|S)P(S|R)P(R)\}. \quad (3)$$

When the probabilities are estimated using the Gibbs distribution, the maximization of the joint posterior distribution is equivalent to minimizing the energy function:

$$E(S, R) = E(D|S) + \alpha E(S|R) + \beta E(R). \quad (4)$$

α and β are the weighting factors. The first term $E(D|S)$ denotes the degree of matching between the serial image S and the observed data D , and it can be calculated by the normalized cross correlation (NCC) between the two image series:

$$E(D|S) = \sum_{k=1}^K -NCC(D_k, S_k). \quad (5)$$

The second term $E(S|R)$ stands for the degree of matching between S and R , with R as the underlying ideal surfaces. Here, $E(S|R)$ is defined by the distance between ideal surface R and the surfaces extracted from S ,

$$E(S|R) = \sum_{k=1}^K \operatorname{dist}(G(S_k), R_k), \quad (6)$$

where the distance $\operatorname{dist}()$ is calculated according to [12]. $G(S_k)$ represents the surface extracted from S_k . The third term of Eq. (4) represents the prior shape constraints of R . In this case, it consists of the spatial and temporal smoothness constraints of the chest surface series. Because it is not necessary to constrain the surface within each axial plane, we only need to consider the smoothness in z -direction as well as in the time-domain (between neighboring phases). Thus, $E(R)$ is calculated as:

$$E(R) = \frac{1}{K} \sum_{k=1}^K \frac{1}{|\Omega|} \sum_{\mathbf{v}} \left(\frac{\partial R_k(\mathbf{v})}{\partial z} \right)^2 + \lambda \frac{1}{K-1} \sum_{k=1}^{K-1} \frac{1}{|\Omega|} \sum_{\mathbf{v}} \left(\mathbf{f}_{k+1}(\mathbf{v} + \mathbf{f}_k(\mathbf{v})) - \mathbf{f}_k(\mathbf{v}) \right)^2, \quad (7)$$

where the first term is the average of the squared surface gradients along z -direction by considering the unequal slice distances, and the second term calculates the average of the temporal smoothness of the deformation field $\mathbf{f}_k, k = 1, \dots, K-1$ across the

image sequence. λ is the tradeoff between them, and Ω is the surface point set of the lung image in phase k .

Compared to the maximizing a posteriori (MAP) formulation, the major difference of the proposed algorithm is that an intermediate ideal surface R is jointly estimated together with S . This helps facilitate the additional spatial and temporal anatomical constraints to the reconstructed 4D-CT images. Finally, after assigning each axial image into their phase by minimizing the energy function defined in Eq. (4), the slices of each phase are arranged according to their table positions. Because of the unequal slice distances, we then use a cubic B-Spline-based interpolation tool to resample them and reconstruct the 3D image sequences with equal slice distance.

2.2 Implementation

The optimization of the energy function in Eq. (4) can be implemented by alternatively calculating R and S . Given a series of N axial images (for lung imaging, N is more than 1000), we can first sort them into K (typically 10) phases based on the surrogate signals, which gives the current data observation D . We use D as the initialization of S and iteratively perform the following two optimization steps:

Step 1. Optimize the ideal serial surfaces R by fixing S . By extracting the surfaces from the images of S and performing longitudinal surface registration [13], we obtain the current surface series R , and their longitudinal deformations $\mathbf{f}_k, k = 1, \dots, K - 1$. Then, R can be optimized using the finite gradient descent method:

$$R \leftarrow R - \xi \partial E(S, R) / \partial R, \quad (8)$$

where ξ is the updating step.

Step 2. Optimize the image sequences S by fixing R . We iterate all the axial images and re-assign each to the i th phase that gives the minimal energy function:

$$i = \operatorname{argmin}_k (E(S, R)) = \operatorname{argmin}_k (E(D|S) + \alpha E(S|R)). \quad (9)$$

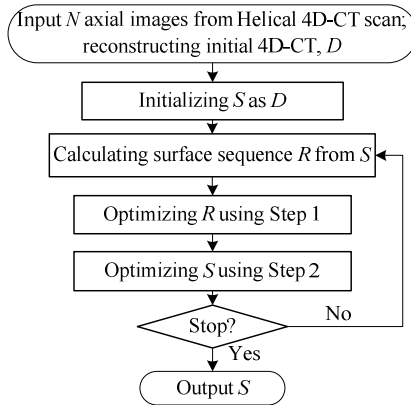


Fig. 1. The framework of the 4D-CT reconstruction algorithm

Notice that because of the nature of helical mode scanning, the distances between neighboring slices within each 3D CT image S_k are different. The surface registration method we used can register two surfaces with different meshes, and thus it can handle this issue. The optimization algorithm stops until the number of phase re-assignment is smaller than a prescribed number (5 in our case), and the algorithm generally stops after 3-4 iterations. Fig. 1 summarizes the process of the algorithm.

3 Results

The datasets of thirty nine patients were used in the experiments. The data were collected using Philips Pinnacle³ in helical mode. The number of slices per scan is around 1330. Slice thickness is 3.0 mm, and pixel spacing in the X and Y directions is $1.17mm \times 1.17mm$. Elastic belt was used for monitoring the breath. Initial respiratory gating-based 4D-CT reconstruction was performed on the Pinnacle machine, which was used as the initialization of our algorithm. The datasets were then transferred to our workstation from PACS, and the proposed Bayesian 4D-CT reconstruction was applied to refine the results using a workstation running Microsoft Windows 7 professional with an Intel i7 CPU (2.30GHz) and 8.00 GB of RAM.

We compared the reconstruction results with two other methods. The first is the one reconstructed by the Pinnacle machine based on respiratory belt gating, and the second is the image matching-based image reconstruction proposed by Carnes *et al.* [10]. The Carnes algorithm first assigns manually the initial axial images into different respiratory phases and then uses slice-by-slice matching to sort the rest axial images. NCC is used as the image similarity measure. To automate this procedure, we used the assignment results of the first 20 axial images from the Pinnacle machine as the initialization of the Carnes algorithm. For our method, α and β were selected as 0.5 divided by the mean value of the corresponding energy functions. λ was selected so that the weight for temporal smoothness was half of the spatial smoothness. After reconstruction, we first visually assessed all the data. For the surrogate method, the artifacts of miss-assignment appeared more frequently, and we can also notice some

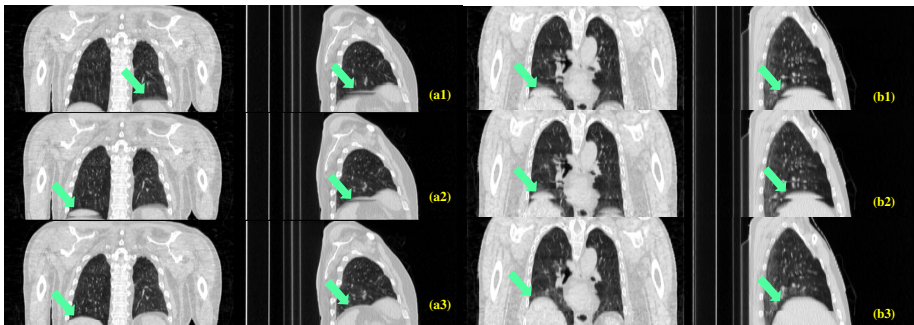


Fig. 2. Visual comparison of 4D-CT reconstruction results. Top: surrogate method; middle: Carnes algorithm; and bottom: the proposed Bayesian 4D-CT reconstruction.

similar discontinuity of the anatomical structures for the Carnes algorithm. Overall, the proposed Bayesian 4D-CT reconstruction preserved the anatomical structure in each 3D CT image much better. Fig. 2 illustrates some examples of the results. The top row shows the results of surrogate method, the middle row shows those of the Carnes algorithm, and the bottom row gives the reconstruction results of the proposed algorithm. Because the areas close to the diaphragm are subject to larger motion, we can notice the artifacts easily for the methods compared, and such motion artifacts have been corrected using the Bayesian 4D-CT reconstruction.

For quantitative comparison, we calculated the spatial and temporal smoothness about the chest surfaces and the lung field surfaces (extracted from the resultant CT images using [14]). Similar to Eq. (7), the spatial smoothness M_s of each subject is defined by the average absolute values of the surface gradients along z-direction,

$$M_s = \frac{1}{K} \sum_{k=1}^K \frac{1}{|\Omega|} \sum_{\mathbf{v} \in \Omega} |\partial R_k(\mathbf{v}) / \partial z|. \quad (10)$$

The temporal smoothness M_T is calculated from the longitudinal deformation fields of the serial surfaces extracted:

$$M_T = \frac{1}{K-1} \sum_{k=1}^{K-1} \frac{1}{|\Omega|} \sum_{\mathbf{v} \in \Omega} |\mathbf{f}_{k+1}(\mathbf{v} + \mathbf{f}_k(\mathbf{v})) - \mathbf{f}_k(\mathbf{v})|. \quad (11)$$

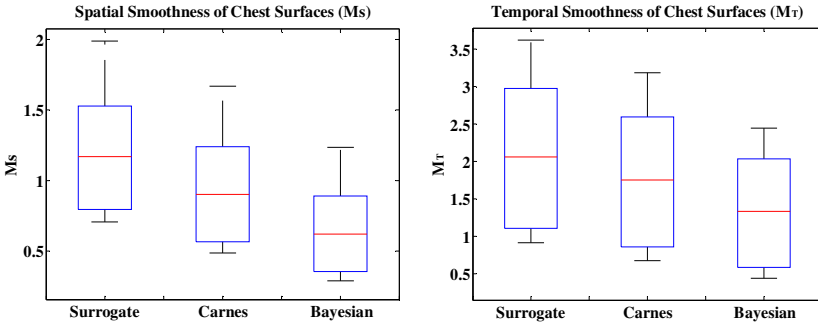


Fig. 3. Comparison of spatial and temporal smoothness of chest surfaces

Fig. 3 is the boxplot of the spatial and temporal smoothness of the chest surfaces for all 39 subjects in the experimental dataset. It can be seen that the proposed Bayesian 4D-CT reconstruction algorithm yielded more spatially and temporally smoother chest surfaces. Because we did not change the original axial images (only cubic B-Spline-interpolation was used), larger average smoothness value may indicate that there are more slices with artifacts in the reconstructed data. Therefore, the quantitative results indicate that there are less sudden jumps of the surfaces or less artifacts as compared to other methods. We also calculated the spatial and temporal smoothness of the lung fields extracted from the experimental results, and similar conclusion can be drawn from the boxplot shown in Fig. 4. It is worth noting that the spatial smoothness for lung field is bigger than that of the chest surface. This may indicate that chest surface is smoother and is suitable for applying the smoothness constraints.

Notice that the chest surface smoothness might be biased because it is also used in the energy function. Since the lung field surfaces were not used in the algorithm, the

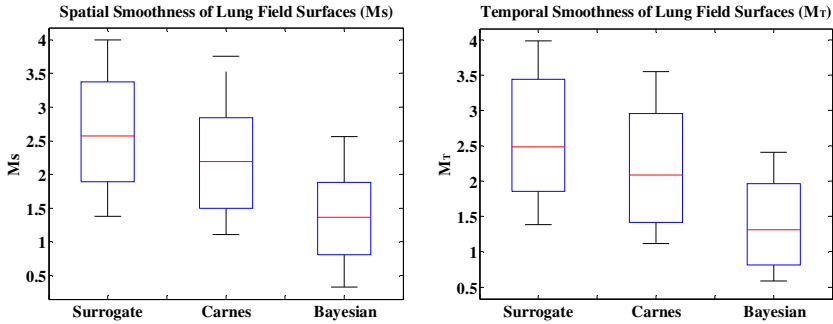


Fig. 4. Comparison of spatial and temporal smoothness of lung field surfaces

spatial and temporal smoothness metrics for the lung field surfaces extracted from the reconstructed images would be more appropriate. Due to the lack of ground truth of the 4D-reconstructed patient data, it is hard to conceive other relevant quantitative metrics at this point. In the future, we would like to further validate the quality of reconstruction using simulated images with known 4D-CT deformation patterns.

To further validate the results, all the reconstructed images were visually evaluated by two expert radiologists. Each image was visually assessed and the number of slices with artifacts (namely with noticeable sudden anatomical jumps) was counted. Fig. 5 illustrates the box plots of such numbers of slices with artifacts. The results confirmed the superiority of the proposed method as compared to others.

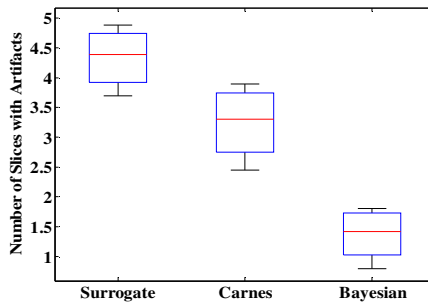


Fig. 5. Average numbers of slices with artifacts of 39 subjects

4 Conclusion

We proposed a Bayesian 4D-CT reconstruction algorithm for helical mode lung scanning. To preserve anatomical structures a joint Bayesian estimation is designed to ensure spatial and temporal smoothness of surfaces in the reconstructed 4D-CT images. Using clinical datasets for patients undergoing radiotherapy planning, we visually and quantitatively compared the performance of the proposed algorithm with the current surrogate and image-matching-based methods. The results showed that the

proposed algorithm yielded much less artifacts. In the future, we plan to incorporate vision-based chest surface monitoring devices in the framework for 4D-CT reconstruction on the fly.

References

1. Wu, G., Lian, J., Shen, D.: Improving image-guided radiation therapy of lung cancer by reconstructing 4D-CT from a single free-breathing 3D-CT on the treatment day. *Medical Physics* 39, 7694–7709 (2012)
2. Wink, N., Panknin, C., Solberg, T.D.: Phase versus amplitude sorting of 4D-CT data. *Journal of Applied Clinical Medical Physics/American College of Medical Physics* 7, 77–85 (2006)
3. Pan, T.: Comparison of helical and cine acquisitions for 4D-CT imaging with multislice CT. *Medical Physics* 32, 627–634 (2005)
4. Lu, W., Parikh, P.J., Hubenschmidt, J.P., Bradley, J.D., Low, D.A.: A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4D CT. *Medical Physics* 33, 2964–2974 (2006)
5. Ehrhardt, J., Werner, R., Saring, D., Frenzel, T., Lu, W., Low, D., Handels, H.: An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing. *Medical Physics* 34, 711–721 (2007)
6. Johnston, E., Diehn, M., Murphy, J.D., Loo Jr., B.W., Maxim, P.G.: Reducing 4D CT artifacts using optimized sorting based on anatomic similarity. *Medical Physics* 38, 2424–2429 (2011)
7. Han, D., Bayouth, J., Song, Q., Bhatia, S., Sonka, M., Wu, X.: Feature guided motion artifact reduction with structure-awareness in 4D CT images. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1057–1064. IEEE (2011)
8. Zeng, R., Fessler, J.A., Balter, J.M., Balter, P.A.: Iterative sorting for 4DCT images based on internal anatomy motion. In: *4th IEEE International Symposium on Biomedical Imaging*, pp. 744–747. IEEE (2007)
9. Li, R., Lewis, J.H., Cervino, L.I., Jiang, S.B.: 4D CT sorting based on patient internal anatomy. *Physics in Medicine and Biology* 54, 4821–4833 (2009)
10. Carnes, G., Gaede, S., Yu, E., Van Dyk, J., Battista, J., Lee, T.Y.: A fully automated non-external marker 4D-CT sorting algorithm using a serial cine scanning protocol. *Physics in Medicine and Biology* 54, 2049–2066 (2009)
11. Gianoli, C., Riboldi, M., Spadea, M.F., Travaini, L.L., Ferrari, M., Mei, R., Orecchia, R., Baroni, G.: A multiple points method for 4D CT image sorting. *Medical Physics* 38, 656–667 (2011)
12. 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)
13. Thomas Yeo, B.T., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical demons: Fast surface registration. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I*. LNCS, vol. 5241, pp. 745–753. Springer, Heidelberg (2008)
14. Zhu, X., Xue, Z., Gao, X., Zhu, Y., Wong, S.T.C.: Voles: Vascularity-Oriented Level Set Algorithm for Pulmonary Vessel Segmentation in Image Guided Intervention Therapy. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009*, pp. 1247–1250. IEEE Press, Boston (2009)

3D Tongue Motion from Tagged and Cine MR Images

Fangxu Xing¹, Jonghye Woo^{1,2}, Emi Z. Murano³, Junghoon Lee^{1,4},
Maureen Stone², and Jerry L. Prince¹

¹ Department of Electrical and Computer Engineering, Johns Hopkins University

² Department of Neural and Pain Sciences, University of Maryland Dental School

³ Department of Otolaryngology, Johns Hopkins School of Medicine

⁴ Department of Radiation Oncology and Molecular Radiation Sciences,
Johns Hopkins School of Medicine, Baltimore, MD, USA

fxing1@jhu.edu

Abstract. Understanding the deformation of the tongue during human speech is important for head and neck surgeons and speech and language scientists. Tagged magnetic resonance (MR) imaging can be used to image 2D motion, and data from multiple image planes can be combined via post-processing to yield estimates of 3D motion. However, lacking boundary information, this approach suffers from inaccurate estimates near the tongue surface. This paper describes a method that combines two sources of information to yield improved estimation of 3D tongue motion. The method uses the harmonic phase (HARP) algorithm to extract motion from tags and diffeomorphic demons to provide surface deformation. It then uses an incompressible deformation estimation algorithm to incorporate both sources of displacement information to form an estimate of the 3D whole tongue motion. Experimental results show that use of combined information improves motion estimation near the tongue surface, a problem that has previously been reported as problematic in HARP analysis, while preserving accurate internal motion estimates. Results on both normal and abnormal tongue motions are shown.

Keywords: Tongue, motion, HARP, registration, 3D, surface.

1 Introduction

The human tongue moves rapidly in complex and incompressible motions during speech [1]. In post-glossectomy patients, i.e., people who have had surgical resection of part of the tongue muscle for cancer or sleep apnea treatment, tongue moving ability and its speech functionality may be adversely affected. Therefore, understanding the tongue motion during speech in both normal and post-glossectomy subjects is of great interest to speech scientists, head and neck surgeons, and their patients.

To capture the tongue's motion during speech, tagged magnetic resonance (MR) images can be acquired over a series of time frames spanning a speech utterance [2,3]. The two-dimensional (2D) motion information carried in these

images can be extracted using the harmonic phase (HARP) algorithm [4]. With a collection of 2D motions from image slices covering the tongue, a high-resolution three-dimensional (3D) motion estimate can be achieved by interpolation with previously reported incompressible deformation estimation algorithm (IDEA) [5].

However, since HARP uses a bandpass filter to extract the harmonic images, object boundaries are blurred and motion estimates near the anatomical surfaces are inaccurate [6,7]. To make matters worse, HARP measurements near the boundaries are sparse because of the sparseness of image plane acquisition. These two problems severely affect 3D motion estimation near anatomical surfaces, as shown in Fig. 1. Zooming in on the back of the tongue (see Fig. 1(a)), 1(b) shows the sparse 2D motion components from HARP and 1(c) is the IDEA reconstruction of 3D motion that shows inaccurate large motion.

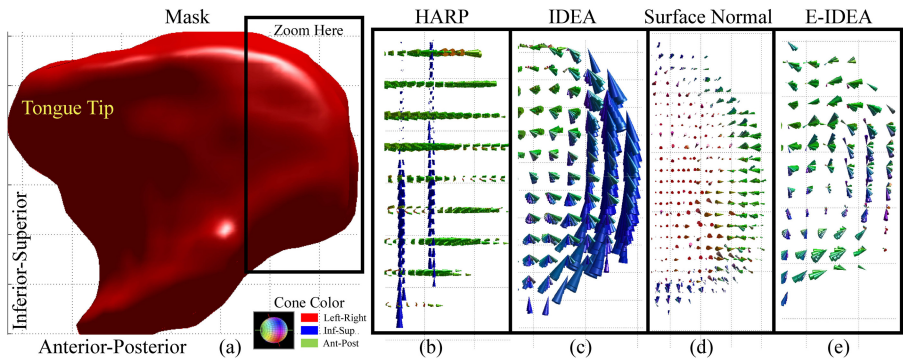


Fig. 1. (a) Tongue mask of a normal control subject (sagittal view). (b) HARP field on axial and coronal slices as input for IDEA, zoomed in at the tongue back. (c) IDEA result at the tongue back. (d) Surface normal deformation component at tongue back surface. (e) Proposed method result. Note: In this paper *cones* are used to visualize motion fields, where cone size indicates motion magnitude and cone color follows conventional DTI scheme (see cone color diagram).

This paper presents a novel approach that combines data from tagged images with surface deformation information derived from cine MR images to dramatically improve 3D tongue motion estimation. At every time frame, the tongue is segmented to achieve a 3D mask, and the deformation between the reference mask at the resting position and the deformed mask is computed using deformable registration. The normal components of surface deformation are then used to augment the HARP measurements within the IDEA estimation framework. Fig. 1(d) shows the additional input and Fig. 1(e) shows the result of proposed method. Comparing with Fig. 1(c), this result is more sensible from a qualitative point of view. Quantitative evaluations provided below also show that this method achieves a more accurate estimate of the whole tongue motion.

2 Methods

2.1 Data Acquisition and HARP Tracking

In this study, subjects repeatedly speak an utterance “a souk” during which tagged and cine MR image sequences are acquired at multiple parallel axial slice locations covering the tongue. The resolution scheme is 1.88 mm in-plane (dense) and 6.00 mm through-plane (sparse). For tagged images, both horizontal and vertical tags are applied on each slice, providing motion components in two in-plane directions (x and y components). To acquire motion components in the through-plane direction (z component), another set of parallel coronal slices orthogonal to axial is also acquired. HARP is then used on every tagged image at every time frame, resulting in a corresponding 2D motion field representing the projection of the 3D motion of every tissue point on the current slice plane. Fig. 1(b) shows such HARP slices for the utterance “a souk” at the moment when /s/ is sounded (current time frame), where the tongue is expected to have moved forward from the /a/ moment (time frame 1) when the tags are applied. Meanwhile, cine images revealing better anatomical structures are going to be used for segmentation and registration to be described in section 2.3.

2.2 IDEA Algorithm

Figs. 2(a) and 2(b) illustrate how HARP data are processed in IDEA [5]. The undeformed tissue at time frame 1 has undeformed reference tag planes. At current time frame, the tag planes have deformed along with the tissue. To each point (pixel location) \mathbf{x}_a on an axial image such as Fig. 2(a), HARP produces two vectors representing components of displacement:

$$\begin{cases} \mathbf{q}_x = q_x \mathbf{e}_x, \\ \mathbf{q}_y = q_y \mathbf{e}_y, \end{cases} \quad (1)$$

where \mathbf{e}_x and \mathbf{e}_y are unit vectors in the x and y directions and \mathbf{q}_x and \mathbf{q}_y are the projections of the 3D motion $\mathbf{u}(\mathbf{x}_a)$ on the current axial plane. Similarly, for each

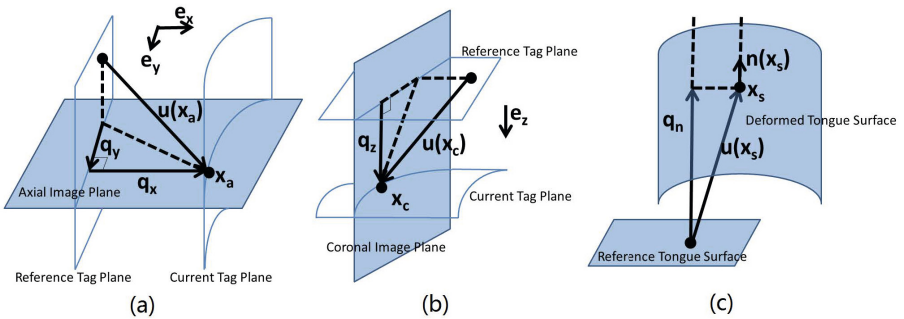


Fig. 2. Relationship between 2D motion components and 3D motion on (a) an axial slice, (b) a coronal slice and (c) the tongue surface

point \mathbf{x}_c on a coronal image such as Fig. 2(b), HARP yields the displacement component vector

$$\mathbf{q}_z = q_z \mathbf{e}_z, \quad (2)$$

where \mathbf{e}_z is the unit vector in the z direction.

IDEA takes such data on all pixels $\{\mathbf{x}_a, \mathbf{q}_x(\mathbf{x}_a), \mathbf{x}_a, \mathbf{q}_y(\mathbf{x}_a), \mathbf{x}_c, \mathbf{q}_z(\mathbf{x}_c)\}$ as input, and estimates an incompressible deformation field $\mathbf{u}(\mathbf{x})$ on a high-resolution grid within the tongue mask. The details are omitted here for lack of space, but are given in [5]. We only note two important aspects. First, IDEA is carried out as a series of smoothing splines, each of which seeks a divergence-free velocity field yielding the deformation field only when integrated. Thus the final field $\mathbf{u}(\mathbf{x})$ is nearly incompressible and its reprojected components at all input points nearly agree with the input measurements. Second, the inputs are observed components of displacements that can arise at any physical position and in any sub-direction of motion. This is the key to utilization of surface deformation measurements within the IDEA framework. In particular, as shown in Fig. 2(c), the tongue surface may deform between time frames, and a point \mathbf{x}_s on the surface at current time frame can be associated with a point on the reference tongue surface. However, like the traditional aperture problem in optical flow, we should not assume to know any tangential information about the surface displacement. This leads to a perfect analogy with HARP data: observations about surface normal deformation, if available, can be used in 3D reconstruction.

2.3 Measuring Tongue Surface Deformation

IDEA requires segmentation of the tongue volume in order to limit the tissue region that is assumed to be incompressible [8]. Cine MR images are used to construct a super-resolution volume [9] at each time frame, which is then manually segmented for the tongue surface mask. We notice that these 3D masks can also be used for deformable registration in order to provide surface deformation information.

The diffeomorphic demons method [10] is applied to the pair of masks between the two time frames where motion is to be computed. Denoting the reference mask at time frame 1 as $I_1 : \Omega_1 \subset \mathbb{R}^3 \rightarrow \{0, 1\}$ and the current deformed mask as $I_t : \Omega_t \subset \mathbb{R}^3 \rightarrow \{0, 1\}$ defined on the open and bounded domains Ω_1 and Ω_t , the deformation field is found and denoted by the mapping $\mathbf{d} : \Omega_t \mapsto \Omega_1$. The estimated displacement field at a point \mathbf{x}_s on the surface of the tongue in current time frame can be denoted as

$$\mathbf{u}(\mathbf{x}_s) = -\mathbf{d}(\mathbf{x}_s). \quad (3)$$

Although diffeomorphic demons generates a whole 3D displacement volume, we take only tongue surface normal components for the reason stated in the previous section. We represent the 3D tongue mask at current time frame by a levelset function $\phi(\mathbf{x})$ that is zero on the surface, positive outside the tongue, and negative inside the tongue. The normal directions of the surface are given by

$$\mathbf{n}(\mathbf{x}_s) = \frac{\nabla \phi(\mathbf{x}_s)}{|\nabla \phi(\mathbf{x}_s)|}. \quad (4)$$

The normal components of motion—serving as additional input to IDEA—are

$$\mathbf{q}_n(\mathbf{x}_s) = (\mathbf{u}(\mathbf{x}_s) \cdot \mathbf{n}(\mathbf{x}_s))\mathbf{n}(\mathbf{x}_s). \quad (5)$$

An example of such a field is shown in Fig. 1(d).

2.4 Enhanced IDEA

With the enhanced input $\{\mathbf{x}_a, \mathbf{q}_x(\mathbf{x}_a), \mathbf{x}_a, \mathbf{q}_y(\mathbf{x}_a), \mathbf{x}_c, \mathbf{q}_z(\mathbf{x}_c), \mathbf{x}_s, \mathbf{q}_n(\mathbf{x}_s)\}$, our proposed method computes the 3D motion over the super-resolution grid points $\{\mathbf{x}_i\}$ and all the surface points $\{\mathbf{x}_s\}$. The algorithm is summarized below.

Algorithm. Enhanced Incompressible Deformation Estimation Algorithm

-
1. Set $\mathbf{u}(\mathbf{x}_i) = \mathbf{0}$ and $\mathbf{u}(\mathbf{x}_s) = \mathbf{0}$.
 2. Set M time steps, **for** $m = 1$ to M **do**
 3. Project currently computed displacement onto input directions by $p_x(\mathbf{x}_a) = \mathbf{u}(\mathbf{x}_a) \cdot \mathbf{e}_x$, $p_y(\mathbf{x}_a) = \mathbf{u}(\mathbf{x}_a) \cdot \mathbf{e}_y$, $p_z(\mathbf{x}_c) = \mathbf{u}(\mathbf{x}_c) \cdot \mathbf{e}_z$, $p_n(\mathbf{x}_s) = \mathbf{u}(\mathbf{x}_s) \cdot \mathbf{n}(\mathbf{x}_s)$.
 4. Compute remaining motion projection by $r_x(\mathbf{x}_a) = q_x(\mathbf{x}_a) - p_x(\mathbf{x}_a)$, $r_y(\mathbf{x}_a) = q_y(\mathbf{x}_a) - p_y(\mathbf{x}_a)$, $r_z(\mathbf{x}_c) = q_z(\mathbf{x}_c) - p_z(\mathbf{x}_c)$, $r_n(\mathbf{x}_s) = q_n(\mathbf{x}_s) - p_n(\mathbf{x}_s)$.
 5. Use part of the remaining motion to approximate velocity: $v_x(\mathbf{x}_a) = r_x(\mathbf{x}_a)/(M - m + 1)$, $v_y(\mathbf{x}_a) = r_y(\mathbf{x}_a)/(M - m + 1)$, $v_z(\mathbf{x}_c) = r_z(\mathbf{x}_c)/(M - m + 1)$, $v_n(\mathbf{x}_s) = r_n(\mathbf{x}_s)/(M - m + 1)$.
 6. Update estimation: $\mathbf{u}(\mathbf{x}_i) = \mathbf{u}(\mathbf{x}_i) + \text{DFVS}\{v_x(\mathbf{x}_a), v_y(\mathbf{x}_a), v_z(\mathbf{x}_c), v_n(\mathbf{x}_s)\}$, $\mathbf{u}(\mathbf{x}_s) = \mathbf{u}(\mathbf{x}_s) + \text{DFVS}\{v_x(\mathbf{x}_a), v_y(\mathbf{x}_a), v_z(\mathbf{x}_c), v_n(\mathbf{x}_s)\}$.
 7. **end for**
-

Here DFVS stands for divergence-free vector spline, which is also the key algorithm “workhorse” of IDEA [5]. M is typically set to 20 which provides a proper trade-off between accuracy and computation time. Enhanced IDEA, which we refer to as E-IDEA below, typically takes about 5 hours on 26 time frames.

3 Results

We evaluated E-IDEA on 50 tongue volumes (25 from a normal control and 25 from a patient) during the utterance “a souk”. Conventional IDEA was also computed for comparison. We computed motion fields relative to time frame 1 which was the /a/ sound, because the resting tongue serves as a good reference configuration, is the natural reference frame for the MR tags, and also fits into continuum mechanics framework for deforming bodies.

Firstly, we visually assessed the motion fields. The results of both subjects are shown in Figs. 1(c), 1(e) and Fig. 3 on two critical time frames: at the /s/, when forward motion is prominent, and at the /k/, when upward motion is prominent (Fig. 1 is for control at time frame /s/). Knowing that the internal muscular structure of tongue prevents its back from performing either too large or zero

motion [1], at tongue’s back, we see E-IDEA has reduced the erroneous large motions for the control, and has captured those small motions where IDEA mistakenly interpolates as zero for the patient. We also see E-IDEA can straighten up the motion at the top of the tongue to better estimate the displacement when the tongue hits the palate vertically (Figs. 3(a), 3(d)). In general, the boundary estimation agrees more with tongue physical mechanics [1].

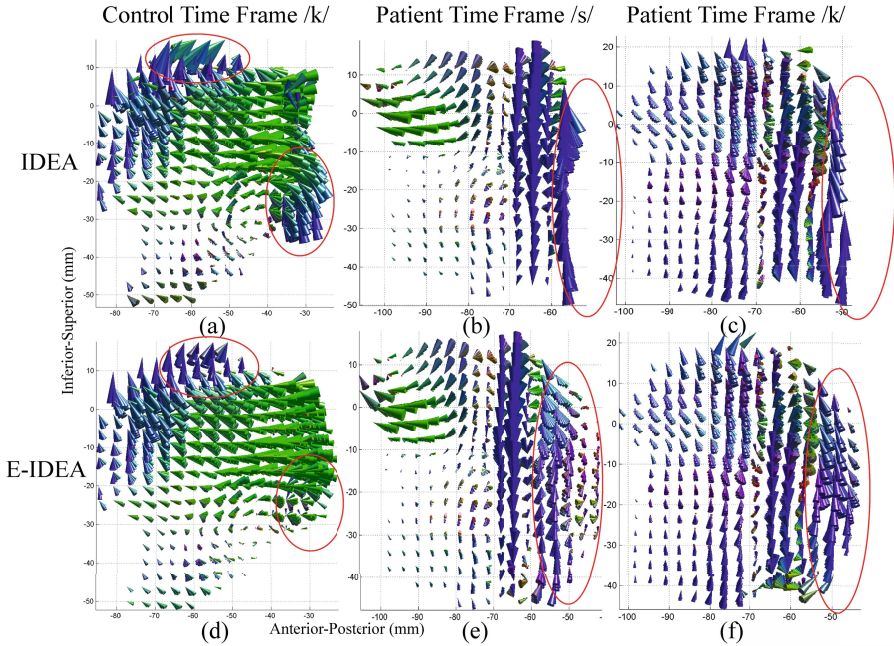


Fig. 3. Visual comparison of conventional IDEA result and E-IDEA result

Secondly, to obtain a numerical comparison, we manually tracked the motions of 15 surface points distributed 5 each on the front, top, and back parts of the tongue (labeled in Fig. 4(a)). We then computed their trajectories with IDEA and E-IDEA motion fields. The tracks of three methods are shown in Fig. 4(a) and errors from manual tracking at each point are shown in Figs. 4(b) and 4(c), boxplotted across all time frames. The error magnitude has been reduced by E-IDEA, especially on the back part of the tongue. Also, the mean error (circles in boxes) is reduced by E-IDEA at all 15 points. The improvement is significant ($p = 0.00003$).

Lastly, we took the estimated 3D motions at input sample locations and reprojected them onto input directions using Eqns. (1) and (5). We then computed a *reprojection error* that gives the error in distance in the input directions between the estimated sample components and the input sample components. This measure assumes input motion components (HARP and surface normal motions)

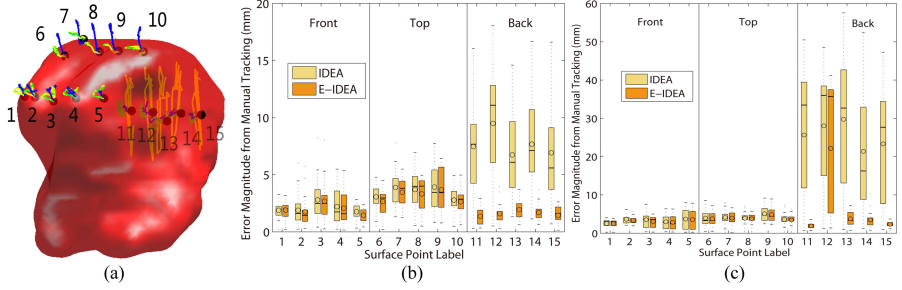


Fig. 4. Comparison of IDEA and E-IDEA with manually tracked surface points. (a) Tracks of the control surface points by manual (*blue*), IDEA (*yellow*), and E-IDEA (*green*). (b) Error magnitude for the control (*bar* is median and *circle* is mean). (c) Error magnitude for the patient.

are the truth. We compare four types of reprojection errors in histograms of Fig. 5: on IDEA internal points, on E-IDEA internal points, on E-IDEA boundary points, and on IDEA boundary points as indicated in the legend. For the control, on a total of 105455 internal points and 108853 boundary points, the mean of the four errors are: 0.32 mm, 0.35 mm, 0.65 mm, and 1.33 mm, respectively. The boundary error has been reduced by 0.68 mm and the internal error has been raised by 0.03 mm. For the patient, on 133302 internal points and 100523 boundary points, the mean of the four errors are: 0.22 mm, 0.24 mm, 0.96 mm and 3.11 mm. The boundary error has been reduced by 2.15 mm and the internal error has been raised by 0.02 mm.

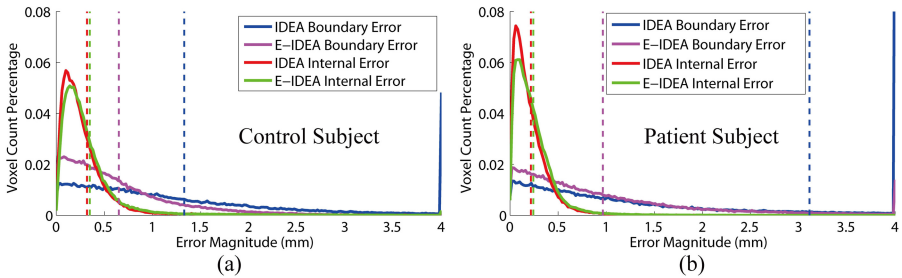


Fig. 5. Regularized histogram of IDEA and E-IDEA's reprojection error on internal and surface points. *Dotted lines* show the mean of four types of reprojection error.

4 Conclusion and Discussion

We have proposed a novel algorithm for estimating the tongue's motion field in 3D. The major innovation is in the incorporation of surface motion as additional information, which compensates for the well-known deficiencies of HARP in

estimating boundary motions. Both qualitative and quantitative improvements are evident using two independent metrics. Especially, from reprojection error, we see that boundary error is substantially reduced while internal error is only minimally increased.

This method is still being improved. Aspects that will be addressed in the future include optimizing the segmentation and registration methods, studying intra-subject volume dependency, and adding data reliability terms to balance HARP and registration information. Also, choice of different reference frames can be explored. And fitting the “internal plus surface motion” idea into other motion estimation frameworks can be an interesting topic.

Acknowledgments. We thank the reviewers for their comments. This work is supported by NIH/NCI 5R01CA133015 and NIH/NIDCD K99/R00 DC009279.

References

1. Kier, W.M., Smith, K.K.: Tongues, Tentacles and Trunks: the Biomechanics of Movement in Muscular-hydrostats. *Zool. J. Linnean Soc.* 83, 307–324 (1985)
2. Zerhouni, E.A., Parish, D.M., Rogers, W.J., Yang, A., Shapiro, E.P.: Human Heart: Tagging with MR Imaging — a Method for Noninvasive Assessment of Myocardial Motion. *Radiology* 169, 59–63 (1988)
3. Parthasarathy, V., Prince, J.L., Stone, M., Murano, E., Nesaiver, M.: Measuring Tongue Motion from Tagged Cine-MRI Using Harmonic Phase (HARP) Processing. *J. Acoust. Soc. Am.* 121(1), 491–504 (2007)
4. Osman, N.F., McVeigh, E.R., Prince, J.L.: Imaging Heart Motion Using Harmonic Phase MRI. *IEEE Trans. Med. Imaging* 19(3), 186–202 (2000)
5. Liu, X., Abd-Elmoniem, K., Stone, M., Murano, E., Zhuo, J., Gullapalli, R., Prince, J.L.: Incompressible Deformation Estimation Algorithm (IDEA) from Tagged MR Images. *IEEE Trans. Med. Imaging* 31(2), 326–340 (2012)
6. Tecelao, S.R., Zwanenburg, J.J., Kuijjer, J.P., Marcus, J.T.: Extended Harmonic Phase Tracking of Myocardial Motion: Improved Coverage of Myocardium and Its Effect on Strain Results. *J. Magn. Reson. Imaging* 23(5), 682–690 (2006)
7. Liu, X., Prince, J.L.: Shortest Path Refinement for Motion Estimation From Tagged MR Images. *IEEE Trans. Med. Imaging* 29(8), 1560–1572 (2010)
8. Xing, F., Lee, J., Murano, E.Z., Woo, J., Stone, M., Prince, J.L.: Estimating 3D Tongue Motion with MR Images. In: *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, California (2012)
9. Woo, J., Murano, E.Z., Stone, M., Prince, J.L.: Reconstruction of High Resolution Tongue Volumes from MRI. *IEEE Trans. Biomedical Engineering* 59(12), 3511–3524 (2012)
10. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic Demons: Efficient Non-parametric Image Registration. *NeuroImage* 45(1), 61–72 (2008)

Learning without Labeling: Domain Adaptation for Ultrasound Transducer Localization

Tobias Heimann¹, Peter Mountney², Matthias John³, and Razvan Ionasec²

¹ Siemens AG, Corporate Technology, Erlangen, Germany

² Siemens Corporation, Corporate Technology, Princeton, NJ, USA

³ Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. The fusion of image data from trans-esophageal echography (TEE) and X-ray fluoroscopy is attracting increasing interest in minimally-invasive treatment of structural heart disease. In order to calculate the needed transform between both imaging systems, we employ a discriminative learning based approach to localize the TEE transducer in X-ray images. Instead of time-consuming manual labeling, we generate the required training data automatically from a single volumetric image of the transducer. In order to adapt this system to real X-ray data, we use unlabeled fluoroscopy images to estimate differences in feature space density and correct covariate shift by instance weighting. An evaluation on more than 1900 images reveals that our approach reduces detection failures by 95% compared to cross validation on the test set and improves the localization error from 1.5 to 0.8 mm. Due to the automatic generation of training data, the proposed system is highly flexible and can be adapted to any medical device with minimal efforts.

1 Introduction

Catheter-based procedures such as trans-aortic valve implantation (TAVI) or paravalvular leak closure are gaining increasing importance for the treatment of structural heart disease. The inherent challenge for the cardiac interventionist is to infer the exact position of the catheter from the available imaging information. X-ray fluoroscopy is the dominant imaging modality for these interventions, increasingly supported by 3D trans-esophageal echography (TEE) [2]. Both modalities show complementary information, but in clinical practice they are controlled and displayed completely independently from each other.

Recently, image fusion was proposed to combine both modalities and to provide the cardiac interventionist with a better overview of the *in situ* conditions. The co-registration can be accomplished by means of electromagnetic tracking (EMT) [3], but this approach requires EMT hardware to be attached to the transducer and is sensitive to EM field distortions. Alternatively, the pose of the transducer can be estimated from its appearance in the X-ray images, either directly [2,6] or supported by fiducial markers attached to the probe head [4]. Since the former approach does not require additional hardware, it is advantageous for integration into the clinical workflow, albeit more challenging to

implement. While 2D-3D registration [2] yields accurate results, it has a limited capture range of < 10 mm, requiring a manual initialization every time a new fluoroscopy sequence is acquired. Discriminative learning (DL) [6] can locate the TEE probe everywhere in the image, but its performance is strongly dependent on quantity and quality of the available training data. In the medical domain, data is generally difficult to acquire, and the required manual labeling is an extremely tedious and time-consuming task. Moreover, trained operators cannot reproducibly annotate images with perfect accuracy, and every variation in ground truth will decrease the performance of the resulting DL system.

In this paper, we propose a novel approach for training a DL system, which is based on *in silico* training data that can be generated automatically in great quantities with perfectly accurate labels. To adapt the system to *in vivo* fluoroscopy data, we employ unsupervised domain adaptation, a technique which is widely used in speech processing and has recently gained attention in the computer vision community [5,1]. In particular, we show how unlabeled data from the target domain (i.e. *in vivo* images) can be used to improve the performance of object localization beyond what is achievable with semi-supervised learning [11]. We start with presenting the basic learning method in the next section and explain our adaptation approach afterwards.

2 Learning from Synthetic Data

2.1 Generation of *in silico* Images

The synthetic training data is based on digitally reconstructed radiographs, which approximate X-ray images from computed tomography (CT) volumes. Source is a high-resolution (0.18 mm/voxel) isotropic C-arm CT of the TEE transducer, which was aligned to the image axes and cropped to contain only the probe head. A binary mask of the transducer was prepared and multiplied with the original volume to remove streak artifacts in the surrounding air.

For each synthetic image, the 3D position and three Euclidean angles of the transducer are randomized with the constraint that the probe is oriented in posterior direction. The flexible shaft of the probe is modeled by a 3D spline originating from a random position at the upper image boundary. Along this spline, a collection of rings is positioned in regular pattern. 2D projections are generated using a composite ray-caster, i.e. every pixel is assigned the sum of all values along the respective ray through the volume. Key to generating realistic-looking images is the transfer function used to calculate the opacities along the ray. Based on the appearance of *in vivo* images, we chose an exponential transfer function with randomized parameters in order to generate sequences with slightly varying appearance and contrast. As background, we used a number of cardiac fluoroscopy sequences (without transducer) and combined them with the generated ray-caster images by additive blending. Annotations were created automatically by storing the 2D position of a fixed point in the center of the transducer together with the respective Euler angles. Figure 1 gives an impression of the look of the generated images compared to *in vivo* data.

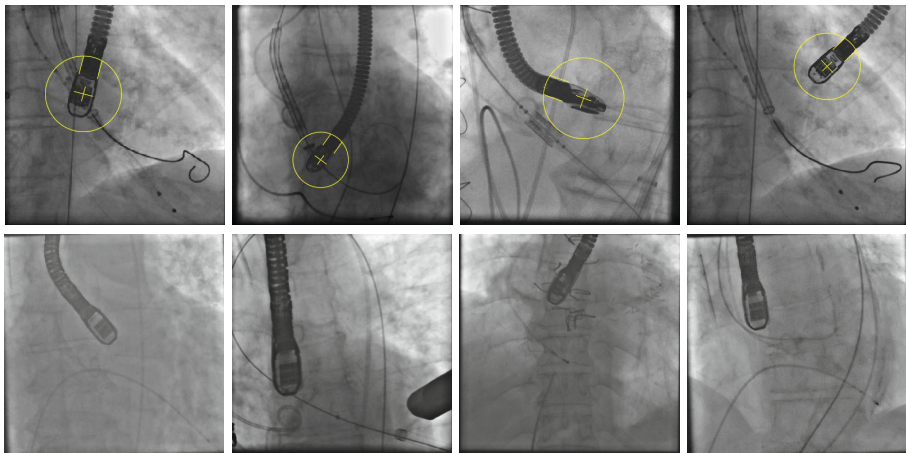


Fig. 1. A selection of generated *in silico* images with automatic labeling (top row) and *in vivo* fluoroscopy images (bottom row)

2.2 Transducer Localization by Discriminative Learning

Following the marginal space learning approach [10], transducer localization is performed in several stages by a pipeline of three discriminative classifiers. The first classifier Φ employs Haar-like features x_H to determine the 2D position of the probe in images rescaled to 1 mm isotropic pixel spacing. All pixels closer than 1 mm to the reference annotation are labeled as $y = Y^+$, all others as $y = Y^-$. During detection, the 50 candidates with the highest classifier output $\hat{p}_\Phi(y = Y^+ | x_H)$ are passed on to orientation detector Θ . Θ is based on steerable features x_S [10] calculated at 0.25 mm isotropic resolution. Possible angles of the transducer are discretized into 6° steps, and all correctly positioned samples deviating $< 4^\circ$ from the annotated angle are labeled as Y^+ . For test images, the 50 candidates with the highest $\hat{p}_\Theta(y = Y^+ | x_S)$ are passed on to scale detector Ψ . Ψ is again based on steerable features x_S with 0.25 mm spacing and selects the most probable size of the transducer from a set of 9 hypotheses, corresponding to feature window sizes from 30–46 mm. Lastly, the 50 highest-ranked candidates are combined by weighted averaging according to their respective $\hat{p}_\Psi(y = Y^+ | x_S)$ and produce the final output. All classifiers of the pipeline are implemented as probabilistic boosting trees (PBTs) [9], which combine high computational efficiency with competitive accuracy.

3 Domain Adaptation

A fundamental assumption in machine learning is that training and test data stem from the same distribution. In our approach, however, the training data originates from the *in silico* source domain S , while the test data comes from the *in vivo* target domain T . Consequently, the above assumption may not hold, in which case the classifiers would work along non-optimal decision boundaries.

Let x represent a feature vector for a sample and $y \in [Y^+, Y^-]$ its label, then the joint probability distribution $P(y, x)$ should be identical for source and target domain. In our case, we know that the marginalized label probabilities are equal, i.e. $P_S(y) = P_T(y)$, since images from both domains show exactly one transducer. Moreover, given a certain feature vector, the question if the corresponding image region shows a probe can also be decided without knowing its domain, which makes it reasonably safe to assume that $P_S(y|x) = P_T(y|x)$. However, the distribution of feature vectors in both domains is probably different, i.e. $P_S(x) \neq P_T(x)$, which leads to a situation called covariate shift [7].

3.1 Learning under Covariate Shift

As described by Shimodaira [7], a classifier can be adapted to different training and test distributions by minimizing its loss function. This is accomplished by assigning each training sample an instance weight according to the ratio of joint probabilities. Under covariate shift, this ratio simplifies to:

$$\frac{P_T(y, x)}{P_S(y, x)} = \frac{P_T(x)P_T(y|x)}{P_S(x)P_S(y|x)} = \frac{P_T(x)}{P_S(x)} \quad (1)$$

Conveniently, this formulation does not include any labels y , i.e. no annotations are required for the target domain in order to adapt the classifier.

There exist a number of approaches to estimate the required density ratio [8]. In this work, we employ the probabilistic classification approach, in which a classifier is trained to differentiate between samples $x_S \in S$ and $x_T \in T$. Among different types of classifiers, logistic regression is especially well suited for this task [8]. During training, all x_S are assigned to $y = 1$ and all x_T to $y = 0$. The density ratio can then be estimated using classifier output \hat{p} by:

$$\frac{P_T(x)}{P_S(x)} = \frac{1}{\hat{p}(y = 1|x)} - 1 \quad (2)$$

3.2 Instance Weighting for Object Localization

While instance weighting has already been employed for a number of different tasks [5], its application to object localization raises two important questions: Which samples should be used to train the logistic regression classifier, and should positive and negative samples be treated equally for weighting? Using all available samples would mean to extract feature vectors for every pixel in every available image multiple times (for different orientation and scale hypotheses). Not only would this result in the impractical amount of 10^{12} feature vectors, but it would also lead to highly unbalanced class labels Y^+ and Y^- . Moreover, as we use a relatively small number of background sequences to generate the *in silico* data, features for Y^- are repeating in the source domain. In summary, this would lead to background samples Y^- completely dominating the logistic regression, while it is the appearance of the transducer (labels Y^+) which should ideally drive the domain adaptation.

We propose a two step approach to solve this problem. In order to draw a subset of samples, we employ a DL pipeline trained on *in silico* data to localize the transducer in another set of synthetic images and unlabeled *in vivo* data. As even an average DL system will detect the transducer with reasonable accuracy on the majority of images, this step effectively reverses the class imbalance in favor of positive samples Y^+ . Feature vectors for the drawn samples are normalized to zero mean and unit variance over the entire set and used to train the logistic regression. As the quality of the density ratio estimation may vary, we relax instance weights w as suggested by Shimodaira [7]:

$$w(x) = \left(\frac{P_T(x)}{P_S(x)} \right)^c \quad (3)$$

with $c \in [0..1]$ as regularization parameter. In this study, we set $c = 0.5$.

The domain adapted classifier is then trained on the *in silico* set used as test data in the first step. For each image of this set, the feature vector x of the drawn sample is used in Eq. 3 to estimate the instance weights for all positive samples. Negative samples remain unweighted.

4 Experiments and Results

4.1 Image Data

Image data originates from two clinical centers and was mostly acquired during standard TAVI procedures. Both centers used an Artis Zeego C-arm system (Siemens AG, Germany) for acquisition of fluoroscopy and an X7-2t 3D transducer (Philips, The Netherlands) for acquisition of TEE. In order to estimate the physical resolution of each fluoroscopy sequence, the pixel spacing of the fluoroscopic detector was divided by the radiologic magnification factor, which accounts for the projection geometry of the C-arm. In order to prevent problems with local feature calculation, we excluded approx. 25% of all frames in which the transducer was too close to the image boundaries. In prospective clinical application, the X-ray window could always be chosen to include the probe entirely, i.e. this data exclusion does not limit the applicability of the proposed approach. In the end, we used 68 sequences from 22 patients for our study, comprising 6280 frames in total. For 37 sequences comprising 1913 frames, the probe head was annotated manually by placing an oriented rectangle over it. We denote this set of annotated *in vivo* images as T_L , while the remaining unlabeled 4367 frames are denoted as T_U . Finally, using the method from Sec. 2.1, we generated two sets S_1, S_2 of 10,000 *in silico* images each. In a small annotation study, the point used for automatic labeling of these sets was selected to best match the center of the rectangle used for manual annotations.

4.2 Selecting the Stages for Domain Adaptation

The first set of experiments was conducted to analyze the effectiveness of domain adaptation (DA) for different stages of the detector pipeline. As baseline system,

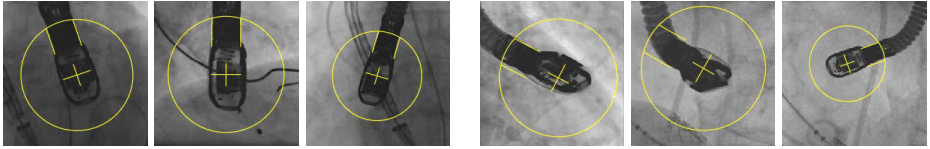


Fig. 2. A selection of *in silico* training samples that received high instance weights (left) and low instance weights (right) for the position detector

we first trained the pipeline presented in Sec. 2.2 on S_1 ($\Phi_0 \Rightarrow \Theta_0 \Rightarrow \Psi_0$). Subsequently, we trained another system on S_2 and used it to draw TEE probe samples from S_1 and T_U . The resulting samples were used to calculate three sets of instance weights for S_1 , using the feature set selected by Φ_0 , Θ_0 , and Ψ_0 , respectively. Some examples for samples that obtained very high and low weights are shown in Fig. 2. Training a position detector on the weighted data from S_1 yielded the first domain-adapted classifier Φ_A , which was integrated into pipeline “DA Pos” ($\Phi_A \Rightarrow \Theta_0 \Rightarrow \Psi_0$). Similarly, weighted orientation (Θ_A) and scale detectors (Ψ_A) were trained and included in pipelines “DA Pos+Ori” ($\Phi_A \Rightarrow \Theta_A \Rightarrow \Psi_0$) and “DA Pos+Ori+Scale” ($\Phi_A \Rightarrow \Theta_A \Rightarrow \Psi_A$).

All systems were evaluated on image set T_L . For a detailed analysis of each system, we looked at the detected candidates before the final averaging step and counted a true positive if one of the candidates had a position error < 1 mm, an orientation error $< 4^\circ$, and a scale error < 3 mm. Plotting these counts against the average number of false positives results in the ROC-style curves shown in Fig. 3. The corresponding areas under the curve (AUCs) are given in Table 1.

As can be seen, domain adaptation on the position detector has the largest impact with an increase of 4.5% AUC relative to the baseline system. Domain adaptation on the orientation detector brings only slight additional improvements (+4.8% AUC relative to baseline), while trying to adapt the scale detector deteriorates the results again (only +1.6% AUC relative to baseline remain).

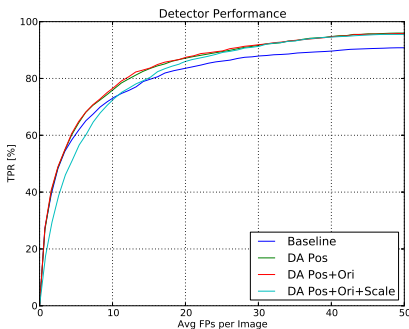


Table 1. Area-under-curve values

System	AUC
Baseline	78.3
DA Pos	81.8
DA Pos+Ori	82.0
DA Pos+Ori+Scale	79.5

Fig. 3. True positive rate (TPR) vs. average number of false positives (FPs)

Table 2. Mean errors with standard deviation for successful detections

	Failed Detections	Position Error	Orientation Error	Scale Error
<i>in vivo</i> Reference	7.34 %	1.5±2.5 mm	3.2±5.4°	3.8±3.0 %
<i>in silico</i> Baseline	2.35 %	0.9±1.1 mm	1.8±1.6°	6.0 (3.2±2.4) %
Domain Adaptation	0.37 %	0.8±0.6 mm	1.7±1.3°	5.7 (3.0±2.3) %
Self Training	1.41 %	0.8±0.8 mm	1.6±1.4°	6.5 (3.0±2.2) %

4.3 Evaluation of Robustness and Accuracy

For the main evaluation, the reference system was trained directly on T_L without any synthetic data (using three-fold cross-validation for evaluation). The baseline system from the previous section, trained exclusively on *in silico* images, came second, and the best-performing domain adaptation (“DA Pos+Ori”) third. For the last system, we used the samples drawn from T_U (as described in Sec. 3.2) to enlarge our synthetic training set and generated another unweighted system from $S_1 \cup T_U$. This is a popular approach in semi-supervised learning and called self-training [11]. For each system, the final output of the pipeline (after candidates are merged) was compared to the reference labels. In case the output was located outside the annotated probe area (circles in Figs. 1 & 2), the localization was counted as failure. For successful detections, average position, orientation and scale errors were computed. The complete results are displayed in Table 2. As it turned out, the labels of *in silico* images had a systematic bias of 5% regarding the scale of the transducer; the scale errors in parenthesis show the bias-corrected results. The complete detection pipeline runs in <40 ms per frame, enabling a real-time localization of the transducer in the operating room.

5 Discussion

Our results clearly demonstrate the dependency of DL systems on the available training data. The reference system in our experiments, although trained on the same domain as the test data, yields the worst overall results. The *in silico* system can compensate its different source domain by an eight times larger training set with perfectly placed labels and reduces the number of failed detections by a factor of three, while at the same time improving on all errors. Given these good results, we were surprised by the large impact of domain adaptation, which managed to reduce misdetections yet considerably further down to 5% of the reference system. Its success is based on up-weighting training samples that appear similarly in the target domain and down-weighting less common samples with e.g. very high contrast or large rotations (see Fig. 2). Obviously, generating *in silico* data with more realistic parameters from the start would have a similar effect, but – as for most applications – the true distribution of parameters in real-world data is not known. Since the largest differences between source and target domain appear in the feature set of the position detector (which has to

cope with different orientations and scales), this stage of the pipeline can benefit most from domain adaptation. In order to gain the complete 3D pose of the transducer, our TEE localization can be combined with 2D-3D registration [2] or template-matching [6] to deliver an automatic, robust, and accurate real-time fusion of TEE and fluoroscopy images.

We believe the combination of automatically generated data and unlabeled real-world images to be a highly promising approach for training DL systems. It resolves the need for thousands of annotated training samples, which is one of the main bottlenecks of machine learning in the medical domain. Moreover, the ability to create large quantities of training data for any X-ray imageable device (e.g. implants or new transducers) within hours offers unmatched flexibility.

References

1. Beijbom, O.: Domain adaptation for computer vision applications. Technical report, University of California, San Diego (June 2012)
2. Gao, G., Penney, G., Ma, Y., Gogin, N., Cathier, P., Arujuna, A., Morton, G., Caulfield, D., Gill, J., Rinaldi, C.A., Hancock, J., Redwood, S., Thomas, M., Razavi, R., Gijssbers, G., Rhode, K.: Registration of 3D trans-esophageal echocardiography to X-ray fluoroscopy using image-based probe tracking. *Med. Image Anal.* 16, 38–49 (2012)
3. Jain, A., Gutierrez, L., Stanton, D.: 3D TEE registration with X-ray fluoroscopy for interventional cardiac applications. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 321–329. Springer, Heidelberg (2009)
4. Lang, P., Seslija, P., Chu, M.W.A., Bainbridge, D., Guiraudon, G.M., Jones, D.L., Peters, T.M.: US - fluoroscopy registration for transcatheter aortic valve implantation. *IEEE Trans. Biomed. Eng.* 59(5), 1444–1453 (2012)
5. Margolis, A.: A literature review of domain adaptation with unlabeled data. Technical report, University of Washington (2011)
6. Mountney, P., Ionasec, R., Kaiser, M., Mamaghani, S., Wu, W., Chen, T., John, M., Boese, J., Comaniciu, D.: Ultrasound and fluoroscopic images fusion by autonomous ultrasound probe detection. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 544–551. Springer, Heidelberg (2012)
7. Shimodaira, H.: Improving predictive inference under covariate shift by weighting the log-likelihood function. *J. Statistical Planning and Inference* 90, 227–244 (2000)
8. Sugiyama, M., Suzuki, T., Kanamori, T.: Density ratio estimation: A comprehensive review. In: Proc. Workshop on Statistical Experiment and Its Related Topics, Kyoto, Japan, pp. 10–31 (March 2010)
9. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: Proc. ICCV, vol. 2, pp. 1589–1596 (October 2005)
10. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Med. Imaging* 27(11), 1668–1681 (2008)
11. Zhu, X.: Semi-supervised learning literature survey. Technical Report 1530, University of Wisconsin-Madison (July 2008)

Segmentation of 4D Echocardiography Using Stochastic Online Dictionary Learning*

Xiaojie Huang¹, Donald P. Dione⁴, Ben A. Lin⁴, Alda Bregasi⁴,
Albert J. Sinusas^{3,4}, and James S. Duncan^{1,2,3}

¹ Departments of Electrical Engineering,

² Departments of Biomedical Engineering,

³ Departments of Diagnostic Radiology,

⁴ Departments of Internal Medicine,

Yale University, New Haven, CT, USA

xiaojie.huang@yale.edu

Abstract. Dictionary learning has been shown to be effective in exploiting spatiotemporal coherence for echocardiographic segmentation. To overcome the limitations of previous methods, we present a stochastic online dictionary learning approach for segmenting left ventricular borders from 4D echocardiography. It is based on stochastic approximations and processes a mini-batch of samples at a time, which results in lower memory consumption and lower computational cost than classical batch algorithms. In contrast to the previous methods, where dictionaries and their weights are optimized only on the most recently segmented frame, our stochastic online learning procedure optimizes the dictionaries and the corresponding weights by aggregating all the past information while adapting them to the dynamically changing data. The rate of updating the past information is controlled and varied according to the appearance scale to seek a balance between old and new information. Results on 26 4D echocardiographic images show the proposed method is more accurate, more robust, and faster than the previous batch algorithm.

1 Introduction

Segmentation of 4D echocardiography plays an important role in the quantitative analysis that provides important cardiac functional parameters such as ejection fraction and strain. Due to gross intensity inhomogeneities, characteristic artifacts, and poor contrast, automatic segmentation of the left ventricle is particularly challenging in echocardiography. The inherent spatiotemporal coherence of echocardiographic data provides useful constraints. The key observation is that the inherent spatio-temporal consistencies regarding image appearance (e.g., speckle pattern) and shape over the sequence can be exploited to guide cardiac border estimation. Statistical models have received considerable attention. Following the seminal work of Cootes et al. on statistical shape modeling [1],

* This work was supported by NIH RO1HL082640.

a number of statistical models [2–5] have been proposed for learning spatiotemporal priors offline from a database. The main limitation of these methods is that the high level spatiotemporal patterns in routine clinical images, especially for disease cases, may deviate from the priors learned from a database.

Exploiting individual data coherence through online learning overcomes this limitation. It is particularly attractive when a database is inapplicable or unavailable. Sparse representation and dictionary learning have recently been successfully applied to modelling local image appearance and segmenting left ventricular borders in 4D echocardiography [6, 7]. Dictionary learning on the fly exploits the spatiotemporal coherence inherent to individual data and achieves promising segmentation results [6]. However, these methods use classical second-order batch procedures for dictionary learning. The batch algorithm assumes a fixed-size dataset and accesses the whole training set at each iteration. It is memory-consuming and computationally expensive. It can be impractical when the training set is large. Every time new data is added to the training set, the dictionary needs to be retrained on the new complete training set in order to incorporate the new information, which makes the batch algorithm inefficient for dynamically changing data and online learning. In [6, 7], the appearance dictionaries are trained only on the last segmented frame rather than all the previous frames. This accelerates error accumulation and compromises the segmentation accuracy and reliability, especially for endocardial borders.

To overcome these limitations, we present a stochastic online dictionary learning approach for segmenting left ventricular borders from 4D echocardiography. It utilizes a stochastic optimization technique and processes a mini-batch of samples at a time, which results in lower memory consumption and lower computational cost than classical second-order batch algorithms. In contrast to the previous methods, our stochastic online learning procedure optimizes the dictionaries and the corresponding weights by aggregating the information of all the past frames while adapting the dictionaries to the latest segmented frame. The past information is carried forward by sufficient statistics. We weight the past information to control the rate at which the past information is updated by the new information. This updating rate varies with appearance scale to maintain a balance between old and new information.

2 Methods

2.1 Segmentation Framework

We employ a frame-by-frame sequential segmentation procedure interlaced with dictionary learning on the fly introduced in [6, 7]. Multiscale appearance dictionaries are dynamically updated each time a new frame is segmented. In a maximum a posteriori (MAP) framework, we estimate the shape S_t in frame I_t given the knowledge of $\hat{S}_{1:t-1}$ and $I_{1:t}$:

$$\hat{S}_t = \arg \max_{S_t} p(S_t | \hat{S}_{1:t-1}, I_{1:t}). \quad (1)$$

It is approximated by a decomposition of information into intensity I_t , local appearance discriminant R_t , and shape prediction S_t^* :

$$\hat{S}_t \approx \arg \max_{S_t} p(S_t^*|S_t)p(R_t|S_t)p(I_t|S_t)p(S_t). \quad (2)$$

The discriminant R_t summarizing multiscale local appearance dominates the estimation. It is predicted by multiscale appearance dictionaries D_t that are derived from $\hat{S}_{1:t-1}$ and $I_{1:t-1}$ through sparse representation and dictionary learning. In [6, 7], the dictionaries D_t are trained only on \hat{S}_{t-1} , I_{t-1} . The knowledge of the previous information is not fully utilized. This paper focuses on computing D_t more efficiently and reliably and achieving more accurate and reliable discriminant R_t . Further details of solving (2) can be found in [6].

2.2 Multiscale Sparse Representation

Let Ω denote the 3D image domain. We describe a pixel $\mathbf{u} \in \Omega$ in frame I_t with a series of appearance vectors $\mathbf{y}_t^k(\mathbf{u}) \in \mathbb{R}^n$ at different appearance scales $k = 1, \dots, J$. $\mathbf{y}_t^k(\mathbf{u})$ is constructed by concatenating orderly the pixels in a local block centered at \mathbf{u} and normalized to unit length. Complementary multiscale appearance information is extracted at different levels of Gaussian pyramid. A shape S_t in I_t is represented by a level set function $\Phi_t(\mathbf{u})$. The regions of interest are two band regions $\Omega_t^1 = \{\mathbf{u} \in \Omega : 0 \leq \Phi_t(\mathbf{u}) < \psi_2\}$ and $\Omega_t^2 = \{\mathbf{u} \in \Omega : 0 > \Phi_t(\mathbf{u}) > -\psi_1\}$ which form two appearance classes. Let $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ denote two dictionaries adapted to appearance classes Ω_t^1 and Ω_t^2 respectively at scale k . Under a sparse linear model, an appearance vector $\mathbf{y} \in \mathbb{R}^n$ can be decomposed as a sparse linear combination of the atoms from a dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$ which encodes the typical patterns of a corresponding appearance class. That is, $\mathbf{y} \approx \mathbf{D}\mathbf{x}$, and $\|\mathbf{x}\|_0$ is small. How well $\mathbf{y}_t^k(\mathbf{u})$ is sparsely represented by the appearance dictionary $\{\mathbf{D}_t^c\}_k$ is measured by the reconstruction residue:

$$\{R_t^c(\mathbf{u})\}_k = \|\mathbf{y}_t^k(\mathbf{u}) - \{\mathbf{D}_t^c \hat{\mathbf{x}}_t^c(\mathbf{u})\}_k\|_2 \quad (3)$$

$\forall k \in \{1, \dots, J\}$ and $c \in \{1, 2\}$, where

$$\{\hat{\mathbf{x}}_t^c(\mathbf{u})\}_k = \arg \min_{\mathbf{x}} \|\mathbf{y}_t^k(\mathbf{u}) - \{\mathbf{D}_t^c\}_k \mathbf{x}\|_2 \text{ s.t. } \|\mathbf{x}\|_0 \leq T, \quad (4)$$

where T is a sparsity factor. The residue indicates the likelihood \mathbf{u} is in class c . Combining the multiscale information, we define the discriminant as

$$R_t(\mathbf{u}) = \sum_{k=1}^J \left[\left(\log \frac{1}{\beta_t^k} \right) \text{sgn}(\{R_t^2(\mathbf{u})\}_k - \{R_t^1(\mathbf{u})\}_k) / \sum_{j=1}^J \left(\log \frac{1}{\beta_t^j} \right) \right], \quad (5)$$

$\forall \mathbf{u} \in \Omega$, where β_t^k 's are the weighting parameters of the J appearance scales.

2.3 Stochastic Online Dictionary Learning

Learning a dictionary $\mathbf{D} \in \mathbb{R}^{n \times K}$ from a finite training set $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_M] \in \mathbb{R}^{n \times M}$ is to solve a joint optimization problem with respect to the dictionary \mathbf{D} and the sparse representation coefficients $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_M] \in \mathbb{R}^{K \times M}$:

$$\min_{\mathbf{D}, \mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \lambda \sum_{i=1}^M \|\mathbf{x}_i\|_q, \quad (6)$$

where $\|\mathbf{x}\|_q$ is a sparsity-inducing regularization that can be ℓ_0 pseudo norm or ℓ_1 norm. Classic algorithms for dictionary learning are second-order iterative batch algorithms such as the K-SVD [8] algorithm that is used in [6, 7]. The batch algorithm accesses the whole training set at each iteration and is memory consuming and computationally expensive. It may become impractical in the case of large training sets. This problem is aggravated when the data is dynamically changing over time like echocardiography, since the dictionary needs to be retrained on the new complete dataset each time new data is available. In [6, 7], the appearance dictionaries are updated each time a new frame is segmented, but they are only optimized on the newly segmented frame rather than all the previous frames. This accelerates accumulation of errors, especially at endocardial borders where there are often large deformations.

Stochastic online learning technique proposed in [9] can be used to overcome these limitations. It has recently been applied to shape modeling [10]. It processes one element of the training set at a time, which particularly suits applications with large training sets or image sequence analysis. It alternates classic sparse coding steps with dictionary update steps where the new dictionary \mathbf{D}_m at m th iteration minimizes a surrogate for the empirical cost (6):

$$\mathbf{D}_m = \arg \min_{\mathbf{D}} \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \lambda \|\mathbf{x}_i\|_1 \right) \quad (7)$$

where sufficient statistics \mathbf{x}_i computed during the previous steps aggregate the past information. The past information is carried forward in matrices:

$$\mathbf{A}_m = \mathbf{A}_{m-1} + \mathbf{x}_m \mathbf{x}_m^T \text{ and } \mathbf{B}_m = \mathbf{B}_{m-1} + \mathbf{y}_m \mathbf{x}_m^T, \quad (8)$$

which enables optimizing dictionaries on the past information without accessing the past data again. Then the dictionary update step (7) is reduced to solving (9) with initialization \mathbf{D}_{m-1} . This procedure leads to faster performance and better dictionaries than classical batch algorithms [9]. It converges almost surely to a stationary point of the cost function and scales up gracefully to large datasets [9]. For dynamic data, the dictionary is dynamically updated by the new data while optimized on the whole dataset. Here we use a variant of [9] as summarized in Algorithm 1. We use a mini-batch extension that accesses a mini-batch of η samples per iteration to accelerate convergence. We assign weights ϱ to the past training data to control the rate of updating out-of-date information.

We introduce a stochastic online learning process supervised in a boosting framework [11] as detailed in Algorithm 2. Algorithm 1 is invoked to enforce the

Algorithm 1. Stochastic Online Dictionary Learning

Require: training set $\mathbf{y} \in \mathbb{R}^n \sim p(\mathbf{y})$, sparsity weight λ , initial dictionary $\mathbf{D}_{m_I} \in \mathbb{R}^{n \times K}$, initial iteration number m_I and terminal iteration number m_T , mini-batch size η , weight ϱ , and initial matrices \mathbf{A}_{m_I} and \mathbf{B}_{m_I} .

for $m = m_I$ to m_T **do**

Draw η samples $\mathbf{Y}_m = \{\mathbf{y}_{m,i}\}_{i=1}^\eta$ from $p(\mathbf{y})$

Sparse coding: $\mathbf{x}_{m,i} = \arg \min_{\mathbf{x} \in \mathbb{R}^K} \|\mathbf{y}_{m,i} - \mathbf{D}_{m-1}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \forall i \in \{1, \dots, \eta\}$

$\mathbf{A}_m = \varrho \mathbf{A}_{m-1} + \frac{1}{\eta} \sum_{i=1}^\eta \mathbf{x}_{m,i} \mathbf{x}_{m,i}^T, \mathbf{B}_m = \varrho \mathbf{B}_{m-1} + \frac{1}{\eta} \sum_{i=1}^\eta \mathbf{y}_{m,i} \mathbf{x}_{m,i}^T.$

Update dictionary: compute \mathbf{D}_m with \mathbf{D}_{m-1} as initialization

$$\mathbf{D}_m = \arg \min_{\mathbf{D}} \left(\frac{1}{2} Tr(\mathbf{D}^T \mathbf{D} \mathbf{A}_m) - Tr(\mathbf{D}^T \mathbf{B}_m) \right). \quad (9)$$

end for

return dictionary \mathbf{D}_{m_T} , and matrices \mathbf{A}_{m_T} and \mathbf{B}_{m_T} .

reconstructive property of the dictionaries. The boosting supervision strengthens the discriminative property and optimizes the weighting of multiscale information. At each time point t , the series of multiscale appearance dictionary pairs $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$, matrices \mathbf{A}_t^k and \mathbf{B}_t^k , and the corresponding weighting parameters $\beta_t^k, k = 1, \dots, J$, are updated by the latest segmented frame $t-1$: training samples of appearance vectors belonging to two classes $\{\mathbf{Y}_{t-1}^1\}_k = \{\mathbf{y}_{t-1}^k(\mathbf{u}) : \mathbf{u} \in \Omega_{t-1}^1\}$ and $\{\mathbf{Y}_{t-1}^2\}_k = \{\mathbf{y}_{t-1}^k(\mathbf{u}) : \mathbf{u} \in \Omega_{t-1}^2\}$. In contrast to [6, 7] where $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ and β_t^k depend only on frame $t-1$, we optimize $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$ and β_t^k by aggregating the information of all the preceding frames (stored in $\mathbf{A}_{t-1}^k, \mathbf{B}_{t-1}^k$, and β_{t-1}^k). If an error occurs in one frame, it can be compensated by the information of the previous frames. The propagation of errors is alleviated. The rate of updating the past information varies with appearance scale. Let l_k be the axial width in millimeter of the local image at scale k , we set $\varrho_k = a l_k^{-2}$ where $a \in \mathbb{R}^+$. Higher ϱ 's are assigned to finer appearance scales to incorporate more past information. Lower ϱ 's are assigned to coarser appearance scales to put more emphasis on the latest information, since the coarse appearance scale is more sensitive to cardiac deformation. The stochastic online learning procedure can be initialized either by offline learning from a suitable database or by a manual tracing.

3 Results

We validated our method on 26 4D canine open-chest echocardiographic images acquired from both healthy and post-infarct animals using Phillips iE33 and an X7-2 array probe. Each image sequence spanned a cardiac cycle and contained about 25 – 30 volumes. The sequential segmentation was initialized with a manual tracing of the end-diastole volume. 100 volumes were randomly selected for expert manual segmentation and quality assessment. We evaluated automatic results against manual tracings using the following segmentation quality metrics: Hausdorff Distance (HD), Mean Absolute Distance (MAD), and Dice coefficient (DICE). We compared the proposed method to [6] that uses the batch dictionary learning technique K-SVD. The two algorithms shared the same set of relevant

Algorithm 2. Boosted Multiscale Online Dictionary Learning

Require: training sets $\{\mathbf{Y}_{t-1}^1\}_k = \{\mathbf{y}_{1,i}^k\}_{i=1}^{M_1}$ and $\{\mathbf{Y}_{t-1}^2\}_k = \{\mathbf{y}_{2,j}^k\}_{j=1}^{M_2}$, initial dictionaries $\{\mathbf{D}_{t-1}^1, \mathbf{D}_{t-1}^2\}_k$, matrices \mathbf{A}_{t-1}^k and \mathbf{B}_{t-1}^k , accumulated # of previous iterations N_{t-1} , weighting parameters $\beta_{t-1}^k, \varrho_k, k = 1, \dots, J$, mini-batch size η , # of iterations N , and sparsity factor T .

$\mathbf{w}_1^1 = \{w_{1,i}^1\}_{i=1}^{M_1} = \mathbf{1}, \mathbf{w}_2^1 = \{w_{2,j}^1\}_{j=1}^{M_2} = \mathbf{1}$.

for $k = 1$ to J **do**

Dictionary Learning: Apply Algorithm 1 for N iterations to adapt $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k$

to $\{\mathbf{Y}_{t-1}^1\}_k \sim \mathbf{p}_1^k = \{p_{1,i}^k\}_{i=1}^{M_1} = \frac{\mathbf{w}_1^k}{\sum_{i=1}^{M_1} w_{1,i}^k}$ and $\{\mathbf{Y}_{t-1}^2\}_k \sim \mathbf{p}_2^k = \{p_{2,j}^k\}_{j=1}^{M_2} = \frac{\mathbf{w}_2^k}{\sum_{j=1}^{M_2} w_{2,j}^k}$. Use $\varrho = \varrho_k$ for the first iteration and $\varrho = 1$ for the rest.

Sparse Coding: $\forall \mathbf{y} \in \{\mathbf{Y}_{t-1}^1, \mathbf{Y}_{t-1}^2\}_k$, solve (4) for sparse representations w.r.t. $\{\mathbf{D}_t^1\}_k$ and $\{\mathbf{D}_t^2\}_k$ and get residues $R(\mathbf{y}, \mathbf{D}_t^1)_k$ and $R(\mathbf{y}, \mathbf{D}_t^2)_k$.

Hypothesis $h_k: \mathbf{y} \in \{\mathbf{Y}_{t-1}^1, \mathbf{Y}_{t-1}^2\}_k \rightarrow \{0, 1\}$: $h_k(\mathbf{y}) = \text{Heaviside}(R(\mathbf{y}, \mathbf{D}_t^2)_k - R(\mathbf{y}, \mathbf{D}_t^1)_k)$. Calculate the error of h_k : $\epsilon_k = \frac{1}{1+\varrho_k} [\sum_{i=1}^{M_1} p_{1,i}^k |h_k(\mathbf{y}_{1,i}^k) - 1| + \sum_{j=1}^{M_2} p_{2,j}^k |h_k(\mathbf{y}_{2,j}^k)|] + \frac{\varrho_k}{1+\varrho_k} (\frac{\beta_{t-1}^k}{1+\beta_{t-1}^k})$. Set $\beta_t^k = \epsilon_k / (1 - \epsilon_k)$.

Weight Update: $w_{1,i}^{k+1} = w_{1,i}^k \beta_t^{k-1 - |h_k(\mathbf{y}_{1,i}^k) - 1|}, w_{2,j}^{k+1} = w_{2,j}^k \beta_t^{k-1 - h_k(\mathbf{y}_{2,j}^k)}$.

end for

return $\{\mathbf{D}_t^1, \mathbf{D}_t^2\}_k, \mathbf{A}_t^k, \mathbf{B}_t^k, \beta_t^k, k = 1, \dots, J$, and $N_t = N_{t-1} + N$.

parameters. We used the following parameter setting: $J = 10, T = 2, K = 1.5n, N = 10(t > 2)$ or $20(t = 2), \eta = 2048, \lambda = 0.8$, and $a = 100$.

Figure 1 shows representative segmentation results for frames at end-systole when it is easiest to access error accumulation. In the top row, the batch method [6] resulted in more errors in end-systolic segmentations, since it learned appearance dictionaries only on the latest segmented frame and did not fully leverage the information carried in all the previous frames. The segmentation error of a frame is likely to propagate to the following frames. Images in the bottom row show the improved segmentation results by employing our new stochastic learning procedure. Since we optimize the dictionaries on all the previous frames, the error in a given frame is compensated by the information of the other frames. Figure 2 presents the quality measure curves from end-diastole to end-systole for the endocardial segmentations of a healthy sequence and a post-infarct sequence. DICE decays and HD and MAD rise from end-diastole to end-systole due to accumulation of errors. Compared to the batch method, our method resulted in flattened curves, which suggests our method effectively alleviates error accumulation and improves segmentation performance for both healthy and post-infarct images. For epicardial segmentation, the improvement was not significant, since the baseline accuracy of [6] was already very high (97% in DICE). Table 1 summarizes the statistics of segmentation quality measures and computational efficiency achieved by the two algorithms in segmenting endocardial borders. The proposed method achieved smaller mean MAD, smaller mean HD, larger mean DICE, and smaller standard deviations of all the measures. The overall segmentation accuracy and robustness were

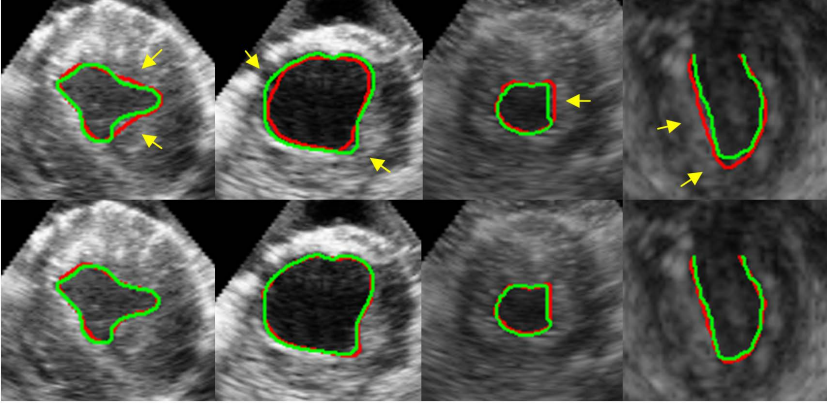


Fig. 1. Comparisons of segmentation results by the batch method (top row) and our method (bottom row). Green: Manual segmentation. Red: Automatic segmentation.

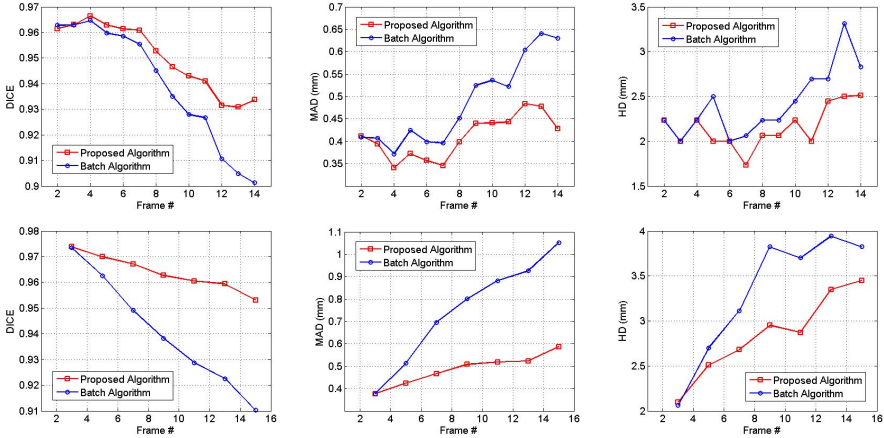


Fig. 2. Segmentation quality measures at different frames of two example sequences (healthy (top row) and post-infarct (bottom row)) from end-diastole to end-systole. Blue: the batch method. Red: the proposed method.

Table 1. Sample means \pm standard deviations of the quality measures and dictionary learning time per frame for the segmentation of endocardial borders

	DICE (%)	MAD (mm)	HD (mm)	Time (s)
Batch Algorithm [6]	93.6 ± 2.49	0.57 ± 0.14	2.95 ± 0.62	~ 45
Proposed Algorithm	94.6 ± 2.17	0.48 ± 0.11	2.83 ± 0.53	~ 25

effectively improved using our stochastic online learning procedure. We tested the two algorithms on a laptop with Intel quad-core 2.2 GHz CPU and 8 GB memory. Both algorithms were implemented with a mixture of MATLAB and C++. The batch algorithm took about 45 seconds per frame for dictionary learning. The proposed algorithm took only about 25 seconds per frame.

4 Conclusion

We have presented an approach for segmenting left ventricular borders from 4D echocardiography using stochastic online dictionary learning. It is based on a stochastic optimization technique resulting in lower memory consumption and computational cost than classical batch algorithms. We optimize the dictionaries and their weights on all the preceding frames while adapting them to the latest segmented frame. The rate of updating the past information is controlled and varies with appearance scale. Our method effectively improved the accuracy and robustness of endocardial segmentation and computational efficiency compared to the previous batch methods. Future work will include automating the dictionary initialization through offline learning. The stochastic learning procedure is suitable for both offline and online learning. A database that is too large for batch methods can be gracefully handled by our method which avoids accessing the database during online learning. Our method can ultimately be used to build an integrated offline and online learning framework.

References

1. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE TPAMI* 23(6), 681–685 (2001)
2. Bosch, J.G., Mitchell, S.C., Lelieveldt, B.P.F., Nijland, F., Kamp, O., Sonka, M., Reiber, J.H.C.: Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE TMI* 21(11), 1374–1383 (2002)
3. Jacob, G., Noble, J.A., Behrenbruch, C.P., Kelion, A.D., Banning, A.P.: A shape-space based approach to tracking myocardial borders and quantifying regional left ventricular function applied in echocardiography. *IEEE TMI* 21(3), 226–238 (2002)
4. Sun, W., Çetin, M., Chan, R., Reddy, V., Holmvang, G., Chandar, V., Willsky, A.S.: Segmenting and tracking the left ventricle by learning the dynamics in cardiac images. In: Christensen, G.E., Sonka, M. (eds.) *IPMI 2005*. LNCS, vol. 3565, pp. 553–565. Springer, Heidelberg (2005)
5. Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: A dynamical shape prior for LV segmentation from RT3D echocardiography. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I*. LNCS, vol. 5761, pp. 206–213. Springer, Heidelberg (2009)
6. Huang, X., Dione, D.P., Compas, C.B., Papademetris, X., Lin, B.A., Sinusas, A.J., Duncan, J.S.: A Dynamical Appearance Model Based on Multiscale Sparse Representation: Segmentation of the Left Ventricle from 4D Echocardiography. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III*. LNCS, vol. 7512, pp. 58–65. Springer, Heidelberg (2012)
7. Huang, X., Lin, B.A., Compas, C.B., Sinusas, A.J., Staib, L.H., Duncan, J.S.: Segmentation of left ventricles from echocardiographic sequences via sparse appearance representation. In: *MMBIA*, pp. 305–312 (2012)
8. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE TSP* 54(11), 4311–4322 (2006)

9. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: ICML, p. 87 (2009)
10. Zhang, S., Zhan, Y., Zhou, Y., Uzunbas, M., Metaxas, D.N.: Shape prior modeling using sparse representation and online dictionary learning. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 435–442. Springer, Heidelberg (2012)
11. 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. 23–37. Springer, Heidelberg (1995)

Atlas Encoding by Randomized Forests for Efficient Label Propagation

Darko Zikic, Ben Glocker, and Antonio Criminisi

Microsoft Research Cambridge

Abstract We propose a method for multi-atlas label propagation based on encoding the individual atlases by randomized classification forests. Most current approaches perform a non-linear registration between all atlases and the target image, followed by a sophisticated fusion scheme. While these approaches can achieve high accuracy, in general they do so at high computational cost. This negatively affects the scalability to large databases and experimentation. To tackle this issue, we propose to use a small and deep classification forest to encode each atlas individually in reference to an aligned probabilistic atlas, resulting in an *Atlas Forest* (AF). At test time, each AF yields a probabilistic label estimate, and fusion is done by averaging. Our scheme performs only one registration per target image, achieves good results with a simple fusion scheme, and allows for efficient experimentation. In contrast to standard forest schemes, incorporation of new scans is possible without retraining, and target-specific selection of atlases remains possible. The evaluation on three different databases shows accuracy at the level of the state of the art, at a significantly lower runtime.

1 Introduction

Labeling of healthy human brain anatomy is a crucial prerequisite for many clinical and research applications. Due to the effort involved in fully manual labeling and increasing database sizes (e.g. ADNI, IXI, OASIS), a lot of research has been devoted to develop automatic methods for this task. While brain labeling is a general segmentation task (with a high number of labels), the standard approach for this task is multi-atlas label propagation (MALP) – see [1] for an overview of the state of the art. With the *atlas* denoting a single labeled scan, MALP methods first derive a set of label proposals for the target image, each based on a single atlas, and then combine these proposals into a final estimate. There are two main strategies for estimating atlas-specific label proposals. The first and larger group of methods non-linearly aligns each of the atlas images to the target image, and then – assuming one-to-one correspondence at each point – uses the atlas labels directly as label proposals, cf. e.g. [2,3,4]. The second group of patch-based methods has recently enjoyed increased attention [5,6,7]. Here, the label proposal is estimated for each point in the target image by a local similarity-based search in the atlas. Patch-based approaches relax the one-to-one assumption, and aim at reducing the computational times by using linear instead

of deformable alignment [5,6], resulting in labeling runtimes of 22-130 minutes per target on the IBSR dataset [6]. However, note that these approaches do not change the actual number of required registrations. The fusion step, which combines the atlas-specific label proposals into a final estimate, aims to correct for inaccurate registration or labellings, and remains an active research topic.

While current state of art techniques can achieve high levels of accuracy, in general they are computationally demanding. This is primarily due to the *non-linear registration between all atlases and the target image*, combined with the long runtimes for the best performing registration schemes for the problem [8]. Current methods state runtimes of 2-20 hours per single registration [1]. Furthermore, sophisticated fusion schemes can also be computationally expensive. State of the art approaches state fusion runtimes of 3-5 hours [9,10,11] on a database of 15 atlases [1]. While the major drawback of high computational costs is the scalability to large and growing databases, they also limit the amount of possible experimentation during the algorithm development phase.

Our method differs from previous approaches in the way how label proposals for a single atlas are generated, and is designed with the goal of low computational cost at test time and experimentation. In this work, we focus on the question of how a single atlas is encoded. From this point of view, methods assuming one-to-one correspondence represent an atlas directly as an image/label-map pair, while patch-based methods encode it by a set of localized patch collections. Variations of the patch-based encoding include use of sparsity [7], or use of label-specific k NN search structures [12]. In contrast to previous representations, we encode a single atlas together with its relation to label priors by a small and deep classification forest – which we call an *Atlas Forest* (AF). Given a target image as input (and an aligned probabilistic atlas), each AF returns a probabilistic label estimate for the target. Label fusion is performed by averaging of the probability estimates. While patch-based methods use a static representation for each image point (i.e. a patch of fixed size), our encoding is spatially varying. In the training step, our approach learns to describe different image points by differently shaped features, depending on the point’s contextual appearance. Compared to current MALP methods, our approach has the following important characteristics:

1. *Only one registration per target is required.* This registration aligns the probabilistic atlas to the target. Since only one registration per target is required, the runtime is independent of the database size in this respect.
2. *Efficient generation of atlas proposals and their fusion.* For proposal generation one AF per atlas is evaluated, and the fusion consists is done by averaging. While both operations scale linearly with database size, they are significantly more efficient than current approaches. For example, for the database with 15 atlases from [1], labeling of a single target takes ca. 4 min.
3. *Efficient Experimentation.* A leave-one-out cross-validation of a standard MALP approach on n atlases requires registration between all images, thus scaling with n^2 . In contrast, the training of the single AFs, which is the most costly component of our approach for experimentation, scales with n (this assumes a given probabilistic atlas which is not part of experimentation).

Besides being efficient, experiments on 3 databases in Sec. 3 indicate that our scheme also achieves high accuracy, comparing favorably to state of the art.

Compared to standard forest schemes (cf. e.g. [13,14,15,16]) which train each tree on data from *all* training images, our model, which trains each tree on a single atlas exemplar, has three advantageous properties for MALP.

1. *Simple incorporation of new atlases into the database.* For standard forest schemes, non-approximative addition of new training data requires complete retraining. In our scenario, a new forest is simply trained on the new atlas exemplar and added to the other, previously trained AFs.
2. *Selection of atlases for target-specific evaluation is straightforward* since every AF is associated with a single atlas. This step seems non-obvious for standard forest schemes. This property allows use of atlas-selection [17], which can reduce the computational cost, improve accuracy.
3. *Efficient experimentation.* For cross-validation, standard schemes have to be trained for every training/testing split of data, which is extremely costly. In our scenario, each AF is trained only once. Any leave- k -out test is performed simply by using the subset of $n-k$ AFs corresponding to the training data.

After presenting the method in Sec. 2, and demonstrating its performance in Sec. 3, we summarize and discuss its properties in Sec. 4.

2 Method - Atlas Forests

An atlas forest (AF) encodes a single atlas by training a classification forest exclusively on the data from the atlas. AFs do not depend on the reference frame of the target image, since every point is described only by its appearance, without considering its location (this can be seen as a further relaxation of the one-to-one assumption). While this allows us to avoid the registration of atlases to the target, a problem with such a location-oblivious approach is that the location carries valuable information about label probabilities (e.g. a point on the far left is unlikely to carry a right-side label). To efficiently integrate this information, we augment the intensity information from the atlas/target image by label priors warped to the image, and AFs operate on this augmented input. For the alignment of the priors, only a *single* registration per image is required.

2.1 Forest Training, and Labeling by Testing and Fusion

We use randomized forests as a classifier since they can efficiently handle a high number of classes, which is important in the MALP setting. Since we use a standard forest type, we keep the description short, and refer for details to e.g. [18,19]. Classification forests consist of a set of trees, and as a supervised learning method, they operate in two stages: training and testing. During training, each tree of the AF a_i is trained on the specific atlas image I_i and the corresponding label map L_i which contains label class values c . Specifically, each tree t learns a label class predictor $p_t(c|f)$ for a high-dimensional feature representation f of

points from I_i . The training is performed by splitting the training examples at each node based on their representation in the feature space. The split functions are computed by maximizing the *information gain* in randomly selected dimensions of the feature space. In this work, we stop the tree growth at a certain tree depth ($d=36$), with the condition that no tree leaf contains less than a certain number of samples ($s_{\min}=8$). Since we are dealing with a high number of classes with extremely varying sizes, we use class re-weighting, i.e. we adjust the probability computation for each class according to its frequency, such as to obtain a uniform distribution at the root node. Without this standard step, small classes would have low influence, resulting in reduced accuracy for these classes. After training, each leaf l contains a class predictor $p^l(c|f)$, which is computed as the re-weighted empirical class distribution of its samples.

At testing, a target image I is labeled by processing its points by the trained AFs. By applying the learned splitting functions to the feature representation f of a point to be labeled, each tree t from a certain AF yields a prediction $p_t(c|f)$. The probabilistic estimate of the AF a with n_t trees is then formed as the average of the single tree predictions $p_a(c|f) = \frac{1}{n_t} \sum_{i=1}^{n_t} p_{t_i}(c|f)$. The fusion of these probabilistic estimates from n_a AFs is done by averaging, i.e. $p(c|f) = \frac{1}{n_a} \sum_{i=1}^{n_a} p_{a_i}(c|f)$, and subsequent maximum selection $\hat{c} = \arg \max_c p(c|f)$.

2.2 Features and Label Priors

We describe the intensity around a certain location by a bank of generic intensity-based parametric features, which are non-local but short-range. Given the point of interest x in image I , offset vectors u, v , cuboids $C_s(x)$ (centered at x with side lengths s, r), and the mean operator μ , we use the following feature types:

1. Local cuboid mean intensity: $\mu(I(C_s(x)))$
2. Difference of local intensity and offset cuboid mean: $I(x) - \mu(I(C_s(x+u)))$
3. Difference of local and offset cuboid means: $\mu(I(C_s(x))) - \mu(I(C_s(x+u)))$
4. Difference of offset cuboid means: $\mu(I(C_s(x+u))) - \mu(I(C_r(x+v)))$

The feature type and the above parameters (u, v, s, r) are drawn randomly during training at each node, thus defining the random feature space dimensions to be explored. Guided by the results from patch-based works [5,6], we use a maximum offset of 10mm, and cuboid side length $s, r < 5$ mm.

Additionally to the random features, we use a set of deterministic features, which are considered at every node. These features are the local intensity $\tilde{I}(x)$ in a multi-channel image \tilde{I} , which is formed by augmenting the atlas image I by the aligned label priors P_L . Next to the priors for the individual labels, we employ further 6 priors, which aggregate priors for left/right, lower/upper and inner/outer labels, thus subdividing the brain in a coarser manner. In a setting with $|L|$ different labels, this results in a $|L|+7$ -dimensional image \tilde{I} . The use of the prior labels allows us to include the available knowledge about the label probabilities at this point in an efficient way, at the cost of a *single registration per target*. For an effect of using the label priors, please see Fig. 1.

In this work, we construct simple label priors ourselves since we deal with varying labeling protocols – for actual applications, a use of carefully constructed, protocol-specific priors would seem beneficial, e.g. [20]. The construction is performed by iterative registration of the training images to their mean [21]. This results in an average intensity image \bar{I} , and applying the computed warps to corresponding label maps followed by averaging yields a set of label priors P_L . To account for potential registration errors at test time, we smooth the prior maps by a Gaussian with $\sigma=2\text{mm}$. We use affine registration, followed by a deformable registration by the FFD-based method from [22], with cross-correlation as data term, and conservative deformable settings with an FFD-grid spacing of 20mm and strong regularization. The registration operates on images down-sampled by a factor of 4, taking less than 30 seconds per image.

At test time, the average intensity image \bar{I} is registered to the target, and the computed transformation is used to align the label priors P_L to the target. Here, the same registration scheme as above is employed.

3 Evaluation

We evaluate our approach on three brain MRI databases. For all tests we perform the standard preprocessing steps: skull-stripping (own implementation), inhomogeneity correction [23], and histogram matching (www.itk.org).

We used the IBSR dataset in this work for the development of the method and the parameter settings. The same settings were then used also for the evaluation on the other two databases. As final settings, we use 5 trees per atlas forest. The single trees are trained down to depth of 36, with the restriction that each leaf contains at least 8 samples. Each node uses 1000 features from a pool of 10000 random features per tree. The training of one tree takes on average ca. 36 minutes on a standard desktop PC (Intel Xeon E5520 2.27GHz, 12 GB RAM). The runtimes reported below are for the label propagation only, and do not include the time for the registration of the probabilistic atlas (ca. 30 seconds), and the preprocessing of the target image.

IBSR Database. The IBSR data (<http://www.cma.mgh.harvard.edu/ibsr/>) contains 18 labeled images with 32 labels. To provide a comparative context, we cite the results from [6], which are shown to compare favorably to average dice scores (DSC) reported previously for the IBSR data. The IBSR data set is used in [6] in a leave-one-out evaluation, and the best performing version of the proposed method (group-wise multipoint) reaches a mean DSC of 83.5%, with a runtime of 130 minutes. A different variant discussed in [6] (fast multipoint), which aims at faster runtimes by performing the search at a reduced number of locations in the image, reaches a DSC of 82.25%, with a labeling runtime of 22 minutes. Our approach with the above settings reaches a DSC of 84.60% with a runtime of 3 minutes per target image. Further, we quantify the influence of some elements of our method on IBSR data (all by leave-one-out experiments):

- Using the proposed AF scheme without the augmentation by label priors significantly reduces the DSC to 77.38%, and introduces noise and extreme

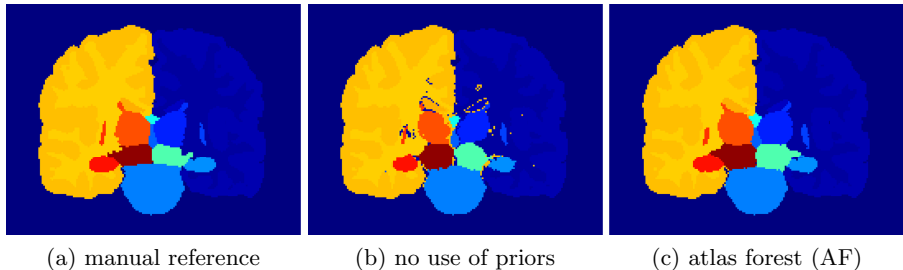


Fig. 1. Labeling example (IBSR): Using intensity-based features only (b) leads to extreme errors, which can be removed by additional use of label priors (c)

errors, as the forest is no longer able to compensate for the missing location information, see Fig. 1(b) for a visualization.

- Using only affine registration for construction and warping of the label priors decreases the DSC to 82.71%, indicating that accuracy improvement through a dedicated registration method might be possible [8].
- A standard forest scheme which uses approximately the same amount of data for training of each tree (i.e. 1/17th of all data), but randomly draws samples from all training images (i.e. performs bagging), reaches a DSC of 84.08%, with otherwise identical settings. This shows that our method does not reduce the quality, while introducing advantages for the MALP setting.

LONI-LPBA40 Database. The LONI-LPBA40 database [20] consists of 40 images of healthy volunteers, with 56 labels, most of them within the cortex. To provide some context, we cite the recent results on this data set from [7], where three methods are evaluated for 54 labels: an implementation of a patch-based scheme as in [5,6] (PBL), and two modifications aiming at sparsity of used patches (SPBL), and spatial consistency (SCPBL). The corresponding reported DSCs for a leave-one-out experiment are 75.06%, 76.46% and 78.04%, with runtimes of 10, 28 and 45 minutes *per class*. Our approach reaches an average DSC of 77.46% with a runtime of 8 minutes per image (for all classes).

MICCAI 2012 Multi-atlas Labeling Challenge. Finally, we apply our approach to the data from [1], consisting of 15 training images and 20 test images from the OASIS project and corresponding label maps as provided by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>) under academic subscription. The evaluation is performed on 134 labels (98 cortical, 36 non-cortical). Here, we train the AFs on the 15 training atlases, and perform the evaluation on the 20 testing target images. With the above settings, our mean DSC is 73.66% over all labels (71.04% for cortical, 80.81% for non-cortical structures) with a runtime of 4 minutes. In the evaluation in [1], this would place our approach in 8th position, out of 25 entries. The approach with the highest DSC in the challenge, PICSL-BC [9], reaches a score of 76.54%. A significant source of error in our approach seems to be a wrong labeling of background labels due to the

used skull stripping – by restricting the evaluation to the reference brain masks, our approach would achieve 76.06%, while PICS-BC would increase to 77.76%.

4 Summary and Discussion

We presented an efficient scheme for encoding of individual atlases for the purpose of multi-atlas label propagation. It represents an atlas by an atlas-specific classification forest, which is in contrast to the currently standard representations as an image/label-map pair, or a set of local patch collections. While previous methods use a static encoding for all points in the image domain, our approach learns a variable representation depending on the local context of the particular points. The major practical advantage of our approach is that only a single registration is required to label a target image. In return, compared to previous approaches, we require a training stage and a probabilistic atlas. However, we show that these additional requirements are not prohibitive. Compared to standard forest schemes, our approach has a number of advantages for label propagation, without loss of accuracy. Overall, our approach achieves accuracy comparable to state of the art at a much lower computational cost, both for the actual use of the system for labeling, as well as for experimentation.

With our approach in an early stage, we see several potential directions for improvement. Use of better atlases [20], registration [8], or skull-stripping might improve results. Early tests indicate that the size of the used feature space can be reduced without loss in accuracy, leading to more efficient training. Finally, adopting existing fusion approaches (e.g. [24]) is an interesting future direction.

References

1. Landman, B., Warfield, S. (eds.): MICCAI Workshop on Multi-Atlas Labeling (2012)
2. 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)
3. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE TMI* 23(7), 903–921 (2004)
4. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33(1), 115–126 (2006)
5. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., Robles, M., Collins, D.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
6. Rousseau, F., Habas, P., Studholme, C.: A supervised patch-based approach for human brain labeling. *IEEE TMI* 30(10), 1852–1862 (2011)
7. Wu, G., Wang, Q., Zhang, D., Shen, D.: Robust patch-based multi-atlas labeling by joint sparsity regularization. In: MICCAI Workshop STMI (2012)

8. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al.: Evaluation of 14 non-linear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46(3), 786–802 (2009)
9. Wang, H., Avants, B., Yushkevich, P.: A combined joint label fusion and corrective learning approach. In: *MICCAI Workshop on Multi-Atlas Labeling* (2012)
10. Asman, A.J., Landman, B.A.: Multi-atlas segmentation using non-local STAPLE. In: *MICCAI Workshop on Multi-Atlas Labeling* (2012)
11. Asman, A., Landman, B.: Multi-atlas segmentation using spatial STAPLE. In: *MICCAI Workshop on Multi-Atlas Labeling* (2012)
12. Wang, Z., Wolz, R., Tong, T., Rueckert, D.: Spatially aware patch-based segmentation (SAPS): An alternative patch-based segmentation framework. In: Menze, B.H., Langs, G., Lu, L., Montillo, A., Tu, Z., Criminisi, A. (eds.) *MCV 2012*. LNCS, vol. 7766, pp. 93–103. Springer, Heidelberg (2013)
13. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: *IEEE Computer Vision and Pattern Recognition, CVPR* (2011)
14. Iglesias, J.E., Konukoglu, E., Montillo, A., Tu, Z., Criminisi, A.: Combining generative and discriminative models for semantic segmentation of CT scans via active learning. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011*. LNCS, vol. 6801, pp. 25–36. Springer, Heidelberg (2011)
15. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of CT images. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011*. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
16. Zikic, D., et al.: Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III*. LNCS, vol. 7512, pp. 369–376. Springer, Heidelberg (2012)
17. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738 (2009)
18. Breiman, L.: Random forests. *Machine Learning* (2001)
19. Criminisi, A., Shotton, J. (eds.): *Decision Forests for Computer Vision and Medical Image Analysis*. Springer (2013)
20. Shattuck, D., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K., Poldrack, R., Bilder, R., Toga, A.: Construction of a 3d probabilistic atlas of human cortical structures. *NeuroImage* 39(3), 1064–1080 (2007)
21. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 23, S151–S160 (2004)
22. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. *MedIA* (2008)
23. Tustison, N., Gee, J.: N4ITK: Nick’s N3 ITK implementation for MRI bias field correction. *The Insight Journal* (2010)
24. Ledig, C., Wolz, R., Aljabar, P., Lötjönen, J., Heckemann, R., Hammers, A., Rueckert, D.: Multi-class brain segmentation using atlas propagation and em-based refinement. In: *IEEE ISBI* (2012)

Robust and Accurate Coronary Artery Centerline Extraction in CTA by Combining Model-Driven and Data-Driven Approaches

Yefeng Zheng, Huseyin Tek, and Gareth Funka-Lea

Imaging and Computer Vision, Siemens Corporate Technology, Princeton, NJ, USA
yefeng.zheng@siemens.com

Abstract. Various methods have been proposed to extract coronary artery centerlines from computed tomography angiography (CTA) data. Almost all previous approaches are data-driven, which try to trace a centerline from an automatically detected or manually specified coronary ostium. No or little high level prior information is used; therefore, the centerline tracing procedure may terminate early at a severe occlusion or an anatomically inconsistent centerline course may be generated. Though the connectivity of coronary arteries exhibits large variations, the position of major coronary arteries relative to the heart chambers is quite stable. In this work, we propose to exploit the automatically segmented chambers to 1) predict the initial position of the major coronary centerlines and 2) define a vessel-specific region-of-interest (ROI) to constrain the following centerline refinement. The proposed prior constraints have been integrated into a model-driven algorithm for the extraction of three major coronary centerlines, namely the left anterior descending artery (LAD), left circumflex artery (LCX), and right coronary artery (RCA). After extracting the major coronary arteries, the side branches are traced using a data-driven approach to handle large anatomical variations in side branches. Experiments on the public Rotterdam coronary CTA database demonstrate the robustness and accuracy of the proposed method. We achieve the best average ranking on overlap metrics among automatic methods and our accuracy metric outperforms all other 22 methods (including both automatic and semi-automatic methods).

1 Introduction

Cardiovascular disease (CVD) is the first leading cause of death in the United States and coronary stenosis (narrowing of the vessel) is the most common CVD. Cardiac computed tomography angiography (CTA) is the primary non-invasive imaging modality to diagnose coronary stenosis thanks to its superior image resolution. To facilitate the diagnosis, coronary centerlines are often extracted before the detection and quantification of the stenosis (*i.e.*, measuring the percentage of the lumen area blocked by plaques). However, automatic centerline extraction is challenging due to the presence of severe calcifications, occlusions, imaging artifacts, and insufficient contrast agent, etc. Large anatomical variations of the coronary tree are another major challenge. For example, depending on the dominance pattern, the posterior descending artery (PDA) and posterolateral branch artery (PLB) can be fed by either the right coronary artery (RCA) or the left circumflex artery (LCX).

Various coronary centerline extraction methods have been proposed in the literature [1,2]. Almost all previous approaches are data-driven, which try to trace a centerline from an automatically detected or manually specified coronary ostium. One advantage of these approaches is the potential to handle anatomical variations. However, since no or little high level prior information is used, the centerline tracing procedure may terminate early at a severe occlusion or an anatomically inconsistent centerline course may be generated. Recently, Kitamura et al. [3] proposed a method to build the coronary shape model composed of 30 discrete nodes sampled from three major coronary arteries and two coronary veins. The coronary shape model is then fitted to the detected coronary candidates via an optimization procedure. However, it is not clear how the coronary anatomical variations are handled with one global shape model.

Though the connectivity of coronary arteries exhibits large variations, the position of major coronary arteries relative to the heart chambers is quite stable. For example, the left anterior descending artery (LAD) runs in the anterior groove between the left and right ventricles, while the LCX and RCA run in the atrial-ventricular groove before extending toward the heart apex. In this work, we propose to exploit the automatically segmented chambers to 1) predict the initial position of the major coronary centerlines and 2) define a vessel-specific region-of-interest (ROI) to constrain the following centerline refinement. The prior knowledge of the relative position of three major coronary arteries w.r.t. heart chambers is encoded in a mean shape model learned from a training set. As the first step of the centerline extraction process, the heart chambers are segmented [4] and coronary ostia are detected [5] automatically in an input volume. The deformation field from the mean shape model to the input volume is estimated using the thin-plate spline (TPS) model [6] with the coronary ostia and heart chambers as anchor points. A centerline in the pre-learned mean shape model is transformed to the input volume as an initial estimate of the coronary path. The centerline is refined using dynamic programming and then further extended to the distal end using a data-driven approach. However, without proper constraints, the refined centerline may be traced to a non-coronary structure. In this work, we propose a vessel-specific region-of-interest (ROI) mask to constrain the tracing of a coronary centerline. Using the ROI mask, the tracing error is reduced and, if it happens, the error is controlled. After extracting the major coronary arteries, the side branches are traced using a data-driven approach [7] since side branches exhibit far more anatomical variations and a data-driven approach can handle such variations.

Different to our previous work [8], here, we propose a new mean centerline generation method to model the full length of a coronary and a vessel-specific ROI mesh to constrain the tracing of centerlines. Previously, we only extracted three major coronary centerlines using a model-driven approach. In this work, we combine it with a data-driven approach to further extract all side branches. The proposed approach has a few advantages compared to the data-driven approaches: 1) It is much more robust under severe occlusions since the prior information is used to help the algorithm cross an occlusion; 2) By combining model-driven and data-driven approaches, it can handle variations in the length and topology of an artery; and 3) We combine the centerline extraction and vessel labeling into the same procedure. Consequently, the extracted centerlines are already labeled once detected.

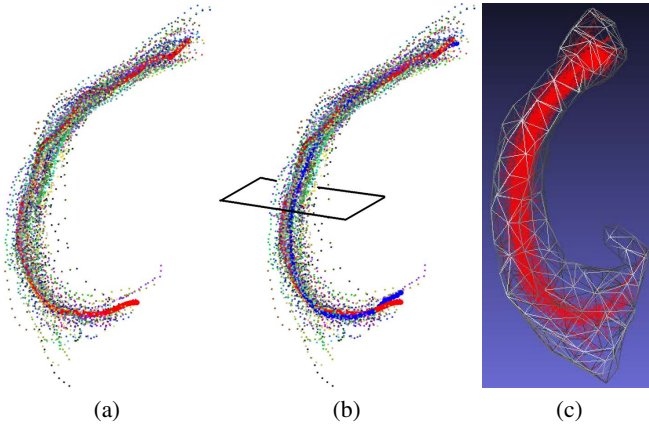


Fig. 1. Generating the centerline mean shape and region-of-interest mesh for the left anterior descending artery from a set of aligned training centerlines. (a) Aligned centerline point cloud and the coarse mean centerline (red line). (b) The refined mean centerline (blue line) (c) Region-of-interest mesh enclosing all the point cloud.

2 Building Prior Coronary Model

The prior knowledge of the coronary artery is embedded in a mean shape model composed of the four heart chambers and the three major coronary arteries. The visible length of a major coronary artery can vary greatly either due to anatomical variations (especially in the distal segment) or insufficient contrast agent inside the vessel. Previously, we proposed a method [8] to truncate the coronary centerlines to the same relative length by discarding the variable distal segments. The truncation based method has two limitations. First, the centerline points may have bad anatomical correspondence across patients due to the variable tortuousness of a coronary artery. Second, the generated mean centerline does not provide a model for the full length of the coronary artery.

In this work, we propose a novel two-step approach to generating a mean coronary centerline. We first align all centerlines in the training set to the same coordinate system by warping them into the space of the mean heart chamber shape, which is generated using a method presented in [4]. Here, we use the nonrigid transformation defined by the TPS model [6]. The deformation field is estimated using the heart chamber mesh points as anchor points and it is then used to warp the annotated coronary centerlines. Fig. 1a shows the aligned LAD centerline points. We then pick a centerline that best represents the whole shape population. That means the average Euclidean distance from other centerlines to this optimal centerline (one-way distance) is the smallest. Suppose two centerlines A and B are given and they are represented as a set of equal-distance points A_0, A_1, \dots, A_{m-1} and B_0, B_1, \dots, B_{n-1} , respectively. For each point A_i on centerline A , we calculate the minimum distance to centerline B as $d(A_i, B)$. The one-way average Euclidean distance from A to B , $D(A, B)$ is defined as

$$D(A, B) = \frac{1}{m} \sum_{i=0}^{m-1} d(A_i, B). \quad (1)$$

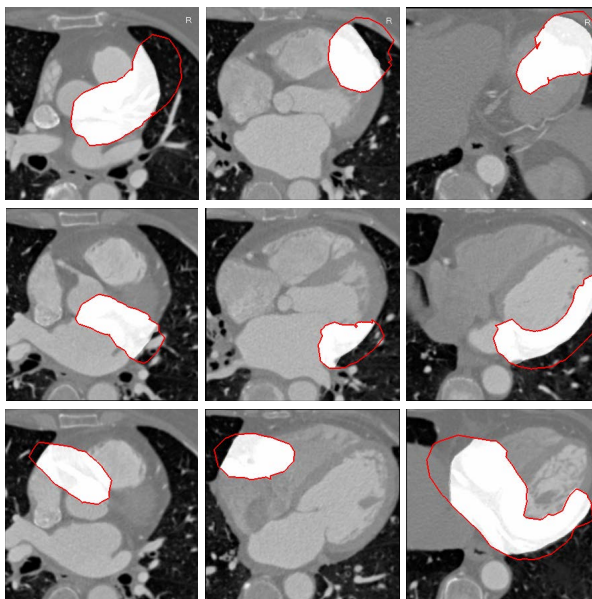


Fig. 2. Region-of-interest (ROI) masks by combining the vessel-specific ROI and pericardium based mask. The LAD, LCX, and RCA masks are shown at the top, middle, and bottom row, respectively. From left to right are the masks of the proximal, middle, and distal segment, respectively. The red contours show the intersection of the ROI meshes with an image slice. The intensity of the masked voxels is increased by 1000 HU for visualization purpose.

Given a set of aligned centerlines C^0, C^1, \dots, C^{k-1} , the coarse mean centerline M^c is picked as the one with the minimum one-way distance from the other centerlines to it,

$$M^c = \operatorname{argmin}_{C^i} \sum_{j=0}^{k-1} D(C^j, C^i). \quad (2)$$

We use the one-way distance to bias toward a longer centerline to model the full length of a coronary artery. The mean centerline picked by the algorithm is always one of the longest, which is shown as the red thick line in Fig. 1a. It has a realistic shape since it is just one example training centerline. Furthermore, it is located roughly at the center of the point cloud of the aligned centerlines. We further refine the mean centerline by moving it even closer to the center of the point cloud. At each point on the coarse mean centerline, we generate a cutting plane perpendicular to the tangent direction at that point, as shown in Fig. 1b. The intersections of all centerlines with the cutting plane are calculated and the mean centerline point is then adjusted to the mean position of intersections. After adjustment, the refined mean centerline (the blue thick line in Fig. 1b) is closer to the point cloud center.

Using the mean shape model composed of the heart chambers and major coronaries, an initial centerline can be generated and is then further refined (refer to Section 3).

However, without proper constraints in the refinement procedure, the extracted centerline may be traced to a non-coronary structure. Sometimes, a distal side branch may be picked as the main trunk, *e.g.*, a long diagonal branch is picked as the LAD. In this work, we propose a vessel-specific coronary mask to constrain the search of each major coronary centerline. Besides reducing centerline tracing leakage and branch labeling errors, the coronary ROI mask can also speed up the computation since irrelevant voxels are excluded. Starting from the aligned point cloud of the training centerlines (Fig. 1a), we generate a surface mesh tightly enclosing all aligned centerlines (Fig. 1c). Since such a surface mesh defines a region-of-interest (ROI) for each major coronary during centerline refinement, we call it an ROI mesh. The ball-pivoting algorithm [9] is used to generate the ROI mesh. Given a ball with a certain size, we roll it around the point cloud. The surface mesh (which may have multiple pieces) is defined by the outer surface of the region that the ball cannot roll into. To generate a single-piece surface mesh, we set the ball size to a relative large value (10% of the maximum distance of any point pairs in the cloud). The generated ROI mesh is tight and it is then expanded a bit (5%) to provide a safety margin. The vessel-specific ROI mesh can be combined with the pericardium based coronary mask [10] to further constrain the search of major coronary centerlines. Fig. 2 shows the ROI mask of three major coronaries. The proposed method can be extended to define an ROI mesh for the whole left or right coronary tree (including all side branches). In [11], a similar ROI is exploited for automatic calcium scoring. However, their ROI is generated by multi-atlas based registration, which is far more time consuming than our approach.

3 Coronary Centerline Extraction

Given an input volume, the heart chambers are segmented [4] and coronary ostia are detected [5] automatically, and they are then used to predict the initial position of the coronary arteries. Since the heart chambers and coronary ostia are available in both the mean shape and the input volume, we use them to estimate a TPS deformation field [6]. The pre-trained mean shape centerline is then transformed to the input volume, using the estimated deformation field, to provide an initial estimate of the centerline. A dynamic programming based optimization is then applied to refine the initial centerline path. The initial centerline is represented as a set of evenly sampled points P_i , for $i = 0, 1, \dots, n - 1$. For each point P_i , we uniformly sample 41×41 candidate positions P_i^j on a plane perpendicular to the centerline path at this point. The candidates P_i^j are sampled on a regular grid of $20 \times 20 \text{ mm}^2$ (with grid spacing of 0.5 mm) centered at the initial centerline point. Now, the problem is how to select the best position for each point P_i . It can be formulated as a shortest path computation problem,

$$\bar{P}_0^{J(0)}, \bar{P}_1^{J(1)}, \dots, \bar{P}_{n-1}^{J(n-1)} = \arg \min_{P_i^{J(i)}} \sum_{i=0}^{n-1} C(P_i^{J(i)}) + w \sum_{i=0}^{n-2} \|P_i^{J(i)} - P_{i+1}^{J(i+1)}\|. \quad (3)$$

The first term is the cost for a single node, measuring how likely this point is at the center of the vessel. Here, a machine learning based vesselness [10] is used as the node cost. The second term is the total length of the path by summing the Euclidean distance between two neighboring points on the path. Free parameter w , which is used to balance

Table 1. Comparison with other automatic centerline extraction methods on the Rotterdam coronary CTA test set (24 datasets) using the overlap metrics. The semi-automatic methods participated in the ranking, but are removed in the table due to the page limit.

Method	Overall Rank	OV			OF			OT		
		%	score	rank	%	score	rank	%	score	rank
Proposed Method	8.91	93.5	53.4	10.98	76.5	54.9	8.22	95.6	70.0	7.54
GFVCoronaryExtractor	9.02	93.7	55.9	10.73	74.2	52.9	9.09	95.9	68.5	7.24
GVFTube'n'Linkage	10.52	92.7	52.3	12.31	71.9	51.4	10.35	95.3	67.0	8.91
SupervisedExtraction	11.28	90.6	53.8	12.75	70.9	49.0	10.52	92.5	61.2	10.56
DepthFirstModelFit	11.86	84.7	48.6	14.26	65.3	49.2	10.19	87.0	60.1	11.14
COR Analyzer	12.92	87.7	50.3	14.53	71.7	47.8	12.00	89.8	59.5	12.22
AutoCoronaryTree	15.07	84.7	46.5	15.88	59.5	36.1	14.23	86.2	50.3	15.11
CocomoBeach	16.23	78.8	42.5	17.66	64.4	40.0	14.19	81.2	46.9	16.83
VirtualContrast	16.44	75.6	39.2	18.26	56.1	34.5	14.80	78.7	45.6	16.27

the two terms, is heuristically tuned on a few datasets and then fixed throughout the experiments. The optimal path can be calculated efficiently using dynamic programming. For patients with short discernible coronaries, the centerline extracted by the above model-driven step may be too long; therefore, the distal centerline may be traced into a non-coronary structure. We shrink the centerline from the end point, one by one, if its vesselness score is less than a threshold. The centerline is then extended to the distal end to extract the full length of the coronary. After the verification-and-extension process, errors in the distal centerline are corrected.

After extracting centerlines of the three major coronary arteries, the algorithm starts to trace side branches. First, the bifurcation of a side branch is detected on a major centerline using region growing based lumen segmentation. Starting from a centerline point, bright voxels connected to the current point are added iteratively. The growing front (composed of the added voxels in the latest iteration) is monitored. If a side branch presents, the region growing procedure will go into the side branch. A side branch is detected when we find a front with a distance to the existing major centerline larger than a threshold. At each detected bifurcation point, a data-driven centerline tracing process is initialized [7]. Please refer to [7] for more details for the tracing of a coronary sub-tree from a given starting point.

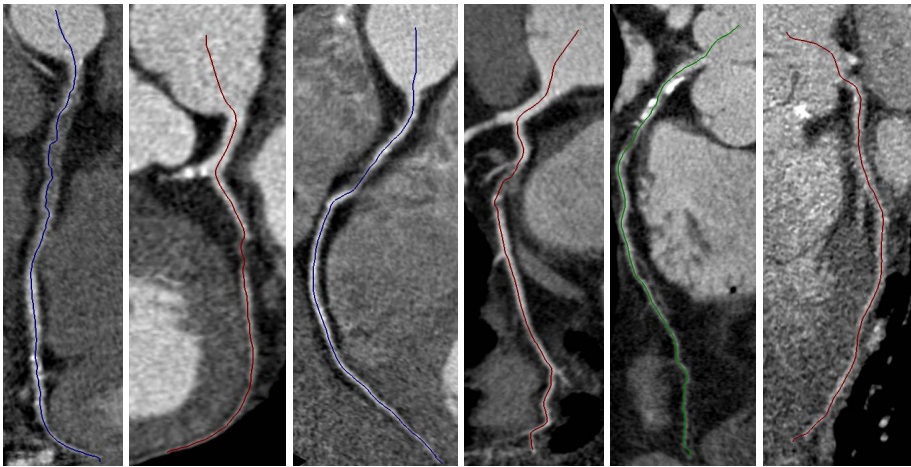
Though in general the extracted centerline is located at or close to the lumen center, sometimes, it may take a shortcut at a tortuous segment or be attracted to calcifications. Further post-processing is exploited to move the centerline to the lumen center.

4 Experiments

The algorithm was trained on our proprietary datasets (108 volumes) and evaluated on the public Rotterdam coronary CTA database [2]. The Rotterdam database contains a training set (eight datasets) and a test set (24 datasets), and each dataset has four manually annotated coronary artery centerlines, namely the RCA, LAD, LCX, and a randomly picked large side branch. An algorithm is evaluated with the overlap and accuracy inside (AI) metrics. The overlap metric is further broken down into three individual measurements, including overlap (OV), overlap until first error (OF), and overlap with the clinically relevant part of the vessel (OT). All measurements are based on point-to-point

Table 2. Comparison with other centerline extraction methods on the Rotterdam coronary CTA test set (24 datasets) using the accuracy inside (AI) metric

Method	Automatic	AI			Method	Automatic	AI		
		mm	Score	Rank			mm	Score	Rank
Proposed Method	Y	0.20	51.6	2.48	VesselTractography	N	0.36	30.7	14.99
ShapeRegression	N	0.23	49.6	3.10	VirtualContrast	Y	0.39	30.6	15.01
MHT	N	0.23	47.9	3.86	GVFTube'n'Linkage	Y	0.37	29.8	15.79
SupervisedExtraction	Y	0.25	47.3	4.28	KnowledgeBasedMinPath	N	0.39	29.2	16.22
Tracer	N	0.26	44.4	6.17	ElasticModel	N	0.40	29.3	16.22
COR Analyzer	Y	0.25	44.8	6.78	TwoPointMinCost	N	0.46	28.0	16.94
VirtualContrast2b	N	0.27	41.6	7.56	AxialSymmetry	N	0.46	26.4	18.07
DepthFirstModelFit	Y	0.28	41.9	7.63	StatisticalTracking	N	0.51	25.1	18.22
BayesianMaxPaths	N	0.29	37.0	10.41	TubSurfGradFlow	N	0.47	24.8	19.14
GFVCoronaryExtractor	Y	0.30	37.1	10.41	3DInteractiveTrack	N	0.51	24.2	19.84
CocomoBeach	Y	0.29	37.7	10.71	CoronaryTreeMorphoRec	N	0.59	20.7	20.75
AutoCoronaryTree	Y	0.34	35.3	11.32					

**Fig. 3.** Examples of the extracted coronary centerlines on the Rotterdam database. The first five patients have occlusions and the last one has low image quality.

correspondence between the detected centerline and the ground truth. A centerline point is claimed to be detected correctly if its distance to the corresponding ground truth point is no more than the radius of the annotated lumen at that point. The AI metric measures the distance between the extracted centerline and the ground truth for the correctly detected centerline part. A score is further assigned based on the inter-observer variability (score 100 for a perfect result and score 50 for an error matching the inter-observer variability). The algorithms are ranked on each vessel and the average ranking is reported. Please refer to [2] for more details about the datasets and evaluation metrics.

On the training set, the proposed method outperforms the other automatic methods on all overlap metrics (OV, OF, and OT). On the test set, as shown in Table 1, our method has the best average ranking (ranking 1st on OF, and 2nd on OV and OT) among all automatic methods (nine submissions in total). Regarding the accuracy measurement (AI), the proposed method clearly outperforms all other 22 algorithms (including both

automatic and semi-automatic methods) on both the training and test sets. Table 2 shows the ranking of all algorithms on the test set using the AI metric. With an average score of 51.6, we achieve an accuracy comparable to inter-observer variability. Please refer to <http://coronary.bigr.nl/centerlines/results/results.php> for more details. Fig. 3 shows a few examples of extracted centerlines using the proposed method.

5 Conclusion

We proposed a novel centerline extraction method combining the advantages of model-driven (robustness under severe occlusions) and data-driven approaches (the capability to handle anatomical variations of the coronary tree). Experiments on the public Rotterdam coronary CTA database demonstrate the robustness and accuracy of the proposed method. We achieve the best average ranking on overlap metrics among automatic methods and our accuracy metric outperforms all other 22 methods (including both automatic and semi-automatic methods).

References

1. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis* 13(6), 819–845 (2009)
2. Schaap, M., Metz, C.T., van Walsum, T., et al.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical Image Analysis* 13, 701–714 (2009)
3. Kitamura, Y., Li, Y., Ito, W.: Automatic coronary extraction by supervised detection and shape matching. In: *Proc. IEEE Int'l Sym. Biomedical Imaging*, pp. 234–237 (2012)
4. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Medical Imaging* 27(11), 1668–1681 (2008)
5. Zheng, Y., Tek, H., Funka-Lea, G., Zhou, S.K., Vega-Higuera, F., Comaniciu, D.: Efficient detection of native and bypass coronary ostia in cardiac CT volumes: Anatomical vs. pathological structures. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part III*. LNCS, vol. 6893, pp. 403–410. Springer, Heidelberg (2011)
6. Bookstein, F.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Machine Intell.* 11(6), 567–585 (1989)
7. Tek, H., Gulsun, M.A., Laguitton, S., Grady, L., Lesage, D., Funka-Lea, G.: Automatic coronary tree modeling. *The Insight Journal*, 1–8 (2008)
8. Zheng, Y., Shen, J., Tek, H., Funka-Lea, G.: Model-driven centerline extraction for severely occluded major coronary arteries. In: Wang, F., Shen, D., Yan, P., Suzuki, K. (eds.) *MLMI 2012*. LNCS, vol. 7588, pp. 10–18. Springer, Heidelberg (2012)
9. Bernardini, F., Mittleman, J., Rushmeier, Silva, C., Taubin, G.: The ball-pivoting algorithm for surface reconstruction. *IEEE Trans. Visualization and Computer Graphics* 5(4), 349–359 (1999)
10. Zheng, Y., Loziczonek, M., Georgescu, B., Zhou, S.K., Vega-Higuera, F., Comaniciu, D.: Machine learning based vesselness measurement for coronary artery segmentation in cardiac CT volumes. In: *Proc. of SPIE Medical Imaging*, vol. 7962, pp. 1–12 (2011)
11. Shahzad, R.K., Schaap, M., van Walsum, T., Klein, S., Weustink, A.C., van Vliet, L.J., Niessen, W.J.: A patient-specific coronary density estimate. In: *Proc. IEEE Int'l Sym. Biomedical Imaging*, pp. 9–12 (2010)

Incorporating Shape Variability in Image Segmentation via Implicit Template Deformation

Raphael Prevost^{1,2}, Remi Cuingnet¹, Benoit Mory¹,
Laurent D. Cohen², and Roberto Ardon¹

¹ Philips Research Medisys, Suresnes, France

² CEREMADE UMR 7534, CNRS, Université Paris Dauphine, Paris, France

Abstract. Implicit template deformation is a model-based segmentation framework that was successfully applied in several medical applications. In this paper, we propose a method to learn and use prior knowledge on shape variability in such framework. This shape prior is learnt via an original and dedicated process in which both an optimal template and principal modes of variations are estimated from a collection of shapes. This learning strategy requires neither a pre-alignment of the training shapes nor one-to-one correspondences between shape sample points. We then generalize the implicit template deformation formulation to automatically select the most plausible deformation as a shape prior. This novel framework maintains the two main properties of implicit template deformation: topology preservation and computational efficiency. Our approach can be applied to any organ with a possibly complex shape but fixed topology. We validate our method on myocardium segmentation from cardiac magnetic resonance short-axis images and demonstrate segmentation improvement over standard template deformation.

1 Introduction

Model-based methods are particularly effective and popular in medical image segmentation. Among them, template deformation has recently been used in various applications [1–4] for its interesting properties (computational efficiency, topology preservation, compatibility with user interactions). This variational method consists in seeking a segmenting implicit function as a deformed implicit template. This template, acting as a shape prior, is therefore of paramount importance. However, in previous works the initial template was either set as a synthetic model (*e.g.* ellipsoid for a kidney [2–4]) or as a segmented organ from a single arbitrary image [1]. Despite the consensus that learning shape priors is a powerful approach to improve robustness [5, 6], this has never been proposed in the context of segmentation by implicit template deformation.

In this paper, our goal is thus to improve implicit template deformation by taking into account learnt shape information while keeping both topology preservation and computational efficiency. Shape learning is often performed through statistical analysis of boundary vertices [6, 7] or implicit functions [8, 9]. However

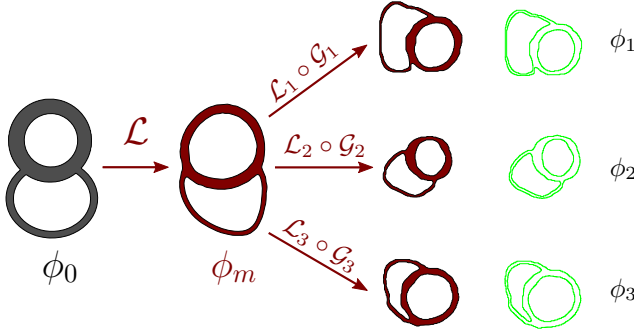


Fig. 1. Given an initial synthetic shape ϕ_0 , a set of shapes $\{\phi_i\}_i$ is simultaneously segmented via implicit template deformation while an intermediate mean shape $\phi_m = \phi_0 \circ \mathcal{L}$ (with the same topology as ϕ_0) is estimated

neither of these two representations guarantees topology preservation after averaging. To do so, other methods represent shapes via diffeomorphisms [10, 11], but they come with a high computational cost that is not compatible with real-time segmentation. We therefore propose a dedicated learning approach by using the template deformation energy as a pre-metric in the shapes space. This yields a co-segmentation process (similar to [12] for registration), within which an optimal template is estimated (see Figure 1). We also capture further information by building a space of main deformations around this template. Finally, we introduce a generalization of the template deformation formulation by using the computed statistics in the regularization term. The proposed framework is generic and can be applied to any organ with a possibly complex and variable shape but a fixed topology. We demonstrate its efficacy and interest by addressing the problem of myocardium segmentation in 2D cine-MR images.

In the following, Section 2 introduces the main notations and recalls the implicit template deformation framework. In Section 3, we develop the learning framework to estimate statistics that will be used to improve segmentation in Section 4. Optimization details are provided in Section 5. Validation results on clinical data are presented in Section 6 and discussion concludes the paper.

2 Segmentation by Implicit Template Deformation

Implicit template deformation [1, 2] is a variational framework for image segmentation. The segmentation is defined through the zero level-set of an implicit function $\phi : \Omega \rightarrow \mathbb{R}$, and ϕ is positive (resp. negative) inside (resp. outside) the segmentation. In this framework, the set of admissible segmentations \mathbb{S} is defined via an implicit template $\phi_0 : \Omega \rightarrow \mathbb{R}$ as the set of all implicit functions with the same topology as ϕ_0 , i.e. $\mathbb{S} = \{\phi : \Omega \rightarrow \mathbb{R} \text{ s.t. } \phi = \phi_0 \circ \psi, \psi \text{ is diffeomorphic}\}$. The unknown is thus the transformation $\psi : \Omega \rightarrow \Omega$ which is sought as a minimum of a region competition energy:

$$\min_{\psi} \left\{ \int_{\Omega} H(\phi_0 \circ \psi) r_{int} + \int_{\Omega} (1 - H(\phi_0 \circ \psi)) r_{ext} + \lambda \mathcal{R}(\psi) \right\}, \quad (1)$$

where H denotes the Heaviside function ($H(a) = 1$ if $a > 0$, 0 otherwise) while r_{int} and r_{ext} are image-based functions such as $r_{int}(\mathbf{x})$ is lower (resp. higher) than $r_{ext}(\mathbf{x})$ if voxel \mathbf{x} seems to belong to the target object (resp. background). $\mathcal{R}(\psi)$ is a constraint term on ψ that prevents the segmentation $\phi = \phi_0 \circ \psi$ to deviate too much from the initial template ϕ_0 ; it is weighted by a parameter λ . In [2], ψ is decomposed into (i) a global transformation $\mathcal{G} \in \mathbb{G}$ (e.g. a similarity) accounting for the pose of the template in the image, and (ii) a diffeomorphism $\mathcal{L} \in \mathbb{D}(\Omega)$ that yields local deformation and does change the shape of the template. This decomposition allows to define the regularization as a function of the deformation only $\mathcal{R}(\psi) = \mathcal{R}(\mathcal{L}) = \frac{1}{2} \|\mathcal{L} - Id\|_2^2$. The problem finally reads

$$\min_{\mathcal{L}, \mathcal{G}} \left\{ \int_{\Omega} H(\phi_0 \circ \mathcal{L} \circ \mathcal{G}) r_{int} + \int_{\Omega} (1 - H(\phi_0 \circ \mathcal{L} \circ \mathcal{G})) r_{ext} + \frac{\lambda}{2} \|\mathcal{L} - Id\|_2^2 \right\}. \quad (2)$$

In such a setting, ϕ_0 not only fixes the topology of the segmentation but also acts as a shape prior, which makes its choice of paramount importance. Moreover, the term \mathcal{R} could be improved by taking into account shape variability of the considered organ. In the next section, we develop a framework to tackle both problems by estimating statistics on a collection of shapes.

3 A Dedicated Learning of Shape Variability

Consider N shapes of a given organ (coming for example from manual expert segmentations) implicitly represented by $\{\phi_i\}_{i=1..N} \subset \mathbb{S}$. From this set we aim to extract useful statistical information in terms of segmentation, that is to say a mean shape and a deformation model.

In order to estimate statistics in \mathbb{S} , we first define an adapted pre-metric in this space. Any shape $\phi_1 \in \mathbb{S}$ can be warped to another shape $\phi_2 \in \mathbb{S}$ via implicit template deformation by solving Problem (2) with $\phi_0 := \phi_1$, and for example $r_{int}^{\phi_2} := \max(-\phi_2, 0)$ and $r_{ext}^{\phi_2} := \max(\phi_2, 0)$. This leads to a dedicated definition of shape dissimilarity:

$$C^2(\phi_1, \phi_2) = \min_{\substack{\mathcal{L} \in \mathbb{D} \\ \mathcal{G} \in \mathbb{G}}} \left\{ \int_{\Omega} H(\phi_1 \circ \mathcal{L} \circ \mathcal{G}) r_{int}^{\phi_2} + \int_{\Omega} (1 - H(\phi_1 \circ \mathcal{L} \circ \mathcal{G})) r_{ext}^{\phi_2} + \frac{\lambda}{2} \|\mathcal{L} - Id\|_2^2 \right\}$$

As in [12], our dedicated notion of mean is defined via a minimization problem :

$$\phi_m = \operatorname{argmin}_{\phi \in \mathbb{S}} \sum_{i=1}^N C^2(\phi, \phi_i) \Leftrightarrow \phi_m = \phi_0 \circ \left\{ \operatorname{argmin}_{\mathcal{L} \in \mathbb{D}(\Omega)} \sum_{i=1}^N C^2(\phi_0 \circ \mathcal{L}, \phi_i) \right\} \quad (3)$$

The right-handside equivalence comes from the constraint $\phi_m \in \mathbb{S}$. Indeed, the mean template has to preserve the topology of the training shapes. Expanding the segmentation costs and neglecting constant terms in Equation (3) yields the following optimization problem to solve:

$$\min_{\substack{\mathcal{L} \in \mathbb{D}(\Omega) \\ (\mathcal{L}_i)_{i \in \mathbb{D}(\Omega)}^N, (\mathcal{G}_i)_{i \in \mathbb{G}^N}}} E_{learn} = - \sum_{i=1}^N \int_{\Omega} H(\phi_0 \circ \mathcal{L} \circ \mathcal{L}_i \circ \mathcal{G}_i) \phi_i + \frac{\lambda}{2} \|\mathcal{L}_i - Id\|_2^2. \quad (4)$$

This can be interpreted as segmenting simultaneously all training shapes $\{\phi_i\}_i$ starting from ϕ_0 while estimating an optimal common intermediate shape $\phi_0 \circ \mathcal{L}$ (see Figure 1). In Eq (4), the energy is minimized (see Section 5) with respect to three kinds of variables

- the global transformations $(\mathcal{G}_i)_i$, called the **poses**, that register all shapes to ϕ_0 with translation, rotation and scaling. As they are part of the optimization process (see Section 5), they do not bias the learning, as a fixed pre-alignment (*e.g.* [8, 9]) would do.
- the **common deformation** \mathcal{L} , which includes the common parts of the deformations from ϕ_0 to all the training shapes.
- the local deformations $(\mathcal{L}_i)_i$, called the **residual deformations**, are the residual components of the deformations from $\phi_0 \circ \mathcal{L}$ to ϕ_i . Unlike \mathcal{L} , their magnitude is penalized so that any deformation which is common to all the training set will be preferably included in \mathcal{L} .

The optimal common deformation \mathcal{L}^* can be used to define the optimal template as $\phi_m = \phi_0 \circ \mathcal{L}^*$. This shape globally minimizes the magnitude of residual deformations to each shape of the dataset. Note that \mathcal{L} is not penalized so the choice of ϕ_0 defines the topology of ϕ_m but does not affects it further.

The optimal residual deformations $(\mathcal{L}_i^*)_i$ are also available and can be used to capture further information on the variability of the training shapes. We build a space of principal deformations \mathbb{L} to constrain future segmentation of new images. As in [13], a principal component analysis (PCA) is applied to the residual deformations $(\mathcal{L}_i^*)_i$ to find a suitable parametrization of such a space. Any deformation $\ell \in \mathbb{L}$ can then be written as a linear combination of the offset $\bar{\ell}$ and $(\ell_k)_{k=1..M}$ the first M modes of variation:

$$\ell[w] = \bar{\ell} + \sum_{k=1}^M w_k \ell_k \quad , \quad w \in \mathbb{R}^M. \quad (5)$$

The space of diffeomorphisms is not stable under linear combinations so elements of \mathbb{L} are not necessarily diffeomorphisms. Nevertheless we show hereafter how this space indirectly in a topology-preserving segmentation framework.

4 Generalized Implicit Template Deformation

The previously estimated statistical information can be used to improve the implicit template deformation described in Eq. (2). A first improvement is achieved by replacing the original template ϕ_0 by the mean template $\phi_m = \phi_0 \circ \mathcal{L}^*$. Secondly, the estimation of the deformation can also be enhanced by using the space of principal deformations \mathbb{L} . In most previous work [7, 9, 14], the learnt variable is directly expressed as a linear combination of modes. Here we rather modify the regularization term \mathcal{R} so that diffeomorphism \mathcal{L} is constrained with respect to the set \mathbb{L} instead of the identity. Thus, only deformations that cannot be explained through the learnt space \mathbb{L} are penalized. The rationale is to use linear

combinations of diffeomorphisms indirectly to preserve both the topology and the computational efficiency. The new segmentation energy therefore reads

$$E_{seg}(\mathcal{L}, \mathcal{G}, w) = \int_{\Omega} H(\phi_m \circ \mathcal{L} \circ \mathcal{G}) r_{int} + (1 - H(\phi_m \circ \mathcal{L} \circ \mathcal{G})) r_{ext} + \frac{\lambda}{2} \|\mathcal{L} - \ell[w]\|_2^2. \quad (6)$$

Minimization of (6) can be performed with an alternate scheme:

Update of the Segmentation. With $\ell[w]$ fixed, the energy is minimized through a gradient descent-like scheme on \mathcal{L} and \mathcal{G} (see Section 5).

Update of the Modes Weights. With \mathcal{L} and \mathcal{G} fixed, the update of $L[w]$ can be seen as a projection of \mathcal{L} onto \mathbb{L} . Indeed the energy minimization comes down to a simple quadratic problem, whose closed-form solution is

$$\forall k \in \{1, \dots, M\}, w_k = \langle \mathcal{L} - \bar{\ell}, \ell_k \rangle / \langle \ell_k, \ell_k \rangle. \quad (7)$$

The first step is similar to standard template deformation [2] and the second one is straightforward. Therefore, the proposed algorithm maintains the efficiency of the original algorithm. Further details on optimization are provided herebelow.

5 Optimization Schemes

Both learning (4) and segmentation (6) energies involve variables, either in \mathbb{G} for poses or in $\mathbb{D}(\Omega)$ for deformations, that are simultaneously updated. Variables in \mathbb{G} can be parameterized by a vector \mathbf{p} in \mathbb{R}^P (translation, rotation angles, scales). Minimization of a given energy E is done through a gradient descent on this vector: $\mathbf{p}^{(n+1)} \leftarrow \mathbf{p}^{(n)} - \Delta t \nabla_{\mathbf{p}} E$, where Δt is the time step. On the other hand, such a process is not suitable in $\mathbb{D}(\Omega)$ as this space is not stable under linear combinations. A more appropriate way is to combine diffeomorphisms via composition since $(\mathbb{D}(\Omega), \circ)$ is a group. Following [1], we therefore update any diffeomorphism \mathcal{L} in the following way: $\mathcal{L}^{(n+1)} \leftarrow (Id - \Delta t \nabla_{\mathcal{L}} E) \circ \mathcal{L}^{(n)}$. The regularity is enforced by a Gaussian filtering of the gradient as in [2].

6 Validation

We validated our method in the context of myocardium analysis and segmentation in cardiac short-axis 2D cine-MR images. This task is particularly challenging for model-based approaches because of the complex topology of the target object, i.e. a band around left and right ventricles.

Our dataset is composed of 245 MR images coming from 61 different patients, which was randomly split into a training set (120 images, 30 patients) and a testing set (125 images, 31 patients). The acquisitions have been synchronised so that each heart is in the same cardiac phase. In every image, a myocardium segmentation has been manually performed by a radiologist. The initial synthetic template ϕ_0 used is shown in Figure 1.

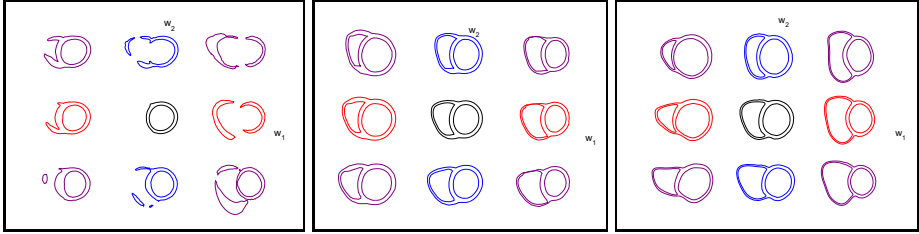


Fig. 2. Mean model and first two modes of the variation of the myocardium learnt on the training dataset using implicit shape model [9] (left), active shape model [7] (middle) and the proposed method (right). For our approach, the visualized shapes are the zero level-sets of $\phi_m \circ (\bar{\ell} + w_1 \ell_1 + w_2 \ell_2)$.

First we compare qualitatively the shape information learnt from the implicit shape model proposed in [9], the active shape framework [7] and our proposed approach. Figure 2 shows the mean shape and first two modes of variation for each method. The implicit method fails at recovering the true topology of the mean shape, as well as its first modes of variations. The explicit method performs better and provides a reasonable mean model. However, the modes of variation are less satisfying than the modes of deformations learnt with our approach (which tends to provide a better topology preservation and seem more realistic).

Then we evaluate how learnt information improves segmentation via implicit template deformation of unseen images. Myocardiums have been segmented in test images using (i) the synthetic model ϕ_0 as template, (ii) the estimated mean model ϕ_m as template, (iii) the new deformation model-based regularization term in addition to the mean model ϕ_m (with 5 modes). The image-based classification functions r_{int} and r_{ext} were negative log-likelihoods of intensity probability distributions inside and outside the myocardium. Performance of each algorithm is quantified using Dice coefficients between the segmentation and the expert ground truth. Results on the whole testing set are summarized in Figure 3. Both the replacement of the template ϕ_0 by ϕ_m and the introduction of the new regularization term improved the robustness of the segmentations (p -value < 0.0001 for a Wilcoxon signed-rank test). To illustrate this improvement, Figure 4 shows some results in three different cases, for the classical regularization term with two values of the shape constraint parameter $\lambda \in \{1, 2\}$ and the new model-based regularization term. In all settings, the template was the mean model ϕ_m . Consider Case #1: since the image term is reliable, a satisfying result is obtained with a small shape constraint. However, the myocardium deviates significantly from the mean shape: using a too strong constraint λ_2 prevents the algorithm to converge towards the right solution. Conversely in Case #2, the image information is much more ambiguous. This provokes some leaks with λ_1 , which shows there is no fixed value that allows a good segmentation in both cases. Yet by introducing the new regularization (fourth column), likely deformations are not penalized. This allows us to widen the capture range while still

avoiding unrealistic leaks. Finally, case #3 illustrates that our method may also improve the result even if no λ was originally successful.

Fig. 3. Boxplot of the Dice coefficients for myocardium segmentation in MR images via implicit template deformation with synthetic model ϕ_0 (left), mean model ϕ_m (middle), mean model ϕ_m and deformation model (right)

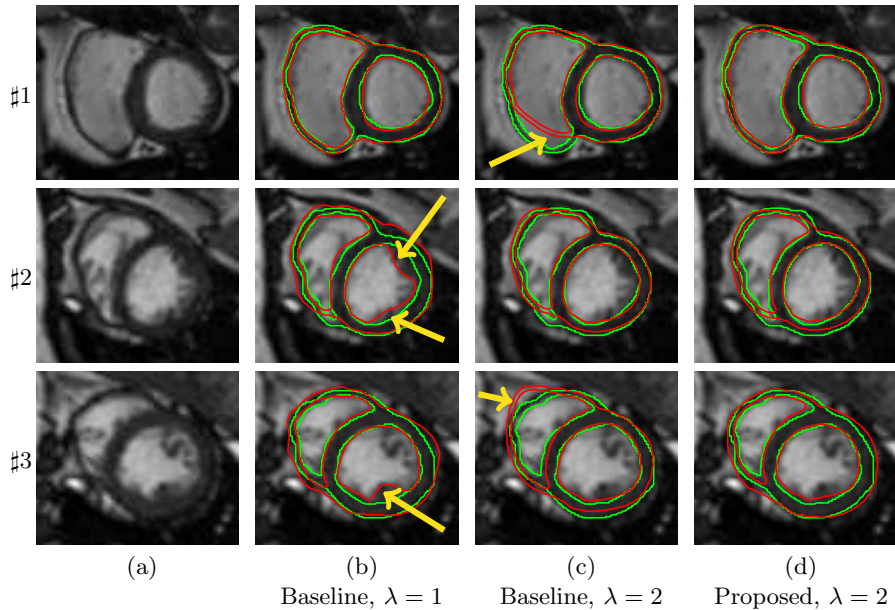
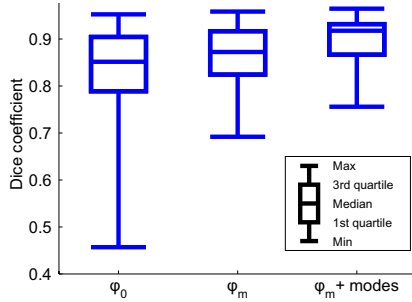


Fig. 4. Segmentation results (red) of different cases versus ground truths (green). Main failures are highlighted by yellow arrows. (a) Original images, (b,c) Baseline method [2] with small ($\lambda = 1$) and high ($\lambda = 2$) shape constraint, (d) Proposed method.

7 Conclusion

In this paper we have presented a method to include organ shape variability in the implicit template deformation framework. A variational approach was proposed to extract statistical information (mean and principal variations) from a collection of shapes. This training method is automatic, does not require landmarks correspondance and relies upon a definition of shape dissimilarity that is

directly derived from the implicit template deformation functional. We also proposed a generalization of the original segmentation algorithm in which the shape prior is automatically adapted to the current image during the deformation process with almost no additional cost (segmentation takes around one second on a standard computer). Quantitative results demonstrated the improvement over implicit template deformation for a 2D application. Our approach is very generic and can be used to segment any object with a complex shape but a fixed topology that shall be preserved. Furthermore, extension in 3D or to multiple objects (*e.g.* brain structures) is straightforward thanks to the implicit representation of shapes. Despite its paramount importance, the image-based term was not investigated as we focused on incorporating shape information on top of any pixelwise classifier. We plan to extend our framework to learn a dedicated appearance model as well.

References

1. Saddi, K., et al.: Region-Based Segmentation via Non-Rigid Template Matching. In: ICCV, vol. 1, pp. 1–7 (2007)
2. Mory, B., Somphone, O., Prevost, R., Ardon, R.: Real-Time 3D Image Segmentation by User-Constrained Template Deformation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 561–568. Springer, Heidelberg (2012)
3. Prevost, R., et al.: Kidney Detection and Real-Time Segmentation in 3D Contrast-Enhanced Ultrasound Images. In: IEEE ISBI, pp. 1559–1562 (2012)
4. Cuingnet, R., Prevost, R., Lesage, D., Cohen, L.D., Mory, B., Ardon, R.: Automatic Detection and Segmentation of Kidneys in 3D CT Images Using Random Forests. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 66–74. Springer, Heidelberg (2012)
5. Cremers, D., et al.: A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. IJCV 72(2), 195–215 (2007)
6. Heimann, T., Meinzer, H.P.: Statistical Shape Models for 3D Medical Image Segmentation: a Review. Medical Image Analysis 13(4), 543 (2009)
7. Cootes, T.F., et al.: Active Shape Models-Their Training and Application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
8. Rousson, M., Paragios, N.: Shape Priors for Level Set Representations. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part II. LNCS, vol. 2351, pp. 78–92. Springer, Heidelberg (2002)
9. Tsai, A., et al.: A Shape-Based Approach to the Segmentation of Medical Imagery using Level Sets. Trans. Med. Imaging 22(2), 137–154 (2003)
10. Vaillant, M., et al.: Statistics on Diffeomorphisms via Tangent Space Representations. NeuroImage 23(1), 161 (2004)
11. 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)
12. Joshi, S., et al.: Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. NeuroImage 23(1), 151 (2004)
13. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic Construction of 3D Statistical Deformation Models using Non-Rigid Registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 77–84. Springer, Heidelberg (2001)
14. Leventon, M., et al.: Statistical Shape Influence in Geodesic Active Contours. In: CVPR, pp. 316–323 (2000)

Automatic 3D Motion Estimation of Left Ventricle from C-arm Rotational Angiocardiology Using a Prior Motion Model and Learning Based Boundary Detector

Mingqing Chen¹, Yefeng Zheng¹, Yang Wang¹,
Kerstin Mueller^{2,3}, and Guenter Lauritsch²

¹ Imaging and Computer Vision, Siemens Corporate Technology, Princeton, NJ, USA

² Healthcare Sector, Siemens AG, Forchheim, Germany

³ Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-University Erlangen-Nuremberg, Germany
yefeng.zheng@siemens.com

Abstract. Compared to pre-operative imaging modalities, it is more convenient to estimate the current cardiac physiological status from C-arm angiocardiology since C-arm is a widely used intra-operative imaging modality to guide many cardiac interventions. The 3D shape and motion of the left ventricle (LV) estimated from rotational angiocardiology provide important cardiac function measurements, e.g., ejection fraction and myocardium motion dyssynchrony. However, automatic estimation of the 3D LV motion is difficult since all anatomical structures overlap on the 2D X-ray projections and the nearby confounding strong image boundaries (e.g., pericardium) often cause ambiguities to LV endocardium boundary detection. In this paper, a new framework is proposed to overcome the aforementioned difficulties: (1) A new learning-based boundary detector is developed by training a boosting boundary classifier combined with the principal component analysis of a local image patch; (2) The prior LV motion model is learned from a set of dynamic cardiac computed tomography (CT) sequences to provide a good initial estimate of the 3D LV shape of different cardiac phases; (3) The 3D motion trajectory is learned for each mesh point; (4) All these components are integrated into a multi-surface graph optimization method to extract the globally coherent motion. The method is tested on seven patient scans, showing significant improvement on the ambiguous boundary cases with a detection accuracy of 2.87 ± 1.00 mm on LV endocardium boundary delineation in the 2D projections.

1 Introduction

With multi-view or rotational angiocardiology, the 3D shape and motion of the left ventricle (LV) can be estimated for intra-operative analysis of cardiac physiology, e.g., ejection fraction (EF) and myocardium motion dyssynchrony. Previous methods in 3D LV shape/motion estimation from angiocardiology include binary voxel reconstruction based on 2D image intensities [1], elastic matching [2], and 4D B-spline model [3], etc. Most of these methods rely on manual editing of the silhouette of the LV endocardium due to the difficulty in automatic boundary detection. Recently, an automatic

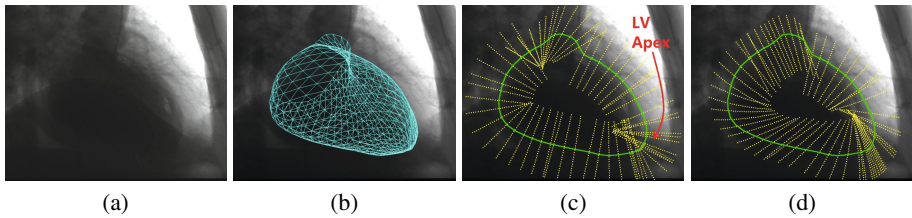


Fig. 1. (a) One projection image; (b) Projected initial mesh; (c) The silhouette (green contour) of the projected mesh and projected trajectory (yellow lines) based on the normal direction; (d) The silhouette of the projected mesh and projected trajectory based on the prior model

method was proposed for 3D motion estimation of the LV from rotational angiocardiology [4]. First, a 3D LV mesh model is segmented from the 3D volume containing motion blur. The mesh is then forward-projected to each 2D X-ray angiogram as an initial estimate of the LV boundary (Fig. 1b). The silhouette of the projected mesh is extracted and deformed to fit the LV boundary observed on the angiogram. Therefore, the 2D motion vector can be estimated for each silhouette point. With the assumption that the 3D LV motion is along the surface normal, the estimated 2D motion vector can be converted into 3D motion.

In this paper, a new method is proposed to overcome several limitations in the previous work. First, the initialization with a static LV model [4] (which is close to the shape of the end-diastolic phase) is not accurate enough, especially for the end-systolic phase, which has a much smaller LV volume. To address this issue, a prior LV motion model is learned from a set of dynamic cardiac CT sequences. Using the previous 3D static LV model generated from the non-gated reconstruction, the prior motion model is aligned to the patient coordinate system to provide individualized initial mesh for each cardiac phase. Second, the previous assumption of motion along the surface-normal is reasonable for the middle and basal LV segments, but not good for the LV apex (Fig. 1c) with many intersections in the trajectories of mesh points around the apex. In this work, for each mesh point, a 3D motion trajectory is also learned from the dynamic cardiac sequences to replace the assumption of motion along the surface normal. Third, instead of using each individual local image feature [4] for the weak classifiers to train an AdaBoost based boundary detector, principal component analysis (PCA) is applied to a local image patch to create a compact feature representation. The new feature pool is composed of the top principal components (PC) and the combination of a pair of PCs, which are more powerful than the previous local image features. Last, instead of estimating the 3D LV shape of each cardiac phase independently, we exploit the recently proposed multi-surface graph optimization approach [5] to extract the whole motion sequence from one single optimization procedure.

2 Method

2.1 Problem Formulation

In rotational angiocardiology, the C-arm X-ray source/detector pair is rotated around the patient over 200 degrees while acquiring 2D projection images. Imaging starts with

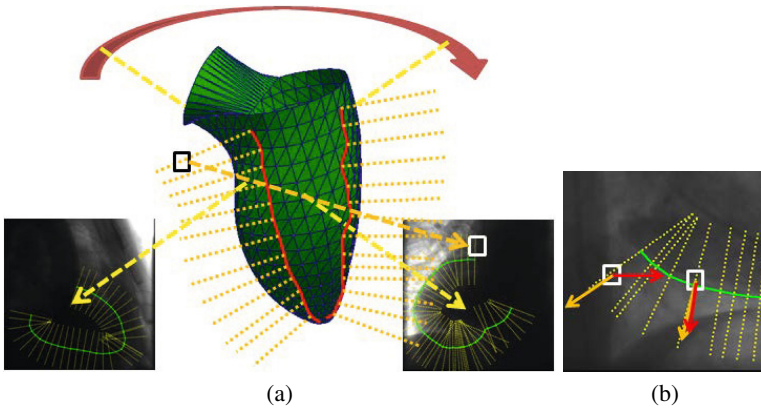


Fig. 2. (a) An illustration of computing cost values along motion trajectory. (b) An illustration of computing the dot product between the 2D motion vector (yellow arrow) and the image gradient (red arrow) in two locations indicated by the white rectangle.

a delay of 1s with respect to the injection of contrast agent into the heart chambers. The electrocardiography (ECG) signal is used to record the cardiac phase synchronous to the acquisition of each projection image. The patients were asked to hold breath during the acquisition. In the proposed framework, The ECG signal is analyzed and a relative phase bin is assigned to each projection image. The LV motion is assumed to be periodic and is represented by a sequence of triangulated meshes $\vec{n}(m, t)$, which indicates the 3D spatial location of a mesh point indexed by $m \in [0, M - 1]$ at phase $t \in [0, T - 1]$. Here, M is the number of mesh points and T is the number of cardiac phase bins. This representation inherently enforces the identical topology of the LV meshes and point correspondence in different phases. The motion detection framework is established on predetermined motion trajectories from an initial location, which provides a close approximation to the detected motion and is used to create the base of the 4D graph. The motion trajectories determine the possible location for each mesh point before the detection. They are sampled by a limited number of points $n_s(m, t)$, where $n_s \in [0, N - 1]$ is the point index. (For an initial estimation based on a static mesh, $n_s(m, t)$ is irrelevant to t .) Each sampled point $n_s(m, t)$ is assigned a cost value, which is used during the optimal detection framework. The optimization aims to minimize the following function,

$$\hat{n}(m, t) = \arg \min_{n_s(m, t)} \sum_{t=0}^{T-1} \sum_{m=0}^{M-1} \text{cost}(n_s(m, t), m, t), \quad (1)$$

where \hat{n} is selected from all the possible n_s . The details of the graph-search-based motion detection framework can be referred to Chen et al. [5], where both spatial and temporal smoothness constraints are incorporated in the optimization framework.

The assignment of the cost value for motion trajectories on the silhouette is illustrated in Fig. 2a, where the black and the white rectangles represent the original 3D sampled location and the corresponding 2D projected locations. For each sampled point the cost

value is computed by the following functions:

$$\text{cost}(n_s, m, t) = - \sum_{p=0}^{P-1} \delta(p, t) \zeta(m, p) w(n_s, m, p), \quad (2)$$

where P is the number of projection images. Function $\delta(p, t)$ equals to one if and only if the projection image p belongs to the phase index t . Otherwise, it equals to zero. Function $\zeta(m, p)$ equals to one if and only if the mesh point m locates on the silhouette in projection p . Otherwise, it equals to zero. Function $w(n_s, m, p)$ computes the cost value for a sampled point n_s of a mesh point m in one projection image p . In [5], the cost function $w(n_s, m, p)$ is assigned based on the dot-product of the projection motion direction and image gradient (Fig. 2b), which works well for lung tumor boundary detection in 2D projections of mega-voltage cone beam CT. In this application, the image boundary of the LV endocardium is much weaker than a nearby confounding boundary of the LV pericardium (the boundary between the heart and lung). Therefore, the deformed silhouette is often attracted to the pericardium. In this work, we propose to use the response of a learning-based boundary detector as the cost function $w(n_s, m, p)$ (Section 2.4).

2.2 Prior Motion Model

The prior LV motion model is derived from a training set containing 12 LV motion sequences extracted from dynamic cardiac CT data sets, where each sequence contains 10 LV meshes (10 phases). The mean LV surface is computed after aligning the 4D (3D+t) shapes based on the Procrustes analysis [6]. Given an input dataset of a patient, a static LV mesh is extracted from the non-ECG-gated reconstruction. The static mesh and the prior motion model are represented by the same LV model with built-in point correspondences. The prior motion model is transformed to the patient coordinate system specified by the static mesh based on the thin-plate-spline (TPS) interpolation approach [7] to provide an personalized initial mesh for each phase. The static mesh is found close to the shape of the end-diastolic (ED) phase, instead of the average shape of the whole cardiac cycle. The displacement between ED phase of the mean LV motion model and the static mesh provides the landmarks/anchor points for the registration technique.

2.3 Motion Trajectory

In order to convert the 2D motion vector estimated from the angiogram to 3D, a previous assumption is that the motion occurs mainly along the surface normal [4]. It is reasonable for most parts of the LV where the motion is dominated by the contraction and relaxation of the LV cavity. However, this assumption does not hold well for the LV apical region, where the LV endocardium forms a sharp wedge. As shown in Fig. 1c, the projected normal directions have severe self-intersections around the LV apex. The intersections in the motion trajectory will eventually result in folds in the estimated 3D mesh.

In this work, the prior motion trajectory is fitted by a 3D quadratic curve for each mesh point (m, t) . The fitting is based on using the transformed prior motion model

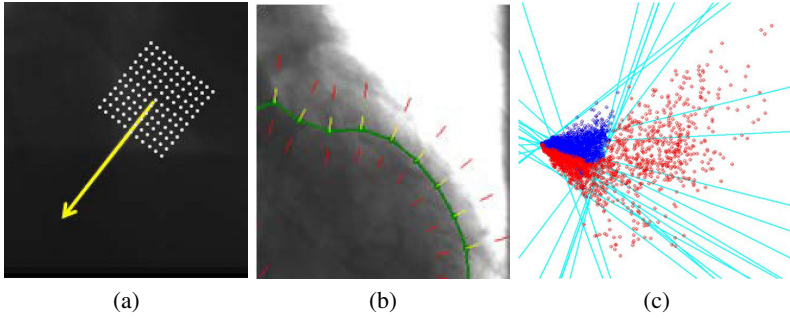


Fig. 3. (a) Sample of an image patch; (b) Along a manual contour (green) in one projection image, positive (yellow arrows) and negative (red arrows) locations are selected to train the boundary detector; (c) Positive (blue) and negative (red) training data visualized in PCA-reduced 2D space. Each cyan line shows the classification boundary of a weak classifier.

introduced in the previous section, where a parametric curve with the following representation is determined by the least-squared errors (LSE) with 10 points (one from each cardiac phase and the whole cardiac cycle contains 10 phases):

$$x(t) = x_0 + a_1t + a_2t^2, \quad (3)$$

$$y(t) = y_0 + b_1t + b_2t^2, \quad (4)$$

$$z(t) = z_0 + c_1t + c_2t^2. \quad (5)$$

(x_0, y_0, z_0) is assigned with the point location of the initial mesh model.

Fig. 1d shows the projected motion trajectory. Compared to the normal direction, the learned motion trajectory has much fewer self-intersections around the LV apex.

2.4 Boundary Detector Based on Supervised Learning

Instead of computing the image gradient for each projected 2D location [5], a binary classifier is trained from a data set containing manually labeled positive and negative data points. The graph node cost $w(n_s, m, p)$ in Equation (2) is computed based on the posterior probability $p(+1|x)$, where x is a feature vector including the intensity of a sampled 11×11 pixel patch centered on the projected location. The orientation of the patch is determined by the tangent direction of the motion trajectory at its center location. Fig. 3a illustrates one image patch example.

The manually annotated LV contours on the projection images are used to create the training data set, which is illustrated in Fig. 3b. The contour points are used as positive data, while the negative data are located at certain distance outside or inside of the contour. The orientation of both positive or negative patches is set to the normal direction of the nearby annotated contour.

The training stage includes a dimensionality reduction step based on the principal component analysis (PCA) and an AdaBoost training step [8]. During the PCA step, the original 121 dimensional space is reduced to a few principal components (PC) with

largest variations. Two types of features are extracted after PCA to train an AdaBoost classifier. The first type of features contains coefficients for the 50 most significant PCs:

$$f_{T1}(X_r) = \sum_i w_i I_{ri}, \quad (6)$$

where w_i and I_{ri} represents the weight and value of the i^{th} PC, respectively. The second type of features contains a linear combination of the first two PCs since the two most significant PCs cover about 98% of all the variance,

$$f_{T2}(X_r) = \cos(\theta)I_{r1} + \sin(\theta)I_{r2}, \quad (7)$$

where I_{r1} and I_{r2} are the value of the training point X_r in the reduced dimension and θ sets the relative weight of these two PCs. Parameter θ is quantized with a step size of 1 degree to generate an finite number of 2D linear features. Fig. 3c shows a few selected type-II features by the AdaBoost classifier. The positive and negative training data are shown as blue and red dots, respectively. A cyan line shows a type-II feature with a specific combination of the first two PCs.

For each training point, the feature function f_{T1} or f_{T2} is able to generate a value used for a weak classifier:

$$h_i(X_r) = \text{sign}(p_i f_i(X_r) - p_i \gamma_i), \quad (8)$$

where p_i is ± 1 , representing the parity of the data, and γ_i is the threshold for the binary classification. The subscript i in those denotations represents the i^{th} feature in the feature pool. A strong classifier is generated by combining a set of selected weak classifiers from the feature pool with a value F :

$$H(X_r) = \text{sign}(F(X_r)), \quad F(X_r) = \sum_{i=1}^K \sigma_i h_i(X_r), \quad (9)$$

where σ_i is a parameter related to the training error of the i^{th} weak classifier and K is the number of selected weak classifiers. The posterior probability of the binary AdaBoost classifier is given by:

$$q(+1|X_r) = \frac{e^{2F(X_r)}}{1 + e^{2F(X_r)}}, \quad q(-1|X_r) = \frac{e^{-2F(X_r)}}{1 + e^{-2F(X_r)}}. \quad (10)$$

3 Experiments

We collected seven sequences of C-arm rotational angiocardiology from two clinical sites. A typical imaging protocol was composed with a rotation of the C-arm X-ray source/detector around the patient for 200 degrees. Each rotation took about five seconds and a total of 133 projection images were recorded. Leave-one-out cross validation is performed to quantitatively evaluate the accuracy of the proposed method by computing the distance between the mesh silhouette and a manual contour. The distance of the two contours is computed by measuring the distance between each silhouette point and its closest location on the manual contour:

$$\text{dist}(C_1, C_2) = \frac{1}{N_p} \sum_{p_1 \in C_1} \min_{p_2 \in C_2} \text{dist}(p_1, p_2), \quad (11)$$

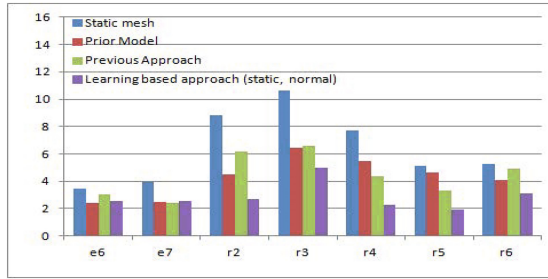


Fig. 4. Mean of the distance (measured in mm) between silhouette derived based on four approaches and manual contour

where p_1 and p_2 are the points belonging to contour C_1 and C_2 , respectively. N_p is the number of points in C_1 that are taken to compute the distance metric. To evaluate the performance of the algorithm for each scan, the mean and variance of the contour distance in all the projection images are computed. Fig. 4 shows the mean distance before and after the motion estimation in seven scans. The bars “Static Mesh” show the initialization error using a static mesh extracted from a 3D volume using non-ECG-gated reconstruction, while the bars “Prior Model” show the initialization error using the pre-learned motion model, which is more accurate than the initialization using a static mesh. The performance of the previous approach [4] is shown as the green bar, which is surpassed by the proposed method (the purple bar). The proposed method achieves a mean contour distance of 2.87 ± 1.00 mm. It should be noted that for a gradient-based approach, the difference to the manual contour is even larger than the initial static model, and it is not shown in this figure.

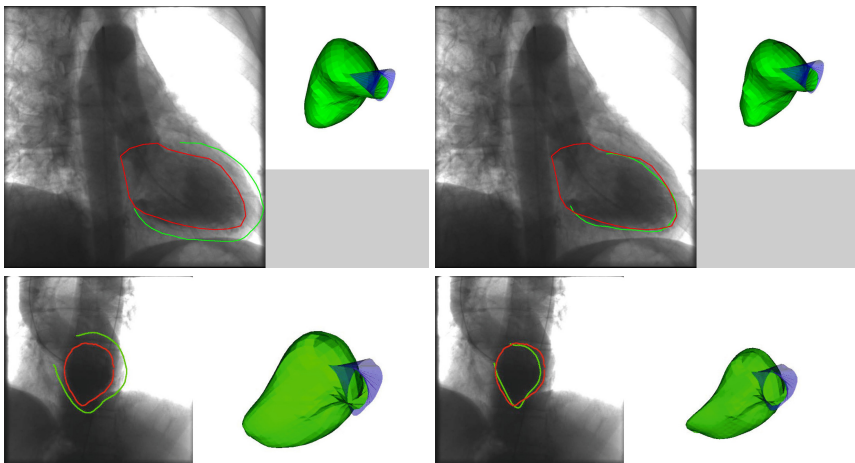


Fig. 5. Detected silhouette (green) using gradient-based boundary detector (left, [5]) and learning-based boundary detector (right) compared with manual contour (red). The 3D mesh of LV (green) and outflow tract (blue) is shown on the right for each 2D projection image.

Fig. 5 compares the detection result of a gradient-based cost function [5] and the proposed learning-based cost function. It can be seen that for the learning-based boundary detector, the detected silhouette is much closer to the manual contour, while the gradient-based approach is attracted to the confounding boundary.

4 Discussions and Conclusions

In this work, a new method was proposed to automatically estimate the 3D motion of the left ventricle from rotational angiocardiology, which incorporated a boundary detector based on the principal component analysis of local image patches. All the components were integrated into a multi-surface graph optimization framework to achieve a coherent 3D motion estimate. The estimated 3D motion can be used for motion compensated reconstruction and intra-operative measurement of the cardiac function, including ejection fraction and dyssynchrony analysis.

References

1. Prause, G.P.M., Onnasch, D.G.W.: 3D reconstruction of the ventricular dynamic shape from the density profiles of biplane angiocardiology image sequences. In: *IEEE Proc. Computers in Cardiology*, pp. 193–196 (1994)
2. Lotjonen, J., Magnin, I., Nenonen, J., Katila, T.: Reconstruction of 3-D geometry using 2-D profiles and a geometric prior model. *IEEE Trans. Medical Imaging* 18(10), 992–1002 (1999)
3. Moriyama, M., Sato, Y., Naito, H., Hanayama, M., Ueguchi, T., Harada, T., Yoshimoto, F., Tamura, S.: Reconstruction of time-varying 3D left ventricular shape from multiview x-ray cineangiocardiology. *IEEE Trans. Medical Imaging*, 773–785 (2002)
4. Chen, M., Zheng, Y., Mueller, K., Rohkohl, C., Lauritsch, G., Boese, J., Funke-Lea, G., Hornegger, J., Comaniciu, D.: Automatic extraction of 3D dynamic left ventricle model from 2D rotational angiocardiology. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part III*. LNCS, vol. 6893, pp. 471–478. Springer, Heidelberg (2011)
5. Chen, M., Bai, J., Zheng, Y., Siochi, R.A.C.: 3D lung tumor motion model extraction from 2D projection images of mega-voltage cone beam CT via optimal graph search. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I*. LNCS, vol. 7510, pp. 239–246. Springer, Heidelberg (2012)
6. Gower, J.C.: Generalized procrustes analysis. *Psychometrika* 40(1), 33–51 (1975)
7. Bookstein, F.: Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Machine Intell.* 11(6), 567–585 (1989)
8. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1), 119–139 (1997)

Interventional Digital Tomosynthesis from a Standard Fluoroscopy System Using 2D-3D Registration

Mazen Alhrishy¹, Andreas Varnavas¹, Tom Carrell²,
Andrew King¹, and Graeme Penney¹

¹ Biomedical Engineering Dept., King's College London, London, UK*

² Vascular Surgery Dept., Guys & St Thomas' NHS Foundation Trust, London, UK

Abstract. Fluoroscopy is the mainstay of interventional radiology. However, the images are 2D and visualisation of vasculature requires nephrotoxic contrast. Cone-beam computed tomography is often available, but involves large radiation dose and interruption to clinical workflow. We propose the use of 2D-3D image registration to allow digital tomosynthesis (DTS) slices to be produced using standard fluoroscopy equipment. Our method automatically produces patient-anatomy-specific slices and removes clutter resulting from bones. Such slices could provide additional intraoperative information, offering improved guidance precision. Image acquisition would fit with interventional clinical workflow and would not require a high x-ray dose. Phantom results showed a 1133% contrast-to-noise improvement compared to standard fluoroscopy. Patient results showed our method enabled visualisation of clinically relevant features: outline of the aorta, the aortic bifurcation and some aortic calcifications.

Keywords: Interventional digital tomosynthesis, 2D-3D image registration, endovascular aneurysm repair.

1 Introduction

The fundamentals behind interventional fluoroscopy have remained largely unchanged since its inception. Large advances have been made in detector sensitivity, however, clinicians still view 2D projective “shadow” images which simply integrate all information along the beam path. This often results in clinically relevant information being obscured by over- or under-lying anatomy.

Enhancement of blood vessels using iodinated contrast is routine, but must be used sparingly as contrast is nephrotoxic. In modern fluoroscopy suites, 3D imaging is often available via semicircular C-arm rotation, i.e. cone beam CT (CBCT). However, the set-up time for CBCT (5~10 min) can cause a large interruption to clinical workflow, especially if multiple acquisitions are required [8]. Repeated CBCT also involves a significant radiation dose [2]. In addition, the

* Thanks go to the King's Overseas Research Studentship scheme for funding.

3D nature of CBCT images requires some interaction from clinicians to scan through 2D sections to find the clinically relevant information. For these reasons, CBCT is not a natural interventional modality, and is unlikely to be used repeatedly during interventions to aid guidance.

Tomosynthesis was the first medical sectional modality, but was largely superseded by computed tomography after its invention in the 1970s. In the last decade, however, digital tomosynthesis (DTS) is being increasingly used for diagnosis of breast lesions and pulmonary nodules as it offers some of the tomographic benefits of CT but at substantially lower dose and shorter acquisition time [3]. Nevertheless, such diagnostic systems require dedicated equipment and suffer from the presence of background “clutter”, caused by high contrast features outside the slice of interest.

Recently, a 3D DTS prototype system, based on a mobile isocentric C-arm, has been proposed for intraoperative guidance of head and neck surgery [1]. The limited DTS arc (e.g. $20^\circ \sim 90^\circ$) enabled a short acquisition time and low radiation dose causing minimal interruption to surgical workflow [2]. However, apart from being modified for intraoperative use, the prototype employs the same technique as diagnostic DTS systems and suffers from the same drawbacks.

In this submission, we propose the use of 2D-3D image registration to facilitate DTS as an interventional modality which allows repeated acquisitions and results in minimal interruption to standard clinical workflow. Moreover, we propose a method which: produces DTS slices using a standard fluoroscopy system; can automatically produce patient-anatomy-specific DTS slices that display the most clinically relevant information; and can automatically remove clutter resulting from bony anatomy.

2 Materials and Methods

DTS slice reconstruction requires a set of 2D intraoperative images to be acquired from a limited range of view directions. These are reconstructed into a sectional slice, commonly using the shift and add method, which combines the fluoroscopy images so that structures in the reconstruction plane line-up, and so appear in-focus, while structures outside the reconstruction plane are not aligned, and so are blurred-out.

2.1 DTS Required Information and Main Limitation

In order to reconstruct a DTS slice, the following is required:

1. Relative view positions of input 2D images.
2. Reconstruction plane position with respect to the imaging device.

Standard diagnostic DTS obtains relative view positions using mechanical tracking. This requires a calibration process, and calibration errors can result in artifacts and reduced image quality [6]. Also in standard diagnostic DTS, a number of slices are reconstructed on planes defined with respect to the imaging

device. Prior to reconstruction it is not possible to define a reconstruction plane to image specific regions of the patient’s anatomy.

DTS attempts to blur-out all structures outside the reconstruction plane, but because of the limited data acquisition, clutter from high contrast structures above and below the reconstruction plane remain. A number of methods have been proposed to reduce the effect of clutter [4], but this remains one of the main problems of DTS [5].

2.2 Using 2D-3D Registration to Facilitate Enhanced DTS Using a Standard Fluoroscopy System

Our interventional DTS method uses an established intensity-based 2D-3D registration algorithm [7]. The novelty of this paper is with respect to the use of 2D-3D registration to facilitate improved DTS reconstruction using standard hardware, on patient-anatomy-specific surfaces and with reduced clutter.

Figure 1 shows an overview of the entire process. This begins at the top with the input images: a C-arm sweep (1) to produce a set (of size n) of intraoperative fluoroscopy images (2), and a preoperative CT scan (3). These images are input into the 2D-3D registration algorithm which calculates the 2D-3D transformations P_i between the CT scan and each of the n fluoroscopy images. The registration provides us with the necessary information to carry out DTS and enables us to greatly reduce clutter from bone. This is illustrated by the three boxes showing the output from the 2D-3D registration in Fig. 1, where the Roman numerals labelling each box correspond to the below descriptions:

- I. Calculate view positions: The transformations P_i enable the relative view positions of the input 2D images to be determined.
- II. Position reconstruction surface: A patient-anatomy-specific plane can be pre-operatively defined in the CT scan. The transformations P_i can position this plane with respect to the fluoroscopy images, enabling reconstruction to occur on a patient-anatomy-specific plane.
- III. Remove bones: Bony details from the CT scan in the form of digitally reconstructed radiographs (DRRs) can be produced for each fluoroscopy image using the transformations P_i . DRRs are produced by casting rays through automatically segmented vertebra from the CT and integrating voxel values above a threshold (200 HU). Therefore, DRR intensities come only from vertebral bone ((4).a). The DRR can be subtracted from the fluoroscopy image ((4).b) to give a “deboned” image ((4).c) leaving just the soft tissue and aortic calcifications (features of clinical interest), and any interventional instruments. This deboning process is carried out on each fluoroscopy image prior to reconstruction to greatly reduce clutter from high contrast bony features.

In addition, we propose the use of curved patient-anatomy-specific reconstruction surfaces. It is rare that structures of clinical interest lie on flat planes. Our aim in interventional DTS is to produce images with enhanced clinically relevant

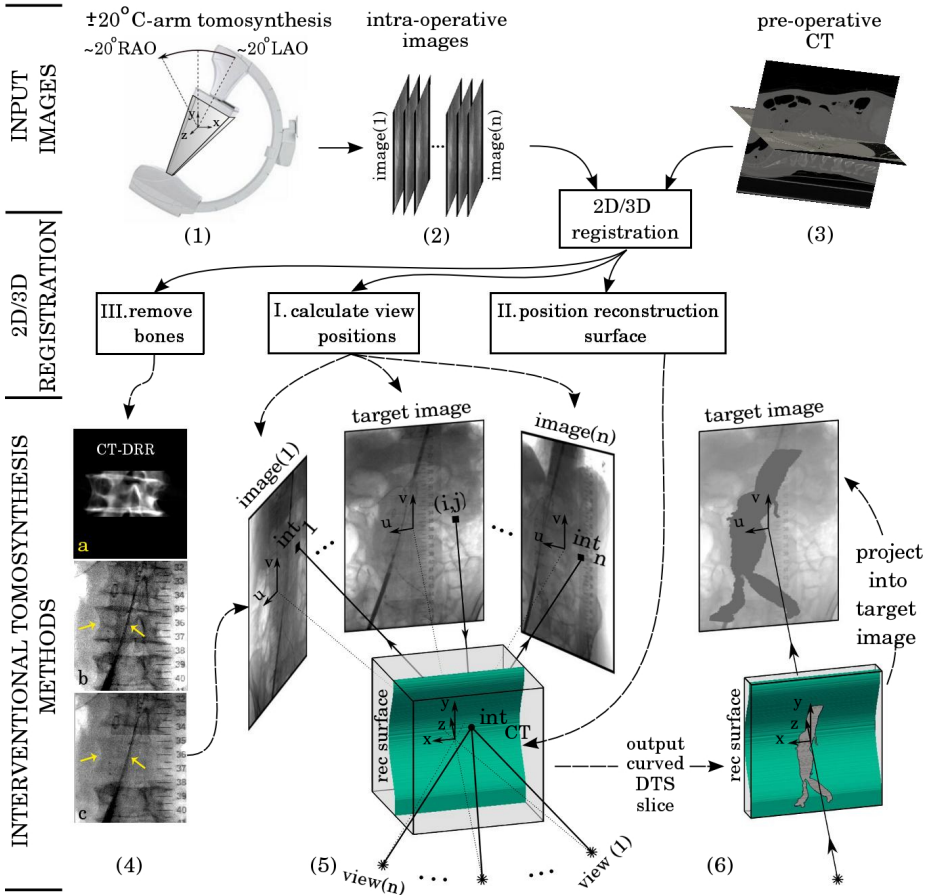


Fig. 1. A flow diagram of overall method. Top shows input images (intraoperative fluoroscopy images (1,2) and preoperative CT (3)). Middle shows the 2D-3D registration which enables bone removal (4), calculation of view directions and positioning of curved patient-anatomy-specific reconstruction surface. Bottom shows interventional DTS process producing the DTS slice containing additional clinical information (5) and then projection of this information into the fluoroscopy image to aid guidance (6).

structures. As shown in Fig. 2, if the structure of clinical interest is the aorta, then only approximately half of its length could be included in a flat reconstruction plane (Fig. 2.(a&b)). The use of a curved surface allows reconstruction of the entire length of the aorta (Fig. 2.(c&d)).

Returning to Fig. 1, the tomosynthesis process is shown in (5). Here, after bone removal and using the transformations P_i , rays are back projected from the target image onto the patient-anatomy-specific reconstruction surface (e.g. pixel (i,j)). Rays are then forward projected from 3D interception positions (e.g. int_{CT}) to each of the other images in turn. The intensity values from each

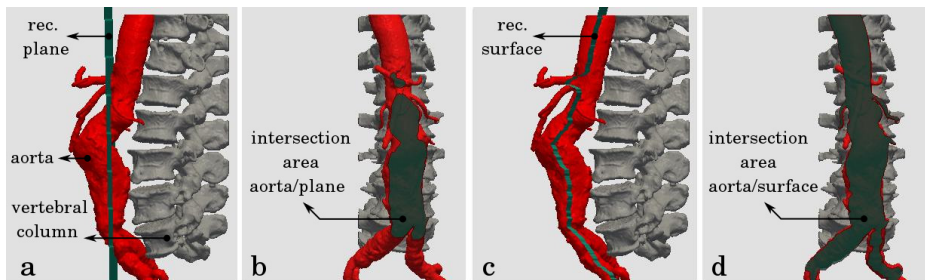


Fig. 2. Illustration of advantage of using curved surface over flat plane for DTS where aorta is feature of clinical interest: (a,c) sagittal views showing (a) flat reconstruction plane intersecting aorta and (c) curved reconstruction surface along aortic centreline; (b,d) anterior views showing (b) flat plane only intersects with roughly half of aorta whereas (d) curved surface can intersect (and therefore can reconstruct) entire aorta

2D interception position (e.g. int_1, \dots, int_n) are summed to produce a curved patient-anatomy-specific DTS slice. In order to allow effective use of this new information, the reconstructed slice is then projected onto the target image being used to guide the operation (6). This automatically produces an enhanced fluoroscopy image which shows additional information of the clinical features of interest, in the view being clinically used.

3 Data and Experiments

Experiments were carried out using data from an abdominal spine phantom and from two patients who underwent endovascular aortic repair (approved by national research ethics committee (09/H0707/64)). Each data set had a pre-operative CT scan; and intraoperative low dose screening images acquired by rotating the C-arm $\sim 20^\circ$ right/left anterior oblique with a frame rate of 30 *fps*, which were resampled to obtain one image per one degree of rotation, i.e. ~ 40 images. Set-up time for the $\sim 20^\circ$ sweep took less than a minute for each data acquisition. For comparison, a series of ~ 40 screening images from an anterior-posterior view were also saved and averaged to produce a high contrast image from a single view direction.

The phantom CT had voxel sizes of $1.094 \times 1.094 \times 1.487 \text{ mm}^3$. Prior to fluoroscopy acquisition, an interventional instrument (a catheter) and three pieces of Blu-Tack (to represent calcium in the aortic wall) were placed on the anterior surface of the phantom. Therefore, the anterior surface of the CT volume was used for reconstruction.

The patients' standard diagnostic CT scans had voxel sizes of approximately $0.75 \times 0.75 \times 0.8 \text{ mm}^3$. The reconstruction surface was defined to intersect the curved aortic centreline and to be perpendicular to the sagittal plane. This surface was chosen to enhance features of interest such as the aortic walls.

DTS slice reconstruction, as described in Sec. 2.2, was carried out for both phantom and clinical data sets to reconstruct two interventional DTS slices, the

first using the standard fluoroscopy images, and the second using the fluoroscopy images after applying the deboning process. After providing a starting position for the first 2D-3D registration the registrations were totally automatic. As real data was used, our experiments included realistic registration errors.

For the phantom data set, the contrast-to-noise ratio (CNR) values were calculated on profile lines as follows: $CNR = (\bar{I}_{FG} - \bar{I}_{BG})/\sigma_{BG}$, where \bar{I}_{FG} and \bar{I}_{BG} were the mean foreground and background intensities, and σ_{BG} was the standard deviation in the background region.

For clinical data, CT segmentations were overlaid onto the DTS slices to provide context to the enhanced features. The first overlay shows the aorta, and the second shows aortic calcification. The overlay was initially positioned using the 2D-3D registration, and then due to intraoperative aortic deformation, the overlay was manually moved to match the aortic outline in the DTS slices.

4 Results

For each data set we show: the target image (TI), the high contrast image (CI), the reconstructed slice (DTS) and the reconstructed slice after deboning, i.e deboned DTS (DDTS). We also show the two CT overlays for the clinical data.

Figure 3 shows the phantom results. The high contrast catheter can be clearly seen in all images; whereas the low contrast synthetic calcium cannot be clearly distinguished in ‘a’ nor ‘b’ from the overlying vertebrae. However, in both DTS reconstructions, ‘c’ and ‘d’, the synthetic calcium is successfully brought into focus (indicated by circles). Significantly more clutter from the underlying vertebrae can be seen in ‘c’, compared to the reconstruction after deboning ‘d’.

Table 1 shows CNR results, and percentage improvement in CNR compared to TI, calculated on the profile lines (PL) shown in Fig. 3. CNR values were calculated on three lines close to the PLs shown (one on the PL and two with the line vertically shifted by ± 3 pixels), and the results were averaged. Foreground region was defined as the region within FWHM calculated using the DDTS image (as features could be most clearly observed). PLs 1,2 and 3 are through synthetic calcium and PL4 is through the catheter. An average improvement of 72% is seen between TI and CI as random noise is averaged. Both DTS and DDTS show much improved CNR for the low contrast synthetic calcium and for the high contrast catheter, and the further improvement due to the deboning method is clearly seen.

Figure 4 shows the patient data results. Comparing the overlay outline in ‘e’ and ‘d’ shows how the DDTS method has been able to show the outline of the aorta. Some calcium deposits (indicated by arrows) were also enhanced, and for patient 1 the aortic bifurcation was visible. Comparison between ‘c’ and ‘d’ clearly show the benefits of the deboning process, and although ‘b’ shows a high contrast image of the instruments and bony anatomy, none of the clinically relevant soft tissue features enhanced by the DTS process are visible.

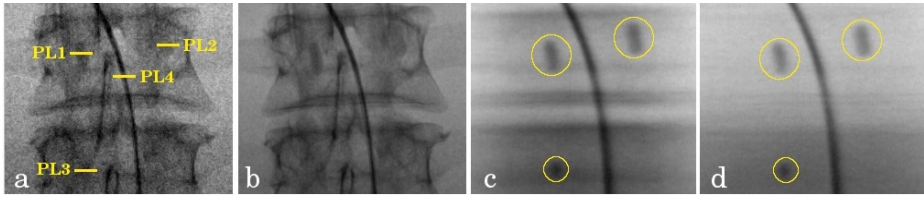


Fig. 3. Phantom results: a) TI with profile lines positions, b) CI, c) DTS and d) DDTS. Circles indicate synthetic calcium.

Table 1. Calculated CNR, with improvement compared to TI in brackets, for the four PLs shown in Fig. 3

	TI (a)	CI (b)	DTS (c)	DDTS (d)
PL1: CNR (Imp.)	0.43 (-%)	1.10 (156%)	4.69 (990%)	4.77 (1009%)
PL2: CNR (Imp.)	0.25 (-%)	0.28 (12%)	3.67 (1368%)	5.06 (1924%)
PL3: CNR (Imp.)	0.56 (-%)	0.63 (12%)	4.50 (703%)	4.61 (723%)
PL4: CNR (Imp.)	0.64 (-%)	1.33 (108%)	6.37 (895%)	6.25 (877%)
Average Imp.	-	72%	989%	1133%

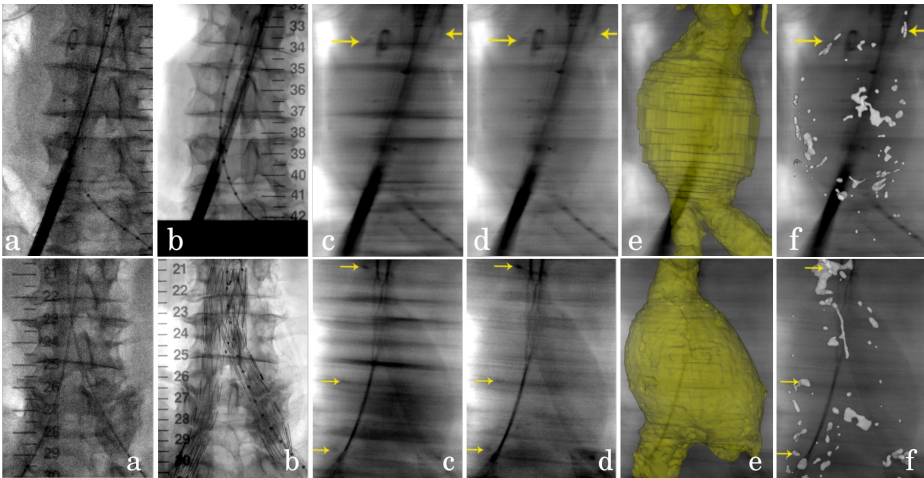


Fig. 4. Patient 1 (top) and patient 2 (bottom) results: (a) TI, (b) CI, (c) DTS, (d) DDTS, (e) AAA overlay and (f) calcification overlay into the DDTS slice. Arrows indicate calcium deposits.

5 Discussion and Conclusions

The development of novel imaging technologies capable of near-real-time visualisation of soft-tissue structures in the interventional suite is challenging. Short acquisition and reconstruction times, low radiation dose and minimal interruption to the clinical workflow are key requirements for an effective interventional

modality. We have presented a novel technique, “interventional digital tomosynthesis”, which can be directly implemented on existing fluoroscopy systems. The small C-arm sweep of $\pm 20^\circ$ takes a fraction of the image acquisition time and radiation dose compared to CBCT, and causes very little disruption to the clinical workflow.

Our method was able to enhance clinically important structures situated on a curved surface. These structures could provide additional spatial information during interventions, offering surgeons an increased guidance precision and confidence. Contrast usage would be reduced compared to the techniques currently used to visualise the aorta, which require injection of iodinated contrast.

The preoperative CT overlays (Fig. 4.(e&f)) needed manual adjustment to accurately match our DDTS images. This was due to anatomical deformation occurring during intervention caused by the stiff interventional instruments [7]. This shows a potential application for our interventional DTS: to provide additional information to update overlays from an image guided surgery system enabling more accurate representation of the intraoperative scene.

In summary, a novel method of interventional DTS has been presented. The method employs a 2D-3D registration algorithm to enable production of DTS slices using standard interventional equipment, with much reduced out-of-plane clutter and on a patient tailored reconstruction surface. Results from a phantom and two patients show the method’s ability to automatically enhance structures of clinical interest.

References

1. Bachar, G., Barker, E., Nithiananthan, S., Chan, H., Daly, M., Irish, J., Siewerdsen, J.: Three-dimensional tomosynthesis and cone-beam computed tomography: an experimental study for fast, low-dose intraoperative imaging technology for guidance of sinus and skull base surgery. *The Laryngoscope* 119(3), 434–441 (2009)
2. Bachar, G., Siewerdsen, J., Daly, M., Jaffray, D., Irish, J.: Image quality and localization accuracy in C-arm tomosynthesis-guided head and neck surgery. *Med. Phys.* 34, 4664 (2007)
3. Dobbins III, J.: Tomosynthesis imaging: at a translational crossroads. *Med. Phys.* 36(6), 1956–1967 (2009)
4. Dobbins III, J., Godfrey, D.: Digital x-ray tomosynthesis: current state of the art and clinical potential. *Phys. Med. Biol.* 48(19), 65–106 (2003)
5. Gang, G., Tward, D., Lee, J., Siewerdsen, J.: Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. *Med. Phys.* 37(5) (2010)
6. Jaffray, D., Siewerdsen, J., Wong, J., Martinez, A., et al.: Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* 53(5), 1337–1349 (2002)
7. Penney, G., Varnavas, A., Dastur, N., Carrell, T.: An image-guided surgery system to aid endovascular treatment of complex aortic aneurysms: Description and initial clinical experience. In: Taylor, R.H., Yang, G.-Z. (eds.) *IPCAI 2011*. LNCS, vol. 6689, pp. 13–24. Springer, Heidelberg (2011)
8. Wallace, M., Kuo, M., Glaiberman, C., Binkert, C., Orth, R., Soulez, G.: Three-dimensional C-arm cone-beam CT: applications in the interventional suite. *J. Vasc. Interv. Radiol.* 19(6), 799–813 (2008)

Calibrationless Parallel MRI with Joint Total Variation Regularization

Chen Chen, Yeqing Li, and Junzhou Huang

University of Texas at Arlington, TX, USA 76019

Abstract. In this paper, a calibrationless method is proposed for parallel magnetic resonance imaging (pMRI). It is motivated by the observation that the gradients of the aliased images are jointly sparse. Therefore, the pMRI problem can be formulated as a joint total variation regularization task. The field of view is finally obtained via a sum of square approach. We develop an iterative algorithm to efficiently solve this problem. Experiments on pMRI datasets demonstrate that our method outperforms the state-of-the-art pMRI methods even when they can achieve sufficient calibration data, and far better than existing calibrationless pMRI algorithms. Clinic MR applications could benefit from this method even when accurate calibration is limited or not possible at all.

1 Introduction

It is routine to accelerate MRI with parallel technique in clinic applications by using multiple receiver coils to acquire undersampled k-space data. The coils are installed around the patient at different locations. Thus, the MR signals are scanned with different spatial sensitivities. If estimated precisely, the sensitivity information could be used to interpolate the missing data due to undersampling.

The data acquired by each receiver coil corresponds to an aliased image. Based on the way to reconstruct each aliased image, existing methods can be classified broadly in two types: a) image domain methods such as SENSE [1] and JSENSE [2], which directly transform the undersampled data to aliased images and then unfold them to the field of view (FOV) via SENSE encoding; b) frequency domain methods such as GRAPPA [3] and SPIRiT [4], which interpolate the missing data to fill the full k-space and then transform them to aliased images. For the image domain methods, the sensitivity map is required to be estimated. Any noise or inaccuracy in the sensitivity map will be amplified significantly and result in visible artifacts in the FOV. For the frequency domain methods, the interpolation weights need to be calibrated with sufficient auto-calibration signals (ACS). However, the efficiency or feasibility of calibration is limited in many applications such as dynamic MRI and non-Cartesian imaging. Without sufficient ACS, the performance of all these methods and their improved versions [5][6] can not be guaranteed.

To avoid calibration, an algorithm CaLMMRI is proposed recently [7]. It is based on compressive sensing MRI (CSMRI) [8] scheme to reconstruct MR image from undersampled data without the need for the sensitivity information. Because all the aliased images represent the same anatomical cross section,

they are jointly reconstructed by exploiting the joint sparsity among them. The FOV is obtained by the sum of square (SoS) approach with the reconstructed aliased images. However, their assumption on the appearance similarity among aliased images does not hold strictly due to the sensing difference caused by the coil locations. This makes their algorithm still hard to be comparable to the state-of-the-art calibration methods such as GRAPPA and SPIRiT. As we are interested in calibrationless pMRI, CSSENSE [5] could also be viewed as a calibrationless method if the SENSE encoding step is replaced by SoS. Compared to CaLMMRI, CSSENSE can only reconstruct each aliased image coil-by-coil. No spatial prior knowledge in the data other than sparsity is utilized to improve the reconstruction. Moreover, all these CS based MRI methods [7][5] strongly rely on the incoherence of the sampling matrix [9][10], which can not be always guaranteed in MR applications.

In this paper, we propose a new pMRI method called Joint Total Variation MRI (JTVMRI), which is an extension of the total variation (TV) [11] regularization for ill-posed problem. Unlike CaLMMRI, we do not assume the images of the different coils have similar appearance but only assume their gradients are similar. Since MR images are often piecewise smooth, most of the gradients are approximate to zeros. The non-zero gradients only appear on the edges, which is consistent in all the aliased images of different coils. An efficient algorithm is developed to efficiently solve the ill-posed pMRI problem. Our method does not require any calibration step and can be applied on arbitrary sampling which is not available for GRAPPA and SPIRiT. Extensive experiments demonstrate that the proposed method is far better than CaLMMRI. And more importantly, it outperforms the state-of-the-art auto-calibration method such as GRAPPA and SPIRiT, which makes calibrationless pMRI much feasible than before.

2 Related Work

2.1 CSSENSE

CSSENSE[5] is a hybrid approach, which apply CSMRI[4] to reconstruct the aliased images and the second step is to unfold them to full FOV with SENSE encoding[1]. Because we are interested in calibrationless pMRI in this paper, we consider it as a calibrationless CSMRI method, where the second step replaced with SoS. The problem to reconstruct one aliased image x_c is:

$$x_c = \arg \min_{x_c} \left\{ \frac{1}{2} \|Fx_c - b_c\|_2^2 + \alpha \|x_c\|_{TV} + \beta \|\Phi x_c\|_1 \right\} \quad c = 1, 2, \dots, C \quad (1)$$

where C denotes the total number of coils; Φ denotes the wavelet transform; F is the undersampled Fourier operator and b_c is the undersampled k-space data for coil c ; α, β are two parameters. $\|x\|_{TV} = \sum_i \sqrt{(\nabla_1 x_i)^2 + (\nabla_2 x_i)^2}$ for an image of N pixels, $i = 1, \dots, N$, where ∇_1 and ∇_2 denote the forward finite difference operators on the first and second coordinates. Conjugate gradient method (CG) [4] is applied to solve this problem. No structured prior information is utilized and each image can only be reconstructed separately.

2.2 CaLMMRI

CaLMMRI[7] is a CSMRI[4] approach for pMRI. It reconstructs the aliased images of all coils simultaneously with joint sparsity and the FOV is obtained via the SoS scheme.

$$X = \arg \min_X \|\Phi X\|_{2,1} \quad s.t. \quad \sum_{c=1}^C \|FX(:, c) - b_c\|_2^2 < \varepsilon \quad (2)$$

where $X = [x_1, x_2, \dots, x_C]$ is the images of all coils. The $\ell_{2,1}$ norm is the summation of each row's ℓ_2 norm. Problem (2) is solved by SPGL1[12] in their method. It is a joint reconstruction method as ours. However, as discussed above, their assumption of the similar appearance among images of all coils can not be strictly satisfied in practice. This results in its relatively lower performance and makes it unable to be comparable to the state-of-the-art auto-calibration methods when sufficient ACS is available.

3 Algorithm

Different from CaLMMRI, we only assume the gradients of aliased images from all the coils are jointly sparse. Figure 1 demonstrates the appearances and gradients of two images from different coils. We could observe that the appearances are only similar at the center area. However, as MR images are often piecewise smooth, the gradients are not only very sparse but also jointly sparse. The non-zero gradients tend to be at the same positions across different images, that is, on the edges. This assumption holds at almost entire image areas. As an extension of TV regularization, the pMRI reconstruction is therefore formulated as a JTV reconstruction:

$$X = \arg \min_X \left\{ \frac{1}{2} \sum_{c=1}^C \|FX(:, c) - b_c\|_2^2 + \alpha \|X\|_{JTV} \right\} \quad (3)$$

where $\|X\|_{JTV} = \sum_{i=1}^N \sqrt{\sum_{c=1}^C ((\nabla_1 X_{ic})^2 + (\nabla_2 X_{ic})^2)}$. The JTV is also known as vectorial TV[13] for color image processing. Problem (3) can be solved by proximal gradient method and accelerate by FISTA [14]. The whole algorithm is summarized in Algorithm 1, where $f(X) = \frac{1}{2} \sum_{c=1}^C \|FX(:, c) - b_c\|_2^2$, $\nabla f(X)$ denotes its gradient and L is its Lipschitz constant.

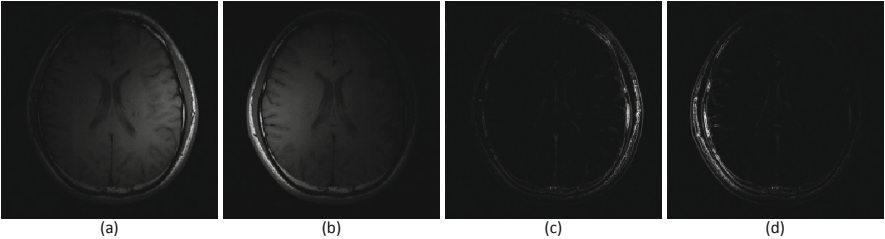


Fig. 1. The appearances and gradients of two aliased images from different coils. (a) and (b): appearances. (c) and (d): magnitude of gradients.

The key problem is thus how to efficiently solve the JTV denoising problem in the second step:

$$X = \arg \min_X \left\{ \frac{1}{2} \|X - B\|^2 + \lambda \|X\|_{JTV} \right\} \quad (4)$$

where $\lambda = \alpha/L$. For convenience of solving (4), X and B are now reshaped to $m \times n \times C$ where m, n denote the number of rows and the number of columns of an image. Following previous works [15], we consider a dual method for (4). Let $P \in \mathbb{R}^{(m-1) \times n \times C}$, $Q \in \mathbb{R}^{m \times (n-1) \times C}$ and they satisfied:

$$\sum_{c=1}^C (P_{i,j,c}^2 + Q_{i,j,c}^2) \leq 1, \quad \forall i, j, \quad |P_{i,n,c}| \leq 1, |Q_{m,j,c}| \leq 1 \quad \forall i, j, c \quad (5)$$

A linear operator is defined as $\mathcal{L}(P, Q)_{i,j,c} = P_{i,j,c} - P_{i-1,j,c} + Q_{i,j,c} - Q_{i,j-1,c}$ and the corresponding inverse operator is defined as $\mathcal{L}^T(X) = (P, Q)$ with $P_{i,j,c} = X_{i,j,c} - X_{i+1,j,c}$, $Q_{i,j,c} = X_{i,j,c} - X_{i,j+1,c}$.

Algorithm 1. JTVMRI

Input: $\rho = \frac{1}{L}$, α , $t^1 = 1$, $Y^1 = X^0$
for $k = 1$ **to** K **do**
 1) $B = Y^k - \rho \nabla f(Y^k)$
 2) $X^k = \arg \min_X \left\{ \frac{1}{2\rho} \|X - B\|^2 + \alpha \|X\|_{JTV} \right\}$
 3) $t^{k+1} = [1 + \sqrt{1 + 4(t^k)^2}] / 2$
 4) $Y^{k+1} = X^k + \frac{t^k - 1}{t^{k+1}} (X^k - X^{k-1})$
end for

Algorithm 2. JTV denoising

Input: λ , B , $(U^1, V^1) = (P^0, Q^0) = (\mathbf{0}_{(m-1) \times n \times C}, \mathbf{0}_{m \times (n-1) \times C})$
for $k = 1$ **to** K **do**
 1) $(P^k, Q^k) = \mathbb{P}_p[(U^k, V^k) + \frac{1}{8\lambda} \mathcal{L}^T(B - \lambda \mathcal{L}(U^k, V^k))]$
 2) $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$
 3) $(U^{k+1}, V^{k+1}) = (P^k, Q^k) + \frac{t^k - 1}{t^{k+1}} (P^k - P^{k-1}, Q^k - Q^{k-1})$
end for
 $X = B - \lambda \mathcal{L}(P^k, Q^k)$

Therefore, the optimal solution for problem (4) is $X = B - \lambda \mathcal{L}(P^*, Q^*)$ [15], where P^*, Q^* is the optimal solution for

$$\min_{P, Q} \left\{ h(P, Q) = \frac{1}{2} \|B - \lambda \mathcal{L}(P, Q)\|_F^2 \right\} \quad (6)$$

Note that the gradient of $h(P, Q) = -\lambda(B - \lambda \mathcal{L}(P, Q))$. Therefore problem (6) can be solved by FISTA and summarized in Algorithm 2.

The projection operator $\mathbb{P}_p(P, Q) = (U, V)$ is used to force (P, Q) satisfy the conditions (5):

$$U_{i,j,c} = \begin{cases} P_{i,j,c} / \max\{1, \sqrt{\sum_{c=1}^C (P_{i,j,c}^2 + Q_{i,j,c}^2)}\} & \forall i, j, c \\ P_{i,n,c} / \max\{1, \sqrt{\sum_{c=1}^C P_{i,n,c}^2}\} & \forall i, c \end{cases} \quad (7)$$

and

$$V_{i,j,c} = \begin{cases} Q_{i,j,c} / \max\{1, \sqrt{\sum_{c=1}^C (P_{i,j,c}^2 + Q_{i,j,c}^2)}\} & \forall i, j, c \\ Q_{m,j,c} / \max\{1, \sqrt{\sum_{c=1}^C Q_{m,j,c}^2}\} & \forall j, c \end{cases} \quad (8)$$

Accelerated by FISTA [14], the convergence rate of Algorithm 2 is $\mathcal{O}(1/k^2)$. The total time complexity of JTVMRI is $\mathcal{O}(CN \log N)$ by applying fast Fourier transform (FFT). There are double loops in the algorithm. Due to the trade off between efficiency and effect, the JTV denoising algorithm only runs 1 iteration in our implementation.

4 Experiments

4.1 Materials and Methods

The experiments are conducted on three brain MRI datasets shown in Figure 2. These images are widely used for validating pMRI methods [4][7][5]. Figure 2(a) shows an image scanned from a GE 3T commercial scanner with an eight-channel head coil using a two-dimensional T1-weighted spin echo protocol (TE = 11ms, TR = 700ms, FOV = 22cm, 256×256 pixels). Figure 2(b) shows a T1-weighted image from spoiled gradient echo (SPGR) sequence, scanned on a GE Signa-Excite 1.5-T scanner with an eight-channel receive coil (TE = 8ms, TR=17.6 ms, FOV = 20cm, 200×200 pixels). The image shown in Figure 2(c) is a T1-weighted image extracted from SRI24 Multi-Channel Brain Atlas Data [16]. It was acquired with a 3D axial IR-prep SPGR sequence, where TR=6.5ms and TE=1.54 ms. Imaging was performed on a 3.0T GE scanner with an 8-channel head coil with 256×256 resolution covering a 24-cm FOV.

Gaussian random sampling masks are used, with more samples are obtained in low frequencies and less samples are obtained in higher frequencies [7]. Our method is compared with the-state-of-the-art auto-calibration methods GRAPPA [3], SPIRiT [4] and the calibrationless CSMRI methods CaLMMRI [7] and CSSENSE [5]. All codes are downloaded from their websites and we carefully follow their experiment setup. For fair comparison, the sampling mask contains a 30×30 fully sampled region to ensure accurate calibration for GRAPPA and SPIRiT (shown in Figure 2(d)). All experiments are conducted in MATLAB with a 3.4GHz CPU. Gaussian white noise with 0.01 standard deviation is added in the original data and Signal-to-Noise Ratio (SNR) is used as the metric for evaluation. The parameters are set $\alpha = 0.04$, $\beta = 0.4$ for our algorithm and CSSENSE. All other algorithms are with their default parameters. SPIRiT runs 10 CG iterations as suggested because the result would be worse with more iterations [4]. All other algorithms run 50 iterations.

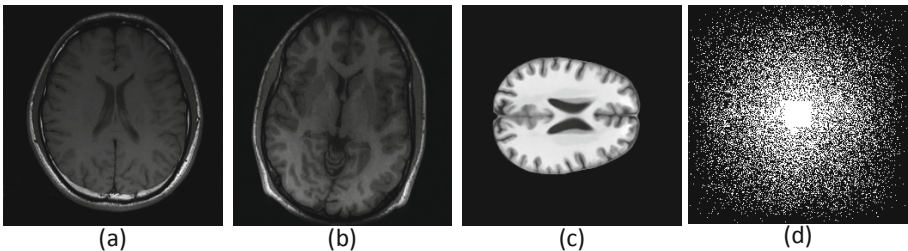


Fig. 2. Three MR brain images and the sampling mask

4.2 Results

Figure 3 shows the reconstruction results on the first brain image at a reduction factor $R = 4$. Compared to all others, the result of our method preserve most details of the original image (as shown in the zoomed region of interest). The results of other algorithms have visible artifacts while ours are the closest to the original one.

For better visualization of the difference errors, all the corresponding error images are shown in Figure 4. The error of our algorithm only appears slightly on the edges while those of all other algorithms significantly distribute over all the image. The SNRs of GRAPPA, SPIRiT, CSSENSE, CaLMMRI and the proposed method are 23.27, 24.73, 23.1., 22.21 and 26.20 respectively. Their CPU time consumptions are 626.2s, 7.8s, 55.7s, 12.5s and 7.4s respectively.

To reduce randomness, each algorithms run 100 times to obtain the average results. The average SNRs and CPU time on the three images are shown in Table 1 and Tabel 2 respectively. On each image, our method always achieve the best results with the least CPU time. GRAPPA is very slow on the random sampling mask, which need more than 10 minutes for a 256×256 image. Unlike other iterative methods, the SNR of SPIRiT does not increase monotonously[4]. How to choose the optimal iteration number to terminate is still unsolved. CaLMMRI is hard to be comparable to the state-of-the-art methods such as CSSENSE and SPIRiT. These results is consistent with the comparisons in previous work[7]. CSSENSE in our implementation is the same as coil-by-coil

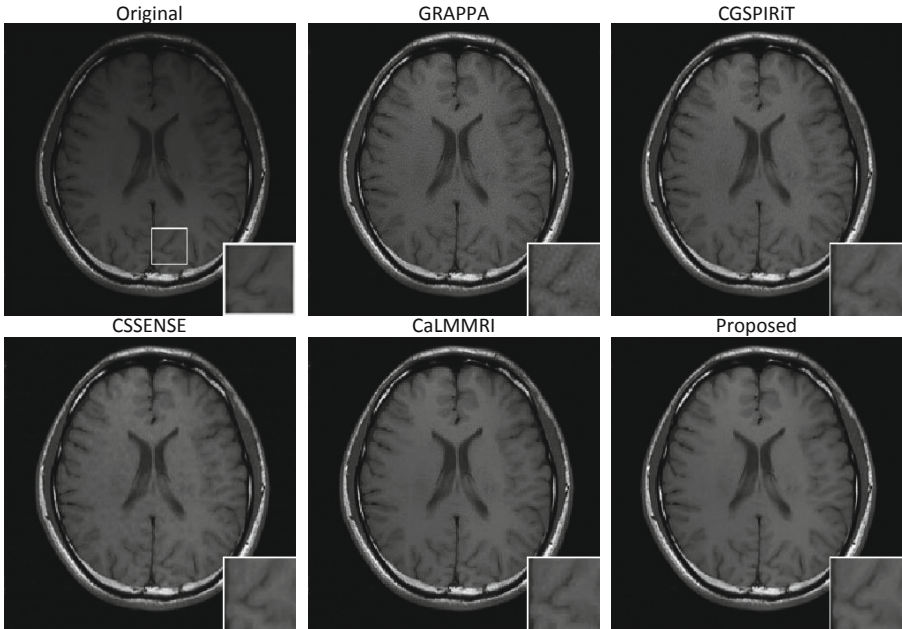


Fig. 3. Visual results of different methods with reduction factor $R = 4$

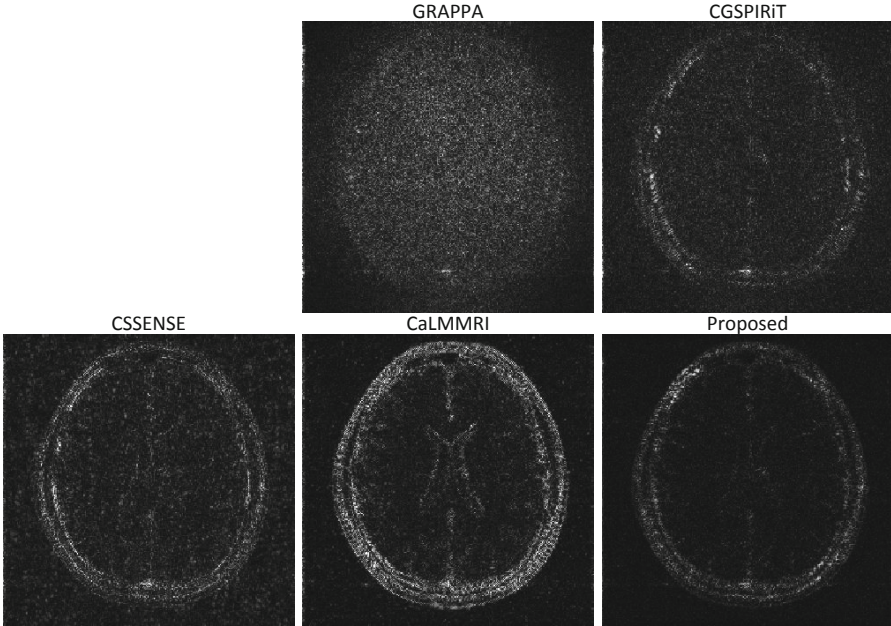


Fig. 4. The error images of different methods with reduction factor $R = 4$

Table 1. The SNRs (dB) of different methods performed on 3 MR images

	GRAPPA[3]	CGSPIRiT[4]	CSENSE[5]	CaLMMRI[7]	Proposed
Brain 1	23.31	24.80	23.10	22.27	26.19
Brain 2	18.39	20.85	21.17	20.28	22.99
Brain 3	36.81	32.00	32.18	28.58	38.13

Table 2. The CPU time (s) of different methods performed on 3 MR images

	GRAPPA[3]	CGSPIRiT[4]	CSENSE[5]	CaLMMRI[7]	Proposed
Brain 1	588.5	7.2	58.4	10.1	6.5
Brain 2	719.1	9.6	71.0	14.3	8.3
Brain 3	723.8	9.4	71.5	14.9	9.1

SparseMRI[8] reconstruction for each individual image. It has stable performance. However, the correlations among the aliased images are not exploited. That is why it always worse than the proposed method. The computational cost of CSSENSE is also expensive compared with other iterative methods, which makes it hard to be applied on large scale data.

4.3 Discussion

In contrast to GRAPPA and SPIRiT, CSSENSE and CaLMMRI are calibration-less methods that do not rely on sufficient ACS. In contrast to GRAPPA and

CSSENSE, SPIRiT and CaLMMRI can perform joint reconstruction on the undersampled data. Compare with CaLMMRI, CSSENSE has utilized the sparsity in gradients to improve reconstruction. Only our method inherit all these benefits in pMRI, which can efficiently, *jointly*, reconstruct MR images by exploiting their sparsity in *gradient* domain *without calibration*.

5 Conclusion

In this paper, we have proposed a novel calibrationless method for calibrationless pMRI. It is motivated by the observation that the images of different coils are jointly sparse in the gradient domain. Compared to existing auto-calibration methods, our algorithms can be widely applied in the cases when calibration is time consuming or impossible. Compared to exiting calibrationless methods, our model is much more applicable on pMRI data and our algorithm is much more efficient. Extensive experiments have demonstrated that our algorithm outperforms the state-of-the-art auto-calibration methods when sufficient calibration can be achieved, and is far better than previous calibrationless methods. All these benefits make calibrationless pMRI much more feasible than before.

References

1. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., et al.: SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* 42(5), 952–962 (1999)
2. Ying, L., Sheng, J.: Joint image reconstruction and sensitivity estimation in SENSE (JSENSE). *Magn. Reson. Med.* 57(6), 1196–1202 (2007)
3. Griswold, M.A., Blaimer, M., Breuer, F., Heidemann, R.M., Mueller, M., Jakob, P.M.: Parallel magnetic resonance imaging using the GRAPPA operator formalism. *Magn. Reson. Med.* 54(6), 1553–1556 (2005)
4. Lustig, M., Pauly, J.M.: SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magn. Reson. Med.* 64(2), 457–471 (2010)
5. Liang, D., Liu, B., Wang, J., Ying, L.: Accelerating SENSE using compressed sensing. *Magn. Reson. Med.* 62(6), 1574–1584 (2009)
6. Zhao, T., Hu, X.: Iterative GRAPPA (iGRAPPA) for improved parallel imaging reconstruction. *Magn. Reson. Med.* 59(4), 903–907 (2008)
7. Majumdar, A., Ward, R.K.: Calibration-Less Multi-coil MR image reconstruction. *Magn. Reson. Imaging* (2012)
8. Lustig, M., Donoho, D., Pauly, J.: Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* 58(6), 1182–1195 (2007)
9. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* 52(2), 489–509 (2006)
10. Donoho, D.: Compressed sensing. *IEEE Trans. Inf. Theory* 52(4), 1289–1306 (2006)
11. Acar, R., Vogel, C.R.: Analysis of bounded variation penalty methods for ill-posed problems. *Inv. Probl.* 10(6), 1217 (1999)
12. Van Den Berg, E., Friedlander, M.: Probing the pareto frontier for basis pursuit solutions. *SIAM J. Sci. Comput.* 31(2), 890–912 (2008)

13. Bresson, X., Chan, T.F.: Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging* 2(4), 455–484 (2008)
14. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2(1), 183–202 (2009)
15. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* 18(11), 2419–2434 (2009)
16. Rohlfing, T., Zahr, N.M., Sullivan, E., Pfefferbaum, A.: The SRI24 multichannel atlas of normal adult human brain structure. *Human Brain Mapping* 31, 798–819 (2010)

Denoising PET Images Using Singular Value Thresholding and Stein’s Unbiased Risk Estimate*

Ulas Bagci** and Daniel J. Mollura

Center for Infectious Diseases Imaging (CIDI), Department of Radiology and Imaging Sciences, National Institutes of Health (NIH), Bethesda, USA

`ulas.bagci@nih.gov`

Abstract. Image denoising is an important pre-processing step for accurately quantifying functional morphology and measuring activities of the tissues using PET images. Unlike structural imaging modalities, PET images have two difficulties: (1) the Gaussian noise model does not *necessarily* fit into PET imaging because the exact nature of noise propagation in PET imaging is not well known, and (2) PET images are low resolution; therefore, it is challenging to denoise them while preserving structural information. To address these two difficulties, we introduce a novel methodology for denoising PET images. The proposed method uses the singular value thresholding concept and Stein’s unbiased risk estimate to optimize a soft thresholding rule. Results, obtained from 40 MRI-PET images, demonstrate that the proposed algorithm is able to denoise PET images successfully, while still maintaining the quantitative information.

Keywords: PET, Denoising, Singular Value Thresholding, Stein Risk Estimate.

1 Introduction

Positron emission tomography (PET) is a molecular imaging technique that has rapidly emerged as an important functional imaging tool; it provides superior *sensitivity* and *specificity*, when combined with anatomical imaging such as computed tomography (CT) or magnetic resonance imaging (MRI). Since PET has a significantly low number of detected photons, constructed images have high amount of noise (i.e., low signal-to-noise ratio (SNR)). Therefore, there is often a need for denoising PET images to determine accurate quantitative measures for evaluating changes in lesion biology.

Current approaches in PET image denoising are limited to Gaussian smoothing and locally adaptive filtering [1–3]. The majority of researchers focused on

* This research is supported by CIDI, the intramural research program of the National Institute of Allergy and Infectious Diseases (NIAID) and the National Institute of Biomedical Imaging and Bioengineering (NIBIB).

** Corresponding author.

additive Gaussian noise models regardless of the image type, and only a few works have considered characteristics of PET images when developing a denoising technique. In these current approaches, [1–3], one has to carefully choose the spatial width of the Gaussian smoothing filter to yield a good compromise between spatial resolution and SNR. However, smoothing can produce a loss of resolution by averaging voxels together and blurring the distinction between two closely adjacent objects. Moreover, the use of edge preserving or Gaussian smoothing filters are not optimal choices when denoising PET images, unlike the common belief, because PET images are low resolution due to low photon counts and consequential statistical noise. In fact, in PET and other images with *low photon counts* (e.g., SPECT, microscopy), the pixel intensities follow a non-Gaussian distribution. More recently, a Poissonian distribution noise model [4] was suggested to be used in PET images because of the following observations: the noise variance in PET images are **multiplicative, asymmetric, and varying** over the image [4]. Therefore, it was shown to be feasible to model noise with Poissonian. It is also important to emphasize that the *invariance of the total photon counts* in a PET images or a region of interest (ROI) in PET images should be preserved within some limits so that it does not affect the diagnostic utility of quantification metrics such as standardized uptake value (*SUV*) which gives physiologically relevant measurements of cellular metabolism.

In [1], locally adaptive filtering (i.e., anisotropic diffusion) was developed to denoise medical images for general purposes; maintaining the edge information, while smoothing the noise was the aim. Although total photon counts were preserved with this approach, the additive noise model, with constant noise variance, violated the non-Gaussian nature of the PET images. In [5], structural information (either from CT or MRI) of matched anatomic images were used in a multi-resolution model to enhance SNR of PET images. However, the presented method is not optimal because one-to-one wavelet parameter exchange between PET images and their anatomical correspondences does not necessarily hold for all subjects. More recently in [2], bilateral filtering was used for denoising PET images. Although improvement in the SNR was observed, substantial change in SUV_{max} was inevitable. To alleviate the challenges described above, we propose to (1) transform PET images with Anscombe’s variance stabilizing transform (VST) in order to Gaussianize the image data and inherent noise, (2) remove Gaussianized noise using Stein’s unbiased risk estimate and a soft thresholding rule within the popular singular value thresholding (SVT) framework, and (3) apply the inverse variance stabilizing transform (IVST) to obtain denoised PET images. In the next section, we present our proposed framework in detail.

2 Methods

Our approach hinges on the basic notion that medical images are mostly low-ranked due to high correlations of its columns (or rows); therefore, it is possible to model the images with simpler components. For a low-rank matrix, with entries perturbed by Gaussian noise, one may recover an estimate of the low-rank matrix

by using techniques based on the singular value decomposition (SVD). However, accuracy of these estimations are often limited. In practice, since the low-rank matrix is not perturbed by Gaussian noise but rather varying magnitudes—as it is the case in PET images—SVD-based methods are not directly applicable to solve such problems. Herein, we describe a method to Gaussianize PET image noise so that SVD-based methods, SVT in particular, can be applicable. We then apply Stein’s unbiased risk estimate formulation to SVT in order to remove noise from the image data.

Anscombe’s Variance Stabilizing Transform. Anscombe’s VST is a special case of Anscombe’s theorem [4], and it allows a Gaussianization of the Poisson distribution. According to the theorem, if S is Poisson distribution (i.e., $S \sim P(\xi)$ where $\sqrt{\xi}$ is the standard deviation of noise), then there is a function VST such that $\mathbf{Y} = VST(S) = 2\sqrt{S} + 3/8$ when $\xi \rightarrow \infty$. Practically, \mathbf{Y} is approximately Gaussian with unit variance for $\xi > 10$ [4], so once the PET images are transformed with Anscombe’s VST, Gaussianized noise can be removed with the proposed methodology. Next, an inverse of Anscombe’s VST is required to transform denoised PET data back into the SUV domain. IVST of the estimated signal is straightforward as VST is explicit.

Singular Value Thresholding. Assume that a 3-D VST transformed PET image is considered as a stack of slices and each slice is denoted as matrix \mathbf{Y} , with a $m \times n$ dimension and the observed PET image is denoted by S . In denoising, we estimate the original data matrix \mathbf{X} from noisy observations: $\mathbf{Y}_{ij} = \mathbf{X}_{ij} + \mathbf{N}_{ij}$, for $i = 1 \dots m, j = 1 \dots n$, where noise \mathbf{N} is modeled as the additive Gaussian, with a constant variance τ : $\mathbf{N} \sim \mathcal{N}(0, \tau^2)$. Since VST is applied to original PET images, the resultant images include residual noise with an approximate constant variance; hence, the Gaussian noise assumption can be used. Conventionally, estimating \mathbf{X} via SVD can simply be done by truncating the singular values of the observed matrix \mathbf{Y} , but this truncation mechanism, known as a hard threshold rule, does not provide continuous estimation [6]:

$$SVT_{hard}(\mathbf{Y}) = \underset{\mathbf{X} \in R^{m \times n}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda \operatorname{rank}(\mathbf{X}), \quad (1)$$

where λ is a positive scalar, and SVD of the \mathbf{Y} can be defined as $\mathbf{Y} = U \Sigma V^* = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^*$. Letting \mathcal{I} be an indicator function denoting the shrinkage of the singular values (i.e., σ_i), then SVT with hard thresholding can simply be written as $SVT_{hard}(\mathbf{Y}) = \sum_{i=1}^{\min(m,n)} \mathcal{I}(\sigma_i > \lambda) u_i v_i^*$. Selecting a positive scalar λ is the core of the estimation problem because Eq.1 is minimized when appropriate λ is found. For this type of estimation problem, Stein, in his seminal paper [7], showed that if $\mathbf{Y}_{ij} \sim \mathcal{N}(\mathbf{X}_{ij}, 1)$ with an estimator $\hat{\mathbf{X}}$ of the form $\hat{\mathbf{X}} = \mathbf{Y} + g(\mathbf{Y})$, where $g_{ij} \in R^{m \times n} \rightarrow R$ is differentiable with respect to \mathbf{Y}_{ij} , then the risk of the estimate is given by $E\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2 = E\{mn + 2\nabla(g(\mathbf{Y})) + \|g(\mathbf{Y})\|_F^2\}$, where ∇ is the first partial derivatives with respect to \mathbf{Y}_{ij} . Since mn is constant, choosing a suitable function g that makes $\nabla(g(\mathbf{Y})) + \|g(\mathbf{Y})\|_F^2$ everywhere negative will yield

a satisfactory estimate for the problem at hand. However, there is no continuous differentiable function, required by Stein's lemma, in SVT methodology based on hard thresholding rule. In order to avoid a discontinuous estimation caused by the indicator function, one may provide a continuous estimation of λ by shrinking the singular values towards zero, namely called soft thresholding rule for SVT, and it can be formulated as

$$SVT_{soft}(\mathbf{Y}) = \sum_{i=1}^{\min(m,n)} (\sigma_i - \lambda)_+ u_i v_i^*. \quad (2)$$

In this equation, since large λ causes a large bias and small λ leads to a high variance, a proper estimation of λ is expected to find the correct trade-off between large bias and high variance in the soft thresholding rule of SVT. As earlier noted, because the soft thresholding rule in SVT formulation (i.e., $(\sigma_i - \lambda)_+$) is continuous, it is differentiable at the singular value matrix (note that derivative of $(\sigma_i - \lambda)_+$ is equivalent to $\mathcal{I}(\sigma_i > \lambda)$ when $\sigma \neq \lambda$) [6]. Hence, SVT_{soft} follows the requirement of the function g stated by the Stein's risk estimation method. Replacing g with SVT_{soft} yields the unbiased risk estimate for SVT as [6]

$$E(SVT_{soft}(\mathbf{Y})) = -mn\tau^2 + \sum_{i=1}^{\min(m,n)} \min(\lambda^2, \sigma_i^2) + 2\tau^2 \nabla(SVT_{soft}(\mathbf{Y})). \quad (3)$$

The solution of Eq.3 gives the optimum λ such that the risk of the estimate cannot be improved at any point through multiplication of λ by a constant factor. For the computation of the $\nabla(SVT_{soft}(\mathbf{Y}))$ operation, a closed-form expression can be obtained by differentiating $SVT_{soft}(\mathbf{Y})$ with respect to the λ as

$$\begin{aligned} \nabla(SVT_{soft}(\lambda)) = & \sum_{i=1}^{\min(m,n)} \left[\mathcal{I}(\sigma_i > \lambda) + |m - n| \left(\frac{\sigma_i - \lambda}{\sigma_i} \right)_+ \right] \\ & + 2 \sum_{i \neq j, i, j=1}^{\min(m,n)} \frac{\sigma_i (\sigma_i - \lambda)_+}{\sigma_i^2 - \sigma_j^2}. \end{aligned} \quad (4)$$

Once λ is found, then denoising is conducted through shrinking singular values of noisy observations \mathbf{Y} towards λ [6]. At the final step, the denoised image is transformed back to the original SUV domain via the Anscombe's $IVST$ method.

3 Experiments and Results

With IRB approval, we collected 40 MRI-PET image sets from 40 different patients. Each patient had either Von-Hippel-Lindau disease, colon cancer, a paraganglioma carcinoid tumor, or hedereditary leiomyomatosis renal cell cancer. PET images had a spatial resolution of $4.17 \times 4.17 \times 2 \text{ mm}^3$. The SNR and

relative contrast of all images were analyzed and compared to other commonly used methods in the literature. The *SUV* analysis—the selected 40 lesions from all the subjects—was conducted to quantify the effect of denoising on the quantitative metrics of PET imaging. For this, we used the SUV_{max} and SUV_{mean} criterion to follow clinical routine. In order to remove any bias towards the proposed method in our evaluations, we transformed PET images via VST prior to conduct experiments for Perona-Malik’s, Bilateral, and Gaussian filtering methods so that the noise properties of the transformed PET images were approximately Gaussian. Otherwise, all these approaches diminished the quantitative indexes of PET images considerably.

We first demonstrate the validation of the proposed risk estimate via Monte Carlo estimation. Herein, z independent samples of \mathbf{Y} were drawn from the model of $\mathbf{Y} = \mathbf{X} + \mathbf{N}$. The data matrix \mathbf{X} was constructed based on predefined singular values in low-rank approximation without having repetitive singular values. Then, we computed the average mean squared error between the Stein’s unbiased risk estimate and the original matrix by $\frac{1}{z} \sum_{u=1}^z \|SVT_{soft}(\mathbf{Y}^u) - \mathbf{X}\|_F^2$. Fig. 1a shows the comparison of the risk estimate using Monte Carlo and Stein’s unbiased risk estimate for $z = 25$ samples. As clearly seen from the Fig. 1a, Stein’s unbiased risk estimate within the SVT framework (solid line), follows the Monte Carlo estimates closely.

Qualitative Evaluation. We qualitatively compared our method with anisotropic diffusion filtering [1], which performed best among the other compared methods. Fig. 1b illustrates the filtered output from the proposed method (ii) and Perona-Malik’s method (iii) with respect to the original PET images (i). Red arrows show reduced noise areas in (ii) and limited success of Perona-Malik filtering in (iii) for the corresponding regions. Similarly, the yellow arrows denote the object of interests, where edge information was preserved with the proposed method (ii); however, the object of interests were over-smoothed by the Perona-Malik filtering (iii). As can be seen in all images (rows show different anatomical level of different subjects) in Fig. 1b, the boundary contrast seems to be higher with the proposed method visually, while Perona-Malik’s method over-smooths the noisy areas. This is because the proposed method is tuned to preserve fine details of the objects of interests better than Perona-Malik’s method.

Quantitative Evaluation. For our quantitative analysis, SNR was used to evaluate the effectiveness of the filtering operation; it is defined as $20 \log_{10}(\frac{M}{\nu})$, where ν and M indicate the standard deviation and mean of the voxel intensities in a ROI. In order to estimate the SNR levels in the image data, several 3-D ROIs were positioned within the liver and lung areas for all image sets, as it is the convention when measuring SNR in PET images. Positioned ROIs were identical for all filtered and original images. Resulting SNR rates were averaged over all subjects and plotted in Fig. 2a. As it is readily seen, improvement in SNR with respect to the original PET images is higher with the proposed method than the rest. Since a measure of SNR alone may not be sufficient enough to

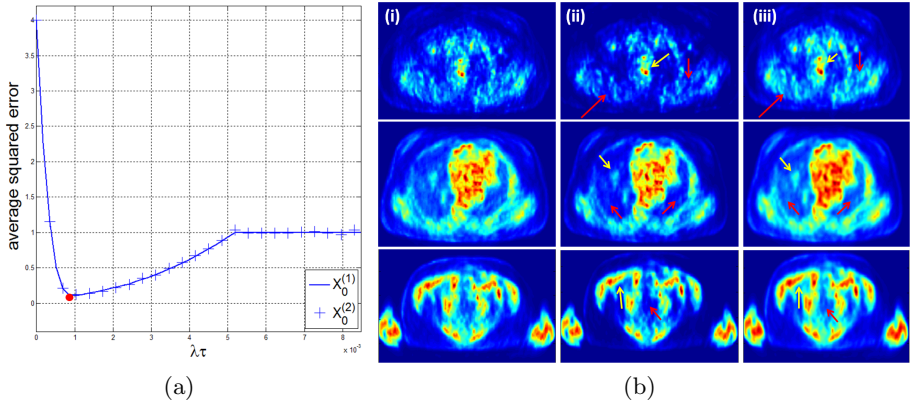


Fig. 1. (a) $X_0^{(1)}$: Stein’s unbiased risk estimate. $X_0^{(2)}$: Monte Carlo estimates. Red point indicates the smallest average squared error (global optima). (b) Each row shows different subjects and different slice levels. (i) Original PET images, (ii) denoised by the proposed method, (iii) denoised by Perona-Malik anisotropic diffusion filtering.

characterize the effectiveness of a filtering method, we also used the concept of *relative contrast* (RC) in our evaluation, where RC was defined as follows [8]:

$$RC = \frac{|M_i^O - M_i^B|}{\sqrt{\nu_i^O \nu_i^B}}, \quad (5)$$

and i implies the specific ROI, M^O and ν^O denote the mean and standard deviation of pixel intensities over the object region O , respectively. Similarly, M^B and σ^B denote the mean and standard deviation of pixel intensities over the background region B , respectively. RC is a measure of object-to-background contrast relative to noise in each region. Object regions were selected from liver and lung areas for each image, as a 3-D ROI, and the resultant RC values are plotted in Fig. 2b. For each experiment, the same set of object and background regions were selected to avoid any bias in the calculations. Note that a higher value of RC indicates a more accurate filtering; hence, the proposed method outperforms all the compared methods, among which Perona-Malik’s anisotropic diffusion filtering [1] was having the second higher accuracy due to its property of preserving structural information.

We also investigated how PET specific quantitative markers such as SUV_{max} and SUV_{mean} change after denoising the PET images. For this purpose, 40 lesions from all subjects were identified manually with a fixed ROI, and the quantitative markers were computed for each lesion from (1) the original PET, (2) the filtered image by Perona-Malik’s method, and (3) the filtered image by our proposed method. As a comparison, only Perona-Malik’s method was used in this evaluation because it was the second best method among other compared methods. The changes of SUV_{max} and SUV_{mean} , for each method and for all

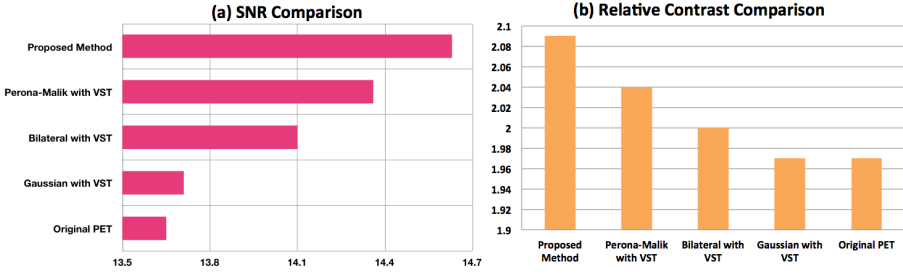


Fig. 2. Proposed method’s SNR (a) and RC (b) comparison with Perona-Malik’s anisotropic filtering, bilateral filtering, and Gaussian smoothing with respect to the original PET images over all subjects are reported

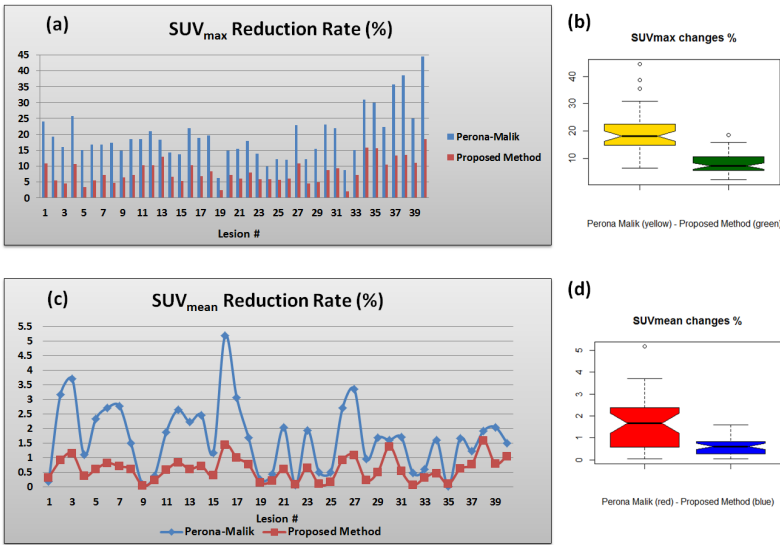


Fig. 3. Comparison of SUV_{max} (a) and SUV_{mean} (c) reduction rates (%) is demonstrated for 40 lesions from different subjects, respectively. Mean (b) and standard deviation (d) of the proposed method and Perona-Malik’s filtering is compared through boxplots

lesions, were computed with respect to the original PET images; the comparative results are reported in Fig. 3a and c. In addition, the mean decrease in SUV_{max} and SUV_{mean} values are shown in Fig. 3b and d, respectively. Notice that *smaller changes* in SUV_{max} and SUV_{mean} are highly desirable in order to have a valid denoising method that does not substantially change quantitative markers of the PET imaging. Our experimental results show that when using the proposed method, the reduction rates in SUV_{max} and SUV_{mean} are much smaller than Perona-Malik’s method [1].

4 Concluding Remarks

In this study, we presented an effective tool for denoising PET images. The proposed algorithm adapts Stein's unbiased risk estimation within the singular value soft thresholding rule. We also emphasized the usefulness of variance stabilizing transform and its inverse before and after the denoising procedure, respectively, in order to Gaussianize the Poisson nature of the noise in PET images. Experimental results demonstrated that the proposed framework respects the tissue boundaries well, reduces the noise considerably without losing quantitative information of PET images including SUV_{max} and SUV_{mean} . As an extension of this work, we are currently comparing the low-rank data representation algorithms (such as [9]) within the same proposed platform and exploring the model-free algorithms for PET denoising.

References

1. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7), 629–639 (1990)
2. Hofheinz, F., Langner, J., Beuthien-Baumann, B., Oehme, L., Steinbach, J., Kotzerke, J., van den Hoff, J.: Suitability of bilateral filtering for edge-preserving noise reduction in pet. *EJNMMI Research* 1(1), 23 (2011)
3. Turkheimer, F.E., Aston, J., Riddell, C., Cunningham, V.: A linear wavelet filter for parametric imaging with dynamic PET. *IEEE Trans. on Medical Imaging* 22, 289–301 (2003)
4. Lee, J.A., Geets, X., Gregoire, V., Bol, A.: Edge-preserving filtering of images with low photon counts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(6), 1014–1027 (2008)
5. Turkheimer, F.E., Boussion, N., Anderson, A.N., Pavese, N., Piccini, P., Visvikis, D.: PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets. *J. Nucl. Med.* 49, 657–666 (2008)
6. Candes, E.J., Sing-Long, C.A., Trzasko, J.D.: Unbiased risk estimates for singular value thresholding and spectral estimators. *Arxiv (arXiv:1210.4139v1)* (2012)
7. Stein, C.M.: Estimation of the mean of a multivariate normal distribution. *Annals of Statistics* 9(6), 1135–1151 (1981)
8. Saha, P., Udupa, J.: Scale-based difusive image filtering preserving boundary sharpness and fine structures. *IEEE Transactions on Medical Imaging* 20(11), 1140–1155 (2001)
9. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 35, 171–184

Super-Resolution Reconstruction Using Cross-Scale Self-similarity in Multi-slice MRI

Esben Plenge*, Dirk H.J. Poot, Wiro J. Niessen, and Erik Meijering

Erasmus MC – University Medical Center Rotterdam
P.O. Box 2040, 3000 CA Rotterdam, The Netherlands
<http://www.bigr.nl/>

Abstract. In MRI, the relatively thick slices of multi-slice acquisitions often hamper visualization and analysis of the underlying anatomy. A group of post-processing techniques referred to as super-resolution reconstruction (SRR) have been developed to address this issue. In this study, we present a novel approach to SRR in MRI, which exploits the high-resolution content usually available in the 2D slices of MRI slice stacks to reconstruct isotropic high-resolution 3D images. Relying on the assumption of local self-similarity of anatomical structures, the method can be applied both to a single slice stack and to the combination of multiple slice stacks that differ in the orientation of their field of view. We evaluate the method quantitatively on synthetic brain MRI and qualitatively on MRI of the lungs. The results show that the method outperforms state-of-the-art MRI super-resolution methods.

Keywords: Super-resolution, reconstruction, MRI, self-similarity, dictionaries, cross-scale, brain, lung, inverse problems, regularization.

1 Introduction

In magnetic resonance imaging (MRI) sequences requiring long repetition times, conventional 3D imaging usually leads to infeasible scan times, and 2D multi-slice imaging is used instead. However, due to hardware induced limitations on gradient strength, requirements on signal-to-noise ratio (SNR), and other factors, the slices are usually relatively thick compared to the in-plane resolution. Such anisotropy negatively affects visualization and hampers analysis. The isotropy and resolution of images may be improved by super-resolution reconstruction (SRR) methods [1], often divided into 1) methods that base the high-resolution (HR) reconstruction on a single image only, and 2) methods that combine multiple low-resolution (LR) images of the same object acquired under varying fields of views. In both cases, the inverse problem of recovering the HR image is ill-posed, and regularization is applied by exploiting prior knowledge of the HR solution, such as that it must be smooth, piecewise smooth, or sparse.

Recently, a quite different prior has been shown to be very powerful: example-based self-similarity. It is based on the observation that small-scale structures

* Corresponding author.

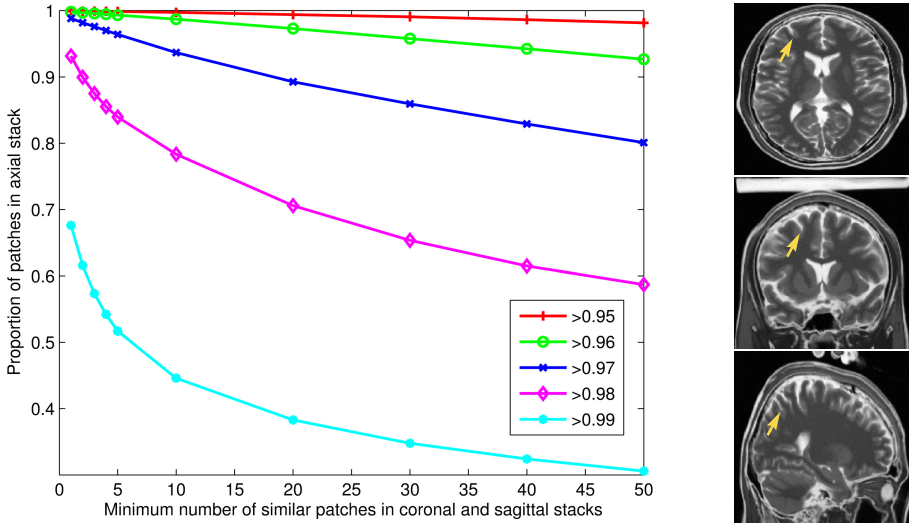


Fig. 1. Left: Quantitative check of self-similarity in a T2-weighted brain MRI from BrainWeb [12]. Similarity between the 10% axial 7×7 patches with the highest variance and all coronal and sagittal 7×7 patches was evaluated. For several similarity thresholds, the fraction of axial patches is plotted as function of the number of coronal and sagittal patches exceeding the threshold. Similarity is measured as the inner product between the patches (after vectorization and unit length normalization). Right: Example of local self-similarity in the cortex. The yellow arrows point at a similar cortex structure in the axial (top), coronal (middle), and sagittal (bottom) slices.

tend to repeat themselves throughout an image. This concept has been widely applied for image compression, denoising, inpainting, and SRR [2], and has spawned methodologies such as image hallucination [3], sparse coding using learned dictionaries [4,5], and non-local means (NLM) [6]. A common feature of these example-based methods is the formulation of a parent-child structure in which the “nearest-neighbor” of an input LR-patch is sought among the LR-parents of HR-child-patches in either a database [3] or in the image itself [7]. In MRI, applications include resolution enhancement of T2-weighted stacks from isotropic HR T1-weighted stacks [8], and upsampling by iteratively applying NLM-denoising to an interpolated version of the image itself [9].

Here we propose a novel approach to SRR in multi-slice MRI based on the concept of *cross-scale* self-similarity [10]. Multi-slice 2D MRI scanning yields two native scales simultaneously: one at in-plane resolution (HR) and one in the slice-selection direction (LR). Local self-similarity of anatomical features occurs both within and across these scales (Fig. 1), which we exploit to achieve SRR. Our method can be applied to both a single image and to multiple images, transcending the methodological division described above. We describe our method and show its potential by comparing it quantitatively and qualitatively with a baseline interpolation scheme and a state-of-the-art SRR algorithm.

2 Method

2.1 From Single-Scale to Cross-Scale Self-similarity

Our method is developed from one of the most successful applications of image self-similarity: the non-local means (NLM) method [6,11]. In NLM, the i th pixel in image \mathbf{u} , defined over the domain Ω , is represented as a weighted average of pixels from similar patches found elsewhere in the same image:

$$\text{NLM}(\mathbf{u}(i)) = \frac{1}{C(i)} \sum_{j \in \Omega} w(i, j) \mathbf{u}(j) \quad (1)$$

with $w(i, j) = \exp(-\|\mathbf{u}(\mathcal{N}(i)) - \mathbf{u}(\mathcal{N}(j))\|_{2,a}^2/h^2)$ and $C(i) = \sum_{j \in \Omega} w(i, j)$. Here, $\mathcal{N}(k)$ is a square patch of a fixed size centered on pixel k , $\|\cdot\|_{2,a}^2$ is the norm weighted by a Gaussian with standard deviation a , and h is a parameter that controls the decay of the weights as a function of the norm.

For our purposes, a HR example image \mathbf{v} is used to modify the above NLM expression, leading to cross-scale NLM [10]:

$$\text{csNLM}(\mathbf{u}(i)) = \frac{1}{C(i)} \sum_{j \in \Omega} \omega(i, j) \mathbf{v}(j) \quad (2)$$

with $\omega(i, j) = \exp(-\|\mathbf{u}(\mathcal{D}(i)) - \mathcal{H}_z(\mathbf{v}(\mathcal{N}(j)))\|_{2,a}^2/h^2)$ and $C(i)$ as defined above. Here, \mathcal{D} maps \mathbf{u} onto a HR lattice, and \mathcal{H}_z is a linear model of the imaging process, with degradation z . In other words, in csNLM, HR patches $\mathbf{v}(j)$ whose degraded versions are similar to an up-scaled LR patch $\mathbf{u}(i)$ are used to update the HR estimate $\text{csNLM}(\mathbf{u}(i))$.

The method we propose is a special case of csNLM. In the limit, when $h \rightarrow 0$, the weight of only one patch (the “nearest neighbor”) dominates. The cross-scale super-resolution expression thus becomes:

$$\text{csSR}(\mathbf{u}(i)) = \frac{1}{\omega(i, j)} \omega(i, j) \mathbf{v}(j) = \mathbf{v}(j) \quad \text{with} \quad j = \underset{l}{\text{argmax}} \omega(i, l). \quad (3)$$

In the experiments we demonstrate that the use of only the nearest neighbor is indeed sufficient to improve resolution and image quality considerably.

2.2 MRI Super-Resolution Using Cross-Scale Self-similarity

From a 2D viewpoint, MR stacks contain both LR images (anisotropic slices containing the slice-selection direction) and HR images (isotropic in-plane slices). Using the knowledge of 2D self-similarity of 3D anatomical structures (Fig. 1), we can apply Eq. (3) as follows: Let \mathbf{u} denote the 2D LR slices of a stack, let \mathcal{D} be an interpolation operator that maps the LR slices onto a 2D HR grid, and \mathcal{H}_z a number of blurring kernels (one of them approximating the point-spread function, PSF, of the upsampled LR slices) applied to the in-plane slices \mathbf{v} . We search for the nearest neighbor of $\mathbf{u}(\mathcal{D}(i))$ over all HR patches j , at all

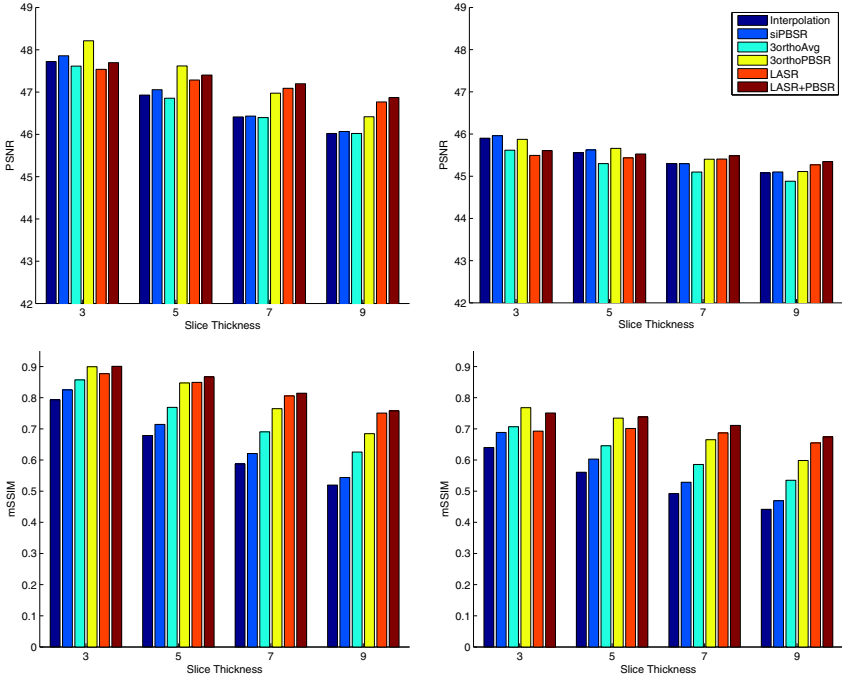


Fig. 2. Quantitative evaluations. Top: PSNR of SRR using six different methods. Bottom: mSSIM of SRR by the six methods. Left: Reconstructions from images of the 3% noise group. Right: Reconstructions from images of the 9% noise group.

blurring levels z , in $\mathcal{H}_z(\mathbf{v}(\mathcal{N}(j)))$. The reason for using multiple blurring levels is to enrich the model and capture cross-scale self-similarity at scales besides the two native ones (relevant for recursive structures such as vessel-trees). In the experiments we used two levels, empirically chosen to achieve good performance.

Multiscale Dictionary. The vectorized versions of all HR patches (or a randomly selected subset thereof) are collected in a matrix. The corresponding vectorized patches of the blurred versions of the HR slices are appended as columns in this matrix, such that each blurring level is represented by a contiguous block of columns. The columns are then normalized and the normalization factors are saved for later rescaling. Adopting the terminology of the sparse coding literature, we shall call this matrix a dictionary.

Reconstruction Process. An initial HR estimate is created by interpolating and aligning the LR images onto a 3D HR grid. The slice stack is traversed, LR slice by LR slice. Around each pixel in each slice, a 2D patch is extracted (of size 7×7 pixels in this study). According to the cross-scale self-similarity assumption, the HR version of this LR patch will be well-approximated by the in-plane patches in the dictionary. For each LR patch, the dictionary is searched for its nearest neighbor (NN), defined as the column that has the maximum

inner product with the vectorized and normalized LR patch. The corresponding HR patch is found in the HR block of the dictionary and rescaled to its original intensity. HR patches are inserted into the grid and averaged where they overlap.

3 Experiments and Results

3.1 Image Data

Simulated Brain MRI. Two isotropic HR T2-weighted images ($1 \times 1 \times 1$ mm voxels) with intensity non-uniformity of 20%, and noise-levels of 3% and 9%, respectively, were downloaded from BrainWeb [12]. To simulate a common case from anatomical neuro-imaging, slice stacks were obtained from each of these HR images by application of a 1D Gaussian PSF in the slice-selection direction. The full-width-half-maximum of the PSF was equal to the downsampling factor applied in the slice-selection direction. By varying this factor as well as the slice-selection direction, three orthogonal slice stacks (axial, coronal, sagittal) were generated for downsampling factors of 3, 5, 7, 9. Thus, a total of eight sets, each containing three orthogonal slice stacks, with resolutions of $1 \times 1 \times 3$ mm, $1 \times 1 \times 5$ mm, $1 \times 1 \times 7$ mm, and $1 \times 1 \times 9$ mm, were generated from the two original images with noise-levels of 3% and 9%.

Lung MRI. Lung MRI data was acquired on a 1.5 Tesla GE scanner. Axial, coronal, and sagittal slice stacks of a test-subject with the lung in expiratory state were acquired during breath hold. A 2D gradient-recalled steady-state sequence (TR = 2.2 s, TE = 0.75 s, flip angle = 35 degrees, number of averages = 1) was used, having a scan time of 15 seconds per slice stack. The field of view was 400×400 mm, the acquisition matrix was 128×160 pixels, yielding an in-plane resolution of 3.125×2.5 mm. The slices were reconstructed by the scanner to a grid of 256×256 pixels with uniform sizes of 1.56×1.56 mm. 40 slices were acquired per stack with a slice thickness of 8 mm. Lung MRI courtesy of H. Tiddens, P. Ciet, and P. Wielopolski, Dept. of Radiology, Erasmus MC.

3.2 Quantitative Evaluation

The proposed method, referred to as patch-based super-resolution (PBSR), was quantitatively evaluated for 1) a single axial slice stack (siPBSR), 2) the combination of three orthogonal stacks (3orthoPBSR), and 3) post-processing of the result of another SRR method (LASR) [13] using three orthogonal stacks (LASR+PBSR). Initial estimates of the HR image were created from the available data: in the case of a single stack, by cubic interpolation, and in the case of three stacks, by either the average of the interpolated and aligned stacks (3orthoAvg) or by the LASR method. Performance was quantified by peak-signal-to-noise ratio (PSNR) and mean structural similarity (mSSIM) [14]. To compute these measures, a noise-free HR T2-weighted image ($1 \times 1 \times 1$ mm voxels) from BrainWeb was used. The results are shown in Fig. 2. As expected, lower noise

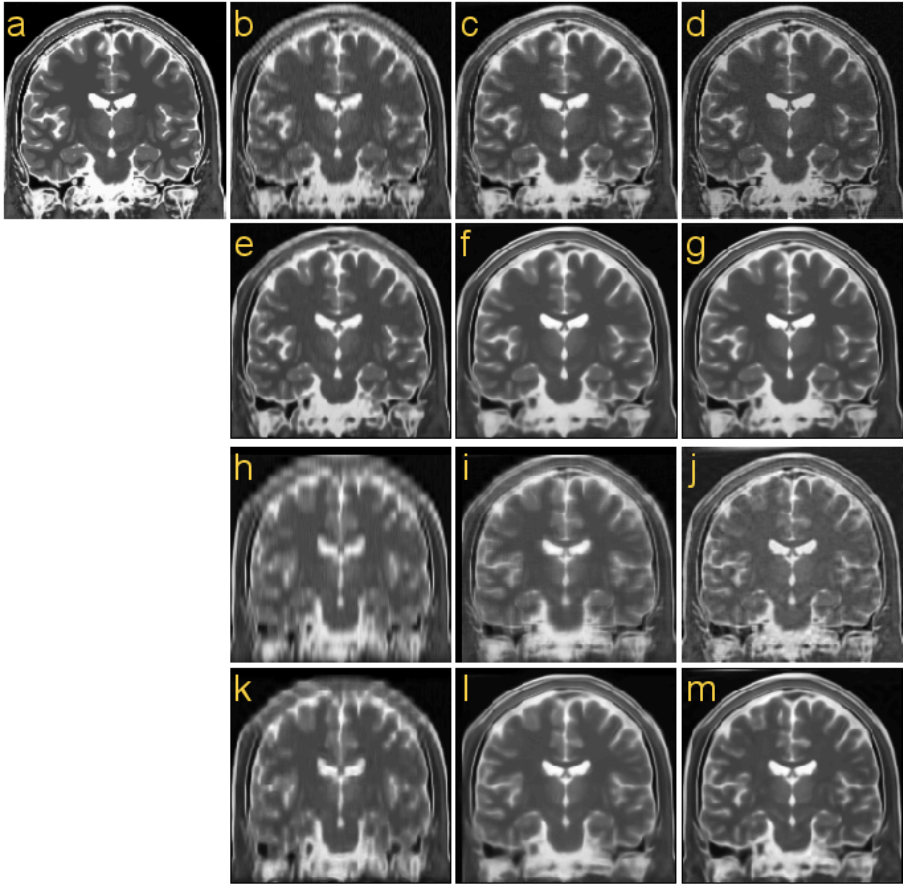


Fig. 3. Coronal slice of the simulated brain images reconstructed using six methods: (a) the noise-free HR reference image, (b) cubic interpolation, (c) 3orthoAvg, (d) LASR, (e) siPBSR, (f) 3orthoPBSR, (g) LASR+PBSR. The reconstruction method of (h)-(m) corresponds to (b)-(g). See Section 3.2 for nomenclature. Reconstructions (b)-(g) and (h)-(m) are, respectively, based on stacks of 3 mm and 7 mm slice thickness.

and thinner slices yielded better reconstructions, and in most cases the reconstructions based on three slice stacks were better than the ones based on a single stack. More importantly, compared to interpolation and LASR, PBSR consistently improved the PSNR and mSSIM of the resulting images. In all cases but one, either 3orthoPBSR or LASR+PBSR performed best.

3.3 Qualitative Evaluation

Figure 3 shows coronal brain MR images reconstructed from either one or three LR stacks with slice thickness of 3 mm or 7 mm generated from the original HR image with a noise level of 3%. The results of the proposed PBSR method

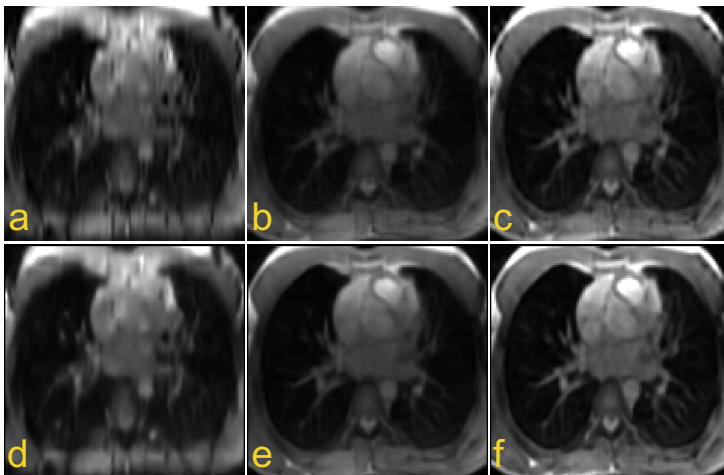


Fig. 4. Axial slice of the lung data reconstructed using the six methods: (a) cubic interpolation, (b) 3orthoAvg, (c) LASR, (d) siPBSR, (e) 3orthoPBSR, (f) LASR+PBSR

(e-g, k-m) appear both sharper and less noisy than the input images (b-d, h-j). Figure 4 shows axial reconstructions of the lung MRI data. The images demonstrate the potential of multi-image SRR for MRI slice stacks: comparing (a), an interpolated coronal slice stack, with (b, c, e, f), where three stacks are combined using SRR methods, the difference is striking. Comparing the SRR images based on three slice stacks with each other, again, the PBSR images (e, f) appear sharper and less noisy than the images they are based on (b, c). PBSR based on a single slice stack (d) also looks sharper and less noisy than the original interpolated image (a), but in this case it is difficult to assess whether (d) becomes a more accurate representation of the underlying anatomy.

4 Discussion and Conclusions

The main contribution of this paper is the idea of exploiting redundancy across scales in an MRI slice stack for SRR. This idea is especially appealing because of the vast amounts of such data that could potentially be enhanced using the proposed method. Cross-scale self-similarity has previously been exploited for SRR in MRI [8,9], but, to the best of our knowledge, the present study is the first that takes explicit advantage of the relation between structural features in the HR slices of a stack and in the orthogonal LR slices.

Our quantitative evaluation revealed that applying PBSR on an initial HR estimate consistently improves the PSNR and mSSIM. The fact that the PBSR improves the HR estimate of another SRR method (LASR) is worth noting. LASR obtains a regularized maximum likelihood estimated after an iterative procedure that takes into account the acquisition process while maintaining global data consistency. The improved performance after applying PBSR indicates that

the image data contain valuable information that is not exploited by the LASR method but can be exploited using a self-similarity prior.

As future work, we plan to study in detail how patch size and the number of patches in the dictionary affect the performance of our method. The number of blurring levels of the dictionary is another free parameter whose effect on the performance will be thoroughly tested. Such results may, however, not generalize easily, since they will be dependent on the specific anatomy, and on the image quality. Also, we have used only the first NNs for our reconstruction. Using an average of the first n NNs may improve the results. Finally we will study how the size of the used dictionary can be optimally chosen to balance computational cost and reconstruction quality.

References

1. Plenge, E., Poot, D.H.J., Bensen, M., Kotek, G., Houston, G., Wielopolski, P., van der Weerd, L., Niessen, W.J., Meijering, E.: Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magn. Reson. Med.* 68, 1983–1993 (2012)
2. Elad, M., Datsenko, D.: Example-based regularization deployed to super-resolution reconstruction of a single image. *Comput. J.* 50, 1–16 (2007)
3. Baker, S., Kanade, T.: Limits on super-resolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 1167–1183 (2002)
4. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* 15, 3736–3745 (2006)
5. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* 19, 2861–2873 (2010)
6. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: *Proc. CVPR*, pp. 60–65 (2005)
7. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: *Proc. ICCV*, pp. 349–356 (2009)
8. Rousseau, F.: A non-local approach for image super-resolution using intermodality priors. *Med. Image Anal.* 14, 594–605 (2010)
9. Manjon, J.V., Coupé, P., Buades, A., Fonov, V., Collins, D.L., Robles, M.: Non-local MRI upsampling. *Med. Image Anal.* 14, 784–792 (2010)
10. Ebrahimi, M., Vrscay, E.R.: Solving the inverse problem of image zooming using “self-examples”. In: Kamel, M., Campilho, A. (eds.) *ICIAR 2007*. LNCS, vol. 4633, pp. 117–130. Springer, Heidelberg (2007)
11. Protter, M., Elad, M., Takeda, H., Milanfar, P.: Generalizing the non-local means to super-resolution reconstruction. *IEEE Trans. Image Process.* 18, 36–51 (2009)
12. Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Evans, A.C.: BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage* 5, S425 (1997)
13. Poot, D.H.J., Van Meir, V., Sijbers, J.: General and efficient super-resolution method for multi-slice MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part I*. LNCS, vol. 6361, pp. 615–622. Springer, Heidelberg (2010)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612 (2004)

Dynamic CT Reconstruction by Smoothed Rank Minimization

Angshul Majumdar^{1,2} and Rabab K. Ward²

¹ Indraprastha Institute of Information Technology, Delhi

² Department of Electrical Engineering, University of British Columbia
angshul@iiitd.ac.in, rababw@ece.ubc.ca

Abstract. We address the problem of dynamic CT reconstruction from parsimoniously sampled sinograms. In this paper we propose a novel approach to solve the aforesaid problem by modeling the dynamic CT sequence as a low-rank matrix. This dynamic CT matrix is formed by stacking each frame as a column of the matrix. As these images are temporally correlated, the dynamic CT matrix would therefore be of low-rank as its columns are not independent. We exploit the low-rank information to reconstruct the CT matrix from its parsimoniously sampled sinograms. Mathematically this is a low-rank matrix recovery problem, and we propose a novel algorithm to solve it. Our proposed method reduces the reconstruction error by 50% or more when compared to previous recovery techniques.

1 Introduction

Traditional knowledge dictates that in order to get a good quality high resolution X-Ray CT image, the sinogram should be densely sampled. Dense sampling of a sinogram requires higher CT dosage than parsimonious sampling. CT reconstruction researchers have been looking for ways to parsimoniously sample the sinogram (thereby reducing the radiation dosage) and to reconstruct a good quality image from it.

Recently, Compressed Sensing (CS) based techniques have shown how transform domain sparsity of the underlying CT image can be exploited in order to recover it from parsimoniously sampled sinogram [1-3]. These studies have shown that CS techniques can indeed be used to cut the CT radiation dose by more than 50% for static CT imaging.

In this work we address the problem of reconstructing dynamic CT images. CS based techniques have also been used in the past to reconstruct the dynamic image sequence for parsimoniously sampled sinograms. In this paper we propose a novel formulation to solve this problem where we model the image sequence as a low-rank matrix. The reconstruction problem is thus recast as a low-rank matrix recovery problem from its parsimoniously sampled sinograms. We also propose a new algorithm to solve the low-rank matrix recovery problem.

The rest of the paper is organized into several sections. The following section briefly reviews the prevalent CS based recovery algorithms in dynamic CT imaging. In section 3, we briefly discuss the similarity between dynamic CT and dynamic magnetic resonance imaging (MRI). We formulate the problem in section 4 and propose a new algorithm to solve it. The experimental results are shown in section 5. Finally the conclusions of this work and future directions of research are discussed in section 6.

2 Compressed Sensing in Dynamic CT

In CT, the data acquisition model can be expressed as follows:

$$y = Ax \quad (1)$$

Here x is the underlying image (to be reconstructed), y is the sampled sinogram and A is the X-ray transform.

For dynamic CT, the sinogram is sampled in an interleaved fashion, so the A matrix changes with time. The data acquisition model for the t^{th} frame is as follows:

$$y_t = A_t x_t \quad (2)$$

This is an inverse problem; one is supposed to reconstruct x_t given A_t and y_t . For non-iterative reconstruction using Filtered Back Projection (FBP), the sinogram needs to be densely sampled; dense sampling translated to higher ionizing radiation for the subject. Researchers in CT reconstruction aim to reconstruct the image from smaller number of sinogram samples. CS based techniques are useful to achieve this goal; CS exploits the sparsity of the image in order to reconstruct it from a smaller number of sinogram samples than was deemed necessary previously [1-3].

Recent papers however have shown how CS techniques can be used for dynamic CT reconstruction [4, 5]. The first step is to generate a static FBP reference image (x_0) from the interleaved projections. Once this reference image is computed, the reconstruction of the t^{th} frame is solved via the following optimization problem,

$$\min_x \alpha \|\Psi_1(x_t - x_0)\|_p^p + (1 - \alpha) \|\Psi_2 x_t\|_p^p \quad \text{subject to } y_t = A_t x_t \quad (3)$$

where Ψ_1 and Ψ_2 are sparsifying transforms (wavelet or gradient). The l_p -norm ($0 < p \leq 1$) is the sparsity promoting objective function. There are two sparsity promoting terms. The first term assumes that the difference between the current frame and the reference image is sparse in Ψ_1 . The second term assumes the t^{th} frame is sparse in Ψ_2 . The scalar α controls the relative importance of the two sparsity promoting terms.

This technique (3) is called Prior Image Constrained Compressed Sensing (PICCS). This was originally developed with convex sparsity promoting l_1 -norm [4] but was later shown to yield even better results with non-convex l_p -norm (NCPICCS) [5]. It should be noted that even though the frames are reconstructed separately, this is an offline technique because the reference image x_0 can only be generated after the full sequence has been collected.

3 Compressed Sensing in Dynamic MRI

In MRI, the data acquisition model is the same as the CT for both static (1) and dynamic (2) scenarios. For MRI, the matrix A is the Fourier transform and the y is the sampled Fourier coefficients (called K-space in MRI). In MRI the challenge is different from CT. For MRI the challenge is to reduce the data acquisition time. Thus A or A_t is not the full Fourier transform (F), it is an under-sampled Fourier transform ($A = RF$); where R is the sampling mask.

Even though the challenges in CT and MRI are different, the fundamental mathematical problem remains the same. When the K-space is parsimoniously sampled, the inverse problems represented by (1) and (2) become under-determined. CS is used to reconstruct the MR images by exploiting their sparsity in a domain such as the wavelet or gradient.

In dynamic MRI reconstruction the main idea is to maximally exploit the spatio-temporal redundancies of the dynamic MRI sequence. The CS based techniques reconstruct the dynamic MRI sequence by solving an optimization problem of the following form [6, 7]:

$$\min_x \|\Psi_1 \otimes \Psi_2(x')\|_1 \text{ subject to } y' = A'x' \tag{4}$$

where y' is the vector formed by concatenating all the acquired vectors y_t 's, similarly x' is the vector formed by concatenating all the unknown x_t 's and A' is a block diagonal matrix formed by A_t 's as the blocks. Ψ_1 is the sparsifying transform along the temporal direction and Ψ_2 is the sparsifying transform along the spatial direction.

CS based sparsity promoting techniques are not the only solution for dynamic MRI reconstruction. In general, one can use a different sampling mask for each frame. But if the sampling mask is the same for all the frames, i.e. if $A_t=A$ for all t 's, then (2) can be expressed as

$$Y = AX \tag{5}$$

where Y is a matrix formed by stacking the y_t 's as columns, similarly X is formed by stacking the x_t 's as columns.

In [8] it is argued that the matrix X is rank deficient; this is because the MRI time-frames are correlated with each other. Thus the columns of X are therefore not independent and thus X can be modeled as a low-rank matrix [8]. Therefore X can be recovered by solving the following problem,

$$\min_x \text{rank}(X) \text{ such that } Y = AX \tag{6}$$

In general, minimizing the rank is a combinatorial problem and it is thus not feasible for large scale systems such as (6). Thus a matrix factorization based approach was proposed in [8] in order to recover X .

4 Proposed Solution

In this work, we propose to solve the dynamic CT reconstruction problem (2) by modeling the sequence of CT images as a low-rank matrix. Our work is motivated by the studies in dynamic MRI reconstruction [8]. In the dynamic CT sequence, the frames are correlated temporally. When the frames from the sequence (x_i 's) are stacked as columns of a matrix (X), the resulting matrix ($X = [x_1 | \dots | x_i | \dots | x_N]$, assuming N frames in all) is rank-deficient; this is because the columns are correlated. We propose to recover this matrix by exploiting its rank-deficiency.

For dynamic CT, the sinograms for different time-frames are sampled in an interleaved fashion. The data acquisition model is expressed as follows:

$$y' = A' x' \quad (7)$$

where y' is the vector formed by concatenating all the (acquired) y_i 's, similarly x' is the vector formed by concatenating all the (unknown) x_i 's and A' is a block diagonal matrix formed by A_i 's as the blocks.

It must be understood by now that x' and X are just two different ways to represent the same group of vectors x_i 's. In x' they are concatenated one after the other, and in X they are stacked as its columns. To exploit the prior information that X is low-rank, we exploit the rank deficiency of X in order to recover it:

$$\min_x \text{rank}(X) \text{ subject to } y' = A' x' \quad (8)$$

This is an NP hard problem. There are two solutions – i) replace the rank by its nearest convex or non-convex surrogate (i.e. nuclear norm or the Schatten-p norm); or, ii) use matrix factorization.

The second approach is computationally faster but does not provide any recovery guarantees. The first approach that recovers the low-rank matrix via nuclear norm minimization [9] provides theoretical recovery guarantees for solving problems like (8). In practice however, it has been found that the non-convex Schatten-p norm minimization yields even better results than the nuclear norm minimization [10].

Our work is motivated by the smoothed l_0 -minimization (SL0) [11] algorithm in CS; SL0 is faster and more accurate than most state-of-the-art l_1 -minimization algorithms. SL0 approximately solves the l_0 -norm minimization, i.e. it does not substitute the NP hard l_0 -norm by its convex (l_1 -norm) or non-convex (l_p -norm) surrogates. In this work we propose to approximately solve the rank-minimization problem (8) by a similar approach.

The matrix X can be expressed in terms of its singular value decomposition (SVD): $X = U \Sigma V^T$, where U and V are the left and right singular vectors and Σ is the diagonal matrix consisting of the singular values σ_j 's. The rank of a matrix is the number of non-zero singular values. We define a function for every singular value,

$$\gamma(\sigma) = \begin{cases} 1 & \text{when } \sigma > 0 \\ 0 & \text{when } \sigma = 0 \end{cases} \quad (9)$$

Based on the above definition, the rank is expressed as $rank(X) = \sum_j \gamma(\sigma_j)$. The function $\gamma(\sigma_j)$ is spiky; i.e. it has the value 0 or 1. Following [11] we replace the spiky function by a smooth zero-mean Gaussian whose spread can be varied by changing its standard deviation (θ) – $f_\theta(\sigma) = e^{(-\sigma^2/2\theta^2)}$

The function is wide when θ is large and becomes narrow when its value reduces. In the limit that the θ is zero, the above function has the following property,

$$\lim_{\theta \rightarrow 0} f_\theta(\sigma) = \begin{cases} 1 & \text{when } \sigma=0 \\ 0 & \text{when } \sigma>0 \end{cases} \tag{10}$$

Therefore, $\lim_{\theta \rightarrow 0} f_\theta(\sigma) = 1 - \gamma(\sigma)$. This allows for approximating the rank by

$\lim_{\theta \rightarrow 0} F_\theta(x) = \sum_{j=1}^n f_\theta(\sigma_j) = \sum_{j=1}^n 1 - \gamma(\sigma_j) = n - rank(X)$, (where n is the minimum of the number of rows or the number of columns in X). Therefore the rank minimization problem (8) can be recast as follows,

$$\max_x F_\theta(x) \text{ subject to } y' = A'x' \tag{11}$$

Since the objective function is smooth, it is easy to solve (11) by gradient based methods. The main idea behind the algorithm proposed below is that at each iteration, (11) is solved for a particular value of θ ; then in the following iteration the value of θ is decreased and (11) is solved again. This continues till the solution converges (i.e. when there is no significant change in the solution).

Algorithm for Smoothed Rank Minimization

Initialization – Obtain the initial solution $\hat{x}^{(0)} = \min_x \|y' - A'x'\|_2^2$. Rearrange $\hat{x}^{(0)}$ in matrix form $\hat{X}^{(0)}$. Compute the SVD, $\hat{X}^{(0)} = U^{(0)}\Sigma^{(0)}V^{(0)T}$

At iteration k – Continue the following steps till solution is reached (i.e. till θ is greater than a specified value)

1. Choose $\theta = c\sigma_1$, where $c > 4$.
2. Maximize (11) for the current value of θ . The Steepest Ascent method is used to achieve this.
 - a. Initialize, $s = diag(\Sigma^{(k-1)})$. Here $diag()$ operator forms a vector from the diagonal elements
 - b. Let $\Delta s = [s_1 \cdot e^{-s_1^2/2\theta^2}, \dots, s_n \cdot e^{-s_n^2/2\theta^2}]^T$.
 - c. Update: $s \leftarrow s - \mu\Delta s$, where μ is a small constant.

Express $\hat{\Sigma}^{(k)} = \text{diag}(s)$. Here $\text{diag}()$ generates a matrix with diagonal elements.

d. Generate the matrix $\hat{X}^{(k)} = U^{(k-1)}\hat{\Sigma}^{(k)}V^{(k-1)T}$ and $\hat{x}^{(k)} = \text{vec}(\hat{X}^{(k)})$.

e. Project the solution back to the feasible set by $x^{(k)} \leftarrow \hat{x}^{(k)} - A^T(A'A^T)^{-1}(A'\hat{x}^{(k)} - y)$.

3. Rearrange $x^{(k)}$ in matrix form and compute the SVD, $X^{(k)} = U^{(k)}\Sigma^{(k)}V^{(k)T}$ and return to step 1 until convergence.

5 Experimental Evaluation

We compared the above proposed technique with the Non-convex Prior Image Constrained Compressed Sensing (NCPICCS) method [7] since this method [7] yields the best reconstruction for dynamic CT when $\alpha = 0.7$ and $p = 0.7$ (refer to problem formulated in (3)). The reconstruction accuracy in our study is measured in terms of Relative Mean Squared Error (RMSE) as this metric has been used previously for the same purpose [5].

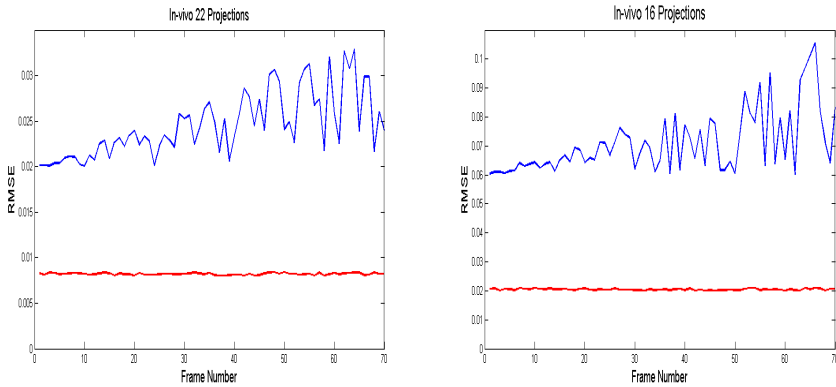


Fig. 1. Variation of RMSE with time. Blue plot represents error from NCPICCS and red plot represents error from proposed method.

We use a portion of the experimental data used in [5]. The reconstructions were carried out on a synthesized Shepp Logan phantom and on an in-vivo animal kidney perfusion CT scans. The Shepp Logan phantom was modified in [5] such that the uppermost ellipses in the simulated original object changed attenuation through time as follows:

$$r = at^b \exp(-t/c) \quad (12)$$

where t is the time and the parameters a , b , and c control the amplitude, width, and speed of decay of the gamma-variate function. The values, $a=.05$, $b=7$, and $c=2$ were used to simulate tissue perfusion. A total of 20 time points were simulated, with time steps $t=0.5$ s.

The in-vivo study scans were performed at 80 kV, using 160 mA s, with 24X1.2 mm collimation and 0 mm table feed. Using a 0.33 sec. gantry rotation time, 70 exposures (images) were acquired with 0.67 sec interval between consecutive frames.

For the simulated Shepp Logan, the reconstruction was carried using 4 and 6 projections (with parallel beam geometry). For the in-vivo experiment the number of projections were 16 and 22. These values were suggested in [5]. The frame-by-frame RMSE's are plotted in Fig. 1. Owing to limitations in space, we only show the results for the in-vivo data.

We see that our proposed method (red plot) reduces the RMSE by 50% or more. Also the variation in error from our proposed method is less compared to NCPICCS. Both of these observations stem from the same fact. NCPICCS and other PICCS based methods reconstruct the images frame-by-frame, whereas our method reconstructs all the frames simultaneously. During reconstruction, our method makes better use of the spatio-temporal redundancy compared to PICCS. That is why our proposed method yields more stable (less variation in time) and better reconstruction results.

To corroborate the numerical results, we show the ground truth, reconstructed and difference (between ground truth and reconstructed) images. Owing to limitations in space, we only show one frame from each of the datasets. The contrast of the difference images is magnified 10 times for visual clarity. From the difference images, it is clearly seen that our proposed method is better than PICCS; the difference images are darker. The improvement from our proposed method is better evident from the phantom. The PICCS reconstructed phantom image clearly shows reconstruction artifacts; the artifacts are absent in our proposed method.



Fig. 2. Left to right: Ground truth, Difference image from NCPICCS, Reconstructed image from NCPICCS, Difference image from proposed method, Reconstructed image from proposed method

6 Conclusion

In this work we propose a novel technique to reconstruct dynamic CT image sequences. The temporal correlation of the CT images allows us to model the entire sequence as a low-rank matrix. We exploit this low-rank structure of the matrix while reconstructing the sequence. The proposed method yields considerably better results than the well known PICCS based technique for reconstructing dynamic CT images from parsimoniously sampled sinograms.

In the future, we want to combine the sparsity promoting reconstruction with the proposed low-rank model to achieve even better reconstruction.

Acknowledgement. This work was supported by NSERC and by Qatar National Research Fund (QNRF) No. NPRP 09 - 310 - 1 - 058.

References

- [1] Song, J., Liu, Q.H., Johnson, G.A., Badea, C.T.: Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-ct. *Medical Physics* 34, 4476–4482 (2007)
- [2] Yu, H., Wang, G.: Compressed sensing based interior tomography. *Physics in Medicine & Biology* 54, 2791–2805 (2009)
- [3] Lee, H., Xing, L., Davidi, R., Li, R., Qian, J., Lee, R.: Improved compressed sensing-based cone-beam CT reconstruction using adaptive prior image constraints. *Physics in Medicine & Biology* 57, 2287 (2012)
- [4] Chen, G.H., Tang, J., Leng, S.: Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Med. Phys.* 35(2), 660–663 (2008)
- [5] Ramirez-Giraldo, J.C., Trzasko, J., Leng, S., Yu, L., Manduca, A., McCollough, C.H.: Nonconvex prior image constrained compressed sensing (NCPICCS): Theory and simulations on perfusion CT. *Med. Phys.* 38(4), 2157–2167 (2011)
- [6] Gamper, U., Boesiger, P., Kozerke, S.: Compressed sensing in dynamic MRI. *Magnetic Resonance in Medicine* 59(2), 365–373 (2008)
- [7] Jung, H., Park, J., Yoo, J., Ye, J.C.: k-t FOCUSS: A General Compressed Sensing Framework for High Resolution Dynamic MRI. *Magnetic Resonance in Medicine* 61, 103–116 (2009)
- [8] Zhao, B., Haldar, J.P., Brinegar, C., Liang, Z.-P.: Low Rank Matrix Recovery for Real-Time Cardiac MRI. In: *IEEE International Symposium on Biomedical Imaging*, pp. 996–999 (2010)
- [9] Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed Minimum Rank Solutions to Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review* 52(3), 471–501 (2010)
- [10] Majumdar, A., Ward, R.K.: Some Empirical Advances in Matrix Completion. *Signal Processing* 91(5), 1334–1338 (2011)
- [11] Mohimani, H., Babaie-Zadeh, M., Jutten, C.: A Fast Approach for Overcomplete Sparse Decomposition Based on Smoothed ℓ_0 Norm. *IEEE Trans. on Signal Processing* 57(1), 289–301 (2008)

Harnessing Group-Sparsity Regularization for Resolution Enhancement of Lung 4D-CT

Arnav Bhavsar, Guorong Wu, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina at Chapel Hill

Abstract. A critical concern with lung 4D-CT is the low superior-inferior resolution, due to the consideration of radiation dose. We propose a resolution enhancement approach that reconstructs missing intermediate slices by exploiting the idea that information lost in one respiratory phase can be found in others, according to the complimentary nature of inter-phase information. Our approach is based on a patch-based framework that explores the role of group-sparsity involving groups of similar neighbouring patches. We discuss the regularizing role of group-sparsity, which helps in reducing the effect of noise and enables better enhancement of anatomical structures. Our results positively demonstrate the potential of group-sparsity for 4D-CT resolution enhancement.

1 Introduction

4D-CT plays a crucial role in radiation therapy for lung cancer. Unlike 3D-CT under free breathing, 4D-CT provides a more accurate estimate of the lung motion across respiratory phases. This helps to better localize the moving structures in the lung.

However, to control radiation dose in 4D-CT, usually a reduced number of slices are acquired, which results in low superior-inferior resolution. This can adversely affect the image quality [1] due to false apparent vessel discontinuities, shape distortions etc., which makes the assessment of tumor and vessel structures difficult [7]. This compromises the potential of 4D-CT for providing accurate structural information. Resolution enhancement here aims at addressing this important limitation.

Resolution enhancement, often known as super-resolution (SR), involves retrieving lost high-frequency sampling information (e.g. fine structures) [2,3]. Typically, SR approaches exploit multiple observations with relative sub-pixel motion [2] or a large dictionary of local patches from an off-line high-resolution (HR) image dataset [3,4].

For 4D-CT, it is difficult to rely on accurate registration due to poor superior-inferior resolution. While there have been attempts for correspondence-based interpolation [5] or super-resolution via motion-estimation [6], in general, registration-based method can be quite inaccurate, as demonstrated in [7]. Also, availability of large off-line HR datasets for constructing dictionaries is infeasible. Nevertheless, one can exploit motion-induced information in a patch-based framework without the need for registration and large HR data. Due to lung motion, local image information across respiratory phases is complimentary, which one can capture in a dictionary of patches [7,8].

We propose a group-sparsity-based approach to integrate such local inter-phase local information. Our approach involves multiple dictionaries that consider spatial neighbourhood and similarity. Our work is closely related to the sparse representation method

in [7], which, however, does not employ any neighbourhood constraints. Accordingly, relatively large-scale patches have to be used in [7], in order to resist noise and other artifacts. However, a large scale results in over-smoothing. On the other hand, a smaller scale to improve localization could result in an error-prone noisy reconstruction.

Such a behaviour is an example of ‘noise-structure trade-off’ [2]. From a clinical view, structural accuracy and localization is important for assessment of shape/extent of tumors and anatomical structures, which is vital in radiation therapy. It is also crucial that noise-effects and artifacts are minimized during reconstruction. Thus, the clinical importance of mitigating the ‘noise-structure trade-off’ can be clearly appreciated.

In this regard, we harness the regularizing potential of group-sparsity over spatially neighbouring patches. This enables a constructive use of smaller patch-scale, which helps in enhancing local structures, while still being robust to noise effects and artifacts.

Thus, our work contributes in: 1) Advancing the recent and clinically important area of 4D-CT resolution enhancement in a different framework than traditional SR methods; 2) Exploring the role of group-sparsity for image reconstruction; 3) Exploiting neighbourhood constraints for better structural enhancement.

2 Role of Group-Sparsity in Regularization

In many estimation problems, multiple factors (tasks) can be related. In such cases, the sparse representation problem can be expressed for a group of tasks as follows:

$$\hat{A} = \arg \min_{\alpha_l} \sum_{l=1}^P \|y_l - D_l \alpha_l\|_2^2 + \lambda \|A\|_{2,1}, \quad \text{where} \quad \|A\|_{2,1} = \sum_{q=1}^m \|\mathbf{a}^q\|_2 \quad (1)$$

Here, y_l is the observation and D_l is the dictionary for the l^{th} task. $A = [\alpha_1, \dots, \alpha_l, \dots, \alpha_P]$ is an $m \times P$ matrix whose columns are coefficient vectors related to the P different tasks. Each vector α_l describes the linear combination of atoms in D_l that best matches y_l . $\|A\|_{2,1}$ is the $l_{2,1}$ -norm, viz. an l_1 -norm over the l_2 -norms of the rows of A , where \mathbf{a}^q denotes the q^{th} row of A . The $l_{2,1}$ -norm enforces coefficient vectors for all tasks to have a similar sparsity structure, i.e., same locations of the zero elements. The intuition is that the related tasks employ the same (or similar) dictionary atoms.

Such an insight helps us to exploit the regularization potential of group-sparsity. We group patches into dictionaries, considering spatial neighbourhood and appearance similarity. The selected patches satisfy closeness with their neighbouring patches, resulting in smoothness, which, however, tends not to adversely affect the salient structures. Thus, group-sparsity plays a regularizing role towards mitigating the noise-structure trade-off. To our knowledge, only [10,11] employ group-sparsity for image reconstruction. While our method is very different, these works support such an intuition.

3 Methodology

Given 4D-CT data $I = \{I_i(s) | i = 1, \dots, L; s = 1, \dots, S\}$ (with L phases and S slices), we estimate a slice between $I_i(s)$ and $I_i(s+1)$ for the i^{th} phase, in a patch-wise manner [7]. We consider patches from $I_i(s)$ and $I_i(s+1)$ to form the observation vectors.

Ours is a two-stage strategy, where the second stage uses the reconstructed slice from the first. We elaborate our method for the first step, and briefly indicate the modifications in the second. Patches are used as lexicographic vectors and, for simplicity, we do not use the subscript denoting the phase in which the slice is reconstructed.

3.1 Group-Based Dictionary Construction and Group-Sparse Representation

Our method involves a group of neighbouring patches. Below, we discuss the dictionary construction for this group, followed by our group-sparse representation.

Dictionary Construction for the Central Patch

The dictionary construction for the central patch is similar to that in [7]. The 2D patches \mathbf{y}_c^U and \mathbf{y}_c^L from $I(s)$ and $I(s+1)$, respectively, and their x and y gradients ($\mathbf{F}^x(\mathbf{y}_c^U)$, $\mathbf{F}^y(\mathbf{y}_c^U)$, $\mathbf{F}^x(\mathbf{y}_c^L)$, $\mathbf{F}^y(\mathbf{y}_c^L)$) are used as observations. We search for candidate patches \mathbf{y}_{p_c} in a 3D region in each phase p (other than the current phase), around the same locations as that of \mathbf{y}_c^U and \mathbf{y}_c^L . We select the K best patches yielding the K lowest costs. Observing that \mathbf{y}_c^U and \mathbf{y}_c^L may be dissimilar, we incorporate a balancing condition in the overall cost E_c^d as defined below.

$$E_c^d = E_c^U + E_c^L \quad \text{if } 1/\epsilon < E_c^U/E_c^L < \epsilon, \quad \text{and } E_c^d = \infty \quad \text{otherwise} \quad (2)$$

where the sub-costs E_c^U and E_c^L , involving \mathbf{y}_c^U and \mathbf{y}_c^L , respectively, are defined as

$$\begin{aligned} E_c^U &= \|\mathbf{y}_c^U - \mathbf{y}_{p_c}\|_2 + \gamma(\|\mathbf{F}^x(\mathbf{y}_c^U) - \mathbf{F}^x(\mathbf{y}_{p_c})\|_2 + \|\mathbf{F}^y(\mathbf{y}_c^U) - \mathbf{F}^y(\mathbf{y}_{p_c})\|_2) \\ E_c^L &= \|\mathbf{y}_c^L - \mathbf{y}_{p_c}\|_2 + \gamma(\|\mathbf{F}^x(\mathbf{y}_c^L) - \mathbf{F}^x(\mathbf{y}_{p_c})\|_2 + \|\mathbf{F}^y(\mathbf{y}_c^L) - \mathbf{F}^y(\mathbf{y}_{p_c})\|_2) \end{aligned} \quad (3)$$

where γ is the weighting of the gradient feature cost. Thus, only those patches which yield a low-cost and are similar to both \mathbf{y}_c^U and \mathbf{y}_c^L are included in the dictionary.

A similar *balancing* as in equation (2) is also used in [7], but during the greedy sparse representation step. We follow a non-greedy optimization for sparse representation, where the balancing term yields non-linearities. Hence, we incorporate it during dictionary construction, to resist imbalanced patches from entering the optimization.

Dictionary Construction for Neighbouring Patches

Having defined the dictionary for the central patch (say D_c), we now construct N neighbourhood dictionaries D_1 to D_N using observed patches $\mathbf{y}_1^U, \dots, \mathbf{y}_N^U$ and $\mathbf{y}_1^L, \dots, \mathbf{y}_N^L$ neighbouring to the central patch \mathbf{y}_c^U and \mathbf{y}_c^L , respectively. This proceeds as follows:

a) Patches in D_1, \dots, D_N are selected such that their spatial relationship with those in D_c is consistent as that of $\mathbf{y}_1^U, \dots, \mathbf{y}_N^U$ to \mathbf{y}_c^U (and $\mathbf{y}_1^L, \dots, \mathbf{y}_N^L$ to \mathbf{y}_c^L). For instance, if \mathbf{y}_1^U is left to \mathbf{y}_c^U , then the patches in D_1 are left to those in D_c . This ensures that the spatial correspondence among neighbouring candidate patches is same as that of neighbouring observed patches. b) To induce a constructive smoothness, the candidate patches that contribute to the dictionaries satisfy a soft-similarity condition (equation (4)), so as to avoid the inclusion of unsuitable patches in the dictionaries.

$$\mathbf{y}_{p_n} \in D_n \quad \text{if } E_n^d < \kappa E_c^d \quad (4)$$

where \mathbf{y}_{p_n} is the candidate patch for dictionary D_n , and E_n^d is the corresponding cost for \mathbf{y}_{p_n} (defined similarly to E_c^d , using proper change in subscripts). κ is a threshold parameter. Note that this condition plays a similar (rather, a softer) role as neighbourhood similarity [4], which advocates that neighbours tend to be similar in appearance. Hence, it is fair to assume that neighbouring patches typically tend to have similar costs. Indeed, the condition helps in excluding those dissimilar neighbouring patches, which may not be suitable in group-sparse representation for inducing a constructive smoothness.

For every patch, we keep the number of atoms in all dictionaries (including D_c) equal. Hence, if the above condition is not satisfied by all the neighbouring patches, we do not include the corresponding patch in any dictionary.

Group-Sparse Coefficient Estimation and Reconstruction

We use the constructed dictionaries to solve for the coefficients that minimizes the group-sparse problem of equation (1), rewritten for our case in the left equation in (5):

$$\hat{A} = \arg \min_{\alpha_k} \sum_{k \in \{1, \dots, N\}, c} \|\tilde{\mathbf{y}}_k - D_k \alpha_k\|_2^2 + \lambda \|A\|_{2,1} \quad \tilde{\mathbf{y}}_c^O = D_c \hat{\alpha}_c \quad (5)$$

where $\tilde{\mathbf{y}}_k$ is the concatenated observation vector for the k^{th} task defined as: $\tilde{\mathbf{y}}_k = [(\mathbf{y}_k^U)^T, (\gamma \mathbf{F}^x(\mathbf{y}_k^U))^T, (\gamma \mathbf{F}^y(\mathbf{y}_k^U))^T, (\mathbf{y}_k^L)^T, (\gamma \mathbf{F}^x(\mathbf{y}_k^L))^T, (\gamma \mathbf{F}^y(\mathbf{y}_k^L))^T]^T$. To maintain dimensional consistency with $\tilde{\mathbf{y}}_k$, the atoms in the dictionary are defined as $[(\tilde{\mathbf{y}}_{p_k})^T, (\tilde{\mathbf{y}}_{p_k}^T)^T]^T$, where $\tilde{\mathbf{y}}_{p_k} = [(\mathbf{y}_{p_k})^T, (\gamma \mathbf{F}^x(\mathbf{y}_{p_k}))^T, (\gamma \mathbf{F}^y(\mathbf{y}_{p_k}))^T]^T$.

The coefficient matrix \hat{A} is estimated via convex optimization of left equation in (5) [9]. Following this, the output vector $\tilde{\mathbf{y}}_c^O$ corresponding to the central patch is reconstructed as in the right equation in (5) using $\hat{\alpha}_c$, the coefficient vector corresponding to the center patch (which is extracted from the matrix \hat{A} and normalized to 1). The top one-sixth of $\tilde{\mathbf{y}}_c^O$ (containing raw intensities) is reshaped and placed into the unknown slice, with overlapping values properly averaged.

Observe that although the coefficients are estimated for all dictionaries, we only reconstruct the central patch, i.e. using only the dictionary D_c and coefficient vector $\hat{\alpha}_c$. It might appear that the group-sparsity is not being exploited during the patch reconstruction. However, we note that group-sparsity regularization guides the coefficient estimation (including $\hat{\alpha}_c$), thus enabling the selection of those dictionary patches from D_c for reconstruction and better inducing the controlled smoothness.

3.2 Two-Stage Strategy

The above approach provides a reasonable reconstruction, but has further scope of improvement. Hence, we employ a second step that improves upon the above reconstruction. We now use only the patches from the reconstructed slice.

The dictionary construction cost for the central patch using the patch \mathbf{y}_c^R in the reconstructed slice, and its gradients $\mathbf{F}^x(\mathbf{y}_c^R)$ and $\mathbf{F}^y(\mathbf{y}_c^R)$, is similar to equation (3) as

$$E_c^d = \|\mathbf{y}_c^R - \mathbf{y}_{p_c}\|_2 + \gamma (\|\mathbf{F}^x(\mathbf{y}_c^R) - \mathbf{F}^x(\mathbf{y}_{p_c})\|_2 + \|\mathbf{F}^y(\mathbf{y}_c^R) - \mathbf{F}^y(\mathbf{y}_{p_c})\|_2) \quad (6)$$

Note that here we do not require any balancing condition. The neighbourhood dictionaries are also constructed similarly as before, except with $\mathbf{y}_1^R, \dots, \mathbf{y}_N^R$ as the patches neighbouring to \mathbf{y}_c^R . Finally, sparse representation and reconstruction is carried out as

$$\hat{A} = \arg \min_{\alpha_k} \sum_{k \in \{1, \dots, N\}, c} \|\tilde{\mathbf{y}}_k^R - D_k \alpha_k\|_2^2 + \lambda \|A\|_{2,1} \quad \text{and} \quad \tilde{\mathbf{y}}_c^F = D_c \hat{\alpha}_c \quad (7)$$

where $\tilde{\mathbf{y}}_k^R = [(\mathbf{y}_k^R)^T, (\gamma \mathbf{F}^x(\mathbf{y}_k^R))^T, (\gamma \mathbf{F}^y(\mathbf{y}_k^R))^T]^T$, and the dictionary atoms for the selected patches are denoted as: $\tilde{\mathbf{y}}_{p_k} = [(\mathbf{y}_{p_k})^T, (\gamma \mathbf{F}^x(\mathbf{y}_{p_k}))^T, (\gamma \mathbf{F}^y(\mathbf{y}_{p_k}))^T]^T$. $\tilde{\mathbf{y}}_c^F$ is the output vector which is used in reconstructing the final output slice.

The need for the second step indicates that the group-sparsity method works better given a ‘good’ estimate from the first step. Hence, inspired by [7], a larger scale is chosen in the first step, which yields smooth but structurally more correct results. We then choose a smaller scale in the second step for better structure localization.

While the method in [7] also reduces scale over iterations, the successive outputs are averaged with a large weight to the output at the highest scale. Our second stage output is constructed strictly at a lower scale, thus yielding better structural enhancement and still maintaining robustness to noise and artifacts due to group-sparsity regularization.

4 Experimental Results

Our experiments involve the DIR-Lab lung data [12], containing 10 cases, each with 10 respiratory phases. The in-plane resolution is 1 mm and the superior-inferior resolution is 2.5 mm. We further subsample this data by removing alternate slices, so as to compare our estimated intermediate slice with the true slice. We provide qualitative and quantitative results, and also compare with the method in [7], a related sparse-representation-based approach. As our work is about resolution enhancement, we select a range of slices with a good amount of vessel structures. The quantitative metrics are computed around vessel regions extracted using the method in [13]. Such an error computation emphasizes structural enhancement, better localization and reduction in artifacts, which is central to SR and an indicator of improvement in spatial resolution. Due to space constraints, we do not provide visual results for interpolation methods.

We use the MALSAR package [14] for group-sparse coefficient computation. Our approach involves parameters γ , ϵ , κ and λ which, for the results below, are as follows: $\gamma = 0.2$, $\epsilon = 1.1$, κ is chosen so that dictionary contains 300 - 400 patches, and $\lambda = 40000$ (sparsity of $\sim 30\%$). The patch-size in the first step is 32×32 and in the second is 16×16 . The 3D search region is $11 \times 11 \times 11$, and we employ $N = 4$ neighbors to the center patch. It is worth noting that our parameter range for best result is narrow, which simplifies tuning, in general.

4.1 Visual Qualitative Results

We first demonstrate some qualitative slice reconstructions in Fig. 1, which shows some typical outputs. The first to third columns depict reconstruction obtained using the approach in [7], our output, and ground truth, respectively. It can be observed that vessel structures in our outputs are better localized and more accurate than those in [7].

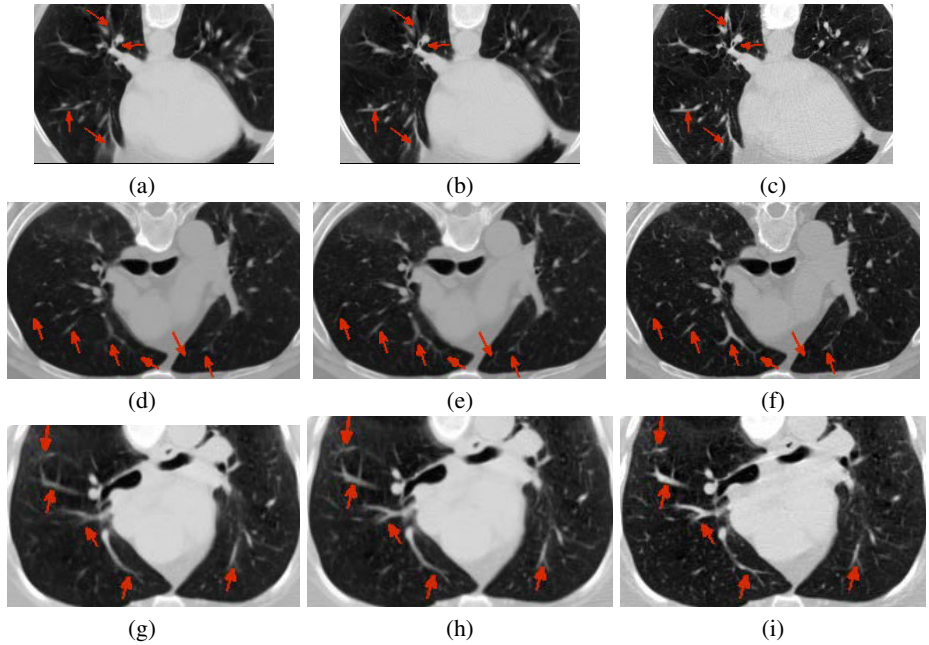


Fig. 1. Examples for reconstruction of some slices: (a,d,g) Outputs using the approach in [7]. (b,e,h) Outputs using the proposed method. (c,f,i) Ground-truth.

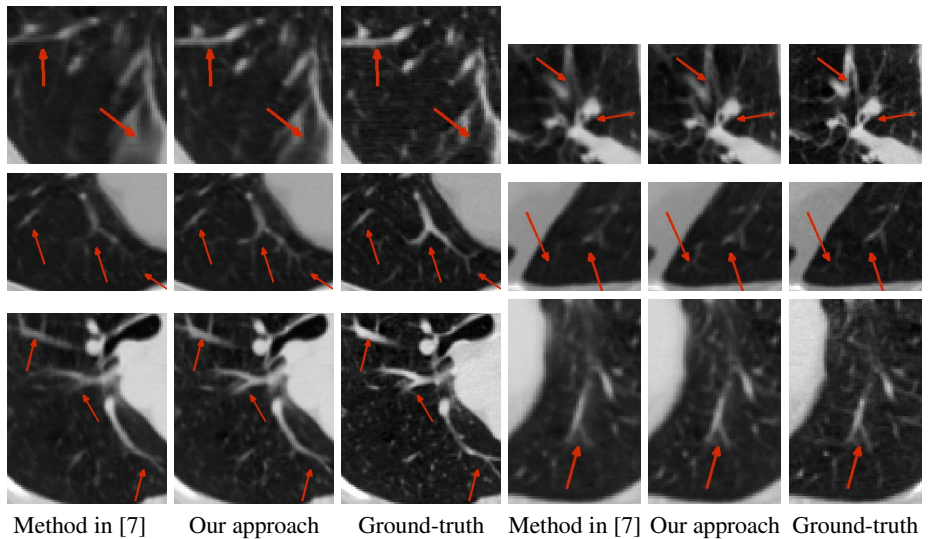


Fig. 2. Close-up views: Each row depicts the close-up views of two regions from slices shown in the corresponding row of Fig. 1. Column pairs {1, 4}, {2, 5}, {3, 6} correspond to the results by the approach in [7], proposed approach and ground-truth, respectively.

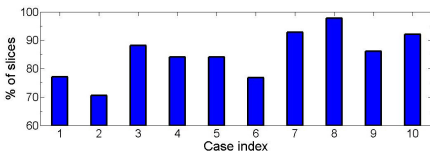
Some of the improvements are marked with red arrows. We zoom into some regions of Fig. 1, and show the close-up views in Fig. 2, where one can further appreciate our method achieving more plausible and better structure reconstruction. While the results are somewhat smooth than ground-truth (which is due to patch averaging), they clearly indicate improved spatial resolution of vessel structures over the state-of-the-art [7].

4.2 Quantitative Results

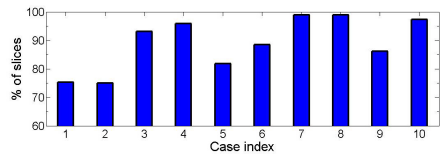
We next provide quantitative results averaged over complete sets of data used in our experiments. Table 1 provides case-wise root-mean-square (RMS) and structural similarity (SSIM) metrics for bi-cubic interpolation, the method in [7] and our approach. Observe that our approach shows a clear improvement with respect to both metrics for all 10 cases (note that such an order of improvement is common in contemporary super-resolution (e.g. [3,15])). The RMS results support our claims for overall mitigation of the noise-structure trade-off, whereas the SSIM results emphasize better structure enhancement. In Fig. 3 we show the percentage of slices across cases, for which our approach favourably compares with that in [7]. Clearly, our approach better reconstructs a vast majority of slices, which highlights that the group-sparsity-based smoothing indeed has a constructive effect. Overall, the RMS improvement is over 85% of the slices and the SSIM improvement is over 89% of the slices.

Table 1. Average RMS and SSIM for 10 cases

Case (No. of Slices)	RMS: Bicubic	RMS: [7]	RMS: Proposed	SSIM: Bicubic	SSIM: [7]	SSIM: proposed
Case 1 (170)	28.97	18.56	17.96	0.5492	0.7069	0.7127
Case 2 (200)	27.83	16.74	16.32	0.5777	0.7445	0.7511
Case 3 (160)	26.82	15.86	15.15	0.5895	0.7589	0.7734
Case 4 (170)	26.95	17.10	16.36	0.5869	0.7372	0.7550
Case 5 (170)	29.00	19.23	18.24	0.5671	0.7184	0.7329
Case 6 (190)	24.81	17.75	17.25	0.5151	0.6255	0.6421
Case 7 (180)	28.12	18.94	17.94	0.5181	0.6622	0.6856
Case 8 (180)	37.38	26.72	24.70	0.4913	0.6137	0.6462
Case 9 (130)	25.34	16.63	16.04	0.5558	0.6980	0.7123
Case 10 (190)	37.10	25.40	23.74	0.4820	0.6148	0.6461
Average (Total: 1740)	29.23	19.29	18.37	0.5433	0.6860	0.7057



(a)



(b)

Fig. 3. % of slices on which our approach performs favourably over [7] for (a) RMS, (b) SSIM

5 Conclusion

We proposed a super-resolution method for 4D-CT, where we discussed the regularizing role of group-sparsity and employed it via careful dictionary construction based on patch neighbourhood and similarity. Our results justify the potential of group-sparsity for noise robustness and structure enhancement. In the future, we aim to explore other group-sparsity frameworks and carry out exhaustive parameter analysis.

References

1. Rit, S., Sarrut, D., Desbat, L.: Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax. *IEEE Trans. on Medical Imaging* 28(10), 1513–1525 (2009)
2. Bose, N., Ahuja, N.: Superresolution and noise filtering using moving least squares. *IEEE Trans. on Image Processing* 15(8), 2239–2248 (2006)
3. Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-Resolution via sparse representation. *IEEE Trans. on Image Processing* 19(11), 2861–2873 (2010)
4. Freeman, W., Jones, T., Pasztor, E.: Example-based super-resolution. *IEEE Computer Graphics and Applications* 22(2), 56–65 (2002)
5. Goshtasby, A., Turner, D., Ackerman, L.: Matching of tomographic slices for interpolation. *IEEE Trans. on Medical Imaging* 11(4), 507–516 (1992)
6. Wu, G., Lian, J., Shen, D.: Improving image-guided radiation therapy of lung cancer by reconstructing 4D-CT from a single free-breathing 3D-CT on the treatment day. *Med. Phys.* 39(12), 7694–7709 (2012)
7. Zhang, Y., Wu, G., Yap, P., Feng, Q., Lian, J., Chen, W., Shen, D.: Hierarchical patch-based sparse representation - A new approach for resolution enhancement of 4D-CT lung data. *IEEE Trans. on Medical Imaging* 31(11), 1993–2005 (2012)
8. Zhang, Y., Wu, G., Yap, P.-T., Feng, Q., Lian, J., Chen, W., Shen, D.: Non-local means resolution enhancement of lung 4D-CT data. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 214–222. Springer, Heidelberg (2012)
9. Liu, J., Ji, S., Ye, J.: Multi-task feature learning via efficient $l_{2,1}$ -norm minimization. In: *Conference on Uncertainty in Artificial Intelligence (UAI 2009)*, pp. 339–348 (2009)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: *International Conference on Computer Vision (ICCV 2009)*, pp. 2272–2279 (2009)
11. Yang, C.-Y., Huang, J.-B., Yang, M.-H.: Exploiting self-similarities for single frame super-resolution. In: *Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part III. LNCS*, vol. 6494, pp. 497–510. Springer, Heidelberg (2011)
12. Castillo, R., Castillo, E., Guerra, R., Johnson, V., McPhail, T., Garg, A., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Phys. Med. Biol.* 54, 1849–1870 (2009)
13. Heng, J., Chen, C., Cole, E., Pisano, E., Shen, D.: Automated delineation of calcified vessels in mammography by tracking with uncertainty and graphical linking techniques. *IEEE Trans. on Medical Imaging* 31(11), 2143–2155 (2012)
14. Zhou, J., Chen, J., Ye, J.: *MALSAR: Multi-tAsk Learning via StructurAl Regularization*. Arizona State University (2012)
15. Takeda, H., Milanfar, P., Protter, M., Elad, M.: Super-Resolution Without Explicit Subpixel Motion Estimation. *IEEE Trans. on Image Processing* 18(9), 1958–1975 (2009)

Improved Myocardial Scar Characterization by Super-Resolution Reconstruction in Late Gadolinium Enhanced MRI*

Oleh Dzyubachyk^{1,**}, Qian Tao^{1,**}, Dirk H. J. Poot³, Hildo Lamb¹,
Katja Zeppenfeld², Boudewijn P. F. Lelieveldt^{1,4}, and Rob J. van der Geest¹

¹ Department of Radiology, Leiden University Medical Center,
Leiden, The Netherlands

² Department of Cardiology, Leiden University Medical Center,
Leiden, The Netherlands

³ Departments of Radiology and Medical Informatics, Erasmus MC — University
Medical Center Rotterdam, Rotterdam, The Netherlands

⁴ Intelligent Systems Department, Delft University of Technology,
Delft, The Netherlands

Abstract. Image resolution is an important factor for accurate myocardial scar assessment from late gadolinium enhanced (LGE) MR. It has been shown that the conventionally used short-axis (SA) LGE acquisition with anisotropic resolution may overestimate the scar size due to partial volume effect, undermining the prognostic and diagnostic accuracy of LGE MRI in critical clinical applications. In this work, we present a method for combining three complementary anisotropic orthogonal LGE sequences of the heart region into a single isotropic volume. Our algorithm is based on the super-resolution reconstruction technique and employs joint localized gradient-correlation-based technique for compensation of breathing motion. The proposed method was validated on the gold standard electroanatomical voltage mapping (EAVM) data of 15 post-infarction patients. The reconstructed myocardial scar image demonstrated improved agreement with the EAVM compared to the conventional SA image, especially at the clinically significant gray zone region.

Keywords: Myocardial scar, late gadolinium enhanced MR, image registration, super-resolution reconstruction.

1 Introduction

Myocardial infarction occurs in more than 7 million patients worldwide annually, and is a frequent cause for cardiac arrhythmia and sudden cardiac death. Accurate characterization of the post-infarction myocardial scar has important

* This research was supported by the Dutch Technology Foundation STW (Stichting Technische Wetenschappen) via grants 10894 and MEDiate.

** Both authors contributed equally to this work.

diagnostic and prognostic value for patient management and treatment [1], in e.g. defibrillator implantation and catheter ablation. In recent years, late gadolinium enhanced (LGE) MR has become the gold standard for myocardial scar imaging in clinical practice [2]. In particular, the size of partially infarcted myocardial tissue, named as *gray zone*, has shown to be an important predictor for risks of death and arrhythmia, outperforming traditional function parameters [1].

However, recent animal studies demonstrated that the LGE-derived myocardial scar size, especially that of the gray zone, is susceptible to partial volume effect (PVE) and changes with MR imaging resolution [3]. These studies underscore the significant importance of image resolution for reliable clinical evaluation of myocardial scar. Unfortunately, current clinical LGE typically has poor through-plane resolution (between 5 to 10 *mm*) compared to in-plane resolution of about 1.5 *mm*, resulting in significant PVE. To alleviate the problem, standard MR protocols often include three LGE acquisitions in orthogonal¹ views: short-axis (SA), two-chamber (2CH), and four-chamber (4CH). To evaluate each view separately is however cumbersome and suboptimal with PVE present in different directions. The newly developed three-dimensional LGE MR technique can be used for isotropic myocardial scar imaging, however, the sequence is still of limited availability and the image quality is currently compromised by motion artifacts [4].

1.1 Related Work

Super-resolution reconstruction (SRR) comprises a group of methods aiming at creating a high-resolution image from multiple low-resolution views of the same scene [5]. During last years, reconstruction of isotropic three-dimensional (3D) MR volumes from a set of anisotropic images has become, probably, the largest consumer of SRR techniques. This is primarily caused by intrinsic limitations imposed on the slice thickness by some MR scanning protocols. Reconstruction of the isotropic volume from three orthogonal views was, in particular, applied to cardiac MR images by using two orthogonal views [6]. However, combining all three views was concluded as not providing any additional benefit compared to averaging due to difficulty in aligning the data.

1.2 Our Contribution

Here we present our algorithm for reconstruction of an isotropic 3D myocardial scar volume from three orthogonal views in LGE MR sequences. To our knowledge, this is the first work to address the issue of low clinical resolution problem in MRI myocardial scar imaging, with validation on *in vivo* clinical measurements. Three orthogonal views are combined by the SRR framework. For correction of the heart position displacements as result of the breath-hold variations, we suggest using a localized gradient-correlation-based registration

¹ Here and further we refer by word “orthogonal” to both orthogonal as well as nearly-orthogonal cases.

algorithm in combination with sub-voxel refinement that is performed simultaneously with the SRR. Estimation of the displacement is performed in a joint manner, using all the available image stacks simultaneously, which guarantees high quality and consistency of the calculated results. In addition, we have developed a method for compensating for possible slice shifts in the SA volumes that are acquired with two separate breath-holds.

2 Method

2.1 MRI Data

Fifteen chronic infarction patients who were referred for MR prior to catheter ablation of ventricular arrhythmia were involved in this study. MR imaging was performed using a 1.5T Gyroscan ACS-NT MRI scanner (Philips Medical Systems, Best, The Netherlands). After obtaining the scout and cine views, a Look-Locker sequence was acquired approximately 15 *min* after bolus injection of gadolinium DPTA (0.15 *mmol/kg*). T₁-weighted LGE images were acquired with an inversion-recovery three-dimensional turbo-field echo sequence with parallel imaging. The heart was imaged with 20 to 24 imaging levels in the SA view, and 5–10 levels in the 2CH and 4CH view. All views were acquired at the same late diastolic phase. Due to the high number of slices in the SA view, it was typically acquired as two stacks in two separate breath-holds. For the other two views, images were acquired in one stack within a single breath-hold. For the SA acquisition, the slice thickness is 10 *mm* with 5 *mm* overlap, with in-plane resolution of 1.56×1.56 *mm*; for the other two acquisitions, the slice thickness is 12 *mm* with 6 *mm* overlap, with in-plane resolution of 1.56×1.56 *mm*.

2.2 Heart Registration

Since each acquisition is performed with a separate breath-hold, the position of the heart may vary in each of the volumes. In particular, this may result in a large discontinuity in the slice direction in the SA volumes that are acquired using two breath-holds.

Registration schemes based on the image gradient, including the one used in this work, are proven to be more robust to the intensity variation in the MR data than those operating on the intensity directly. For alignment of the heart in different images, we used the FFT-based scale invariant image registration with gradient correlation [7], which allows immediate recovery of the unknown translation, rotation, and scaling between two image stacks. This method can also estimate sub-pixel image displacements, but, unfortunately, this methodology was not yet extended to 3D.

Before performing the registration, all the image stacks have to be transformed into a common coordinate system. We achieve it by bringing all three image stacks into the coordinate space of the SA view. At this stage, we aim at rough (with voxel precision) compensation of the inter-scan heart displacement. Thus,

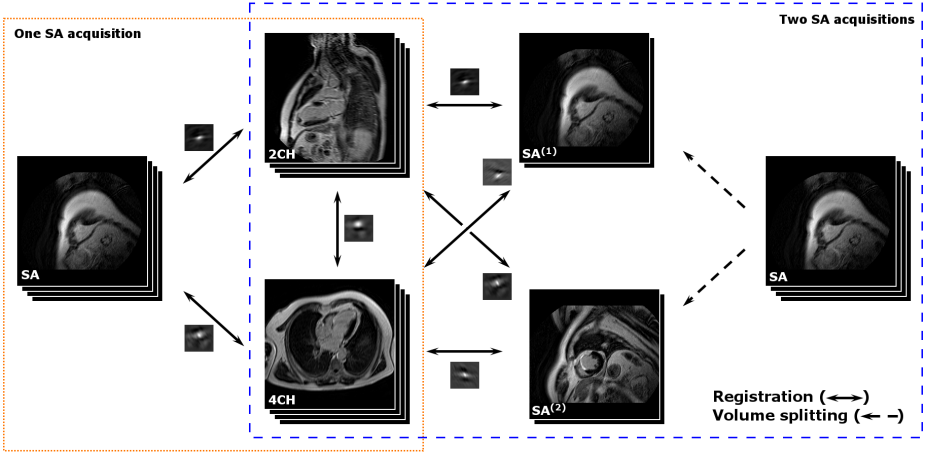


Fig. 1. Joint gradient-correlation-based registration of three orthogonal heart views

we restrict ourselves to estimation of the unknown displacement. As the region around the heart is the volume of interest, the registration method is expected to perform reliable alignment inside it while neglecting possible misalignment outside. To constrain the registration, we calculate the region where all three views intersect. Then, we find the largest right rectangular parallelepiped R that fits entirely in the described overlap region. This step is necessary for performing the correlation, which expects rectangular images as input. Finally, each of the views is cropped to the region R .

The gradient of each view in the region R is used to estimate the unknown displacements $\mathbf{d}_{(V_1, V_2)}$ between each volume pair $V_1, V_2 \in \{\text{SA}, 2\text{CH}, 4\text{CH}\}$. The best triple of the displacement vectors $\mathcal{D} = \{\mathbf{d}_{(\text{SA}, 2\text{CH})}; \mathbf{d}_{(\text{SA}, 4\text{CH})}; \mathbf{d}_{(2\text{CH}, 4\text{CH})}\}$ should maximize each of corresponding gradient-correlation maps

$$\mathbf{d}_{(V_1, V_2)} = \arg \max_{\mathbf{x}} GC(\mathbf{x}; V_1, V_2),$$

and satisfy the vector addition rule

$$\mathbf{d}_{(\text{SA}, 2\text{CH})} + \mathbf{d}_{(2\text{CH}, 4\text{CH})} = \mathbf{d}_{(\text{SA}, 4\text{CH})}.$$

The gradient-correlation maps are scaled beforehand by dividing them by their largest value. Consequently, we calculate the set \mathcal{D} by jointly analyzing all three correlation maps (see left panel of the Fig. 1) for all possible combinations of displacements. In order to limit the search space, for each correlation map we select only the points with the intensity higher than 99.99 percentile of all the intensity values.

2.3 Slice Shift Correction in SA Volume

For correction of a possible slice shift in the SA view, we developed a method that utilizes information from the other two views. More precisely, we split the

entire SA volume into two sub-volumes $SA^{(1)}$ and $SA^{(2)}$, each acquired within a separate breath-hold. Consequently, both sub-volumes are registered to the 2CH and the 4CH volumes. This procedure, illustrated in the right panel of Fig. 1, is similar to the one described above. The only difference is that in this case joint estimation of the displacements is performed on five gradient correlation maps

$$\mathcal{D} = \left\{ \mathbf{d}_{(SA^{(1)},2CH)}, \mathbf{d}_{(SA^{(1)},4CH)}, \mathbf{d}_{(SA^{(2)},2CH)}, \mathbf{d}_{(SA^{(2)},4CH)}; \mathbf{d}_{(2CH,4CH)} \right\}$$

with the following constraints

$$d_{(SA^k,2CH)} + d_{2CH,4CH} = d_{(SA^k,4CH)}, \quad k = 1, 2.$$

The unknown shift $\mathbf{d}_{(SA^{(1)},SA^{(2)})}$ between the $SA^{(1)}$ and the $SA^{(2)}$ is calculated based on the estimated displacements using the vector addition rule

$$\mathbf{d}_{(SA^{(1)},SA^{(2)})} = \mathbf{d}_{(SA^{(1)},2CH)} - \mathbf{d}_{(SA^{(2)},2CH)} = \mathbf{d}_{(SA^{(1)},4CH)} - \mathbf{d}_{(SA^{(2)},4CH)}. \quad (1)$$

Allowing arbitrary slice shifts may result in an unwanted gap in the image stack in case the estimated displacement $\mathbf{d}_{(SA^{(1)},SA^{(2)})}$ has a positive component in the slice direction z . To avoid such scenarios, we assume that the slice shift is in-plane. Mathematically this means that $d_{(SA^{(1)},SA^{(2)}),z} = 0$, which adds two extra constraints

$$d_{(SA^{(1)},2CH),z} = d_{(SA^{(2)},2CH),z}, \quad d_{(SA^{(1)},4CH),z} = d_{(SA^{(2)},4CH),z}. \quad (2)$$

After the slice shift $\mathbf{d}_{(SA^{(1)},SA^{(2)})}$ is calculated, the corresponding displacement is applied to $SA^{(2)}$ and both parts are recombined into a single SA volume.

2.4 Orthogonal Super-Resolution Reconstruction

For calculating the high-resolution image, we use the method of Poot *et al.* [8]. Intensities of all three views were normalized by histogram equalization beforehand. The SSR was implemented with total variation smoothing. During reconstruction, all the affine transformation parameters: translation, rotation, and scale, are iteratively updated to allow sub-voxel precision of image alignment.

3 Experiments and Results

3.1 Myocardial Scar Characterization

SA LGE sequence is conventionally used in clinical studies to assess myocardial scar. To evaluate the improvement of the proposed SRR over SA, we interpolated the original SA volume to the same isotropic resolution as SRR by cubic spline and used it as a reference in all experiments. From both SRR and SA volume, we performed myocardial scar characterization using a validated scar segmentation method [9]. The myocardial scar region was identified using the Otsu method to differentiate normal and scar tissue. Within this region, the scar core zone and gray zone were further differentiated by the full-width-half-maximum method as validated on histological data [10]. It is worth mentioning that the scar was not necessarily at the same location among subjects.

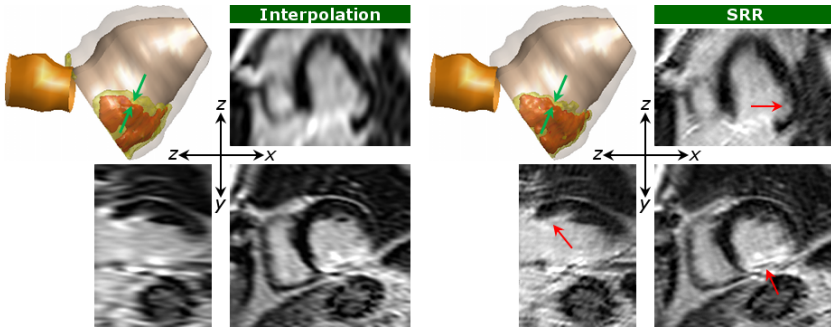


Fig. 2. Comparison between the interpolated SA and the SRR volumes. A 3D reconstruction of the myocardial scar and one slice of each orthogonal view are shown. The locations of improvement in scar definition are indicated by the red and green arrows.

3.2 SRR Reconstruction Results

Compared to the interpolated SA volume, similar core zone size (percent to left ventricle volume) was identified from the SRR volume: $12.0 \pm 9.2\%$ vs $11.6 \pm 8.7\%$, $p = NS$ (absolute size: $17.2 \pm 16.7\text{cm}^3$ vs. $16.5 \pm 15.6\text{cm}^3$, $p = NS$) by the nonparametric Kolmogorov-Smirnov test, but reduced gray zone size was identified: $6.2 \pm 4.7\%$ vs $9.8 \pm 6.5\%$, $p < 0.05$ (absolute size: $8.4 \pm 5.2\text{cm}^3$ vs. $13.2 \pm 8.0\text{cm}^3$, $p < 0.05$). The results were in line with previous observation that low imaging resolution does not significantly influence the scar core zone estimation, but may lead to over-estimation of scar gray zone, since PVE is pronounced in the latter zone [3]. Fig. 2 a illustrates such phenomenon in 3D, where the reduction of gray zone is obvious in the through-plane direction. The increase of image resolution is further observed in all three orthogonal views, especially in the z-x and y-z plane (see red arrows in Fig. 2 b).

In the interpolated SA volume, the image gradients were anisotropic (Fig. 3 b). Subtle difference was further observed between the two in-plane gradients in the SA volume. The x -direction is smoother than the y -direction, as can be attributed to the non-symmetric acquisition matrix (256×104) during MR image acquisition: more frequency components were encoded in the y -direction than in the x -direction. In the SRR reconstructed volume, the gradient was more homogenous in all directions as a result of integrating two orthogonal views.

3.3 Electroanatomical Mapping

All patients underwent catheter ablation of ventricular arrhythmias. During the procedure, high-density electroanatomical voltage mapping (EAVM) was performed on the endocardial surface to measure the electrophysiological (EP) characteristics in and around the myocardial scar region. The amplitude of the bipolar voltage reflects the local EP activities *in vivo*. Previous clinical studies have validated the empirical voltage thresholds for myocardial tissue characterization: normal $v > 1.5 \text{ mV}$, scar gray zone $0.5 \text{ mV} < v < 1.5 \text{ mV}$, and scar

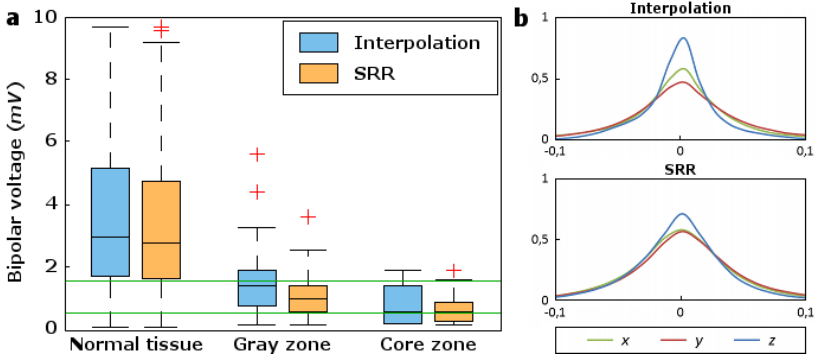


Fig. 3. Distribution of the bipolar voltages in identified normal tissue, scar gray zone, and core zone, from the interpolated SA and SRR volume (a). Green lines indicate 0.5 mV and 1.5 mV respectively, while the gray zone is supposed to have voltages between these two empirical thresholds. Distribution of image gradient in three directions (b).

core zone $v < 0.5\text{ mV}$. In absence of histological data, the EAVM measurement provides the best available gold standard for characterizing the underlying myocardial tissue composition. Registration between the EAVM and MRI data was performed by minimizing the point-to-surface distance under the assumption of a fixed heart orientation [11]. To evaluate the myocardial scar characterization from the SRR and the SA volumes, the acquired bipolar voltage was labeled according to their identified location on the endocardial surface. Fig. 3 a shows the box plot of the voltages in three regions: normal myocardium, scar gray zone, and scar core zone.

No significant difference was observed between the bipolar voltages in the normal tissue and scar core zone identified from the two volumes. The difference was however pronounced at the gray zone: the bipolar voltages from the gray zones of the interpolated SA volume were higher from those from the SRR reconstructed volume, $2.14 \pm 1.38\text{ mV}$ vs $1.07 \pm 0.66\text{ mV}$ with $p < 0.05$ by the nonparametric Kolmogorov-Smirnov test. This indicates that with the conventional SA volume, overestimation of scar gray zone can indeed occur due to PVE. In comparison, the identified gray zone from the SRR volume showed better agreement with clinical EAVM value (between the two horizontal lines of EAVM voltage), indicating improved gray zone characterization.

4 Conclusion

By integrating image information from the SA, the 2CH, and the 4CH LGE sequences, we are able to reconstruct a high-resolution myocardial scar representation. Our method uses the SRR framework to reconstruct the isotropic volume from the three anisotropic volumes. For compensation of the inter-scan motion of the heart, whose precise alignment is a crucial prerequisite for success

of the SRR reconstruction, we have developed a method that jointly estimates the optimal and consistent positioning of the volumes with respect of each other. An important property of this method is that it is very well suited for the MR data as it is based on the gradient information. The proposed method enables more accurate characterization of myocardial scar compared to conventional SA LGE volume, especially with respect to the clinically significant gray zone, allowing improved risk profiling and procedure planning in individual patients.

References

1. Yan, A.T., Shayne, A., Brown, K.A., Gupta, S., Chan, C., Luu, T., Carli, M.D., Reynolds, H.G., Stevenson, W.G., Kwong, R.Y.: Characterization of the perinfarct zone by contrast-enhanced cardiac magnetic resonance imaging is a powerful predictor of post-myocardial infarction mortality. *Circulation* 114, 32–39 (2006)
2. Kim, R.J., Fieno, D.S., Parrish, T.B., Harris, K., Chen, E.L., Simonetti, O., Bundy, J., Finn, J.P., Klocke, F.J., Judd, R.M.: Relationship of MRI delayed contrast enhancement to irreversible injury, infarct age, and contractile function. *Circulation* 100(19), 1992–2002 (1999)
3. Schelbert, E.B., Hsu, L.Y.Y., Anderson, S.A., Mohanty, B.D., Karim, S.M., Kellman, P., Aletras, A.H., Arai, A.E.: Late gadolinium-enhancement cardiac magnetic resonance identifies postinfarction myocardial fibrosis and the border zone at the near cellular level in ex vivo rat heart. *Circulation. Cardiovascular Imaging* 3(6), 743–752 (2010)
4. Goetti, R., Kozerke, S., Donati, O.F., Sürder, D., Stolzmann, P., Kaufmann, P.A., Lüscher, T.F., Corti, R., Manka, R.: Acute, subacute, and chronic myocardial infarction: Quantitative comparison of 2D and 3D late gadolinium enhancement MR imaging. *Radiology* 259, 704–711 (2011)
5. Shilling, R.Z., Robbie, T.Q., Bailloeu, T., Mewes, K., Mersereau, R.M., Brummer, M.E.: A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI. *IEEE Trans. Med. Imag.* 28(5), 633–644 (2009)
6. Ur Rahman, S., Wesarg, S.: Upsampling of cardiac MR images: Comparison of averaging and super-resolution for the combination of multiple views. In: ITAB 2010, University of Ioannina, pp. 1–4. IEEE Inc., New York (2010)
7. Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., Stathaki, T.: Robust FFT-based scale-invariant image registration with image gradients. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1899–1906 (2010)
8. Poot, D.H.J., Van Meir, V., Sijbers, J.: General and efficient super-resolution method for multi-slice MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 615–622. Springer, Heidelberg (2010)
9. Tao, Q., Milles, J., Zeppenfeld, K., Lamb, H., Bax, J., Reiber, J., van der Geest, R.J.: Automated segmentation of myocardial scar in late enhancement MRI using combined intensity and spatial information. *Magn. Reson. Med.* 64, 586–594 (2010)
10. Amado, L., Gerber, B., Gupta, S., Rettmann, D., Szar, G., Schock, R., Nasir, K., Kraitchman, D., Lima, J.: Accurate and objective infarct sizing by contrast-enhanced magnetic resonance imaging in a canine myocardial infarction model. *J. Am. Coll. Cardiol.* 12, 2383–2389 (2004)
11. Tao, Q., Milles, J., van Huls van Taxis, C., Lamb, H., Bax, J., Reiber, J., Zeppenfeld, K., van der Geest, R.J.: Toward magnetic resonance-guided electroanatomical voltage mapping for catheter ablation of scar-related ventricular tachycardia: A comparison of registration methods. *J. Cardiovasc. Electr.* 23, 74–80 (2012)

Direct Parametric Image Reconstruction of Rapid Multi-tracer PET

Xiaoyin Cheng¹, Nassir Navab², Sibylle I. Ziegler¹, and Kuangyu Shi¹

¹ Department of Nuclear Medicine, Technische Universität München, Germany

² Computer Aided Medical Procedures, Technische Universität München, Germany

Abstract. The separation of multiple PET tracers within an overlapped scan based on intrinsic difference of pharmacokinetics is challenging due to the limited SNR of PET measurements and high complexity of fitting models. This study developed a novel direct parametric reconstruction method by integrating a multi-tracer model with reduced number of fitting parameters into image reconstruction. To incorporate the multi-tracer model, we adopted EM surrogate functions for the optimization of the penalized log-likelihood. The algorithm was validated on realistic simulation phantoms and real rapid [¹⁸F]FDG and [¹⁸F]FLT PET imaging of mice with lymphoma mouse tumor. Both results have been compared with conventional methods and demonstrated evident improvements for the separation of multiple tracers.

1 Introduction

The complexity of tumor microenvironment reflects multiple physiological features, such as glycolysis, angiogenesis, proliferation, and hypoxia. These properties can be captured by positron-emission tomography (PET) imaging with different radiolabeled metabolites (tracers). The clinical value of combining multiple tracers in oncological detection, staging, localization and the consequent individualization of cancer therapy has been confirmed in several clinical studies [1]. For example, the combination of [¹⁸F]FDG and [¹⁸F]FLT has significantly improved the sensitivity and specificity in the diagnosis of lung nodules [2].

However, radioactive signals of different PET tracers cannot be physically differentiated [3] and the typical practice of multi-tracer PET imaging needs to wait for the full decay and clearance of each tracer, leading to scans in consecutive days. This imposes additional dose due to multiple CT scans on PET/CT as well as increasing labor and financial costs. The possible physiological and anatomical discrepancy due to separate scans may even reduce the expected clinical value. These limitations hamper a wider application of multi-tracer imaging.

Rapid multi-tracer PET imaging aims to differentiate physically identical signals of different tracers based on the intrinsic difference of their pharmacokinetics. This enables the acquisition of multiple tracers with overlaps. In practice, tracers are injected with a short interval of 10-15 mins and activities of different tracers are separated by fitting the superposed pharmacokinetic models of the

corresponding tracers [4,5]. However, this method suffers from the low signal-to-noise ratio (SNR) of the measured time-activity curves (TACs) and the high complexity of multi-tracer model. The poor quality of model fitting makes the separation of tracers unstable and failure prone. Previous efforts have been made to improve image reconstruction for better kinetic estimation [6], or to reduce the number of fitting parameters for improving stability [1,7].

One way to improve the pharmacokinetic estimation is to integrate the model fitting into reconstruction [8,9]. Direct parametric image reconstruction (DPIR) methods utilize the complete spatial and temporal information and highly improves SNR for reconstruction as well as kinetic model fitting. However, conventional DPIR is limited to simple models due to the increased complexity of the objective function for nonlinear model with high parameter dimension.

In this study, we propose a novel DPIR algorithm to reliably separate multi-tracer signals within a rapid overlapped scan. This is achieved by integrating a multi-tracer model with reduced number of parameters into reconstruction. EM surrogate functions are employed for the optimization of the penalized log-likelihood [10,11]. By incorporating the spatiotemporal consistency in model fitting, this method results in reconstructed parametric images with higher SNR and less statistical errors. We evaluated our algorithm both in realistic simulation phantoms and micro-PET scans of mice with lymphoma tumor using [^{18}F]FDG and [^{18}F]FLT. The results have been compared with conventional indirect methods and demonstrated improved quality for the separation of multiple tracers.

2 Methods

2.1 Multi-tracer Kinetic Model

A measured TAC $C(t)$ of an image voxel for a dynamic multi-tracer PET scan describes the temporal development of the mixed uptakes of L investigated tracers. It can be modeled as superposition of the L pharmacokinetic models [12]:

$$\hat{C}(t) = \theta_B \bar{C}_P + \sum_{l=1}^L [\theta_{Pl} \tilde{C}_{Pl} + \theta_{al} S_{al}(t; \nu_{al}, C_{Pl}) + \theta_{bl} S_{bl}(t; \nu_{al}, \nu_{bl}, C_{Pl})] \quad (1)$$

$$\left\{ \begin{array}{l} \bar{C}_P \triangleq \sum_{l=1}^L C_{Pl} \\ \tilde{C}_{Pl}(t) \triangleq \int_0^t e^{-(t-\tau)} C_{Pl}(\tau) d\tau \\ S_{al}(t; \nu_a, C_P(t)) \triangleq \int_0^t e^{-\nu_{al}(t-\tau)} C_{Pl}(\tau) d\tau \\ S_{bl}(t; \nu_a, \nu_b, C_P(t)) \triangleq \int_0^t e^{-\nu_{bl}(t-\tau)} - e^{-\nu_{al}(t-\tau)} C_{Pl}(\tau) d\tau \end{array} \right. \quad (2)$$

where $[\theta_B, \theta_{Pl}, \theta_{al}, \theta_{bl}]^T$ and $[\nu_{al}, \nu_{bl}]^T$ are intermediate parameters of tracer l ; $C_{Pl}(t)$ is the arterial input function (AIF); Intermediate parameters are combinations of conventional kinetic constants $\mathbf{k}_l = [K1_l, k2_l, k3_l, k4_l, \theta_B]^T$. θ_B is the fractional blood volume. The modeled $\hat{C}(t)$ can be expressed as:

$$\hat{C}(t) = \boldsymbol{\theta}^T \mathbf{y}(\boldsymbol{\nu}) \quad (3)$$

where $\boldsymbol{\theta}$ denotes the linear parameters $[\theta_B, \theta_{P1}, \theta_{a1}, \theta_{b1}, \dots, \theta_{PL}, \theta_{aL}, \theta_{bL}]^T$ and \mathbf{y} represents the nonlinear parts $[\bar{C}_P, \tilde{C}_{P1}, S_{a1}, S_{b1}, \dots, \tilde{C}_{PL}, S_{aL}, S_{bL}]^T$, which is a function of nonlinear parameters $\boldsymbol{\nu} = [\nu_{a1}, \nu_{b1}, \dots, \nu_{aL}, \nu_{bL}]$ and AIFs;

Reduced parameter space [7] can be then formulated through calculating the weighted sum square error (WSSE):

$$WSSE = \sum_{t=1}^{N_t} w_t (C_t - \hat{C}_t)^2 \quad (4)$$

where N_t is the number of acquisition time points; w_t denotes the weights of measurements [12]. To minimize Eqn. 4, we set $\partial WSSE / \partial \boldsymbol{\theta} = 0$ and derive:

$$\underbrace{\sum_t^{N_t} w_t \begin{bmatrix} (\bar{C}_{Pt})^2 & \tilde{C}_{P1t} \bar{C}_{Pt} & S_{a1t} \bar{C}_{Pt} & S_{b1t} \bar{C}_{Pt} & \dots \\ \cdot & (\tilde{C}_{P1t})^2 & S_{a1t} \tilde{C}_{P1t} & S_{b1t} \tilde{C}_{P1t} & \dots \\ \cdot & \cdot & (S_{a1t})^2 & S_{b1t} S_{a1t} & \dots \\ \cdot & \cdot & \cdot & (S_{b1t})^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} \theta_B \\ \theta_{P1} \\ \theta_{a1} \\ \theta_{b1} \\ \vdots \end{bmatrix}}_{\boldsymbol{\theta}} = \underbrace{\sum_t^{N_t} w_t C_t \begin{bmatrix} \bar{C}_{Pt} \\ \tilde{C}_{P1t} \\ S_{a1t} \\ S_{b1t} \\ \vdots \end{bmatrix}}_{\mathbf{x}} \quad (5)$$

where $\mathbf{M} \in \mathbb{R}^{(3L+1) \times (3L+1)}$ is a symmetric matrix.

Substituting the linear parameters $\boldsymbol{\theta}$ in Eqn. 3 with $\mathbf{M}^{-1} \mathbf{x}$, the multi-tracer model $\hat{C}(t)$ can be reduced with only non-linear parameters $\boldsymbol{\nu}$:

$$\hat{C} = [\mathbf{M}^{-1} \mathbf{x}]^T \mathbf{y} \quad (6)$$

Applying Eqn. 6 back in Eqn. 4, $\boldsymbol{\nu}$ can be estimated by minimizing WSSE using numerical algorithms. $\boldsymbol{\theta}$ is calculated analytically and the original kinetic parameters \mathbf{k} can be obtained from $\boldsymbol{\theta}$ and $\boldsymbol{\nu}$ [12].

2.2 Direct Parametric Reconstruction Using Optimization Transfer

The expectation of projected PET measurements \mathbf{y}_k in the k^{th} frame of sinogram can be expressed as an affine transform of the current image $x_k(\boldsymbol{\nu})$:

$$\bar{\mathbf{y}}_k(\boldsymbol{\nu}) = \mathbf{P} \mathbf{x}_k(\boldsymbol{\nu}) + \mathbf{r}_k + \mathbf{s}_k \quad (7)$$

$\mathbf{P} \in \mathbb{R}^{N_i \times N_j}$ is the system matrix where $p_{i,j}$ measures the probability that an event emits from the j^{th} voxel being detected by the i^{th} detector pair; N_j and N_i the total number of image voxels and line-of-responses (LORs); \mathbf{r} and \mathbf{s} are randoms and scatters respectively. The relationship between image intensity $x_k(\boldsymbol{\nu}_j)$ and activity concentration $C(t; \boldsymbol{\nu}_j)$ of voxel j at frame k is given by:

$$x_k(\boldsymbol{\nu}_j) = \int_{t,s}^{t,e} C(\tau, \boldsymbol{\nu}_j) d\tau \quad (8)$$

where t, s and t, e are the start and end times of frame k .

To estimate the parametric images $\boldsymbol{\nu}$ from \mathbf{y} directly, we are seeking the $\hat{\boldsymbol{\nu}}$ that maximizes the following penalized log-likelihood function $\Phi(\boldsymbol{\nu})$:

$$\hat{\boldsymbol{\nu}} = \arg \max_{\boldsymbol{\nu}} \Phi(\boldsymbol{\nu}) = \arg \max_{\boldsymbol{\nu}} \{L(\mathbf{y}|\boldsymbol{\nu}) - \beta U(\boldsymbol{\nu})\} \quad (9)$$

$$L(\mathbf{y}|\boldsymbol{\nu}) = \sum_{k=1}^{N_k} \sum_{i=1}^{N_i} y_{ik} \log \bar{y}_{ik}(\boldsymbol{\nu}) - \bar{y}_{ik}(\boldsymbol{\nu}) \quad (10)$$

where $L(\mathbf{y}|\boldsymbol{\nu})$ is the log-likelihood function with respect to $\boldsymbol{\nu}$ [8]; N_k the total number of time frames; $U(\boldsymbol{\nu})$ the smoothness penalty term and β the regularization parameter controlling the tradeoff between image resolution and noise. Here a common quadratic penalty is used as described in [11].

To include multi-tracer model Eqn. 6 into reconstruction, maximization of Eqn. 9 is achieved by in each iteration optimizing Expectation Maximization (EM) surrogate functions Q_L , Q_U designed for $L(\mathbf{y}|\boldsymbol{\nu})$ and $U(\boldsymbol{\nu})$ [10,11]:

$$\left\{ \begin{array}{l} Q_L(\boldsymbol{\nu}; \boldsymbol{\nu}^{(n)}) = \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} p_{ij} \left(\sum_{k=1}^{N_k} \hat{x}_{jk}^{(n)} \log x_k(\boldsymbol{\nu}_j) - x_k(\boldsymbol{\nu}_j) \right) \\ Q_U(\boldsymbol{\nu}; \boldsymbol{\nu}^{(n)}) = \frac{1}{2} \sum_{k=1}^{N_k} \sum_{j=1}^{N_j} \sum_{l \in \mathcal{N}_j} \alpha_{jl} \left(\hat{x}_{jk}^{reg,(n)} - x_k(\boldsymbol{\nu}_j) \right)^2 \\ \hat{x}_{jk}^{(n)} = \frac{x_k(\boldsymbol{\nu}_j^{(n)})}{\sum_{i=1}^{N_i} p_{ij}} \sum_{i=1}^{N_i} p_{ij} \frac{y_{ik}}{\bar{y}_{ik}(\boldsymbol{\nu}^{(n)})} \\ \hat{x}_{jk}^{reg,(n)} = \frac{1}{2 \sum_{l \in \mathcal{N}_j} \alpha_{jl}} \sum_{l \in \mathcal{N}_j} \alpha_{jl} \left(x_k(\boldsymbol{\nu}_j^{(n)}) + x_k(\boldsymbol{\nu}_l^{(n)}) \right) \end{array} \right. \quad (11)$$

where \mathcal{N}_j is the set of neighborhood voxels centering around voxel j ; α_{jl} the weighting factors set to be the inverse of distance between voxels j and l ; $\hat{\mathbf{x}}^{(n)}$ and $\hat{\mathbf{x}}^{reg,(n)}$ are intermediate reconstructed images and smoothed images respectively at the n^{th} iteration.

The overall surrogate function is then $Q \triangleq Q_L - \beta Q_U$. Thus the maximization of Eqn. 9 is transferred to maximize $Q(\boldsymbol{\nu}; \boldsymbol{\nu}^{(n)})$. As both Q_L and Q_U are separable for voxels, the maximization can be further transferred into a voxel-wise scale:

$$\boldsymbol{\nu}_j^{(n+1)} = \arg \max_{\boldsymbol{\nu}_j} q(\boldsymbol{\nu}_j; \boldsymbol{\nu}_j^{(n)}) = \arg \max_{\boldsymbol{\nu}_j} q_L(\boldsymbol{\nu}_j; \boldsymbol{\nu}_j^{(n)}) - \beta q_U(\boldsymbol{\nu}_j; \boldsymbol{\nu}_j^{(n)}) \quad (12)$$

Eqn. 12 resembles a 1D curve fitting problem. Here we used the modified Levenberg-Marquardt (LM) algorithm as described in [11] and combined with the Hooke and Jeeves pattern search.

Overall, the proposed algorithm consists of three main steps. Given $\boldsymbol{\nu}^{(0)}$, the dynamic dual-tracer images \mathbf{x} is initialized in Eqn. 8. Each iteration consists of a frame-wise reconstruction to obtain Q_L , a frame-wise regularization to calculate Q_U in case of $\beta > 0$, and a voxel-wise fitting step to update $\boldsymbol{\nu}$. As loops end, linear parameters $\boldsymbol{\theta}$ are calculated from Eqn. 5 analytically. Separated dynamic images of each tracer are retrieved from corresponding pharmacokinetic models.

3 Results

To validate the algorithm, a realistic phantom was generated using computational simulations following the procedure in [11]. The combination of $[^{18}\text{F}]\text{FDG}$ and $[^{18}\text{F}]\text{FET}$ were simulated with an interval of 15 mins for 45 minutes of 65 frames on the Zubal brain phantom, which includes gray matter (2 types), white matter, caudate, putamen, skin, skull, thalamus and an additional tumor. The pharmacokinetic parameters of $[^{18}\text{F}]\text{FDG}$ and $[^{18}\text{F}]\text{FET}$ were derived from clinic data. To test the separation of tracers, exclusive uptake was assumed on the tumor area and therefore there is only contrast in FET and the tumor is not differentiable in FDG. The PET measurements were simulated based on the geometry of real clinical scanner (FWHM 4.0mm) with attenuation, scattering and random effects. Poisson noise was added to each generated sinogram bin.

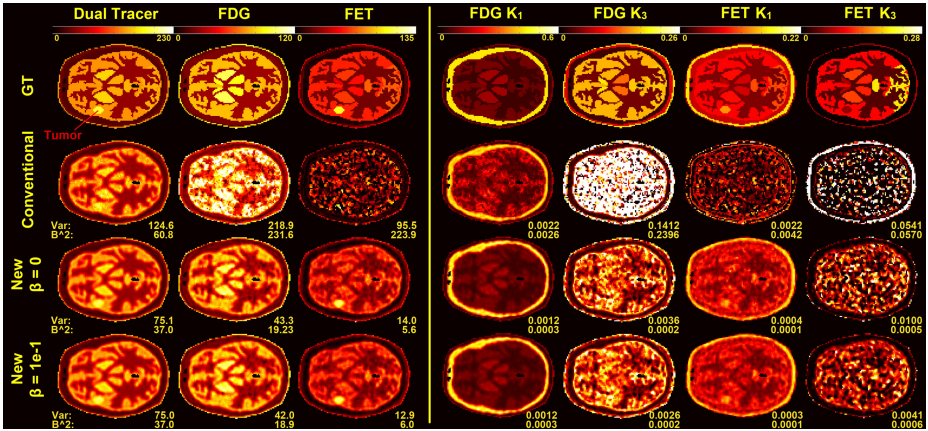


Fig. 1. The results of the separated tracer uptakes and the kinetic parametric images using the conventional and the proposed method on a phantom

Fifty datasets were simulated with different noise realizations. The separation of two tracers were tested with the proposed algorithm without regularization and the conventional indirect method (kinetic modeling after image reconstruction). For both methods, 128 iterations were used for EM reconstruction to keep the balance between bias and noise. For the conventional method, kinetic parameters were estimated using nonlinear least square regression. For fair comparisons, uniform weights were assigned here for the proposed method. Initial values of all parameters in both methods were set to 0.001 for all voxels. The lower and upper bounds were set to 0 and 2.0 respectively. The last frame ($t=45$ min) of the overlapped and separated tracer uptake maps as well as the parametric images were compared with the ground truth (GT) as shown in Fig. 1. The conventional method was able to recover some voxels of single tracers but it was in general unstable and noisy. With the proposed method, parametric

images and individual tracer uptakes were well recovered with relative less mean bias (B^2) and variations over all datasets. Tumor is clearly visible in the recovered FET images and no contrast in FDG. Although k_3 is sensitive to noise, the new method has better result than the conventional method. Overall, the new algorithm has reduced up to 87.8% variance and 85.7% bias for the separated tracer uptake maps and up to 97.8% variance and 98.4% bias for the parametric images.

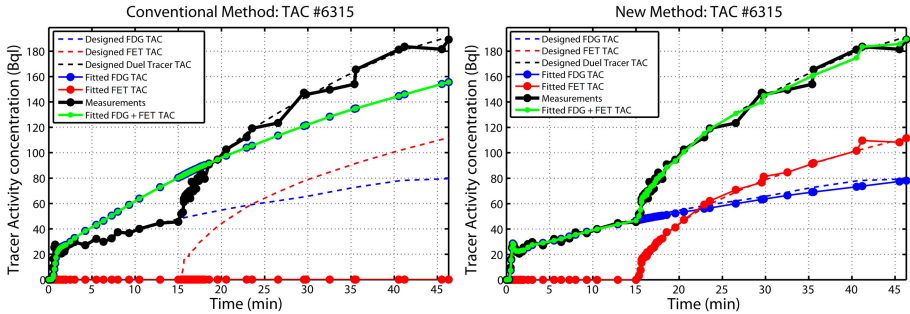


Fig. 2. The fitting results of the conventional (left) and the proposed method (right) for an example TAC in the tumor area

The effect of the regularization term β of the proposed algorithm was further evaluated on the phantom. The total square bias and variance of intermediate parameter ν with β from 0 to $4e - 1$ were computed and the variance-bias² curves were plotted in Fig. 3. An optimized $\beta = 1e - 1$ was chosen for the tradeoff between bias and variance for ν^{FET} . The results were shown in Fig. 1. With the regularization, the variance can be further reduced up to 1.1% for the separated tracer uptake and 10.1% for the parametric images.

The conventional method failed to separate the tracer uptake in the tumor area. An example TAC is shown in Fig. 2. The left plot demonstrates the fitting result of the conventional method. It falsely ended up with a zero uptake for FET, which conflicts with the expectation. The right plot shows the result for the proposed algorithm, the temporal information of each tracer were well recovered and reflected the true designed TACs.

We further evaluated the proposed algorithm with the real micro-PET data. Four mice with SUDHL-1 tumor model were scanned for 70 min after injection of approximately 6 MBq $[^{18}\text{F}]\text{FDG}$ and $[^{18}\text{F}]\text{FLT}$ with an interval of 10 min. The dual-tracer AIF was derived from left ventricle data at early time and 2 venous blood samples in the end of the scan [13]. Two single tracer AIFs were separated by fitting the combined AIF models [14]. This lymphoma tumor model has been reported to have higher specificity for FLT [15]. Thus the FLT image has higher tumor to muscle contrast than FDG images. The better consistency of the results

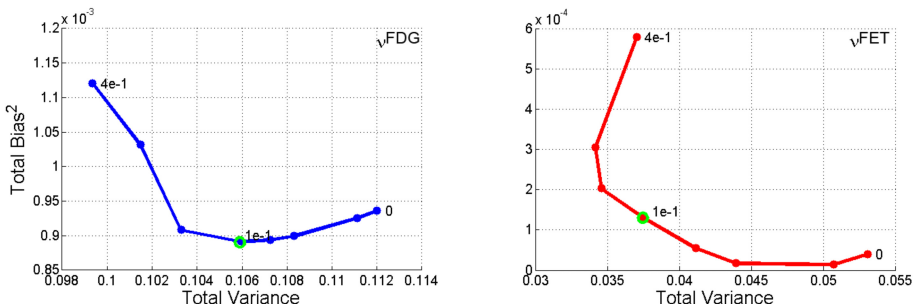


Fig. 3. Total squared bias v.s. variance tradeoff of ν images

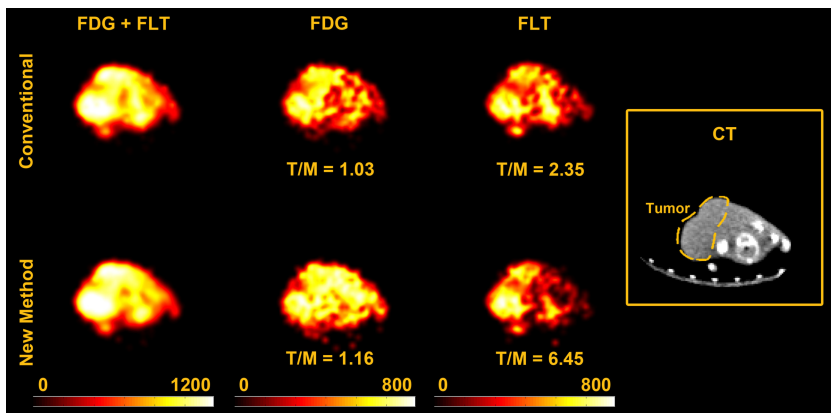


Fig. 4. An example of the results of the conventional and the proposed methods for a mouse with lymphoma tumor model scanned using $[^{18}\text{F}]\text{FDG}$ and $[^{18}\text{F}]\text{FLT}$ PET

of the new method (an example shown in Fig. 4) with the know properties of the tumor model confirms further that the new method can separate the two tracers better.

4 Conclusion

This study proposed a novel direct rapid multi-tracer PET reconstruction algorithm which can robustly retrieve single tracer images from overlapped acquisitions. In particular, we integrated a multi-tracer model with reduced number of parameters into parametric image reconstruction and incorporate EM surrogate functions for the optimization of the penalized log-likelihood. Both phantom and real data has proved that this new algorithm can separate the overlapped signals significantly better than the conventional methods. Further improvements may be done by smoothing noisy TAC for better model fitting concerning low

measurement statistics of dynamic PET acquisitions. The reliable and less noisy separation of the overlapped PET signal using the proposed method demonstrated a great potential to promote the application of multi-tracer in clinical practice, which could influence the state of the art of the individualized medicine.

Acknowledgement. DFG SFB 824 Z3, BFS, Zhoulei Li (providing tumor cell line), Sybille Reder (technical assistant) and Dr. Guobao Wang (discussion).

References

1. Kadrmas, D.J., Rust, T.C., Hoffman, J.M.: Single-scan dual-tracer FLT+FDG PET tumor characterization. *Phys. Med. Biol.* 58, 429–449 (2013)
2. Tian, J., Yang, X., Yu, L., Chen, P., et al.: A multicenter clinical trial on the diagnostic value of dual-tracer PET/CT in pulmonary lesions using 3'-deoxy-3'-18F-fluorothymidine and 18F-FDG. *J. Nucl. Med.* 49, 186–194 (2008)
3. Andreyev, A., Celler, A.: Dual-isotope PET using positron-gamma emitters. *Phys. Med. Biol.* 56, 4539–4556 (2011)
4. Kadrmas, D., Rust, T.: Feasibility of rapid multitracer PET tumor imaging. *IEEE Trans. Nucl. Sci.* 52, 1341–1347 (2005)
5. Verhaeghe, J., Reader, A.J.: Accelerated PET water activation acquisition with signal separation methodology. *Med. Phys.* 40 (2013)
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. Oktay, M., Kadrmas, D.: Reduced parameter space formulations for fast and robust kinetic modeling. *SNM* 53, 2292 (2012)
8. Matthews, J., Bailey, D., Price, P., et al.: The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction. *Phys. Med. Biol.* 42, 1155–1173 (1997)
9. Yan, J., Planeta-Wilson, B., Gallezot, J.D., et al.: Initial evaluation of direct 4D parametric reconstruction with human PET data. *IEEE NSS* 1, 2503–2506 (2009)
10. De Pierro, A.: A modified EM algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. Med. Imaging* 14, 132–137 (1995)
11. Wang, G., Qi, J.: An optimization transfer algorithm for nonlinear parametric image reconstruction from dynamic PET data. *IEEE Trans. Med. Imaging* 31, 1977–1988 (2012)
12. Gunn, R.N., Gunn, S.R., Cunningham, V.J.: Positron emission tomography compartmental models. *J. Cereb. Blood. Flow. Metab.* 21, 635–652 (2001)
13. Ferl, G.Z., Zhang, X., Wu, H.M., et al.: Estimation of the 18F-FDG input function in mice by use of dynamic small-animal PET and minimal blood sample data. *J. Nucl. Med.* 48, 2037–2045 (2007)
14. Feng, D., Huang, S.C., Wang, X.: Models for computer simulation studies of input functions for tracer kinetic modeling with PET. *Int. J. Biomed. Comput.* 32, 95–110 (1993)
15. Li, Z., Graf, N., Herrmann, K., Junger, et al.: FLT-PET is superior to FDG-PET for very early response prediction in NPM-ALK-positive lymphoma treated with targeted therapy. *Cancer. Res.* 72, 5014–5024 (2012)

Non-rigid 2D-3D Medical Image Registration Using Markov Random Fields

Enzo Ferrante and Nikos Paragios

Center for Visual Computing, Ecole Centrale de Paris, France
{enzo.ferrante, nikos.paragios}@ecp.fr
<http://cvc.centrale-ponts.fr/>

Abstract. The aim of this paper is to propose a novel mapping algorithm between 2D images and a 3D volume seeking simultaneously a linear plane transformation and an in-plane dense deformation. We adopt a metric free locally over-parametrized graphical model that combines linear and deformable parameters within a coupled formulation on a 5-dimensional space. Image similarity is encoded in singleton terms, while geometric linear consistency of the solution (common/single plane) and in-plane deformations smoothness are modeled in a pair-wise term. The robustness of the method and its promising results with respect to the state of the art demonstrate the extreme potential of this approach.

Keywords: 2D-3D registration, medical imaging, markov random fields, discrete optimization.

1 Introduction

2D-3D image registration is an important problem in medical imaging and it can be applied in multiple medical procedures. Depending on the technology used to capture the 2D image, it can be a projective (e.g. X-Ray) or sliced (e.g. Ultrasound) image; in this work we focused on sliced images. Image guided surgeries, as laparoscopic or endoscopic [1], and brain surgeries [2] use such images. In those surgeries, pre-operative 3D images (e.g. Computed Tomography (CT) or Magnetic Resonance Images (MRI)) and intra-operative 2D images are used to guide surgeons during the procedure. 2D-3D registration plays a key role in this process because it allows doctors to guide surgery using the 3D pre-operative high resolution annotated data. Tissue shift, as well as breathing and heart motion during the surgery, causes elastic deformation in the images and makes the registration process an extremely challenging problem.

Several methods to deal with slice-to-volume registration have been proposed. [1] proposes a method to register endoscopic and laparoscopic ultrasound (US) images with pre-operative computed tomography volumes that potentially could work in real time. It is based on a new phase correlation technique called LEPART and it manages only rigid registration. [3] presents a flexible framework for intensity based slice-to-volume nonrigid registration algorithms used to register histological section images to human brain MRI. A feature based method that performs slice to volume registration is presented in [4]. It uses several slices in order to improve the quality of the results. [5] tracks intra-operative MRI slices of prostate images with a pre-operative MRI volume.

This monomodal registration is designed to provide patient tracking information for prostate biopsy performed under MR guidance.

Discrete optimization of Markov Random Fields (MRFs) has been widely used to solve the problem of non-rigid image registration in recent years [6, 7]. However, to the best of our knowledge, most of those works focus on 2D-2D or 3D-3D registration instead of 2D-3D registration. [8] presents a method based on MRFs to perform 2D-3D registration, but it estimates only rigid transformations and works with projective images.

The main contributions of this paper consists of a local pair-wise method to register 2D and 3D images using MRFs and discrete optimization techniques capable of capturing the plane and the in-plane dense deformations. It is intensity based and independent of the metric that is being used, so it can be adapted to different image modalities.

The remainder of the paper is organized as follows: the MRF based formulation of 2D-3D deformable registration is described in Section 2. Section 3 provides the validation tests and results, while Section 4 concludes our paper and provides some ideas on relevant future directions.

2 2D-3D Non-rigid Registration Using MRFs

The problem of non-rigid 2D-3D image registration can be formulated as an optimization problem. Given a 2D source image I and a 3D target volume J , we seek the 2D-2D in-plane local deformation field \hat{T}_D and the plane $\hat{\pi}[J]$ (i.e. a bi-dimensional slice from the volume J) which in the most general case minimize the following objective function:

$$\hat{T}_D, \hat{\pi} = \underset{T_D, \pi}{\operatorname{argmin}} \mathcal{M}(I \circ T_D(\mathbf{x}), \pi[J](\mathbf{x})) + \mathcal{R}(T_D, \pi), \quad (1)$$

where \mathcal{M} represents the data term and \mathcal{R} the regularization term. The data term \mathcal{M} measures the matching quality between the deformed 2D source image and the corresponding 3D slice. The regularization term \mathcal{R} imposes certain constraints on the final solution that can be used to render the problem well posed and imposes certain expected geometric properties on the deformation field. The plane, $\hat{\pi}$, that minimizes the equation indicates the location of the 3D volume slice that is most similar to the deformed source image. The deformation field \hat{T}_D represents the in-plane deformations that must be applied to the source image in order to minimize the energy function.

Our MRF based formulation of the 2D-3D non-rigid registration problem consists of an undirected pair-wise graph $G = \langle V, E \rangle$ super-imposed to the 2D image with a set of nodes V and a set of cliques E . The nodes (a regular lattice) are interpreted as control points of a bi-dimensional quasi-planar grid that models at the same time the in-plane deformations and the current position of the 2D image into the 3D volume. In order to represent the in-plane deformations, the grid is interpreted as a Free Form Deformation model (FFD) where each control point has local influence on the deformation.

Such a coupled estimation problem can be expressed using graphical models of varying complexity. The most natural one is to consider a two layer graph, one modeling the global linear mapping as done in [8] and another modeling the in-plane deformation as done in [6] where interconnection between them will produce coupling and global

consistency on the obtained solution. Despite the theoretical soundness of such an approach, it is not suitable simply because the linear mapping is a global variable. This is due to the fact that global linear mapping variables are to be associated with all control nodes resulting in a densely connected graph, while at the same time the parameter space is high-dimensional and the associated variables are not bounded resulting in rather complex discrete sampling requirements. The aforementioned limitations make such an approach problematic in terms of the quality of the obtained solution (non-submodular terms) and inefficient from a computational view point.

We propose to overcome these limitations through a local pair-wise over-parameterized graphical model. In our formulation, the energy is formed by data terms $\mathbf{g} = \{g_i(\cdot)\}$ (unary potentials) associated to each graph vertex and regularization terms $\mathbf{f} = \{f_{ij}(\cdot, \cdot)\}$ (pairwise potentials) associated to the edges. The first ones are typically used for encoding some sort of data likelihood, whereas the later ones act as regularizers and thus play an important role in obtaining high-quality results [6]. The minimization energy problem in the context of a discrete MRF is thus defined as:

$$MRF(\mathbf{g}, \mathbf{f}) = \min \sum_{i \in V} g_i(\mathbf{u}_i) + \sum_{(i,j) \in E} f_{ij}(\mathbf{u}_i, \mathbf{u}_j), \quad (2)$$

where $\mathbf{u}_i, \mathbf{u}_j \in L$ are the labels assigned to the vertices $v_i, v_j \in V$ respectively.

Vertices $v_i \in V$ are moved by assigning them different labels $\mathbf{u}_i \in L$ (where L is the label space) until an optimal position is found. To reach such an optimal position, we need to define an energy term that will be minimized using an optimization algorithm. We adopt the FastPD algorithm [9] for the optimization of the aforementioned MRF due to good trade-off between complexity and quality of the obtained minimum in the context of non-submodular MRFs (this is our case due to the definition of the pair-wise terms).

Label Space. It includes all the possible values that a vertex label can take to deform the graph. Our label space L consists of 5-tuples $\mathbf{u} = (d_x, d_y, d_z, \phi, \theta)$, where the first three parameters define a displacement vector \mathbf{d}_i in the cartesian coordinate system, and the angles (ϕ, θ) define a vector \mathbf{N}_i over a unit sphere, expressed using spherical coordinates. Let us say we have a control point $\mathbf{p}_i^t = (p_{xi}^t, p_{yi}^t, p_{zi}^t)$ at optimization step t and we assign the label $\mathbf{u}_i = (d_{xi}, d_{yi}, d_{zi}, \phi_i, \theta_i)$ to this point. So, the new point position at optimization step $t + 1$ will be calculated using the displacement vector resulting in $\mathbf{p}_i^{t+1} = (p_{xi}^t + d_{xi}, p_{yi}^t + d_{yi}, p_{zi}^t + d_{zi})$. Additionally, we define a plane π_i containing the displaced control point \mathbf{p}_i^{t+1} and whose unit normal vector (expressed in spherical coordinates and with constant radius $r = 1$) is $\mathbf{N}_i = (\phi_i, \theta_i)$. One of the most important constraints to be considered is that our graph should have a quasi-planar structure, i.e. it should be similar to a plane; the plane π_i associated to every control point \mathbf{p}_i will be used by the energy term to take this constraint into account.

Unary Potentials. The formulation of the unary potentials that we propose is independent of the similarity measure. It is calculated for each control point given any intensity based metric δ capable of measuring the similarity between two bi-dimensional images

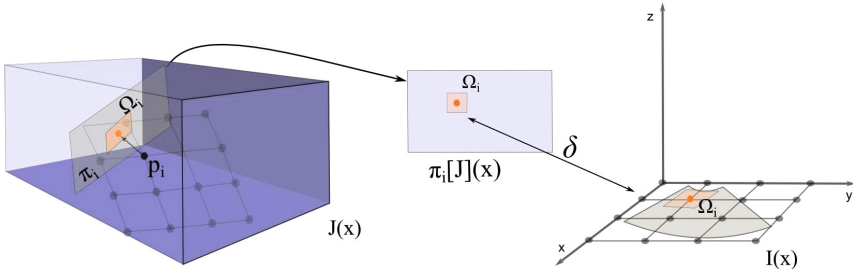


Fig. 1. Data term (Unary potential). The points $\mathbf{x} \in \Omega_i$ are used to calculate the unary potential. $\pi_i[J](\mathbf{x})$ returns the intensity of the point in the 2D slice corresponding to the plane π_i in the 3D image, whereas $I(\mathbf{x})$ returns the 2D image intensity. δ represents the similarity measure.

(e.g sum of absolute differences, mutual information, normalized cross correlation). This calculation is done for each control point \mathbf{p}_i , using its associated plane π_i in the target image J and the source 2D image I . An oriented patch Ω_i over the plane π_i (centered in \mathbf{p}_i) is extracted from the volume J , so that the metric δ can be calculated between that patch and the corresponding area over the source 2D image (see Figure 1):

$$g_i(\mathbf{u}_i) = \int_{\Omega_i} \delta(I(\mathbf{x}) - \pi_i[J](\mathbf{x}))d\mathbf{x}. \tag{3}$$

In monomodal scenarios, where two images of the same modality are compared, the simplest and the most used similarity measure is the Sum of Absolute Differences (SAD). In multimodal scenarios, where different modalities are compared (e.g. CT with US images), statistical similarity measures such as Mutual Information (MI) are generally used since we can not assume that corresponding objects have the same intensities in the two images. So, depending on the type of image that we want to register, the framework can be adapted using any similarity measure defined over two bi-dimensional images.

Pairwise Potentials. Generally, these terms are used to encode the regularization of the displacement field. In our formulation, the pairwise potentials are defined by two terms: the first one (F_1) controls the grid deformation assuming that it is a plane, whereas the second one (F_2) maintains the plane structure of the mesh. Those terms are weighted by a coefficient $\alpha \in [0, 1]$ resulting in $f_{ij}(\mathbf{u}_i, \mathbf{u}_j) = \alpha F_1(\mathbf{u}_i, \mathbf{u}_j) + (1 - \alpha)F_2(\mathbf{u}_i, \mathbf{u}_j)$.

The in-plane grid deformation is thus controlled using a distance preserving approach: it tries to preserve the original distance between the control points of the grid. Since this metric is based on the euclidean distance between the points, it assumes that they are coplanar. So, the equation that regularizes the in-plane deformations is:

$$F_1(\mathbf{u}_i, \mathbf{u}_j) = 1 - \frac{\|(\mathbf{p}_i + \mathbf{d}_i) - (\mathbf{p}_j + \mathbf{d}_j)\|}{\|(\mathbf{p}_{o,i}) - (\mathbf{p}_{o,j})\|}, \tag{4}$$

where $\mathbf{p}_{o,i}$ and $\mathbf{p}_{o,j}$ are the original positions of the control points. Regarding the plane preservation regularization term, it penalizes the average distance between the control

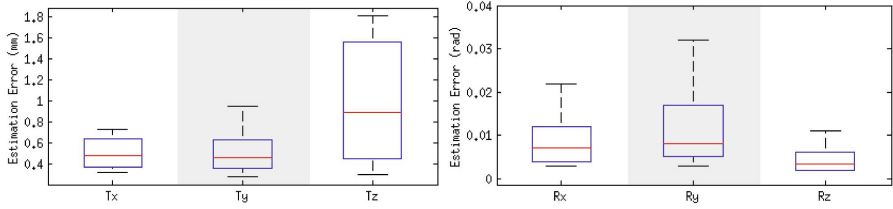


Fig. 2. Average of absolute differences between the ground truth and estimated plane translation (T_x , T_y and T_z) and rotation (R_x , R_y and R_z) parameters for 10 sequences of 20 images each one. The average error is less than $0.013rad$ for rotations and less than $1mm$ for translations.

points and the plane corresponding to the neighboring one. The aim is to maintain the quasi-planar structure of the grid. Given that the distance between a point and a plane is zero when the point is contained in the plane, this term will be minimum when both of the control points are on the same plane. The term F_2 is then defined using the distance between a point $\mathbf{p} = (p_x, p_y, p_z)$ and a plane π given by $D_\pi(\mathbf{p})$. So, we calculate the average of $D_{\pi_j}(\mathbf{p}_i + \mathbf{d}_i)$ and $D_{\pi_i}(\mathbf{p}_j + \mathbf{d}_j)$:

$$F_2(\mathbf{u}_i, \mathbf{u}_j) = \frac{1}{2}(D_{\pi_j}(\mathbf{p}_i + \mathbf{d}_i) + D_{\pi_i}(\mathbf{p}_j + \mathbf{d}_j)). \quad (5)$$

3 Validation and Results Discussion

A dataset was created in order to validate both the resulting 2D-2D deformation field and the final plane location using a temporal series of 3D heart MRI. The monomodal dataset consists of a temporal series of twenty 2D slices, I_i , each one extracted from its corresponding volume M_i . Starting from a random initial translation $T_0 = (T_{x_0}, T_{y_0}, T_{z_0})$ and rotation $R_0 = (R_{x_0}, R_{y_0}, R_{z_0})$, we extracted a 2D slice I_0 from the initial volume M_0 . Gaussian noise was added to every parameter in order to generate the position used to extract the next slice from the next volume. We used $\sigma_r = 3^\circ$ for the rotation and $\sigma_t = 5mm$ for the translation parameters. Those parameters generate maximum distances of around 25mm between the current and its succeeding plane. In that way, we generated a series that corresponds to a trajectory into the volume. Since the series consists of temporally spaced volumes of the heart, there are local deformations between them due to the heartbeat.

We tested it over 10 sequences of 20 images to validate the estimated plane locations, giving a total of 200 registration cases, using SAD similarity measure. The MRI resolution was $192 \times 192 \times 11$ voxels and the voxel size was $1.25 \times 1.25 \times 8mm^3$. The resulting position of the slice I_i was used to initialize the slice I_{i+1} . As shown in Figure 2, the average error is less than $0.013rad$ (0.74°) for rotation and less than $1mm$ for translation parameters. Given that the image resolution in z axis is lower than in x and y , we can recognize a bigger error in the estimated translation for z coordinate.

In order to validate the in-plane deformations, we created a set of manual segmentations \mathcal{S} of the left endocardium from a set of 20 slices (extracted from the volumes M_i).

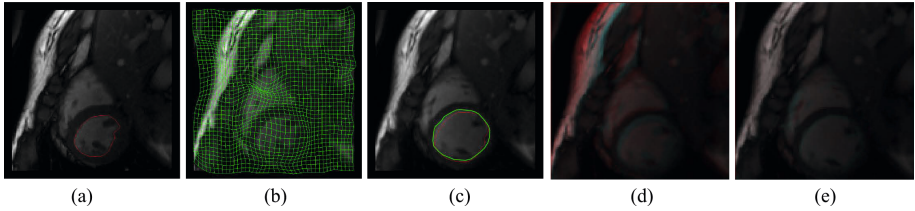


Fig. 3. One of the slices used in the validation test of the in-plane deformations: (a) Source image. The red line is the initial segmentation of the left endocardium. (b) Deformation field obtained after the registration process. (c) Slice corresponding to the estimated plane and extracted from the target volume M_0 . The red line corresponds to the deformed initial segmentation and the green line is the ground truth. (d) Overlapping between source image (red) and target initial plane (cyan). (e) Overlapping between deformed source image (red) and estimated plane (cyan).

Each slice was registered with the volume M_0 starting from a random position around the ground truth. Positions were generated adding gaussian noise with $\sigma_r = 4.5^\circ$ and $\sigma_t = 5mm$ to every translation (T_x, T_y, T_z) and rotation (R_x, R_y, R_z) parameters respectively. Those parameters generate maximum distances of about 25mm between the initialization and the ground truth. The estimated deformation field T_{D_i} was applied to the corresponding initial segmentation $s_i \in \mathcal{S}$ and it was compared with the ground truth using DICE coefficient. The average DICE coefficient for the 20 test cases was 0.93 and the average distance between the initial and the estimated parameters was $\hat{R} = (0.011, 0.007, 0.003)rad$, $\hat{T} = (0.503, 0.302, 0.578)mm$, showing that our method can capture in-plane deformations at the same time as it looks for the optimal plane location (see Figure 3). In all these cases (both plane estimation and in-plane deformation tests) we use 13122 labels, $\alpha = 0.9$, 3 levels of grid refinement and final grid resolution of 16×16 nodes; the execution time was about 4min for every case. Another dataset was used to test our approach over multimodal registration. Since it was performed over images of different modalities, we used Mutual Information as similarity measure instead of SAD. The dataset consists of a preoperative brain MRI volume (voxel size of $0.5 \times 0.5 \times 0.5mm^3$ and resolution of $394 \times 466 \times 378$ voxels) and 6 series of 10 US images extracted from the patient 01 of the database MNI BITE presented in [2]. The intraoperative US images were acquired using the prototype neuronavigation system IBIS NeuroNav. We generated 6 different sequences of 10 2D US images of the brain ventricles, with a size of $48 \times 38mm$ and resolution of $0.3 \times 0.3mm$. The ventricles were manually segmented in both modalities. The estimated position of the slice n was used to initialize the registration process of slice $n + 1$. Slice 0 was initialized in a position near to the Ground Truth using the rigid transformation provided together with the dataset. The DICE coefficient and Contour Mean Distance (CMD) were calculated before and after registration. Figure 4 summarizes the average DICE and CMD coefficients for every series. It shows that the DICE increases after the registration process an average of 0.05 while CMD decreases an average of $0.4mm$. Note that average DICE coefficients are always greater than 0.7. Given that we are dealing with highly challenging images of low resolution being heavily corrupted from speckle, those results are extremely promising. In all the registration cases an initial grid size of

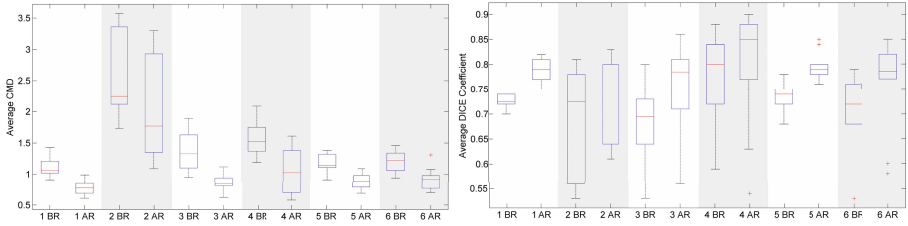


Fig. 4. Average DICE Coefficient (a) and Average CMD (Contour Mean Distance) (b) of all the slices, for every brain series, before (BR) and after (AR) registration process. The average DICE increment after registration of all the series is 0.05 and the average CMD decrement is 0.4mm.

4mm, 6174 labels, $\alpha = 0.9$, 3 grid levels and 4 iterations of the optimization process were used, giving an average time of around 10min per case.

It is also important to analyze results presented by other authors performing slice-to-volume registration. [5] worked with monomodal registration of prostate MRI images and reported average target registration errors below 1mm. [1] tackled the problem of multimodal registration (US and CT images) reporting results with an error of 1.56 ± 0.78 mm, using initializations with uniformly random shifts in the range -5 to 5mm. Both of them model only the rigid transformations without taking into account the in-plane deformations. Our method achieved results below 1mm in case of monomodal registration and it maintained the DICE coefficient greater than 0.7 in case of multimodal registration, but was able to deal with the in-plane deformations.

4 Conclusion

In this paper we proposed a novel method for deformable 2D-3D registration using a single shot optimization method that involves plane selection and in plane dense-deformation. This was achieved through an over-parameterized graphical model (5-dimensional representation) that is metric free, can cope with arbitrary deformation models and encodes different in-plane regularization constraints. Clinical validation using real scenarios and examples where 2D acquisitions have been simulated demonstrated the potentials of our method, proved its efficiency in terms of precision and, compared to other methods that tackle a similar problem (like [1] and [5]), seems to achieve state of the art results.

The proposed formulation from theoretical view point inherits two limitations, one related with the dimensionality of the label space (that can be handled due to the limited 2D grid size), and a second related with the coplanarity constraint that is approximately imposed through the suggested over-parameterization. Both of them could be alleviated through the use of third-order potentials. Given a 3D label deformation space and triples of neighborhood control points, the plane and the in-plane deformation can be automatically determined. This will introduce certain additional complexity in terms of optimization that could be easily dealt with higher-order to pair-wise MRF reduction methods [10] or dual decomposition [11]. Such an approach will be more precise in terms of data/regularization term definition and of comparable complexity.

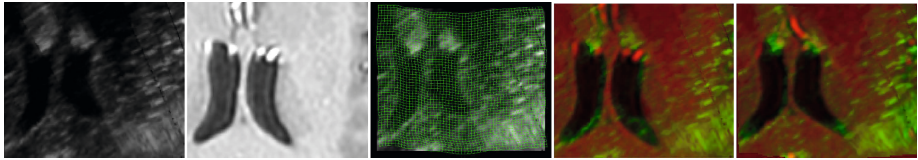


Fig. 5. Registration of a 2D US image and a MRI scan of the brain. (a) 2D US source image. (b) Slice extracted from the MRI corresponding to the initial position of the plane. (c) Deformation field obtained after the registration process. (d) Overlapping between images *a* (red) and *b* (green). (e) Overlapping between the deformed source image (red) and the MRI slice corresponding to the estimated plane (green).

References

1. San José Estépar, R., Westin, C., Vosburgh, K.: Towards real time 2d to 3d registration for ultrasound-guided endoscopic and laparoscopic procedures. *IJCARS* 4(6), 549–560 (2009)
2. Mercier, L., Del Maestro, R.F., Petrecca, K., Araujo, D., Haegelen, C., Collins, D.L.: Online database of clinical mr and ultrasound images of brain tumors. *Medical Physics* 39 (2012)
3. Osechinskiy, S., Kruggel, F.: Slice-to-volume nonrigid registration of histological sections to mr images of the human brain. *Anatomy Research International* 2011 (2010)
4. Dalvi, R., Abugharbieh, R.: Fast feature based multi slice to volume registration using phase congruency. In: *EMBS*, pp. 5390–5393 (2008)
5. Gill, S., Abolmaesumi, P., Vikal, S., Mousavi, P., Fichtinger, G.: Intraoperative prostate tracking with slice-to-volume registration in MRI. In: *SMIT*, pp. 154–158 (August 2008)
6. Glocker, B., Sotiras, A., Komodakis, N., Paragios, N.: Deformable medical image registration: setting the state of the art with discrete methods. *Annu. Rev. Biomed. Eng.* 13 (2011)
7. Mahapatra, D., Sun, Y.: Nonrigid registration of dynamic renal MR images using a saliency based MRF model. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 771–779. Springer, Heidelberg (2008)
8. Zikic, D., Glocker, B., Kutter, O., Groher, M., et al.: Linear intensity-based image registration by markov random fields and discrete optimization. *Med. Image Anal.* 14(4), 550–562 (2010)
9. Komodakis, N., Tziritas, G., Paragios, N.: Fast, approximately optimal solutions for single and dynamic MRFs. In: *CVPR*, pp. 1–8 (2007)
10. Ishikawa, H.: Transformation of general binary mrf minimization to the first-order case. *IEEE Transactions on PAMI* 33(6), 1234–1249 (2011)
11. Komodakis, N., Paragios, N.: Beyond pairwise energies: Efficient optimization for higher-order MRFs. In: *CVPR*, pp. 2985–2992 (2009)

Learning Nonrigid Deformations for Constrained Multi-modal Image Registration

John A. Onofrey¹, Lawrence H. Staib^{1,2,3}, and Xenophon Papademetris^{1,3}

¹ Departments of Biomedical Engineering

² Electrical Engineering

³ Diagnostic Radiology

Yale University, New Haven CT 06520, USA

{john.onofrey,lawrence.staib,xenophon.papademetris}@yale.edu

Abstract. We present a new strategy to constrain nonrigid registrations of multi-modal images using a low-dimensional statistical deformation model and test this in registering pre-operative and post-operative images from epilepsy patients. For those patients who may undergo surgical resection for treatment, the current gold-standard to identify regions of seizure involves craniotomy and implantation of intracranial electrodes. To guide surgical resection, surgeons utilize pre-op anatomical and functional MR images in conjunction with post-electrode implantation MR and CT images. The electrode positions from the CT image need to be registered to pre-op functional and structural MR images. The post-op MRI serves as an intermediate registration step between the pre-op MR and CT images. In this work, we propose to bypass the post-op MR image registration step and directly register the pre-op MR and post-op CT images using a low-dimensional nonrigid registration that captures the gross deformation after electrode implantation. We learn the nonrigid deformation characteristics from a principal component analysis of a set of training deformations and demonstrate results using clinical data. We show that our technique significantly outperforms both standard rigid and nonrigid intensity-based registration methods in terms of mean and maximum registration error.

Keywords: nonrigid registration, multi-modal, statistical deformation model, principal component analysis, image-guided surgery.

1 Introduction

Nonrigid, multi-modal image registration is a challenging task. The ability of nonrigid registration algorithms to successfully find a globally optimal deformation is made difficult by the high dimensionality of the deformations being modeled. In multi-modal registration tasks, nonlinear intensity relationships exacerbate the problem by causing similarity metrics to have many local minima. With an unconstrained search space, algorithms fail to escape the local minima and inaccurately register the images. In this paper, we present a solution to this problem that uses a training set to learn a low-dimensionality parameterization

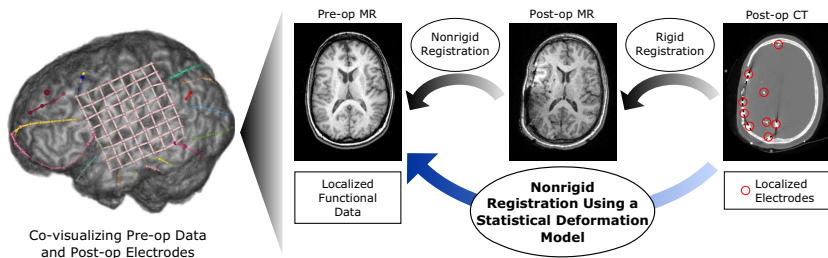


Fig. 1. To co-visualize icEEG electrodes with pre-op imaging data, the currently practiced registration framework rigidly registers the post-op CT image and the localized electrodes to a post-op MRI, and then nonrigidly registers the post-op MRI to the pre-op MRI. Our proposed method directly registers the post-op CT image to pre-op MRI using a learned statistical deformation model. Without having a constrained model of deformation, missing anatomical correspondences, *e.g.* removal of the skull during surgery, as well as imaging artifacts caused by the presence of the electrodes make direct nonrigid registration of the CT to the pre-op MRI inaccurate, even less accurate than rigid registration (see Figure 2).

of the deformation space for a specific task. We present results showing our initial application to nonrigidly map post-surgical electrode implantation CT images to pre-operative MR images acquired as part of epilepsy treatment.

For epilepsy patients whose seizures do not adequately respond to medication, surgical treatment is often an effective method to reduce or to eliminate seizure activity. Intracranial electroencephalography (icEEG), in which surgeons perform a craniotomy and implant (typically over 200) electrodes in suspected regions the brain, is the current “gold-standard” for localizing the focus of seizure activity [8]. Following electrode implantation, clinicians constantly monitor the icEEG for several days to identify, if possible, the electrodes nearest to the seizure focus. The surgeons localize the physical locations of these electrodes using post-op CT imaging. Successful electrode localization in combination with pre-operatively acquired functional brain images can be used to determine the feasibility of surgical tissue resection. Therefore, accurate spatial registration of electrodes with respect to functionally eloquent areas of the brain is critical.

The best, current approach to co-register the icEEG electrodes with the pre-op MR data involves the acquisition of a post-op MR image. The post-op MRI, in this case, serves as an intermediate registration step between the post-op CT and pre-op MR images. This registration framework projects the electrodes to the pre-op MRI space by first rigidly registering the CT image to the post-op MRI, and then nonrigidly registering the post-op MRI to the pre-op MRI to account for post-surgical deformations. Once the electrodes have been transformed to pre-op imaging space, they may be co-visualized with any other functional imaging studies. However, not all institutions are capable of acquiring the post-op MR image, and instead rely upon rigid registration of the post-op CT and pre-op MR images, which leads to inaccurate electrode localization. Given that post-op

MRIs are available at our institution, we propose to use this information to learn the nonrigid deformations from the post-op CT space to pre-op MR space, as Figure 1 illustrates.

Our proposed method leverages the currently used mono-modal post-op MRI to pre-op MRI nonrigid registrations to learn a statistical deformation model (SDM). The SDM is surgical-site dependent because larger brain deformations generally occur ipsilateral to the craniotomy site [1], thus we assume that patients with craniotomies in similar locations will experience similar deformation characteristics. We perform a principal component analysis (PCA) of the non-rigid deformations to construct our SDM [6]. While PCA has difficulty with high-dimensional data, we hypothesize that the post-surgical brain deformation is of low enough dimension that our SDM can capture the gross, intra-subject deformations observed after surgery. Other authors have made use of PCA SDMs to register mono-modal medical images for inter-subject registration [10,4] and for intra-subject motion compensation [2]. In contrast, we use our SDM trained on MR images to directly register each patient’s post-op CT image to their pre-op MRI. Our SDM models intra-subject deformations that result from surgical intervention, and *not* inter-subject anatomical variability. To the best of our knowledge, training a SDM on a mono-modality registration task and using that SDM to perform a multi-modality registration is a novel application. By doing so, the SDM can model subcortical deformations in the CT image that would otherwise not be possible with standard, intensity-only registration to the pre-op MR. Our results show our approach significantly reduces both mean and maximum registration error compared to standard rigid and nonrigid intensity-only MR-CT registration methods. We emphasize that standard intensity-based non-rigid methods perform worse than rigid ones, and our approach is the first, to our knowledge, to successfully directly register pre-op MRIs and post-electrode implantation CT images.

2 Methods

2.1 Training the Deformation Model

Given a database of surgical epilepsy patients, we select N patients with craniotomies at similar locations to train our SDM. Each patient’s dataset consists of a pre-op MR image $I_{\text{MR}}^{(1)}$, a post-op MR image $I_{\text{MR}}^{(2)}$, and a post-op CT image $I_{\text{CT}}^{(2)}$, where $I^{(t)}$ denotes pre-op images acquired at time $t = 1$ and post-op images at time $t = 2$. To create our SDM, we need to transform all N images into a common reference space. As per current practice, we first rigidly register the post-op MR and CT images by maximizing their normalized mutual information (NMI) [9], and obtain the transformation $T_{\text{CT} \rightarrow \text{MR}}$ (here, $i \rightarrow j$ denotes rigid transformation from space i to j). Next, we rigidly register the pre- and post-op MR images using NMI to produce a transformation $T_{\text{MR} \rightarrow \text{MR}}$. Finally, we nonrigidly register all pre-op MR images to the MNI Colin 27 brain, I_{MNI} , using a free-form deformation (FFD) [7] with NMI and write this transformation

$T_{\text{MR} \rightsquigarrow \text{MNI}}$ (we denote nonrigid transformation from space i to j as $i \rightsquigarrow j$). Once we have computed all transformations, we reslice all images from each of the N patients into MNI space ($181 \times 217 \times 181$ volume with 1mm^3 resolution) by concatenating the transformations. Spatial normalization to MNI space is necessary for defining a common reference space for intra-subject nonrigid deformations.

In MNI space, the post-op MR has yet to be nonrigidly registered to the pre-op MR. In current practice, this registration task uses skull-stripped brains to mitigate the effects of missing anatomical correspondences after surgery. However, since we are planning to register the CT images directly to the pre-op MR, the skull is actually one of the most informative structures to register, even if it is lacking some correspondence. Therefore, to accurately register the two non-skull stripped MRIs, we first create brain surface masks with some manual refinement, and then utilize an integrated intensity and point-feature registration algorithm [5]. This algorithm uses a FFD transformation model with 15mm control point spacing and minimizes the NMI similarity metric. With points weighting parameter set to 0.1, the brain surface points constrain the algorithm to align the cortical surface that otherwise has difficulty being accurately registered using intensity registration by itself. We denote the resulting transformations $T_{\text{MR} \rightsquigarrow \text{MR}}$. It is these transformations that we use to train our SDM.

For each patient $i = 1, \dots, N$, we rewrite the transformation $T_{\text{MR} \rightsquigarrow \text{MR}}$ as a column vector of P concatenated FFD control point displacements in 3D, $\mathbf{d}_i \in \mathbb{R}^{3P}$. Using PCA, we linearly approximate the deformation distribution [6]

$$\mathbf{d} = \bar{\mathbf{d}} + \Phi \mathbf{w} \quad (1)$$

where $\bar{\mathbf{d}} = \frac{1}{N} \sum_{i=1}^N \mathbf{d}_i$ is the mean deformation of the N training registrations, Φ is the matrix of orthogonal principal components, and \mathbf{w} is the vector of model variation coefficients. We compute the principal components from the eigensystem decomposition of the covariance matrix $\mathbf{C} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{d}_i - \bar{\mathbf{d}})(\mathbf{d}_i - \bar{\mathbf{d}})^T$. Using this formulation, $\Phi = (\phi_1 | \phi_2 | \dots | \phi_K) \in \mathbb{R}^{3P \times K}$ is a matrix of $K \leq \min\{N, 3P\}$ eigenvectors $\phi_k \in \mathbb{R}^{3P}$ with corresponding eigenvalues in decreasing order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$, and $\mathbf{w} \in \mathbb{R}^K$. The eigenvectors with the k largest eigenvalues define a SDM using k principal modes of variation, with $1 \leq k \leq K$.

2.2 Nonrigid SDM MR-CT Registration

Given a previously unseen pre-op MR and post-op CT image pair for a new patient, we use the learned SDM in Equation 1 to drive the nonrigid registration of post-op CT images to pre-op MR without the use of a post-op MR image. Before we nonrigidly register the two images, we must first transform the images into our model reference space. We do so by rigidly registering $I_{\text{CT}}^{(2)}$ to $I_{\text{MR}}^{(1)}$ and then nonrigidly registering $I_{\text{MR}}^{(1)}$ to I_{MNI} . In both cases we maximize NMI, and use a FFD transformation for nonrigid registration. We use the resulting transformations to reslice both images into MNI space such that $I_{\text{MR}}^{(1)'} = T_{\text{MR} \rightarrow \text{MNI}} \circ I_{\text{MR}}^{(1)}$

and $I_{\text{CT}}^{(2)'} = T_{\text{CT} \rightarrow \text{MR}} \circ T_{\text{MR} \rightarrow \text{MNI}} \circ I_{\text{CT}}^{(2)}$, where \circ is the transformation operator and I' indicates a resliced image.

With the images now transformed to our SDM reference space, we nonrigidly register $I_{\text{CT}}^{(2)'}$ to $I_{\text{MR}}^{(1)'}$. The SDM model coefficients \mathbf{w} in Equation 1 are a low-dimensional parameterization of a high-dimensional FFD \mathbf{d} , and we denote this transformation $T_{\text{SDM}}(\mathbf{x}; \mathbf{d})$ for points \mathbf{x} in the reference image domain $\Omega_{\text{MNI}} \subset \mathbb{R}^3$. We register $I_{\text{CT}}^{(2)'}$ and $I_{\text{MR}}^{(1)'}$ by optimizing the cost function

$$\hat{T}_{\text{SDM}} = \arg \max_{\mathbf{w}} J(I_{\text{MR}}^{(1)'}, T_{\text{SDM}}(\mathbf{x}; \bar{\mathbf{d}} + \Phi \mathbf{w}) \circ I_{\text{CT}}^{(2)'}, \quad \forall \mathbf{x} \in \Omega_{\text{MNI}} \quad (2)$$

where J is the NMI similarity metric. We use a multi-resolution image pyramid and conjugate gradient optimization to solve Equation 2.

3 Results and Discussion

From the database of surgical epilepsy patients available at our institution, we manually identified 18 patients with lateral craniotomies (10 on the right side, 8 on the left). In order to increase our dataset size, we flipped the left-side craniotomy images to be right-side craniotomies under the assumption that the direction of gross brain deformation correlates to craniotomy location. For each patient in the database, we have images $I_{\text{MR}}^{(1)}$ ($256 \times 256 \times 106$ at $0.977 \times 0.977 \times 1.5\text{mm}$ resolution), $I_{\text{MR}}^{(2)}$ ($256 \times 256 \times 110$ at $0.977 \times 0.977 \times 1.5\text{mm}$ resolution), and $I_{\text{CT}}^{(2)}$ ($512 \times 512 \times 137$ at $0.488 \times 0.488 \times 1.25\text{mm}$ resolution).

We performed *leave-one-out* testing to demonstrate our approach. For each patient $i = 1, \dots, 18$, we trained the SDM as described in Section 2.1 by omitting the i -th patient from the training set, which consists of the $N = 17$ remaining samples. We then register the i -th patient’s post-op CT to their pre-op MR using the SDM in MNI space as described in Section 2.2. We repeated our registration method using different numbers of modes of variation in \mathbf{w} , *i.e.* different values for k . We compared our method to rigid registration, which in our case $\hat{T}_{\text{R}} = I$, and intensity-only FFD \hat{T}_{FFD} with 15mm control point spacing. We also compared to FFD using a lower-dimensional 30mm control point spacing, but the results were similar to 15mm. Thus, for brevity, we reported only the results using 15mm FFD. We implemented and ran our algorithm as part of BioImage Suite [3].

To evaluate registration performance, we treated the nonrigid transformations $T_{\text{MR} \rightarrow \text{MR}}$ found during training with the post-op MR image, and as used in current practice, as a ground-truth. For an estimated transformation \hat{T} , we calculated the magnitude of transformation error

$$\varepsilon(\mathbf{x}) = \| \hat{T}(\mathbf{x}) - T_{\text{MR} \rightarrow \text{MR}}(\mathbf{x}) \|, \quad \mathbf{x} \in \Omega_{\text{MNI}}.$$

While this method evaluates performance against the current standard, we assessed our MR-CT registration results with the rigid registration \hat{T}_{R} in mind.

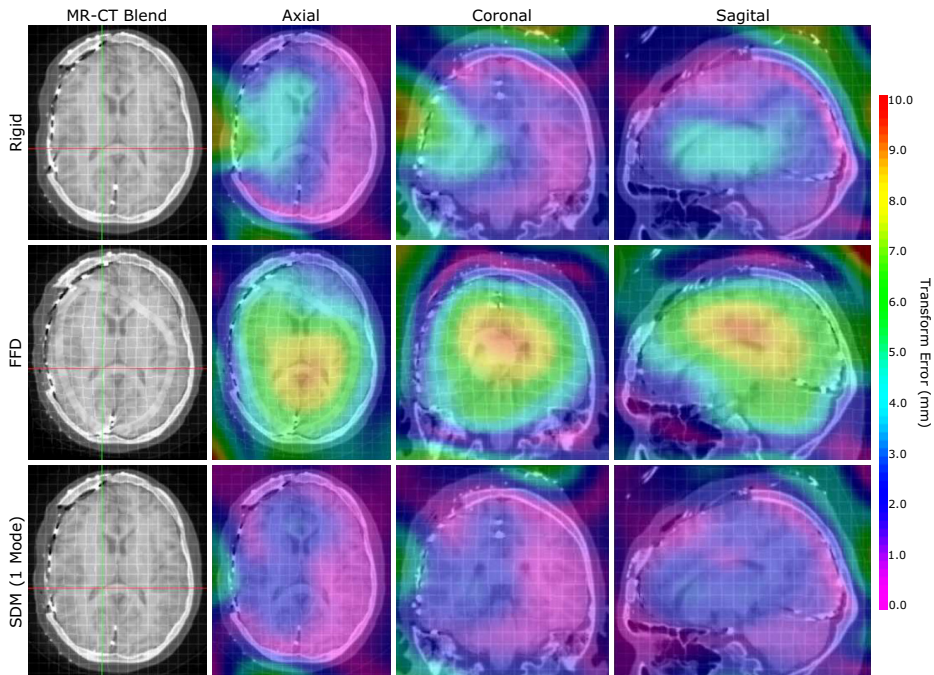


Fig. 2. Our nonrigid SDM registration method using a single mode of variation has lower transformation error throughout much of the brain, particularly in the areas around the craniotomy and around the ventricles, than both standard rigid and FFD intensity-only registration methods. We spatially visualize voxel-wise transformation error using a colormap overlay for an exemplar patient with a craniotomy on right side of the skull (left side of the axial images in radiological convention). The left column shows intensity-blended MR-CT images that highlight electrodes and skull location with respect to the skull-stripped MRI brain, along with a deformation grid.

It is our experience that surgeons generally only trust rigid registration for volumetric MR-CT registration. Furthermore, we evaluated maximum error because surgeons are most interested in quantifying worst-case performance.

We spatially visualized $\varepsilon(\mathbf{x}), \forall \mathbf{x} \in \Omega_{\text{MNI}}$ using a colormap overlay, as shown in Figure 2, to compare our estimates for \hat{T}_{R} , \hat{T}_{FFD} , and \hat{T}_{SDM} . In comparison to rigid registration, our method reduced error throughout the brain, and particularly so around the craniotomy and ventricles. We highlight the poor performance of the intensity-only FFD MR-CT registration. Due to the poor soft-tissue contrast in the CT, the FFD failed to accurately register the interior of the brain. Even though both FFD and SDM used the same NMI similarity metric, the SDM constrained the transformation to accurately mimic the interior deformations.

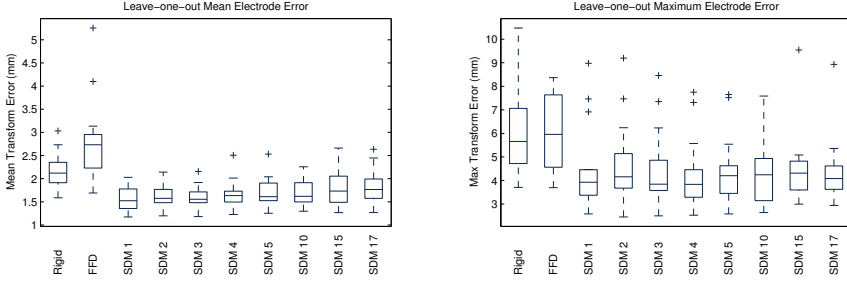


Fig. 3. Our proposed nonrigid SDM registration method significantly reduced transformation error compared to standard rigid and FFD intensity registration methods. SDM k denotes registration using the first $k = 1, 2, 3, \dots$ principal modes of variation. We plot the distributions of both mean (left plot) and maximum (right plot) transformation error at electrode locations for the 18 leave-one-out MR-CT registrations. The boxplots show median, inner quartile, extremes, and outlier values.

We next quantified $\varepsilon(\mathbf{x}_e)$ at each patient’s electrode locations $\mathbf{x}_e \in \Omega_E \subset \Omega_{MNI}$, where the mean number of electrodes for a patient was 197. Figure 3 summarizes patient mean error $\bar{\varepsilon}$ and maximum error ε_{\max} over all $\mathbf{x}_e \in \Omega_E$. Our proposed SDM registration significantly ($p \leq 0.05$, paired t-test) reduced both $\bar{\varepsilon}$ and ε_{\max} for all modes of variation with respect to rigid registration. The mean and maximum ε across all 18 patients using SDM with 1 mode was $1.58 \pm 0.24\text{mm}$ and $4.39 \pm 1.70\text{mm}$, respectively, which compared to $2.12 \pm 0.37\text{mm}$ and $6.08 \pm 1.84\text{mm}$ for rigid registration and $2.75 \pm 0.85\text{mm}$ and $5.95 \pm 1.64\text{mm}$ for FFD (all reported values are mean \pm std). FFD significantly increased $\bar{\varepsilon}$ with respect to rigid registration.

Additionally, we computed mean error throughout 4 different volumes of interest (VOIs): the right brain hemisphere Ω_{RB} , the right skull Ω_{RS} , the left brain hemisphere Ω_{LB} , and left skull Ω_{LS} , such that $\Omega_i \subset \Omega_{MNI}$ and $\Omega_i \cap \Omega_j = \emptyset, i \neq j$. The Ω_{RB} and Ω_{RS} VOIs were of particular interest since they were ipsilateral to the craniotomy. Compared to rigid registration, our method significantly ($p \leq 0.05$) reduced $\bar{\varepsilon}$ in Ω_{RB} using 1-5 modes of variation and in Ω_{RS} using 1 and 3 modes. SDM registration significantly increased $\bar{\varepsilon}$ in Ω_{LB} when using 15 and 17 modes, but otherwise there were no significant increases in $\bar{\varepsilon}$ in Ω_{LB} or Ω_{LS} compared to rigid registration. We noted that $\bar{\varepsilon}$ generally increased as we used more modes of variation in our SDM, which is most likely explained by the PCA over-fitting to the training set.

4 Conclusion

Our proposed method models post-surgical nonrigid deformations to significantly reduce both mean and maximum transformation errors in multi-modality nonrigid MR-CT registration compared to standard rigid and unconstrained

nonrigid registrations. As shown in Figures 2 and 3, standard nonrigid registration methods perform worse than rigid methods, at least in this dataset where the CT images had both deformation and significant artifacts. Although we present results for training a SDM at only a single craniotomy location, in the future, we aim to create craniotomy site-specific SDMs to model the corresponding deformations at different locations. Furthermore, we could improve our training set registrations by including additional labeled anatomical structures, *e.g.* the ventricles, to improve PCA model construction. The ideas presented in this paper constitute a general framework to effectively register multi-modal volumetric images nonrigidly. In particular we demonstrate how we can use a small subset of high quality training data (in this case the rare availability of a post-electrode implantation MRI) to learn about the properties of the deformation model in a given case of nonrigid deformation, and to subsequently use this knowledge to solve the nonrigid registration problem in the more general case with lesser quality data (*i.e.* the direct nonrigid multimodal registration of pre-op MRI to post-implantation CT). Similar principles could be applied to nonrigidly register, for example, interventional ultrasound images to pre-operative MRI.

References

1. Hartkens, T., Hill, D., Castellano-Smith, A., Hawkes, D., Maurer Jr., C.R., Martin, A., Hall, W., Liu, H., Truwit, C.: Measurement and analysis of brain deformation during neurosurgery. *IEEE Trans. on Medical Imaging* 22(1), 82–92 (2003)
2. He, T., Xue, Z., Xie, W., Wong, S.T.C.: Online 4-D CT estimation for patient-specific respiratory motion based on real-time breathing signals. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part III*. LNCS, vol. 6363, pp. 392–399. Springer, Heidelberg (2010)
3. Joshi, A., Scheinost, D., Okuda, H., Belhachemi, D., Murphy, I., Staib, L., Papademetris, X.: Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* 9, 69–84 (2011)
4. Kim, M.J., Kim, M.H., Shen, D.: Learning-based deformation estimation for fast non-rigid registration. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008*, pp. 1–6 (June 2008)
5. Papademetris, X., Jackowski, A.P., Schultz, R.T., Staib, L.H., Duncan, J.S.: Integrated intensity and point-feature nonrigid registration. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3216, pp. 763–770. Springer, Heidelberg (2004)
6. Rueckert, D., Frangi, A., Schnabel, J.: Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Trans. on Medical Imaging* 22(8), 1014–1025 (2003)
7. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Trans. on Medical Imaging* 18(8), 712–721 (1999)
8. Spencer, S.S., Sperling, M., Shewmon, A.: Intracranial electrodes. In: *Epilepsy, A Comprehensive Textbook*, pp. 1719–1748 (1998)
9. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition* 32(1), 71–86 (1999)
10. Xue, Z., Shen, D., Davatzikos, C.: Statistical representation of high-dimensional deformation fields with application to statistically constrained 3d warping. *Medical Image Analysis* 10(5), 740–751 (2006)

Non-Rigid 2D-3D Registration Using Anisotropic Error Ellipsoids to Account for Projection Uncertainties during Aortic Surgery

Alexis Guyot^{1,*}, Andreas Varnavas¹, Tom Carrell², and Graeme Penney¹

¹ King's College London, Division of Imaging Sciences, London, UK

² Guy's & St Thomas' NHS Foundation Trust, Dept. of Vascular Surgery, London, UK

Abstract. Overlay of preoperative images is increasingly being used to aid complex endovascular aortic repair and is obtained by rigid 2D-3D registration of 3D preoperative (CT) and 2D intraoperative (X-ray) data. However, for tortuous aortas large non-rigid deformations occur, thus a non-rigid registration must be performed to enable an accurate overlay. This article proposes the use of Thin-Plate Splines (TPS) to perform non-rigid 2D-3D registration. Intraoperative X-ray data contain no spatial information along the X-ray projection direction. Our approach accounts for this lack of spatial information by the use of an approximating TPS with non-isotropic error ellipsoids, where the major ellipsoid axis is aligned with the X-ray projection direction. Experiments are carried out using 1D-2D and 2D-3D simulated data and 2D-3D interventional data. Simulated results show that our proposed method is 1.5 times more accurate than interpolating TPS based registration. Interventional data results show how large rigid registration errors of 9mm can be reduced to 4mm using our proposed method.

Keywords: Non-Rigid, 2D-3D registration, thin plate spline, non-isotropic errors.

1 Introduction

Complex endovascular aortic repair (EVAR), where the aneurysm extends over visceral vessels, is increasingly being used instead of open surgery. Complex EVAR is a technically challenging procedure requiring accurate 3D placement of often seven or more stent-grafts. The standard imaging protocol includes a contrast enhanced preoperative CT, showing 3D bony and vascular anatomy. The procedure is carried out using fluoroscopy guidance which provides 2D images of bony structures and interventional instruments, and 2D Digital Subtraction Angiography (DSA) images that are used to visualise the blood vessels.

Due to the complex nature of these procedures, computer assistance has been proposed [1, 2], which involves rigid registration of the 3D preoperative data with the 2D intraoperative data, to enable 3D information from the preoperative

* We are grateful to Philips Healthcare for funding this research.

CT to be overlaid onto the intraoperative fluoroscopy (as seen in Fig. 5(a)). However, in some cases (e.g. aortas with a neck angulation greater than 30° [3]), inserting instruments leads to non-rigid deformation of the patient’s aorta, which can result in large errors if a rigid registration method is used.

Several authors have proposed non-rigid 2D-3D registration to address this concern. Two such methods [4, 5] assume constant vessel length, and that deformations minimise a smoothness criterion. Liao *et. al.* ’s method uses a 3D CT angiographic scan and proposes automatic segmentation of blood vessels from 2D DSA images, which is difficult to achieve robustly on clinical images. In addition, the deformations caused by stiff instruments inside the aorta are likely to violate the assumption of a minimum bending energy made by these two approaches.

We present a novel approach for 2D-3D non-rigid registration: Thin-Plate Splines (TPS) plus Projection Uncertainties (which we refer to as TPS+PU). After an initial rigid registration, aortic deformation is based on preoperatively determined fixed landmarks and a small number (less than 10) of manually identified moving landmarks in two fluoroscopy images. Unlike standard back-projection methods, our approach does not require corresponding landmarks to be identified in both fluoroscopy images. Such corresponding landmarks are difficult to find in clinical images. Instead our method models the lack of information perpendicular to the 2D fluoroscopy images using non-isotropic error ellipsoids, where the major axis lies along the back-projection lines.

2 Theory

Spline-based deformation algorithms have been widely used (see [6–8] for examples) for non-rigid registration when the source and target data have the same dimensionality, and are based upon the matching of source and target landmarks and minimisation of a smoothness criterion.

We propose a TPS-based registration method to register data sets with different dimensionality: a 3D preoperative CT to 2D intraoperative fluoroscopy. This configuration leads to projection uncertainties as the 3D position of landmarks selected from a 2D image cannot be determined within the 3D space. When two views are used, the 3D position of a landmark can be deduced from its corresponding 2D positions in both views. However, due to the nature of the images used during EVAR, corresponding landmarks are only rarely observed in both views, thus 3D reconstruction is generally not feasible.

Our method uses approximating TPS with non-isotropic errors [8] in a two-view scenario to cope with the aforementioned issue. As shown in Fig. 1, 3D source points $\{s_i\}$ and 2D target points $\{t_i^{2D}\}$ are picked on the aorta surface and in the fluoroscopy images. 3D target points $\{t_i^{3D}\}$ are calculated from $\{t_i^{2D}\}$ as the closest point to $\{s_i\}$ on the back-projection line $\{l_i\}$ that joins $\{t_i^{2D}\}$ to the X-ray source. Covariance matrices $\{\Sigma_i\}$ that represent the uncertainty in matching points s_i to t_i^{3D} along lines l_i are computed as follows: the major axis of the error ellipsoid is set to the direction of l_i and to have length 1000

(approximate fluoroscopy focal length) while the two other axes are set to zero length. Using these parameters for the ellipsoids and setting a weighting function λ to 1 allows loose matching of landmarks along the back-projection lines. The following energy function, which weights the smoothness of the deformation $u = (u_1, u_2, u_3)$ against its ability to match source and target points, is minimised:

$$J_\lambda(u) = \frac{1}{n} \sum_{i=1}^n (t_i^{3D} - u(s_i))^T \Sigma_i^{-1} (t_i^{3D} - u(s_i)) + \lambda \sum_{k=1}^3 J_2^3(u_k)$$

$$\text{where } J_2^3(u_k) = \sum_{\alpha_1 + \alpha_2 + \alpha_3 = 2} \frac{2}{\alpha_1! \alpha_2! \alpha_3!} \int_{R^3} \left(\frac{\partial^2 u_k}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \partial x_3^{\alpha_3}} \right)^2 dX$$

and n is the number of source landmarks. The analytical solution is:

$$u_k(s) = \sum_{v=1}^4 a_{k,v} \phi_v(s) + \sum_{i=1}^n w_{k,i} U(s, s_i) \text{ for all } k \text{ in } \{1, 2, 3\}$$

with $s = (s_x, s_y, s_z)^T$ a point from the source image, $\phi_1(s) = 1$, $\phi_2(s) = s_x$, $\phi_3(s) = s_y$, $\phi_4(s) = s_z$ and U a suitable radial basis function. The $a_{k,v}$ and $w_{k,i}$ are sets of coefficients that respectively account for the rigid and the non-rigid part of u_k . Solution is obtained by solving the system:

$$(K + n\lambda W^{-1})w + Sa = v$$

$$S^T w = 0$$

where $K_{ij} = U(s_i, s_j)$, $S_{ij} = \phi_j(s_i)$, $W^{-1} = \text{diag}(\Sigma_1, \dots, \Sigma_n)$, $v = (t_1^T, \dots, t_n^T)$, with $t_i^T = (t_{i,x}^{3D}, t_{i,y}^{3D}, t_{i,z}^{3D})$, $a = (a_1^T, \dots, a_4^T)$, with $a_v^T = (a_{x,v}, a_{y,v}, a_{z,v})$ and $w = (w_1^T, \dots, w_n^T)$, with $w_i^T = (w_{x,i}, w_{y,i}, w_{z,i})$.

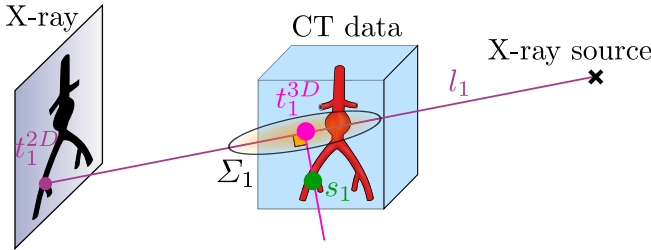


Fig. 1. Approximating TPS can be used to account for the uncertainty along the back-projection lines: 3D-3D TPS registration is performed using landmarks $\{s_i\}$ and $\{t_i^{3D}\}$ with associated error ellipsoids with their major axes along the back-projection lines

The solution of this system is computed and applied to the aorta surface. We refer to this novel method as the ‘single warp’ strategy (see Fig. 2(left)). We compare it to a ‘sequential warp’ strategy which registers to each 2D image in turn and does not use non-isotropic error ellipsoids (see Fig. 2(right)).

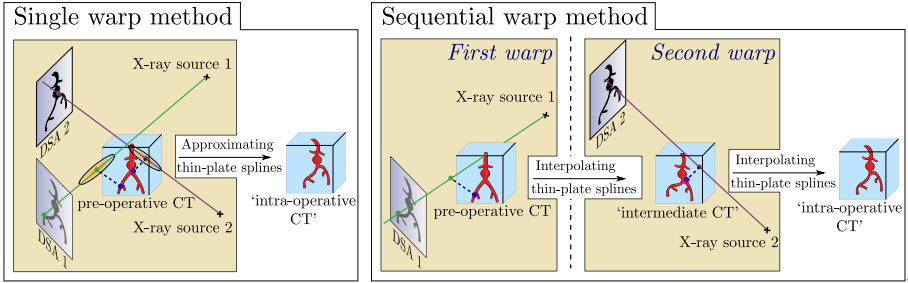


Fig. 2. Single warp strategy: registration is performed using target points from both views. Sequential warp strategy: a first registration is performed using target points from the first view, yielding an intermediate surface. A second registration is then performed using target points from the second view to warp the intermediate surface.

3 Experiments

3.1 Data

Three sets of data were used: (1) simulated 1D-2D synthetic data, (2) simulated 2D-3D data generated from interventional data, (3) 2D-3D interventional data.

Simulated 1D-2D synthetic data was produced to resemble an axial 2D slice of the aorta at the site of the renal arteries. This 2D slice is the simulated preoperative image (see Fig. 3 (a)), which was then warped to generate intraoperative datasets as follows: TPS transformations were carried out using five fixed points (the corners of a square around the aorta to anchor the deformation and the ostium of the left renal artery) and a single moving point at the right renal ostium which was displaced by 10mm at an angle of $i \times 45^\circ$, $i = 1, \dots, 8$ for each i^{th} deformation. From each of these warped images two 1D simulated intraoperative datasets were generated by a) projecting the ostia of the renal arteries onto one 1D plane (rr and lr in Fig. 3 (b)) and then b) projecting the edges of the aorta onto a second, 30° rotated, 1D plane (ae and pe in Fig. 3 (b)).

Simulated 2D-3D data generated from interventional data was produced using a patient CT as the preoperative data. Intraoperative data sets were generated by warping the CT image, eight times, using a TPS transformation. Fixed points were the corners of a large bounding box surrounding the abdominal aorta, the centres of vertebrae L1 to L5 and the bifurcations of the iliac arteries. Moving points were placed on the renal artery ostia and anterior and posterior edges of the aorta level with the renal arteries. These points were displaced 10mm. From each simulated 'intraoperative' CT scan, two intraoperative DSA images, with 30° difference in view direction, were synthesised.

2D-3D interventional data were used from a complex EVAR procedure. The data consisted of a preoperative CT scan, two intraoperative DSA images acquired after insertion of stent delivery device (views differed by 20° rotation).

3.2 Methods

For each experiment, the aorta was deformed with either a single warp (TPS+PU, Fig 2(left)) or a sequential warp (interpolating TPS, Fig 2(right)). Each experiment used a set of fixed points (picked preoperatively) and moving points (picked from intraoperative images). The fixed points were as follows: for the 1D-2D synthetic experiment, corners of a square surrounding the aorta were fixed. For the 2D-3D experiments, corners of a large cube surrounding the aorta, centre of lumbar vertebrae and bifurcation of common iliac arteries were fixed.

Moving points for 1D-2D synthetic data were picked as illustrated in Fig. 3 (c). As in the case with real DSA images different information is visible in different views. We assume that the renal ostia are visible, and so can be picked, in plane 1 (equivalent to an anterior-posterior DSA), whereas only the outline of the aorta is likely to be visible from the rotated view, plane 2.

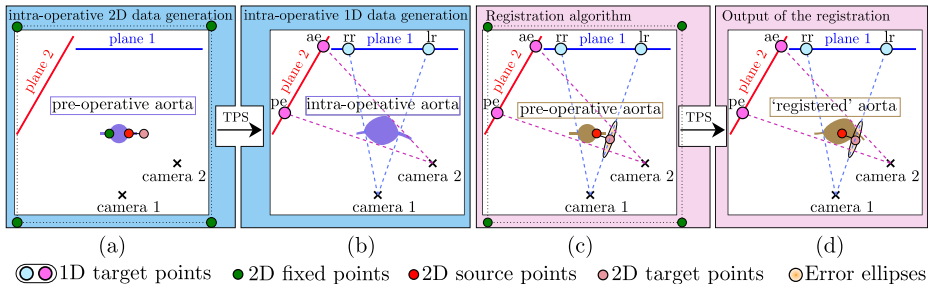


Fig. 3. (a) interpolating TPS is applied to the preoperative aorta to obtain the intraoperative aorta (b) 1D intraoperative points: left, right renals (lr, rr) and anterior, posterior edges (ae, pe), are generated by projecting the intraoperative aorta onto the 1D planes. (c) 1D target points are selected on the 1D intraoperative planes and the points on the relevant aortic area closest to the back-projection lines are selected as corresponding 2D source points. The points on the back-projection lines closest to the 2D source points are selected as corresponding 2D target points. For the single warp, error ellipses are produced along the back-projection lines. (d) Result of applying non-rigid TPS + PU registration on the aorta using 2D source and 2D target points.

1D target points are selected on the 1D intraoperative planes and the corresponding points on the aorta are selected as 2D source points. For the renal positions corresponding points were the 2D renal artery ostia (see Fig. 3 (c)). For aortic edge positions corresponding points were the closest 2D edge points to back projection lines from the 1D point positions.

Moving points for the simulated 2D-3D data were picked at the ostia of the renal arteries and the Superior Mesenteric Artery (SMA) on the first DSA image and at the outline of the aorta on the second DSA image. The location of corresponding 3D source and 3D target points was found in a process equivalent to the one presented in Fig. 3 (c).

Moving points for the interventional data were picked at the ostia of the vessels when visible, generating 3D source and 3D target points as in the case of simulated 2D-3D data. The rigidly-registered aortic CT surface was then overlaid onto the first DSA image and the location of largest displacement with respect to the aortic outline was located. An additional moving point was picked at this location, non-rigid registration was performed using all previously picked points and the displayed aortic overlay was updated. This process was repeated, generating one additional moving point at each iteration until the aortic overlay closely matched the DSA image (i.e. within $\approx 2\text{mm}$). Moving points were selected in the second DSA image using the same procedure. Overall, 7 moving points were picked within approximately 2 minutes.

3.3 Validation

For simulated data experiments Target Registration Errors (TRE) were calculated to the known ground truth using the following regions of interest: *ROI 1* was $26 \times 26\text{mm}^2$ or $30 \times 30 \times 30\text{mm}^3$ region centred on the aorta at the level of the renal arteries for the 1D-2D and 2D-3D data respectively. *ROI 2* was a $2 \times 2\text{mm}^2$ or $2 \times 2 \times 2\text{mm}^3$ region centred on the right renal artery ostium (for the 2D-3D data *ROI 3* denotes the left renal ostium).

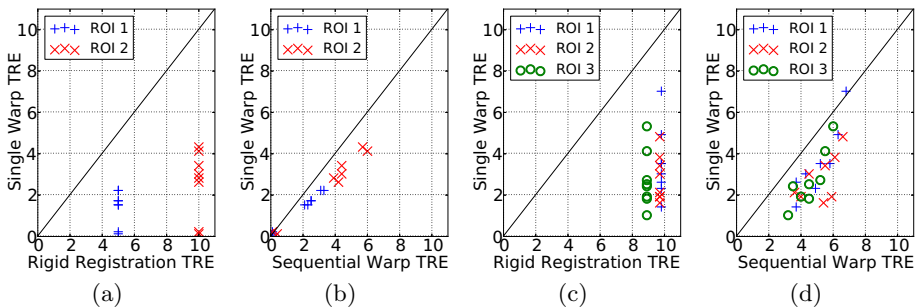


Fig. 4. All TRE values are expressed in mm. (a) 1D-2D synthetic data: Single Warp vs. Rigid Registration TRE (b) 1D-2D synthetic data: Single Warp vs. Sequential Warp TRE (c) 2D-3D synthetic data: Single Warp vs. Rigid Registration TRE (d) 2D-3D synthetic data: Single Warp vs. Sequential Warp TRE.

No ground truth was available for the interventional data. Instead the position of the CT overlay was compared to a DSA image acquired later in the intervention. Reprojection error distances were calculated (as described in [2]) at the left (*ROI 1*) and right (*ROI 2*) renal artery ostia.

4 Results

Results comparing TRE values for rigid registration, single warp and sequential warp for the 1D-2D and 2D-3D synthesised data are shown in Fig. 4. The single warp results show a major improvement compared to rigid registration: registration accuracy is on average 3.8 times higher for the 1D-2D data and 3.1 times higher for the 2D-3D data. The single warp results also show an improvement compared to the sequential warp: registration accuracy is on average 1.4 times higher for the 1D-2D data and 1.6 times higher for the 2D-3D data.

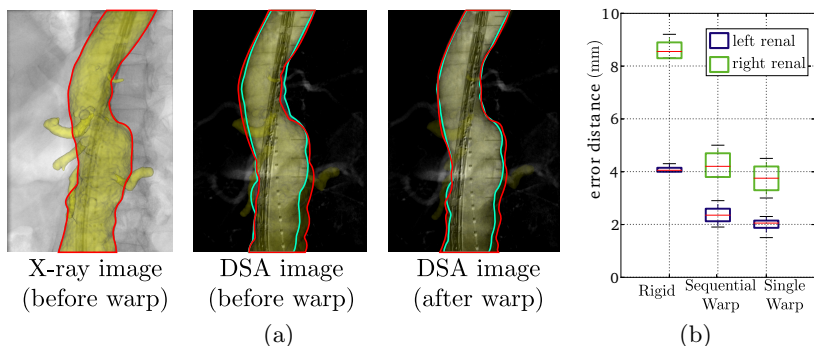


Fig. 5. (a) Comparison of CT overlay with DSA image before and after warp. Aorta outline from overlay (red) matches more accurately to DSA (cyan) after warp. (b) Registration accuracy at renal ostia using rigid registration, sequential and single warp.

Fig 5(a) visually shows the improvement in using the single warp compared to rigid registration on clinical data. Numerical results are presented in Fig. 5(b) where applying the single warp halved the rigid registration error.

5 Discussion and Conclusions

Two approaches (sequential warp with interpolating TPS and single warp with TPS+PU) to deform the aorta were described and compared. Results on simulated data show a clear improvement in accuracy when using TPS+PU. A recent study [3] reported a range of up to 11mm (mean 4.5mm) for aortic movement. Despite applying deformations of 10mm, almost at the top of this reported range, we managed to achieve the required registration accuracy less than 3mm

(as proposed in [3]) for half the single warp registrations. No sequential warp registration achieved $< 3\text{mm}$ accuracy (see Fig. 4(d)).

Our method fits well within the standard clinical workflow. It requires knowledgeable manual input for point picking, but only a few points are required, which are usually located in standard clinical images within a few minutes (compared to the procedure time of 4+ hours). Automated landmark identification would be preferred, however this is very difficult to achieve with guaranteed 100% robustness, and subsequent required checks on automated landmark selection are likely to require knowledgeable manual visual inspection. 2D-3D rigid registration is performed regularly to account for rigid patient, and fluoroscopy gantry, movement during complex EVAR. Our current approach assumes that only one major non-rigid deformation (which occurs after inserting a stiff wire) takes place during the procedure. Therefore, the non-rigid registration algorithm would only need to be applied once per procedure. This assumption remains to be clinically verified.

In conclusion, we present a new non-rigid 2D-3D registration method using a TPS with non-isotropic error ellipsoids to model projection uncertainties. Results show improved accuracy using TPS+PU, which has the potential to increase the number of EVAR patients who can benefit from computer-assisted surgery.

References

1. Kaladji, A., Dumenil, A., Castro, M., Haigron, P., Heautot, J.F., Haulon, S.: Endovascular aortic repair of a postdissecting thoracoabdominal aneurysm using intraoperative fusion imaging. *J. Vasc. Surg.* 57, 1109–1112 (2013)
2. Penney, G., Varnavas, A., Dastur, N., Carrell, T.: An image-guided surgery system to aid endovascular treatment of complex aortic aneurysms: Description and initial clinical experience. In: Taylor, R.H., Yang, G.-Z. (eds.) *IPCAI 2011*. LNCS, vol. 6689, pp. 13–24. Springer, Heidelberg (2011)
3. Carrell, T., Modarai, B., Brown, J., Penney, G.: Feasibility and limitations of an automated 2D-3D rigid image registration system for complex endovascular aortic procedures. *J. Endovasc. Ther.* 17(4), 527–533 (2010)
4. Groher, M., Baust, M., Zikic, D., Navab, N.: Monocular deformable model-to-image registration of vascular structures. In: Fischer, B., Dawant, B.M., Lorenz, C. (eds.) *WBIR 2010*. LNCS, vol. 6204, pp. 37–47. Springer, Heidelberg (2010)
5. Liao, R., Tan, Y., Sundar, H., Pfister, M., Kamen, A.: An efficient graph-based deformable 2D/3D registration algorithm with applications for abdominal aortic aneurysm interventions. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) *MIAR 2010*. LNCS, vol. 6326, pp. 561–570. Springer, Heidelberg (2010)
6. Bazen, A.M., Gerez, S.H.: Fingerprint matching by thin-plate spline modelling of elastic deformations. *Pattern Recogn.* 36(8), 1859–1867 (2003)
7. Johnson, H.J., Christensen, G.E.: Landmark and intensity-based, consistent thin-plate spline image registration. In: Insana, M.F., Leahy, R.M. (eds.) *IPMI 2001*. LNCS, vol. 2082, pp. 329–343. Springer, Heidelberg (2001)
8. Rohr, K., Stiehl, H., Sprengel, R., Buzug, T., Weese, J., Kuhn, M.: Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging* 20(6), 526–534 (2001)

A Variational Formulation for Discrete Registration

Karteek Popuri, Dana Cobzas, and Martin Jägersand

Department of Computing Science, University of Alberta, Edmonton, Canada

Abstract. We present a novel variational formulation of discrete deformable registration as the minimization of a convex energy functional that involves diffusion regularization. We show that a finite difference solution (FD) of the variational formulation is equivalent to a continuous-valued Gaussian Markov random field (MRF) energy minimization formulation previously proposed as the random walker deformable registration method [1]. A computationally efficient solution using the finite element method (FEM) method has been proposed to solve the variational minimization problem. Our proposed method obtained competitive results when compared with 14 other deformable registration methods on the CUMC12 MRI dataset.

1 Introduction

The task of non-rigid or deformable image registration refers to the process of finding meaningful dense correspondences between two images. It is required in a wide variety of medical imaging applications such as longitudinal studies on disease progression, multi-modal image fusion, statistical studies on anatomical variability and atlas-based segmentation. More formally, the problem of deformable image registration is posed as the estimation of an optimal displacement field that maps points in an image to the corresponding points in another image such that a similarity criterion is minimized. This minimization is inherently ill-posed due to the large number of degrees of freedom (DOF) involved and therefore it relies on regularization or smoothing constraints.

Over the years, numerous regularization approaches have been proposed for deformable registration (see [2] for a detailed review). These approaches can be broadly categorized into two groups, namely non-parametric and parametric approaches. In non-parametric approaches, explicit regularization terms such as elastic [3], diffusion [4] and curvature [5] are added to the data term and a variational minimization of the combined energy functional is performed. Alternatively, a demons minimization strategy can be used, where the smoothing of the displacement field is decoupled from the minimization of the data term [6], [7], [8]. In parametric approaches, regularization is enforced in an implicit manner through the parametrization of the displacement field using a finite set of basis functions, such as radial basis functions (RBF) [9], B-spline based free form deformations (FFD) [10], finite element method (FEM) basis functions [11].

The various non-parametric and parametric deformable registration approaches mentioned above attempt to determine the unknown displacement field through the continuous optimization of energy functionals. As an alternative to continuous formulations, there has been some interest in formulating the deformable registration task as a discrete labeling problem. Here, the space of possible displacements is discretized and

a Markov random field (MRF) energy corresponding to the registration objective is minimized. However, as the MRF energy minimization is in general NP-hard, existing methods on MRF-based discrete deformable registration are only able to find good quality approximate solutions using graph cuts [12] and linear programming [13].

Previous discrete formulations of deformable image registration were inherently discrete in both the image domain and displacement space. In this work, we seek a formulation that keeps the image domain continuous resulting in a variational formulation for discrete deformable registration. Our formulation results in a convex functional which can be conveniently minimized to obtain a unique solution. Specifically, we associate a continuous prior probability map to each of the possible displacement values using image similarities. The prior probability maps denote how likely the corresponding displacement value is at particular spatial location in the continuous image domain. The discrete deformable registration task is then posed as a variational problem corresponding to the diffusion-based smoothing of prior probability maps. For solving this variational problem, we propose a finite element method (FEM) that employs a non-uniform mesh well adapted to the salient image features. This significantly reduces the number of DOFs involved in the minimization compared to a traditional finite difference (FD) discretization and hence leads to a highly computationally efficient solution. Further, we show that a FD discretization of our variational formulation is equivalent to the the random walker (RW) solution of discrete registration recently proposed by Cobzas *et al.* [1].

To summarize, we make the following contributions:

- (1) We develop a novel variational formulation of discrete deformable registration on the continuous image domain. This formulation results in a convex energy functional that involves diffusion regularization.
- (2) We show that a finite difference solution (FD) of the variational formulation is equivalent to a continuous-valued Gaussian MRF energy minimization formulation.
- (3) We propose a computationally efficient FEM solution.
- (4) We validate our method on the publicly available CUMC12 MRI dataset [14] and show a competitive performance of our proposed method in comparison to 14 other deformable registration methods.

2 Methods

2.1 Variational Formulation of Discrete Deformable Registration

Given a source image $I_S : \Omega_S \rightarrow \mathbb{R}$, $\Omega_S \subset \mathbb{R}^\nu$ and a target image $I_T : \Omega_T \rightarrow \mathbb{R}$, $\Omega_T \subset \mathbb{R}^\nu$, $\nu = 2$ or 3 , the goal of non-rigid or deformable registration is to estimate a displacement field $\mathbf{U} : \Omega_T \rightarrow \mathbb{R}^\nu$ such that the warped source image $I_S(\mathbf{x} + \mathbf{U}(\mathbf{x}))$ is similar to the target image I_T . A similarity map $\Psi : \Omega_T \rightarrow \mathbb{R}$ can be defined to measure the similarity between the warped source and target images as:

$$\Psi(\mathbf{x}; \mathbf{U}, I_T, I_S) = (I_S(\mathbf{x} + \mathbf{U}(\mathbf{x})) - I_T(\mathbf{x}))^2. \quad (1)$$

In discrete deformable registration, the space of possible displacement values is discretized or quantized such that it corresponds to a finite set of K vectors, i.e.,

$\mathbf{U} : \Omega_T \rightarrow \mathcal{D}$, where $\mathcal{D} = \{\mathbf{d}^k\}_{k=1}^K$, $\mathbf{d}^k \in \mathbb{R}^\nu$. Based on the similarity map, similar to [1] we can define a prior probability map $\lambda^k : \Omega_T \rightarrow [0, 1]$ for each of the displacement values $\mathbf{d}^k \in \mathcal{D}$ as:

$$\lambda^k(\mathbf{x}; \mathbf{d}^k) = \frac{\exp(-\alpha\Psi(\mathbf{x}; \mathbf{d}^k, I_T, I_S))}{\sum_{r=1}^K \exp(-\alpha\Psi(\mathbf{x}; \mathbf{d}^r, I_T, I_S))}. \quad (2)$$

It can be seen that the prior $\lambda^k(\mathbf{x}; \mathbf{d}^k)$ essentially encodes the belief that the corresponding displacement value \mathbf{d}^k is more likely at a spatial location $\mathbf{x} \in \Omega_T$, when the similarity measure between the warped source image and the target image at $\mathbf{x} \in \Omega_T$ is small. The discrete deformable registration task is next formulated as a variational energy minimization problem that estimates an unknown probability map $L^k : \Omega_T \rightarrow \mathbb{R}$ corresponding to each of the displacement values $\mathbf{d}^k \in \mathcal{D}$:

$$\mathbf{L}^* = \underset{\mathbf{L}}{\operatorname{argmin}} E_D[\mathbf{L}; I_T, I_S, \mathcal{D}] + \beta E_R^{\text{diff}}[\mathbf{L}], \quad (3)$$

$$\text{with } E_D[\mathbf{L}; I_T, I_S, \mathcal{D}] = \int_{\Omega_T} \|\mathbf{L}(\mathbf{x}) - \boldsymbol{\lambda}(\mathbf{x}; \mathcal{D})\|^2 d\mathbf{x},$$

$$E_R^{\text{diff}}[\mathbf{L}] = \sum_{k=1}^K \int_{\Omega_T} \nabla L^{kT}(\mathbf{x}) W(\mathbf{x}) \nabla L^k(\mathbf{x}) d\mathbf{x},$$

where $\boldsymbol{\lambda} = [\lambda^k]_{k=1}^K$, $\mathbf{L} = [L^k]_{k=1}^K$, $\mathbf{L}^* = [L^{*k}]_{k=1}^K$ and β is a regularization constant. The above energy minimization formulation essentially corresponds to the anisotropic diffusion-based smoothing of the prior probability maps $\{\lambda^k\}_{k=1}^K$. In other words, the minimization attempts to find a smooth probability map L^k associated with the displacement value \mathbf{d}^k such that it is as ‘‘close’’ as possible to the corresponding prior probability map λ^k . The smoothness is enforced through the anisotropic diffusion-based regularization term E_R^{diff} , where W is a $\nu \times \nu$ symmetric matrix denoting the diffusivity or the stiffness field. Note that in the above we do not need an explicit constraint to make sure that the unknown probability maps $\{L^k\}_{k=1}^K$ sum to 1. This because the unique minimizer of (3) should naturally satisfy this constraint, as the prior probability maps $\{\lambda^k\}_{k=1}^K$ were defined such that they sum to 1 (like in Grady [15]). The estimated displacement field $\mathbf{U}(\mathbf{x})$ at a spatial location $\mathbf{x} \in \Omega_T$ is obtained by choosing the displacement value \mathbf{d}^k with the highest optimal probability at that spatial location:

$$\mathbf{U}(\mathbf{x}) = \mathbf{d}^k \quad \forall \mathbf{x} \in \Omega_T, \quad (4)$$

$$\text{where } k = \underset{r \in \{1, 2, \dots, K\}}{\operatorname{argmax}} L^{*r}(\mathbf{x}).$$

However, it should be noted that even though the above displacement field is constructed from the optimal probability maps $\{L^{*k}\}_{k=1}^K$, this estimated displacement field by itself does not necessarily minimize any formal registration energy.

2.2 Equivalence to the Random Walker (RW) Registration Method

We next show that a finite difference solution (FD) of the variational formulation in (3) is equivalent to the continuous-valued Gaussian MRF energy minimization

formulation referred as the random walker registration in [1]. For this, consider the Euler-Lagrange equations corresponding to the energy functional in equation (3) for each $k \in \{1, 2, \dots, K\}$:

$$\nabla_{L^k}(E_D + \beta E_R^{\text{diff}}) = (L^k - \lambda^k) - \beta \nabla(W \nabla L^k) = 0. \quad (5)$$

Rearranging we get for each $k \in \{1, 2, \dots, K\}$:

$$L^k - \beta \nabla(W \nabla L^k) = \lambda^k. \quad (6)$$

Now, consider a uniform discretization $\mathcal{G} = (\{\mathbf{x}_i\}_{i=1}^R, \mathcal{N})$ of the template image domain Ω_T , where $\{\mathbf{x}_i\}_{i=1}^R$ is the set of pixels (voxels) and \mathcal{N} denotes neighborhoods. Further, choose an image dependent stiffness field $W(\mathbf{x}) = \exp(-\gamma(\text{diag}(\nabla I_T))^2)$. Writing the FDM approximation of the equations in (6) we obtain the following set of equations for each $k \in \{1, 2, \dots, K\}$:

$$\left(\text{Id}_N + \beta \Delta_{\mathcal{G}} \right) \mathbf{L}^k = \boldsymbol{\lambda}^k. \quad (7)$$

Here, Id_N is a $N \times N$ identity matrix and $\Delta_{\mathcal{G}}$ is the discrete Laplacian operator defined on the uniform mesh \mathcal{G} as:

$$(\Delta_{\mathcal{G}})_{ij} = \begin{cases} \sum_{r \in \mathcal{N}(i)} w_{ir} & \text{if } i = j \\ -w_{ij} & \text{if } j \in \mathcal{N}(i) \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

where $w_{ij} = \exp(-\gamma(I_T(\mathbf{x}_i) - I_T(\mathbf{x}_j))^2)$. It is easy to observe that the FD discretization on the uniform grid \mathcal{G} of the Euler-Lagrange equations in (5) corresponds to the set of linear equations in (7) that arise from the minimization of the continuous-valued MRF based discrete registration problem in [1].

2.3 FEM-Based Solution for Variational Discrete Deformable Registration

A more efficient solution of the proposed variational formulation for discrete registration (3) can be obtained using the FEM method. Consider a non-uniform discretization $\mathcal{M} = (\{P_n\}_{n=1}^N, \mathcal{T})$ of the domain Ω_T , where $\{P_n\}_{n=1}^N$ denotes the nodes of the mesh and \mathcal{T} is the set of elements (triangles in 2D, tetrahedra in 3D) (see Figure 1). Then using nodal basis functions $\{\phi_n\}_{n=1}^N$ we can discretize the probability maps as:

$$\mathbf{L}(\mathbf{x}) = \sum_{n=1}^N \mathbf{L}_n \phi_n(\mathbf{x}; \mathcal{M}) \quad \forall \mathbf{x} \in \Omega_T, \quad (9)$$

where $\mathbf{L}_n = \mathbf{L}(P_n)$ which represents the value of the probability map at the node P_n of the mesh \mathcal{M} . The task is now to find the nodal probabilities $\boldsymbol{\Theta} = [\mathbf{L}_n]_{n=1}^N$. The FEM approximation of the energy in (3) can be written as:

$$\Theta^* = \operatorname{argmin}_{\Theta \in \mathbb{R}^{NK}} E_D(\Theta; I_T, I_S, \mathcal{D}) + \beta E_R^{\text{diff}}(\Theta), \quad (10)$$

$$\text{with } E_D(\Theta; I_T, I_S, \mathcal{D}) = \int_{\Omega_T} \left\| \sum_{n=1}^N \mathbf{L}_n \phi_n(\mathbf{x}) - \lambda(\mathbf{x}; \mathcal{D}) \right\|^2 d\mathbf{x},$$

$$E_R^{\text{diff}}(\Theta) = \sum_{k=1}^K \sum_{n=1}^N L_n^k L_m^k \int_{\Omega_T} \nabla \phi_n^T(\mathbf{x}) W(\mathbf{x}) \nabla \phi_m(\mathbf{x}) d\mathbf{x},$$

where $\mathbf{L}_n = [L_n^k]_{k,n=1}^{K,N}$.

Setting the gradient of (10) to zero w.r.t each $\{\mathbf{L}_n\}_{n=1}^N$ and by re-writing $\Theta = [\Theta^k]_{k=1}^K$, where $\Theta^k = [L_m^k]_{m=1}^N$ is as a $N \times 1$ vector of nodal probabilities, we obtain the following system of linear equations for each $k \in \{1, 2, \dots, K\}$:

$$\mathbf{S}\Theta^k - \mathbf{F}^k = 0, \quad (11)$$

where $\mathbf{S} = [S_{m,n}]_{m,n=1}^N$ is a $N \times N$ matrix and $\mathbf{F}^k = [F_m^k]_{m=1}^N$ is a $N \times 1$ vector defined as:

$$S_{m,n} = \int_{\Omega_T} \phi_m(\mathbf{x}) \phi_n(\mathbf{x}) d\mathbf{x} + \beta \int_{\Omega_T} \nabla \phi_m^T(\mathbf{x}) W(\mathbf{x}) \nabla \phi_n(\mathbf{x}) d\mathbf{x},$$

$$F_m^k = \int_{\Omega_T} \lambda^k(\mathbf{x}; \mathbf{d}^k) \phi_n(\mathbf{x}) d\mathbf{x}. \quad (12)$$

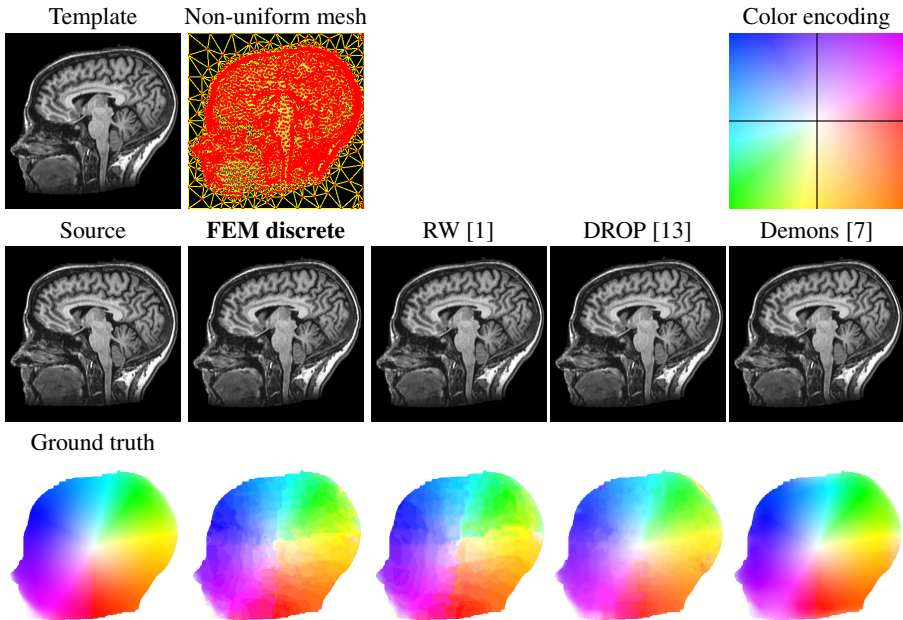
The system of independent $N \times K$ linear equations in (11) is solved to obtain the optimal nodal probabilities $\Theta^* = [\mathbf{L}_m^*]_{m=1}^N = [L_m^{*k}]_{k,m=1}^{K,N}$. Now, the optimal probabilities $\mathbf{L}^*(\mathbf{x}) = [L^{*k}(\mathbf{x})]_{k=1}^K$ at any spatial location $\mathbf{x} \in \Omega_T$ can be found through interpolation of the nodal probabilities $[\mathbf{L}_m^*]_{m=1}^N$ based on the finite element approximation in equation (9). Then, the estimated displacement field $\mathbf{U}(\mathbf{x})$ at $\mathbf{x} \in \Omega_T$ is obtained by choosing the displacement value \mathbf{d}^k with the highest optimal probability at that spatial location, i.e., $\mathbf{U}(\mathbf{x}) = \mathbf{d}^k$ where $k = \operatorname{argmax}_{r \in \{1, 2, \dots, K\}} L^{*r}(\mathbf{x})$.

3 Experiments

In this section, we evaluate our proposed FEM-based variational discrete deformable registration method on synthetic and real medical images. We implemented our method in a multi-resolution framework with 4 levels. The range of displacements used in each of the 4 levels are as follows: $[0, \pm 0.125, \dots, \pm 0.5]^\nu$, $[0, \pm 0.25, \dots, \pm 1.0]^\nu$, $[0, \pm 0.5, \dots, \pm 1.5]^\nu$, $[0, \pm 1.0]^\nu$, where $\nu = 2$ or 3 . We chose the Perona-Malik [16] diffusivity $W(\mathbf{x}) = 1/(1 + \exp(\|\nabla I_T\|^2/\gamma))$ with $\gamma = 0.05$ and set $\alpha = 1.0$ in the experiments below. An image-adaptive meshing strategy proposed by Yang *et al.* [17] was used to generate the non-uniform FEM mesh. We coded our method in MATLAB using the MEX facility. We ran all experiments on a Intel i7 3.60 GHz machine with 32GB RAM.

3.1 Comparison with Random Walker (RW), DROP and Demons Registration

In Figure 1, we show the performance of our FEM-based discrete registration method in comparison to the random walker (RW) [1], DROP [13] discrete registration methods and also the continuous diffusion-based demons [7] registration method on synthetic 2D brain MRI images with a known ground truth displacement field. The RW, DROP and demons methods all use a uniform mesh, while our FEM-based discrete registration method employs a non-uniform mesh well adapted to the image features. The number of DOFs associated with a mesh is given as twice the number of nodes in the mesh. Further, the computational times reported for the various algorithms correspond to the time taken by their respective optimization steps. In the case of our proposed method the optimization step involves solving the equation system in (11). It can be clearly seen that all the four registration methods successfully recover the displacement field between the two images. But, our FEM-based discrete registration method achieves



Method	AAE	SSDE	DOF	Time(sec)
Source (before registration)	—	4866.67	—	—
FEM discrete (proposed)	4.06°	102.73	11264	0.02
Random walker (RW) [1]	4.72°	142.96	331588	1.27
DROP [13]	2.48°	11.63	331062	54.74
Demons [7]	1.93°	31.35	331588	122.20

Fig. 1. Results on synthetic 2D MRI data. Average angular error (AAE), sum of squared differences error (SSDE) and degrees of freedom (DOF) are shown along with the computational times (we set $\beta = 50$ in our proposed method).

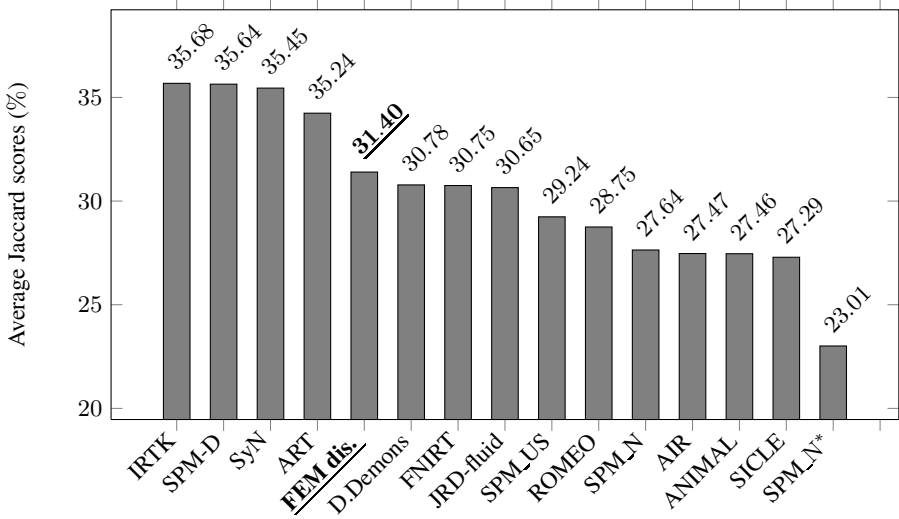


Fig. 2. Results on the CUMC12 MRI brain database (we set $\beta = 200$ in our proposed method)

this with a considerably lower computational effort (~ 30 times lower DOFs and > 60 times faster) when compared to the other three registration methods.

3.2 Validation on the CUMC12 MRI Database

The CUMC12 dataset [14] consists of 3D brain MRI scans from 12 subjects taken at $256 \times 256 \times 124$ resolution with a $0.86\text{mm} \times 0.86\text{mm} \times 1.5\text{mm}$ voxel spacing. Manual segmentations of 128 anatomical regions were provided for each of these images. We performed $12 \times 12 - 12 = 132$ pair-wise registrations between the images using our proposed FEM-based discrete registration method. The overall computational time for each registration is about 15.5 minutes of which 4.5 minutes are taken for the optimization step (solving equation system in (11)). For evaluating registration accuracy, anatomical labels on the source image were mapped to the template using the estimated displacement field. The Jaccard overlap measure was computed between the warped source labels and the template labels. In Figure 2, we report the Jaccard score obtained by our proposed method averaged over 128 anatomical regions and 132 pair-wise registrations. Klein et al. [18] reported average Jaccard scores obtained by 14 popular deformable registration methods on the CUMC12 dataset. We can see that our methods ranks in the top 5 among these methods.

4 Conclusion

We developed a computationally efficient FEM-based discrete deformable registration method using the squared differences similarity measure and diffusion-based regularization. A salient aspect of our method was the use of an image-adaptive discretization

of the problem domain. This resulted in our method being multiple orders of magnitude faster than the existing registration methods that are implemented using uniform meshes. A limitation of our method is that the estimated displacement field is not diffeomorphic. This is because the displacement field is not explicitly regularized and the regularization is only implicit through the smoothing of prior probability maps. In future, we want to explore the use of an additional explicit regularization step that would yield diffeomorphic displacement fields. Further, we intend to extend our method for multi-modality registration using more complex similarity measures.

References

1. Cobzas, D., Sen, A.: Random walks for deformable image registration. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 557–565. Springer, Heidelberg (2011)
2. Sotiras, A., Paragios, N., et al.: Deformable image registration: A survey. INRIA Research report (2012)
3. Broit, C.: Optimal registration of deformed images. PhD thesis, Univ. of Penn. (1981)
4. Modersitzki, J.: Numerical methods for image registration. Oxford Univ. Press (2004)
5. Fischer, B., Modersitzki, J.: Curvature based image registration. *Journal of Mathematical Imaging and Vision* 18(1), 81–85 (2003)
6. Pennec, X., Cachier, P., Ayache, N.: Understanding the demon’s algorithm: 3D non-rigid registration by gradient descent. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 597–605. Springer, Heidelberg (1999)
7. Stefanescu, R., Pennec, X., Ayache, N.: Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis* 8(3), 325–342 (2004)
8. Cahill, N., Noble, J., Hawkes, D.: Demons algorithms for fluid and curvature registration. In: ISBI, pp. 730–733 (2009)
9. Rohde, G., Aldroubi, A., Dawant, B.: The adaptive bases algorithm for intensity-based non-rigid image registration. *IEEE Tran. on Medical Imaging* 22(11), 1470–1479 (2003)
10. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Tran. on Medical Imaging* 18(8), 712–721 (1999)
11. Jordan, P., Socrate, S., Zickler, T., Howe, R.: Constitutive modeling of porcine liver in indentation using 3d ultrasound imaging. *Journal of the Mechanical Behavior of Biomedical Materials* 2(2), 192–201 (2009)
12. Tang, T.W.H., Chung, A.C.S.: Non-rigid image registration using graph-cuts. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 916–924. Springer, Heidelberg (2007)
13. Glocker, B., Komodakis, N., Tziritas, G., et al.: Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis* 12(6), 731–741 (2008)
14. Columbia univ. medical centre mri dataset, www.mindboggle.info/papers/evaluation_NeuroImage2009/data/CUMC12.tar.gz (accessed: March 1, 2013)
15. Grady, L.: Multilabel random walker image segmentation using prior models. In: CVPR, vol. 1, pp. 763–770 (2005)
16. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Tran. on Pattern Analysis and Machine Intelligence* 12(7), 629–639 (1990)
17. Yang, Y., Wernick, N., Brankov, G.: A fast approach for accurate context-adaptive mesh generation. *IEEE Transactions on Image Processing* 12(8), 866–881 (2003)
18. Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M., Christensen, G., Collins, D., Gee, J., Hellier, P., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage* 46(3), 786–802 (2009)

FLOOR: Fusing Locally Optimal Registrations

Dong Hye Ye¹, Jihun Hamm², Benoit Desjardins¹, and Kilian M. Pohl¹

¹ Department of Radiology, University of Pennsylvania, Philadelphia, PA, 19104

² Department of Computer Science, Ohio State University, Columbus, OH, 43210

Dong.Ye@uphs.upenn.edu

Abstract. Most registration algorithms, such as Demons [1], align two scans by iteratively finding the deformation minimizing the image dissimilarity at each location and smoothing this minimum across the image domain. These methods generally get stuck in local minima, are negatively impacted by missing correspondences between the images, and require careful tuning of the smoothing parameters to achieve optimal results. In this paper, we propose to improve on those issues by choosing the minimum from a set of candidates. Our method generates such candidates by running the registration algorithm multiple times varying the setting of the smoothing and the image domain. We iteratively refine those candidates by fusing them with the outcome of alternative approaches and locally adapting the smoothing parameters. We implement our algorithm based on Demons [1] and find alternative minima via manifold learning [2]. Compared to those two methods, our 600 pairwise registrations of cardiac MRIs significantly better handle the large shape variations of the heart and the different field of views captured by scans.

Keywords: Image Registration, Manifold Learning, Cardiac MRI.

1 Introduction

General-purpose registrations, such as [1], model deformable image alignment as a problem of finding the (local) minimum of an energy function. The accuracy of the search is sensitive to the setting of the optimization method, which commonly assumes one-to-one correspondence between moving and target scan. Scans, such as cardiac MRI (see Fig. 1), often vary with respect to the anatomy captured in their field of view as well as the shape of the anatomy, such as the heart. It is thus difficult to find the optimal setting for methods off-line. In this paper, we address this shortcoming of existing tools by refining their solution to the minimization problem through iteratively fusing the results of multiple searches.

Applications relying on non-rigid registration, such as atlas-based segmentation, contain various coping mechanisms for these issues. The most common technique is to crop the scan to a specific region, such as in skull stripping, before refining registration. Accurate cropping is only reliable for certain areas, such as brain MRI, but even then generally requires manual editing. Alternatively, researchers register scans multiple times or from multiple subjects and then fuse the aligned label maps [3].

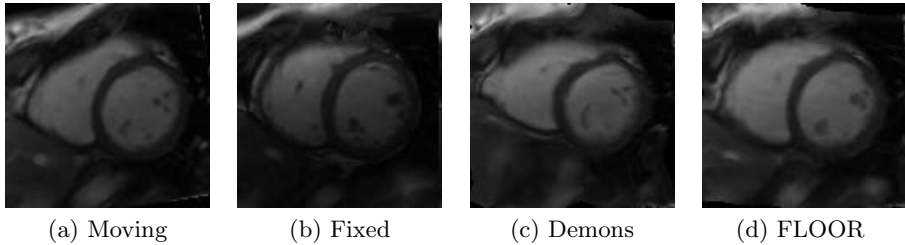


Fig. 1. (c) shows the result of Demon aligning (a) to (b). The artifacts shown in (c) are caused by missing correspondences. FLOOR’s output (d) much better matches (b).

These mechanisms are less beneficial when directly analyzing deformations, such as in [4]. Here, the optimal registration setting is often determined via cross validation [5] hoping the results during training translate to new scans. This assumption generally does not hold for heterogeneous populations. Alternatively, graphical models find the global instead of a local minimum [6]. The accuracy of these methods suffers from the discretization of the deformation map. Finally, manifold learning [2,7] improves registration by decomposing the problems into a set of simpler registration tasks. The accuracy of these methods depends on the size of the training data as well as the accuracy of the solutions to those simpler tasks. To address the first issue, manifolds are created for image parcels instead of the entire domain[8,9]. Our proposed method, called Fusing Locally Optimal Registrations (FLOOR), is a first attempt to advance these learning techniques with respect to the second issue by iterating between improving the tasks’ solutions and refining the corresponding manifolds. To the best of our knowledge, FLOOR is also the first method to improve deformation maps (and not aligned segmentations as in [3]) by fusing the outcome of several registrations.

FLOOR improves the accuracy of a registration by iteratively reducing the corresponding minimum based on a set of local optima, called candidates. At each iteration, FLOOR generates candidates by applying the registration with different smoothing settings to various parcels of the image and including the outcome of alternative methods, such as manifold learning. It improves the minimum by fusing the candidates, i.e. selecting the deformation at each voxel that minimizes the energy function. Note, all candidates are local minima of the energy function but differ in the degree of flexibility, the parcel they are defined on, and the way they were generated. The approach converges to a smooth deformation map after the selection process does not improve candidates.

FLOOR addresses many of the previous issues as it draws from multiple searches, contains the influence of missing correspondences to parcels instead of the entire domain, and adapts the stiffness locally and online. If we include Manifold Learning based Deformation maps (MLD)[2] in the candidate set, the corresponding manifolds are also likely to be improved at each iteration as they are trained on refined pairwise registrations. Our implementation of FLOOR demonstrates this by fusing the results of Demons[1] with MLD. We register 600 pairs of cardiac MRIs. FLOOR not only significantly improves the result of Demons, such as in Fig. 1, but also of MLD trained on the entire domain.

2 Fusing Deformation Maps

Non-linear registration approaches generally define the best mapping between two images F and M as the spatial transformation \mathbf{s} that minimizes $\text{Sim}(F, M \circ \mathbf{s})$, the dissimilarity between the aligned images, while being smooth according to the regularizer $\text{Reg}(\mathbf{s})$. In other words, \mathbf{s} minimizes the energy function

$$E(\mathbf{s}) \triangleq \text{Sim}(F, M \circ \mathbf{s}) + \frac{1}{\sigma^2} \text{Reg}(\mathbf{s}), \quad (1)$$

where σ controls the flexibility of the deformation and $E(\cdot)$ is defined with respect to the image domain Ω . We now briefly review the Demons algorithm [1], which finds a local minimum for a specific instance of this equation. We then derive FLOOR, which iteratively refines those results by spatially varying σ , confining $E(\cdot)$ to parcels and fusing them with alternative deformation maps.

Demons: The method defines both terms of Eq. 1 with respect to L2-norm $\|\cdot\|$, where $\text{Sim}(F, M \circ \mathbf{s}) \triangleq \|(F - M \circ \mathbf{s})\|^2$ and $\text{Reg}(\mathbf{s}) \triangleq \|\nabla \mathbf{s}\|^2$. The algorithm determines the local minimum \mathbf{s} of $E(\cdot)$ by iteratively minimizing for \mathbf{s} and an approximation of \mathbf{s} , called \mathbf{c} , according to

$$E(\mathbf{s}, \mathbf{c}) \triangleq \sum_{x \in \Omega} \left(\|F(x) - M(x + \mathbf{c}(x))\|^2 + \|\mathbf{c}(x) - \mathbf{s}(x)\|^2 + \frac{1}{\sigma^2} \|\nabla_x \mathbf{s}\|^2 \right). \quad (2)$$

$\nabla_x \mathbf{s}$ is the gradient of \mathbf{s} at x . Given \mathbf{s} , it first updates \mathbf{c} at each voxel $x \in \Omega$:

$$\mathbf{c}(x) \leftarrow \arg \min_{\mathbf{t}(x) \in \mathbb{R}^3} \|F(x) - M(x + \mathbf{t}(x))\|^2 + \|\mathbf{t}(x) - \mathbf{s}(x)\|^2, \quad (3)$$

which has a closed form solution. Keeping \mathbf{c} fixed, the method updates \mathbf{s} by applying \mathbf{c} to a Gaussian kernel $K(\sigma)$ with the width depending on σ , i.e.

$$\mathbf{s} \leftarrow K(\sigma) \star \mathbf{c}. \quad (4)$$

The algorithm iterates between Eq. 3 and Eq. 4 until convergence.

FLOOR: Demons assumes the degree of stiffness of the deformation controlled by σ and the error measured by $\text{Sim}(\cdot, \cdot)$ is uniformly defined across the domain Ω . The method also does not improve the results once it converges to a local optimum. These assumptions generally reduce Demons accuracy when, for example, missing correspondences in one region impact the deformation in the remaining image domain. FLOOR drops these assumptions by iteratively fusing and improving the results of Demons based methods (see Algorithm 1).

FLOOR drops the assumption of the uniform stiffness across Ω by replacing σ in Eq. 2 with $\sigma(\cdot)$, a smoothly varying map. At each voxel $x \in \Omega$, $\sigma(\cdot)$ is confined to a small range \mathbb{S} , such as $\mathbb{S} = [0.5, 4.5]$. Given that $\sigma(\cdot)$ should be specific to the corresponding image pair, we assume that our approach is initialized with a set of candidates $\mathbb{C} \triangleq \{C_1, \dots, C_M\}$ where each candidate $C_i \triangleq (\mathbf{c}_i, \sigma_i)$ consists of a deformation map \mathbf{c}_i that minimizes Eq. 2 for the specific smoothing map σ_i . We furthermore assume that all of those candidates are close to the true

deformation \mathbf{s} and smoothing map $\sigma(\cdot)$. Thus, $\|\mathbf{c}_i(x) - \mathbf{s}(x)\|^2 \approx 0$ in Eq. 2 so that we omit the term in Eq. 3 and improve the estimate of $\mathbf{c}(\cdot)$ at $x \in \Omega$ via

$$(\mathbf{c}'(x), \sigma'(x)) \leftarrow \arg \min_{\mathbf{C}_i \in \mathbb{C}} \|F(x) - M(x + \mathbf{c}_i(x))\|^2. \quad (5)$$

If we assume that $\sigma'(x)$ is constant in close proximity of x , i.e. is locally smooth, then $\mathbf{c}'(x)$ is equal to $\mathbf{s}(x)$. To enforce smoothness, we first compute $\sigma'(\cdot)$ over Ω and then smooth $\sigma'(\cdot)$ similar to Eq. 4 via $\hat{\sigma} \leftarrow K(\gamma) \star \sigma'$ with γ being the smoothing parameter specific to the smoothing map. Next, we replace the Gaussian kernel in Eq. 4 with one whose width at $x \in \Omega$ is defined by $\hat{\sigma}(x)$:

$$\mathbf{s}(x) \leftarrow K(\hat{\sigma}(x)) \star_x \mathbf{c}'. \quad (6)$$

where \star_x is the convolution at x . We call this process *fusing* deformation maps.

Next, we drop the assumption of misalignment being uniformly accounted for across the domain Ω by fusing candidates, who are optimal solutions with respect to specific image parcels. Specifically, we first parcellate Ω into overlapping regions $\mathbb{O} \triangleq \{\Omega_1, \dots, \Omega_N\}$ so that $\Omega_i \subset \Omega$, $\Omega_i \cap (\mathbb{O} \setminus \{\Omega_i\}) \neq \emptyset$, and $\cup_{i=1}^N \Omega_i = \Omega$. If we now replace Ω in Eq. 2 with Ω_i then the deformation map \mathbf{s} of Eq. 4 is also only defined for Ω_i and not impacted by the mismatches between the images outside of Ω_i . We then create the candidates of \mathbb{C} by applying Demons to Eq. 2 confined by samples taken from \mathbb{O} and the smoothing interval S . Now fusing those candidates according to Eq. 5 and Eq. 6 results in a deformation map over Ω that is the minimum solution with respect to maps defined on overlapping parcels instead of the entire image domain. Unlike the original outcome of Eq. 4, the impact of missing correspondences on \mathbf{s} are thus locally confined.

Finally, we incorporate in FLOOR deformation maps from other registration algorithms, which are initialized by the current set \mathbb{C} . An example of such an algorithm is MLD[2]. We apply MLD to each collection of candidates, which register a group of images using the same smoothing setting and parcel. For each set, MLD encodes the manifold as a kNN graph with each node representing an image and the weight of the edges between nodes defined by the similarity measure of Eq. 5. MLD's solution to Eq. 2 is then based on the manifold's geodesic path from M to F . The "geodesic" deformations are added to \mathbb{C} and \mathbf{s} is updated according to Eq. 5 and Eq. 6. We then learn the manifold based on the new set of \mathbf{s} defined over the group of images. We repeat the process of updating \mathbf{s} and the manifold until convergence. Note, the algorithm has to converge as each iteration reduces the minimum of Eq. 5.

By improving the pairwise registrations, we also expect the accuracy of the manifolds to be improved. FLOOR thus address an issue of current manifold learning techniques, whose accuracy depend on the initial, pairwise registrations, by iteratively fusing pairwise with manifold-based deformation maps. This idea applies to many other registration algorithms. The fusing of deformation maps as proposed by FLOOR is simple as long as one can first find the minimum locally and then modify the corresponding deformation map to comply with global constraints.

Algorithm 1. FUSING LOCALLY OPTIMAL REGISTRATIONS(\mathbb{S}, \mathbb{O})

Create a set of candidates \mathbb{C} by running Demons varying $\sigma \in \mathbb{S}$ and $\Omega \in \mathbb{O}$

repeat

STEP 1: Train MLDs for a set of images based on \mathbb{C} . Add the results to \mathbb{C}

STEP 2: Find the best candidate (c', σ') according to Eq. 5 and smooth σ'

STEP 3: Update s via Eq. 6 and add to \mathbb{C}

until s converges

3 Experiments on Cardiac MRI Dataset

We analyze the accuracy of FLOOR on the pairwise registrations of 25 cardiac MRIs at end-diastole. We measure the accuracy for the 600 registrations (25×24) by the Dice score [10] between the aligned and corresponding manual segmentations of the right ventricle blood pool (RV), the left ventricle myocardium (MY) and blood pool (LV) (see Fig. 2(a)). RV and LV have similar intensity patterns and MY is relatively thin. Thus, the Dice score with respect to these structures is a good indicator for the quality of deformations. For comparison, we measure the accuracy of different implementations of FLOOR.

Data Set: The set consists of short-axis MRIs at end-diastole of 15 healthy and 10 Tetralogy of Fallot (TOF) subjects. Registering these scans is difficult due to missing correspondences, low image quality, and the irregular ventricular shapes of TOF hearts (see Fig. 1). The resolution of each scan is $1.25 \times 1.25 \times 8mm^3$. To speed up registration, we crop scans to 12 slices showing the ventricles and 120×120 inplane voxels centered at the heart. Scans are bias corrected [11] and linearly aligned to a template [12]. Finally, an expert segmented LV, RV and MY in each scan via [13]. Note, these maps are only used for the Dice computation.

Parameters: Parameters of FLOOR are the parcellation \mathbb{O} , the set of smoothing parameters \mathbb{S} , and the size K of the neighborhood of the kNN graph generated by MLD. Out of convenience, we parcellate the scans into $3 \times 3 \times 2$ overlapping parcels as shown in Fig. 2(b). Each parcel is of size $55 \times 55 \times 7$ voxels with the overlapping regions shown in red and blue. The set of smoothing widths is $\mathbb{S} = \{0.5, 1.5, 2.5, 3.5, 4.5\}$ as Demons achieves the highest overall accuracy for $\sigma = 2.5$. Note, we now refer to σ as the actual width of the Gaussian kernel of Eq. 4 and not the weight of Eq. 2, which indirectly defines σ . Thus, the larger σ the less flexible the resulting deformation. Finally, we choose $K = 10$, which is double the neighborhood size of the thinnest connected kNN graph.

Implementations: As a baseline, we apply Demons [1] to the entire image with fixed $\sigma = 2.5$ and train MLD on the resulting 600 pairwise registrations (called GMLD). Next, we run Algorithm 1 omitting Step 1, i.e. not generating candidates via MLD. Specifically, we fuse the outcome of Demons applied to the entire image domain with the different smoothness parameter σ of \mathbb{S} (sDemon), applied to the different parcels of \mathbb{O} but with fixed $\sigma = 2.5$ (pDemon), and by

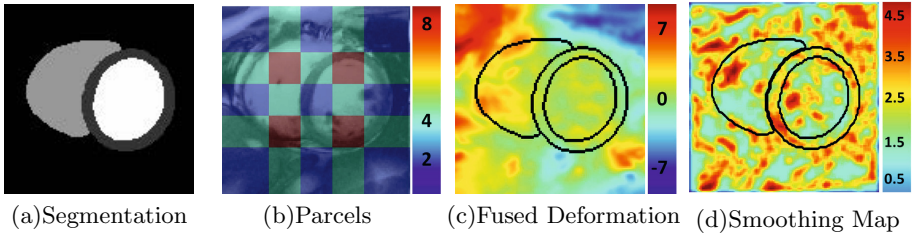


Fig. 2. (a) Expert segmentation (LV: white, RV: light grey, MY: dark grey) of Fig. 1(b). (b) The image is divided into $3 \times 3 \times 2$ parcels with the colors indicating the number of overlapping parcels. (c) shows the y-displacement of the fused deformation map and (d) the smoothing map ($\sigma(\cdot)$) generated by FLOOR. Both maps are smooth.

running Demons with different σ as well as on different parcels (fDemon). Finally, we run FLOOR as specified in Alg. 1. pFLOOR applies Demons and learns a MLD for each of the parcels of \mathbb{O} with fixed $\sigma = 2.5$. fFLOOR applies Demons and learns a MLD for different smoothing setting as well as parcels. It also learns a new MLD based on the updated (600) deformations until convergence.

Comparison: We now review Table 1, which lists each implementation’s average Dice score (*Avg*) and standard deviation (*std*), denoted by $Avg \pm std\%$, across all pairwise registrations for each structure and across structures, i.e. 600×3 scores. Note, we call improvements significant if the p-value is below 1% for the unpaired t-test of the scores of two implementations across all image pairs.

Out of all implementations, GMLD received the lowest score (72.9%), which was slightly lower than Demons. The decrease in accuracy indicates that the neighborhood size K was too small. Increasing K did not greatly improve the score until we set it close to 25, which means that every image pair is directly connected, i.e. the results are equivalent to Demons. An alternative explanation for the low score of GMLD is the number of training samples (25), which was too small to properly learn the manifold over the entire image domain. This is a common problem in medical imaging [8,9] and a motivation behind FLOOR.

The score increases by simply fusing the results of Demons, when run with different smoothing parameters (sDemon). There is significant improvement for pDemon and fDemon, who both apply Demons to different parcels. These results indicate that for our data set the overall accuracy of Demons is mostly impacted by missing correspondences in the background and not the uniform smoothness setting of the deformation map. The way our approach addresses this problem is by running Demons on different parcels so that not all of the results are corrupted by the same missing correspondences during the fusion process (Step 2 in Algorithm 1). The corruption free maps most likely will be chosen over the corrupted ones resulting in a more accurate registration.

Finally, the scores further improve when including MLDs such as for pFLOOR vs. pDemon. Unlike for GMLD, whose manifold is based on the entire image domain, the training set is now large enough to properly learn the manifold for specific parcels. fFLOOR receives the highest average scores with 77.9 % after the first iteration and converges to 79.2 % after three iterations. This score is

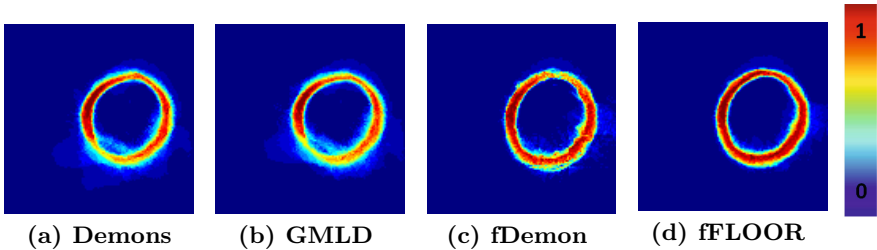


Fig. 3. Average across 24 binary maps of MY aligned to Fig. 1(b) with (d) being best

Table 1. Average and standard deviation of Dice scores of 600 registrations for each individual structure and across all 3 structures. ffLOOR achieves the highest scores

	GMLD	Demon	sDemon	pDemon	fDemon	pFLOOR	ffLOOR
LV	84.9 ± 0.07	85.3 ± 0.06	86.5 ± 0.06	88.3 ± 0.04	88.6 ± 0.04	88.7 ± 0.04	89.7 ± 0.04
MY	58.6 ± 0.11	59.7 ± 0.10	58.0 ± 0.09	65.1 ± 0.08	64.4 ± 0.08	65.8 ± 0.08	68.1 ± 0.08
RV	75.2 ± 0.09	75.8 ± 0.09	76.5 ± 0.09	78.5 ± 0.08	78.9 ± 0.08	79.1 ± 0.08	79.7 ± 0.08
ALL	72.9 ± 0.14	73.6 ± 0.13	73.7 ± 0.13	77.3 ± 0.11	77.6 ± 0.11	77.9 ± 0.12	79.2 ± 0.11

significantly better than fDemon. Thus, manifold learning really improves the results of FLOOR as it provides alternative candidates that locally minimize Demons energy function. Furthermore, ffLOOR produces smooth deformation and smoothing maps as also shown in Fig. 2(c-d). The smooth deformation maps indicate that our method correctly fuses the different candidates produced by ffLOOR. The variation in the smoothing width map support our claim that optimal regularization parameter differ across the image. We end the comparison noting that ffLOOR was significantly better than the state of the art represented by 5.6 % improvement over Demons and 6.3 % over GMLD.

Our findings based on the average Dice score also translates to the scores of the individual structures. Most noticeable are the differences of the implementations with respect to MY, for which the average score of ffLOOR was 68.1% which is 3.7% higher than fDemon, 8.4% higher than Demons, and 9.5% higher than GMLD. The MY is smaller than RV and LV so that the Dice score is most sensitive to deformation errors. This is also reflected by the average maps shown in Fig. 3. The maps are the results of aligning 24 cases to the scan of Fig. 1 (b) and then applying the deformations to the corresponding MY segmentations. The map produced by ffLOOR shows the least amount of variation across the 24 aligned segmentations. The map of fDemon varies a little bit more but far less than the maps of Demons and GMLD. These results further confirm our intuition behind FLOOR that iteratively fusing the deformation maps of pairwise registration with parcel-specific manifolds further improves registration.

4 Conclusion

We derived FLOOR, an algorithm for improving registrations by fusing multiple solutions to the corresponding minimization problem. FLOOR generated

such solutions, called candidates, by applying known methods, such as Demons, varying the regularization and registration regions. FLOOR can also include candidates from other methods, such as MLD, in an iterative fashion. Compared to those original methods, the 600 pairwise registrations of cardiac MR scans of FLOOR were much more accurate in dealing with the large shape variations of the heart and the difference in the field of views across scans. We expect similar results when adapting FLOOR to other registration methods and scans.

Acknowledgment. The project was supported in part by the NIH Grant UL1RR024134 and in part by the Institute for Transnational Medicine and Therapeutics'(ITMAT) Trans-disciplinary Program in Transnational Medicine and Therapeutics.

References

1. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. *NI* 45, S61–S72 (2009)
2. Hamm, J., Ye, D.H., Verma, R., Davatzikos, C.: GRAM: A framework for geodesic registration on anatomical manifolds. *MedIA* 14(5), 633–642 (2010)
3. Rohlfing, T., Russakoff, D., Maurer, C.R.J.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *TMI* 23(8), 983–994 (2004)
4. Ashburner, J., Friston, K.J.: Voxel-based morphometry - the methods. *NI* 11, 805–821 (2000)
5. Yeo, B.T.T., Sabuncu, M.R., Desikan, R., Fischl, B., Golland, P.: Effects of registration regularization and atlas sharpness on segmentation accuracy. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 683–691. Springer, Heidelberg (2007)
6. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. *MedIA* 12(6) (2008)
7. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S.C., Whitaker, R.T.: Manifold modeling for brain population analysis. *MedIA* 14(5), 643–653 (2010)
8. Ye, D.H., Hamm, J., Kwon, D., Davatzikos, C., Pohl, K.M.: Regional manifold learning for deformable registration of brain MR images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 131–138. Springer, Heidelberg (2012)
9. Bhatia, K.K., Rao, A., Price, A.N., Wolz, R., Hajnal, J., Rueckert, D.: Hierarchical manifold learning. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 512–519. Springer, Heidelberg (2012)
10. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* 26(3), 297–302 (1945)
11. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: improved N3 bias correction. *TMI* 29(6), 1310–1320 (2010)
12. Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M.: Bayesian analysis of neuroimaging data in fsl. *NI* 45(suppl. 1), S173–S186 (2009)
13. Heiberg, E., Ugander, M., Carlsson, M., Engblom, H., Arheden, H.: Design and validation of segment–freely available software for cardiovascular image analysis. *BMC Med. Imaging* 10, 1 (2010)

Particle-Guided Image Registration

JooHWi Lee¹, Ilwoo Lyu¹, İpek Oğuz³, and Martin A. Styner^{1,2}

¹ Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

² Department of Psychiatry, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

³ Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52246, USA

Abstract. We present a novel image registration method based on B-spline free-form deformation that simultaneously optimizes particle correspondence and image similarity metrics. Different from previous B-spline based registration methods optimized w.r.t. the control points, the deformation in our method is estimated from a set of dense unstructured pair of points, which we refer as *corresponding particles*. As intensity values are matched on the corresponding location, the registration performance is iteratively improved. Moreover, the use of corresponding particles naturally extends our method to a group-wise registration by computing a mean of particles. Motivated by a surface-based group-wise particle correspondence method, we developed a novel system that takes such particles to the image domain, while keeping the spirit of the method similar. The core algorithm both minimizes an entropy based group-wise correspondence metric as well as maximizes the space sampling of the particles. We demonstrate the results of our method in an application of rodent brain structure segmentation and show that our method provides better accuracy in two structures compared to other registration methods.

1 Introduction

The study of brain changes in rodent models of neuropathology and drug exposure has been of increasing interest to the neuroscience community. In contrary to human studies, rodent models have several advantages, such as a well controlled environments and access to genetic modifications as well as shorter lifespan. Magnetic Resonance Imaging (MRI) has emerged as an important modality to study such rodent brain morphological changes. Non-rigid registration is a crucial tool to process such MRIs providing structural segmentations and enabling the analysis of group differences.

Several methods have been proposed for the study of rodent brains. Among those atlas-based registration methods are popularly used. However, a single atlas-based method has a disadvantage of the introduction of bias that might cause poor segmentation and dilute the difference between groups [6]. Group-wise registration method which deals with every subject together can be an

alternative to reduce the effects of template selection [2]. Since a group-wise registration method does not require the choice of a template or reference, it is expected to produce consistent results which means a consistent comparison of groups.

Motivated by a particle correspondence algorithm [3], a non-parametric and group-wise surface correspondence method, we propose a novel group-wise image registration method guided by particles. These particles are distributed inside a particular region and directly optimized so that each particle will be placed at corresponding positions across subjects minimizing a group-wise intensity metric. During this optimization, a B-spline free-form deformation is estimated for each subject to constitute a common reference frame. The contributions of the proposed method are following:

1. **Unbiased Group-Wise Registration with Implicit Mean:** Instead of choosing a specific template, a common reference frame is estimated from dynamic particles distributed inside ROIs, *i.e.* a brain mask. Using the Euclidean mean of those particles, each subject is efficiently registered into a common space.
2. **Computational Efficiency with Particles:** Since the number of particles is fewer than the number of voxels, a common reference frame is more efficiently computed than other methods [2]. We compensate this sparsity by considering a local patch for each particle, which also provides robust performance than single voxel random sampling strategy.
3. **Flexibility in Adaptive Processing:** In contrary to a regular control point grid, particles are unstructured and independent each other so that it is easy to adopt adaptive strategy depending on local context. For example, particles can be easily placed more densely in salient areas, *i.e.* edges, by controlling a single parameter.

As our work is in an early stage, we demonstrate preliminary results of rodent brain structure segmentation with comparison to two different registration methods for humans: the spline-based FFD available in Slicer and to SyN available in ANTS. We show that our group-wise algorithm performs better in different sizes as well as produces statistically indifferent results with the comparing methods otherwise.

2 Methods

We propose a group-wise image registration method guided by dynamic particles. The structure of our method is similar to the surface-based particle correspondence algorithm [3] and can be thought as an extension of the algorithm. The application of particles in an image domain, however, has never been attempted. Our method is also uniquely different from the previous one in that we introduce B-spline free-form deformation to associate different subject spaces as well as deal with local patch information for robust performance.

2.1 Particle Correspondence with Local Similarity

The main goal of our method is to drive each particle toward a corresponding position that satisfies two conditions in the mean space: 1) overlapping of particles and 2) local intensity similarity. The particles are governed by two forces: a positional coherence force and a force from local intensity similarity.

To describe the motion of particles, we define the particle system \mathbf{P} that comprises N number of subject volumes $\mathbf{V} = \{V^1, V^2, \dots, V^N\}$. For each subject j , we sample the same n number of particles $P^j = \{p_1^j, p_2^j, \dots, p_n^j\}$ in which $p_i = \{p_i^1, p_i^2, \dots, p_i^N\}$ is the corresponding particles from each subject. From these correspondences, an implicit mean space \bar{V} is estimated from the mean of particles $\bar{P} = \frac{1}{N} \sum_{j=1}^N P^j$. Our group-wise registration process is formulated to find an optimal particle configuration \check{p}_i^j that minimizes a positional coherence metric $H_P(p_i)$ and a local intensity similarity metric $H_I(p_i)$. The final dense deformation field \check{T}_j that maps V^j to \bar{V} is derived by taking P^j and \bar{P} as a set of correspondences.

Correspondence Formulation. By the transform T_j , a particle p_i^j is mapped to $q_i^j \in \bar{V}$. Ideally, it is assumed that $q_i^1 \simeq q_i^2 \simeq \dots \simeq q_i^N$. Therefore, each particle of p_i should move to the direction where the variance of q_i is minimized as depicted in Figure 1a. In the mean time, if there is local differences in intensity values, the particles are allowed to deviate from the overlapping position so that the local variance of intensity values are minimized such that $I^1(q_i^1) \simeq I^2(q_i^2) \simeq \dots \simeq I^N(q_i^N)$ where $I^j = V_j(T_j^{-1}(q_i))$. To compare similarity among a group, [3] and [2] approach in similar using entropy. The entropy of a random variable q with a given p.d.f $f(q)$ is minimized when there is less information in q and formulated $H(q) = -\int f(q) \log(f(q)) dq$. Denoting the random variable as q_i and \mathcal{I}_i respectively for q_i^j and $I^j(q_i^j)$, the goal is to find the optimal particle configuration $\hat{\mathbf{P}}$ such that

$$\hat{\mathbf{P}} = \arg \min_{P^j \in \Omega_j} S(P) = \lambda_P \sum_{i=1}^n H_P(q_i) + \lambda_I \sum_{i=1}^n H_I(\mathcal{I}_i). \quad (1)$$

Since the number of particles is much smaller than the number of voxels, we sample a local patch near by a particle q_i^j so that $\mathcal{I}_i^j = \{I_i^j(1), I_i^j(2), \dots, I_i^j(M)\}$, where M is the number of neighborhoods of q_i^j .

Correspondence Optimization. Given covariance matrices of Σ and Λ that follows N , we derive H_P and H_I analytically [7] so that

$$H_P(q_i) = -r + \frac{1}{2} \ln(2\pi)^r |\Sigma|, \quad H_I(\mathcal{I}_i) = -M + \frac{1}{2} \ln(2\pi)^M |\Lambda|, \quad (2)$$

where r is the dimension of q and $|\Sigma|$ and $|\Lambda|$ are the determinants. The gradient of H_P and H_I in the space of V are given $\frac{\partial H_P}{\partial p} = \mathbf{J}_{T_j^{-1}} \check{p}^j (\Sigma + \alpha I)^{-1}$ and

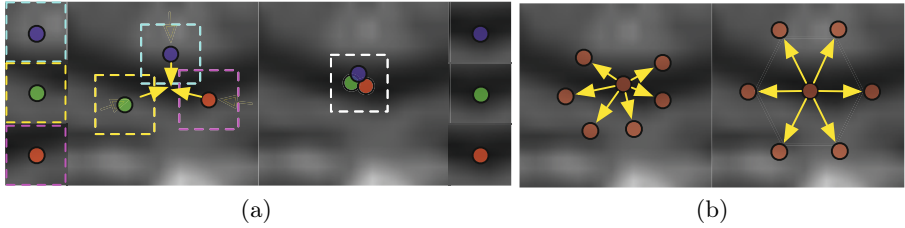


Fig. 1. Schematic diagram of (a) overlapping particles and local intensity similarity in correspondence across subjects. Colored in blue, green, and red, each particle has correspondence across subjects and attracts together minimizing H_P . At the same time, the entropy of local intensities sampled in colored squares is also minimized so that the particles stay at a locally similar position. (b) a repulsion force uniformly distributes in-subject particles to fill a given region.

$\frac{\partial H_I}{\partial p} = \mathbf{J}_{T_j^{-1}} I' (\Lambda + \beta)^{-1} \frac{\partial I}{\partial p}$, respectively. where \tilde{p}' and I' are displacement from the mean and α, β is a relaxation factor to avoid degenerative cases.

From the particle perspective, the negative gradient direction, $-\frac{\partial H_P}{\partial p}$ and $-\frac{\partial H_I}{\partial p}$ can be interpreted as two different forces: a positional coherence force and an intensity force as depicted in Fig. 1.

2.2 Particle Sampling in a Volume

Corresponding particles across subjects are attracted together to be overlapped at a locally similar position. Without an appropriate repulsion force, the particles would degenerate to a single point. Moreover, since we sample local intensity values nearby a particle, a repulsion force is required to uniformly sample a given image domain. In order for that, we extend the surface-based particle correspondence algorithm [3] to the image domain to uniformly sample a set of particles in a volume. In the algorithm, each particle position is rendered as a random variable with regard to a particular region and iteratively optimized to maximally contain the spatial information of the region.

Problem Definition. Given a bounded region of interest Ω in a volume V , we sample n number of points $X = (x_1^T, x_2^T, \dots, x_n^T)^T \in \Omega^n$ where $x_i = (x, y, z)$. By letting \mathcal{X} be a random variable of X , the goal is to find an instance \tilde{X} such that

$$\tilde{X} = \arg \max_{\tilde{X} \in \Omega^n} H_S[\mathcal{X}] = \arg \min_{\tilde{X} \in \Omega^n} \sum_{i=1}^n \int_{x_i \in \Omega} p(x_i) \log p(x_i) dx_i, \quad (3)$$

where H_S is the differential entropy

$$H_S(\mathcal{X}) = - \int_{X \in \Omega^n} p(X) \log p(X) dX, \quad (4)$$

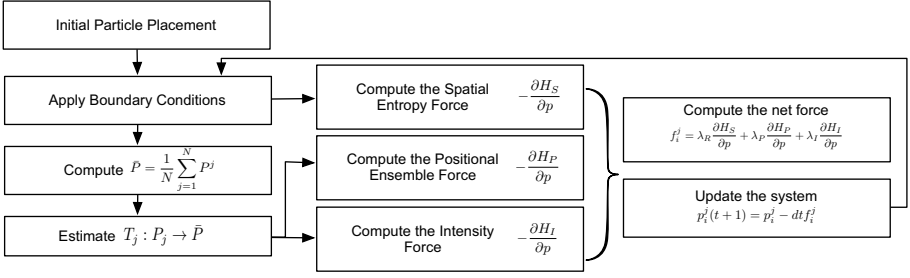


Fig. 2. Overall algorithm flow. The registration process is finished when the system stabilizes, and images are registered with the estimated T_j .

and $p(X)$ is the p.d.f of X . Assuming x_i is i.i.d, $H_S(\mathcal{X})$ can be decomposed into the sum of the spatial entropy $H_S(x_i)$. From the definition, \tilde{X} contains maximal information of Ω .

Sampling Optimization. A key step to compute $H_S[\mathcal{X}]$ is the density estimation of $p(x)$. The density for a particle is estimated as $p(x_i) = \frac{1}{n(n-1)} \sum_{j=1, j \neq i}^n G(x_i - x_j, \sigma_i)$ using a nonparametric, Parzen windowing estimation[3] with the assumption of Gaussian. The negative gradient of $H_S[\mathcal{X}]$ to maximize the cost function is

$$-\frac{\partial H_S}{\partial x_i} = \frac{1}{\sigma_i^2} \frac{\sum_{j \neq i}^n (x_i - x_j) G(x_i - x_j, \sigma_j)}{\sum_{j \neq i}^n G(x_i - x_j, \sigma_j)} = \sigma_i^{-2} \sum_{j \neq i}^n (x_i - x_j) w_{ij}. \quad (5)$$

For the optimization, we employ a standard gradient descent optimization via Euler scheme, $x_{t+1} = x_t - \alpha \frac{\partial H_S}{\partial x}$. The control of adaptivity is achieved by assigning different σ_j for each particle [5].

2.3 B-spline Deformation Driven by Corresponding Particles

An improved FFD B-spline is proposed by [9]. In [9], the authors show that the straightforward optimization of B-spline control points is suboptimal and propose a fitting-based strategy that directly manipulates free-form deformations. In the same regard, we estimate the deformation T_j directly from the set of corresponding particles interpolating B-spline deformation in Least Squares sense. [8] gives a solution for the interpolation generalized to n -dimensional scattered data. The overall algorithm flow of our method is shown in Fig. 2.

2.4 Particle Initialization

Since the registration is performed by iterative particle optimization, the initial particle placement is important to achieve good registration results. Assuming that a basic preprocessing such as the rigid or affine registration is performed, we compute the initial particle placement as following:

1. Compute the intersection Ω_M of a set of given ROIs $\Omega_1, \Omega_2, \dots, \Omega_N$
2. Choose random particle samples \mathbf{X} , inside of the intersection Ω_M
3. Uniformly distribute the sampled particles \mathbf{X} inside Ω_M
4. Transfer \mathbf{X} into each subject i and distribute \mathbf{X}_i inside Ω_i

We rely on this heuristic to set up particles. By gradually distributing particles, the corresponding particles will be located at similar position inside each mask Ω . This strategy is specifically useful for rodent brain where the volume of each subcortical structure is proportional to whole brain.

3 Experimental Results

3.1 Data Set

The data set acquired post mortem, at 3 age groups across adolescence (postnatal days 28 through 80). MR images of each animal using a Bruker BioSpec 9.4T horizontal bore MRI system (Bruker, Billerica, MA). Images were acquired using a 4-channel phase-array surface coil with the rat in supine position. 3D MDEFT sequence was used for T1-weighted image acquisition with the following parameters: TE=6.7 ms, TR=4000 ms, NEX=4; matrix size of 320×210 , and the voxel size of 0.1mm isotropic, and acquisition time was 6 hours. To improve signal-to-noise-ratio (SNR), two images were acquired immediately following each other for each animal, and these two were averaged together following rigid registration. Total imaging time was 12 hours.

3.2 Evaluation

For the preliminary results of our method, we compared the results of our method with two popularly used non-rigid registration methods using cross correlation as a similarity metric: the non-rigid FFD B-spline image registration method packaged in Slicer3, and the SyN image registration method implemented in ANTS [1]. To study the performance of our group-wise registration, we warped manual regions of interests of brain structures, Thalamus and Cerebellum, of each subject to every other subject with each method. We included all 17 subjects ranging from postnatal days 28 to postnatal days 72 and computed total 272 pairs. We then computed Dice overlap ratios ($2|A \cap B|/(|A| + |B|)$) between the manual and automatic structural segmentations. For the proposed method, we sampled 2048 particles from each volume and used $7 \times 7 \times 7$ intensity regions per each particle. For B-spline displacement field interpolation, we used $8 \times 8 \times 8$ control points grid with the order of 3 splines. Each compared method was applied with its default settings except the number of B-spline control points matched with ours. The average Dice ratios of two ROIs for post-mortem rat images for each method are shown in Table 1.

From the results, the proposed method showed higher Dice coefficients than other two methods. Our method showed better performance in Dice coefficients than the FFD B-spline implementation and ANTS tool in the manually segmented regions.

Table 1. Overall Dice coefficients and its standard deviation of Thalamus and Cerebellum, by the proposed group-wise method, Symmetric Diffeomorphic Mapping in ANTS, and FFD B-spline registration

Methods	Thalamus	Cerebellum
The proposed method	86% ($\pm 8\%$)	87.8% ($\pm 6\%$)
ANTS	81% ($\pm 6\%$)	84% ($\pm 14\%$)
B-spline	81% ($\pm 6\%$)	79% ($\pm 11\%$)

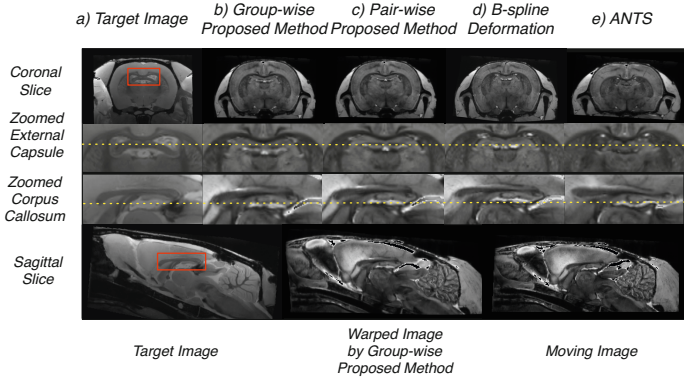


Fig. 3. Visual comparison of segmentation results. From left to right, the moving, fixed, result of proposed method, B-spline, and ANTS respectively in the first three rows. The bottom row shows sagittal slices of the fixed image, the result of the proposed image, and the moving image. The intensity scale was inverted during the acquisition but corrected in the experiments.

4 Conclusion

We proposed a novel image registration method that is guided by dynamic particles. Having correspondences each other, those particles are driven to locally similar positions in the mean space. By computing an implicit mean rather than an explicit image, our method was efficiently performed group-wise image registration in a linear time with respect to the number of subjects. Our method can be immediately applied to for example the multi-atlas joint registration/segmentation, the detection of outliers in a large data study, the inclusion of statistical shape information during registration, etc. Since the proposed method stays at a very early stage of research, future work will include thorough validation for its accuracy and robustness as well as comparison to other group-wise registration method [2,4,10].

Acknowledgements. The following grants are acknowledged for financial support: P01 DA022446, P30 HD03110, R41 NS059095, U01 AA020022, A020023, A020024, AA06059, AA019969, U54 sEB005149.

References

1. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12(1), 26–41 (2008)
2. Balci, S.K., Golland, P., Shenton, M., Wells, W.M.: Free-form b-spline deformation model for groupwise registration. In: *Medical Image Computing and Computer-Assisted Intervention: MICCAI... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 10, p. 23. NIH Public Access (2007)
3. Cates, J.E., Fletcher, P.T., Styner, M.A., Shenton, M.E., Whitaker, R.T.: Shape modeling and analysis with entropy-based particle systems. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007. LNCS*, vol. 4584, pp. 333–345. Springer, Heidelberg (2007)
4. Joshi, S., Davis, B., Jomier, M., Gerig, G., et al.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23(1), 151 (2004)
5. Meyer, M.D., Georgel, P., Whitaker, R.T.: Robust particle systems for curvature dependent sampling of implicit surfaces. In: *2005 International Conference on Shape Modeling and Applications*, pp. 124–133. IEEE (2005)
6. Nie, J., Shen, D.: Automated segmentation of mouse brain images using multi-atlas multi-roI deformation and label fusion. *Neuroinformatics*, 1–11 (2013)
7. Oguz, I., Niethammer, M., Cates, J., Whitaker, R., Fletcher, T., Vachet, C., Styner, M.: Cortical correspondence with probabilistic fiber connectivity. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) *IPMI 2009. LNCS*, vol. 5636, pp. 651–663. Springer, Heidelberg (2009)
8. Tustison, N.J., Gee, J.C.: Generalized n -D C^k B-spline scattered data approximation with confidence values. In: Yang, G.Z., Jiang, T.-Z., Shen, D., Gu, L., Yang, J. (eds.) *MIAR 2006. LNCS*, vol. 4091, pp. 76–83. Springer, Heidelberg (2006)
9. Tustison, N.J., Avants, B.A., Gee, J.C.: Improved ffd b-spline image registration. In: *IEEE 11th International Conference on Computer Vision, ICCV 2007*, pp. 1–8. IEEE (2007)
10. Wu, G., Yap, P.-T., Wang, Q., Shen, D.: Groupwise registration from exemplar to group mean: extending hammer to groupwise registration. In: *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 396–399. IEEE (2010)

Contour-Driven Regression for Label Inference in Atlas-Based Segmentation

Christian Wachinger^{1,2}, Gregory C. Sharp², and Polina Golland¹

¹ Computer Science and Artificial Intelligence Lab. MIT

² Massachusetts General Hospital, Harvard Medical School

Abstract. We present a novel method for inferring tissue labels in atlas-based image segmentation using Gaussian process regression. Atlas-based segmentation results in probabilistic label maps that serve as input to our method. We introduce a contour-driven prior distribution over label maps to incorporate image features of the input scan into the label inference problem. The mean function of the Gaussian process posterior distribution yields the MAP estimate of the label map and is used in the subsequent voting. We demonstrate improved segmentation accuracy when our approach is combined with two different patch-based segmentation techniques. We focus on the segmentation of parotid glands in CT scans of patients with head and neck cancer, which is important for radiation therapy planning.

1 Introduction

Atlas-based segmentation extracts information from image collections with manually labeled images to facilitate the automatic segmentation of new images. Methods that use atlas information can be broadly classified into two groups. The first group employs deformable registration to align atlas images to the novel scan [6,10]. The estimated deformation fields propagate labels from the atlas to the new image. The second group searches for image patches most similar to the voxel neighborhood [4,11]. Since similar patches tend to share the segmentation label, weighted voting based on patch similarity promises to produce accurate segmentation.

High anatomical variability presents a serious challenge for atlas-based segmentation. Registration approaches often fail to warp structures that vary significantly in shape due to regularization constraints. Such inaccuracies cause segmentation errors at the boundaries. Patch-based approaches also experience difficulties in correctly segmenting regions close to the boundaries. Fig. 1 illustrates this problem for a patch-based segmentation of the left parotid gland. To further investigate the source of errors, we examine patches in the atlas that are the most similar to the one example patch in the image. According to the manual labeling, the selected patch belongs to the left parotid gland. However, all of the closest patches vote for background, yielding a wrong result. Such errors are not surprising because it is possible that patches have a very similar appearance

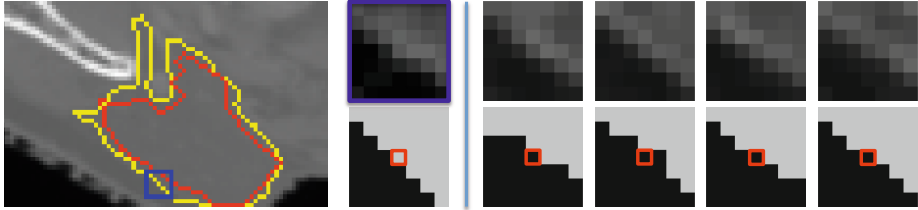


Fig. 1. Left: CT image with segmentation of left parotid (yellow: manual, red: patch-based). Right: Magnification of the blue patch (top) with manual segmentation (bottom). The four most similar patches in the repository vote for background (black at the center location), although the patch belongs to the left parotid. Intensity values of patches are normalized for visualization.

overall but vary slightly in the center. Such variations are especially problematic close to organ boundaries, where they can cause segmentation errors.

We present a new probabilistic approach to atlas-based segmentation to incorporate image contour information into the decision on segmentation labels. We achieve this by defining an image-specific distribution over label maps based on Gaussian processes. We employ the concept of intervening contours [1] to construct contour-driven covariance functions. A robust contour estimation is obtained by calculating image and texture gradients on multiple scales. Conditioning the distribution over label maps on the atlas information results in label maps that are consistent with image contours while also accommodating the label maps proposed by the atlas. We experiment with two patch-based segmentation approaches to obtain the initial label maps that serve as input to our algorithm.

We evaluate our approach by segmenting parotid glands in CT scans of patients with head and neck cancer. Radiation therapy motivates our work. Radiation therapy planning aims to maximize the dose in the target region while minimizing the radiation dose in the surrounding tissue. Intensity modulated radiation therapy allows the more effective administration of the radiation dose to reduce the damage to healthy cells. During the planning phase, experts delineate most critical structures, also called organs at risk, to ensure low radiation in these regions. The parotid glands are organs at risk for head and neck cancer treatment because they are the most important salivary glands. Irradiation of the parotid glands can lead to xerostomia, resulting in difficulties for mastication, deglutition, and speech of the patients. Automatic segmentation is challenging due to low soft-tissue contrast in CT images and high anatomical variability.

1.1 Related Work

Our work builds on previously mentioned atlas-based segmentation methods and is related to algorithms for label refinement. Spectral label fusion [14] extracts

superpixels from the image to perform region-based voting. It further relates to an approach for the refinement of atlas propagation with graph cuts [12]. Regression has been previously used to estimate correlations of errors for atlas-based segmentation [15]. Our probabilistic approach uses Gaussian processes, which arise in numerous fields of machine learning [9]. In [13], Gaussian processes were applied for image segmentation of natural images. In contrast to our work, the identity covariance function was used, samples from the process are thresholded, and no atlas information is available.

Atlas-based segmentation of parotid glands with deformable registration was demonstrated in [5,8]. In [3], the atlas images are used for training an active shape model of parotid glands. The refinement of head and neck segmentations based on classification with features was proposed in [7].

2 Method

Given a novel image I , we aim to infer segmentation S based on an atlas that contains images $\mathcal{I} = \{I_1, \dots, I_n\}$ and segmentations $\mathcal{S} = \{S_1, \dots, S_n\}$. A probabilistic label map $L = \{L^1, \dots, L^m\}$ specifies the likelihood for each label $l \in \{1, \dots, m\}$, i.e., $L^l(x) = p(S(x) = l | I, \mathcal{I}, \mathcal{S})$ and serves as an intermediate segmentation result. The estimated segmentation \hat{S} at voxel x is obtained by choosing the label with highest probability at voxel x . A perfect label map assigns probability one to the correct label for each location. Atlas-based methods produce label map L_o , which might be susceptible to errors, motivating the model $L_o^l = L^l + \varepsilon$, where L^l is the underlying true label map for label l . Under the assumption of independent and identically distributed noise, we have $L_o^l(x) = L^l(x) + \varepsilon(x)$ for all locations x in the image, with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. The assumption of independent Gaussian noise may interfere with the normalization requirements ($\sum_l L_o^l(x) = 1$ and $0 \leq L_o^l(x) \leq 1$), which can be satisfied with a subsequent normalization step. In our application, this is not necessary because we decide on the segmentation based on the maximal value across label maps. We drop the label index l in the following discussion, to simplify notation.

2.1 Atlas-Based Segmentation

We briefly review two atlas-based segmentation methods we use to obtain the initial label map L_o . We focus on patch-based approaches because they are well suited to handle the high variability of parotid glands. Further, standardized intensity values of CT images make patches comparable across subjects. Prior to segmentation, we define regions of interest (ROI) that surround the parotid glands to restrict the search. Such regions could be obtained from a coarse registration. We exploit the knowledge that the parotid glands are adjacent to the mandible bone, which we detect with a simple template matching method.

The first baseline method is the non-local means (NLM) segmentation [4,11]. For each location x within the ROI, we create the surrounding patch P_x of

size $7 \times 7 \times 3$ and retrieve the $N = 10$ closest patches \mathcal{P} with corresponding labels \mathcal{L} from the repository. The label map is obtained as a weighted sum [4]:

$$L_o(x) = \frac{\sum_{i=1}^N w_{i,x} \mathcal{L}_i}{\sum_{i=1}^N w_{i,x}} \quad \text{with} \quad w_{i,x} = \exp\left(-\frac{\|\mathcal{P}_i - P_x\|_2^2}{\min_j \|\mathcal{P}_j - P_x\|_2^2}\right). \quad (1)$$

The second approach uses a random forest classifier [2] to predict the segmentation label $L_o(x)$ for each location in the ROI. In contrast to NLM labeling, the classifier has to be trained first. In our experiments, we randomly select six patients for training. We train different classifiers for left and right parotid glands on patches selected from the ROI. We choose 500 trees per random forest with 12 predictors sampled for splitting at each node.

2.2 Gaussian Process Regression for Label Inference

Our approach to inferring the latent label map L from L_o employs a distribution over label maps $p(L)$. In contrast to most atlas-based methods that make decisions at each voxel separately and do not consider contour information, we choose a label distribution that models the relationship between locations, exploiting the contour information in image I . Stochastic processes offer a versatile framework to model interactions between possibly infinite number of random variables. We view label maps as realizations from a Gaussian process, $L \sim \mathcal{GP}(m, k)$, with mean m and covariance k . Gaussian processes are entirely characterized by mean and covariance functions and have the property that every finite subset is distributed according to a multivariate Gaussian distribution [9].

To obtain the posterior distribution over label maps, we condition the distribution of label maps L on the labels L_o implied by the atlas:

$$p(L|L_o) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (2)$$

with mean and covariance

$$\boldsymbol{\mu} = \mathbf{m} + K \cdot [K + \sigma^2 \mathbf{I}]^{-1} \cdot (L_o - \mathbf{m}), \quad (3)$$

$$\boldsymbol{\Sigma} = K - K \cdot [K + \sigma^2 \mathbf{I}]^{-1} \cdot K, \quad (4)$$

where \mathbf{I} is the identity matrix and σ^2 is the noise variance. The mean vector \mathbf{m} and the kernel matrix K are constructed from the mean function m and kernel function k , respectively. We use the Cholesky factorization for the matrix inversion. The maximum a posteriori label map coincides with the mean label map for Gaussian distributions, $L^{\text{MAP}} = \arg \max_L p(L|L_o) = \arg \max_L \mathcal{N}(L; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\mu}$. Performing this estimation for all labels yields segmentation:

$$\hat{S}(x) = \arg \max_l \boldsymbol{\mu}^l(x). \quad (5)$$

The mean function m causes a constant additive shift of all label maps $\boldsymbol{\mu}^l$ and therefore does not influence the segmentation result \hat{S} , motivating the choice of $m = 0$. Fig. 2 illustrates the key steps of the segmentation process.

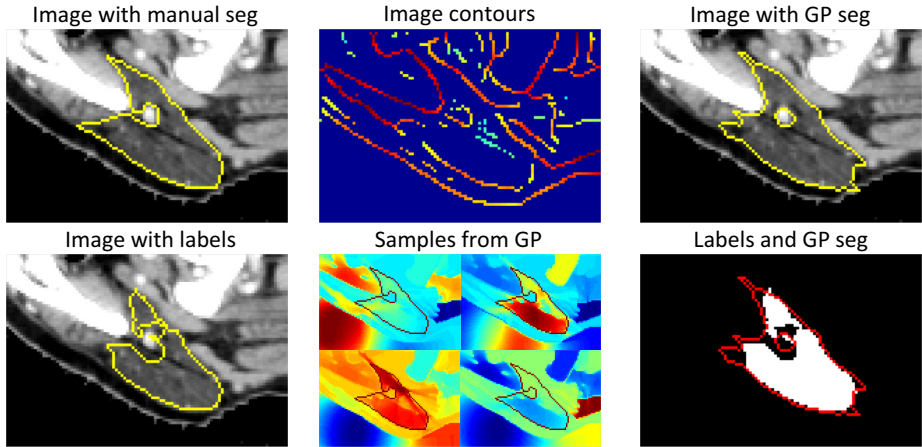


Fig. 2. Gaussian process segmentation of parotid gland. The initial label from the atlas-based segmentation only partially agrees with the manual segmentation. We extract contours from the image and use them in the kernel function k that allows us to sample label maps $L \sim \mathcal{GP}(0, k)$, supported by the image. Conditioning these on the atlas labels results in an improved segmentation.

2.3 Contour-Driven Distributions over Label Maps

The distribution over label maps $p(L)$ is determined by the covariance or kernel function k . We seek label maps that are supported by intensity and texture features in the input image I . In the first step, we estimate image and texture gradients per slice with the oriented gradient signal, following closely the construction in [1]. This method calculates the χ^2 distance between the histograms of two half-discs at each location for various orientations and at multiple scales. Textons are calculated to quantify the texture by convolving the image with 17 Gaussian derivative and center-surround filters and subsequently clustering with K-means into 64 classes [1]. Image and texture gradients of multiple scales are added to yield the multi-scale contour Γ . We use the contour information to calculate weights between in-plane points x and x' , following the concept of intervening contours [1] by identifying the maximum response along the line $\overline{xx'}$:

$$k(x, x') = \exp\left(-\max_{y \in \overline{xx'}}\{\Gamma(y)\}/\rho\right). \quad (6)$$

We set the scale parameter to $\rho = 0.1$ and only consider locations within the ROI that are at most 20 pixels away from each other, giving rise to sparse kernel matrices. High weights are assigned to pairs of points that are not separated by a contour and these points are subsequently encouraged to share the same label.

Fig. 2 shows samples drawn from the prior distribution $p(L)$, where we have overlaid the manual segmentation for reference. We observe that the prior

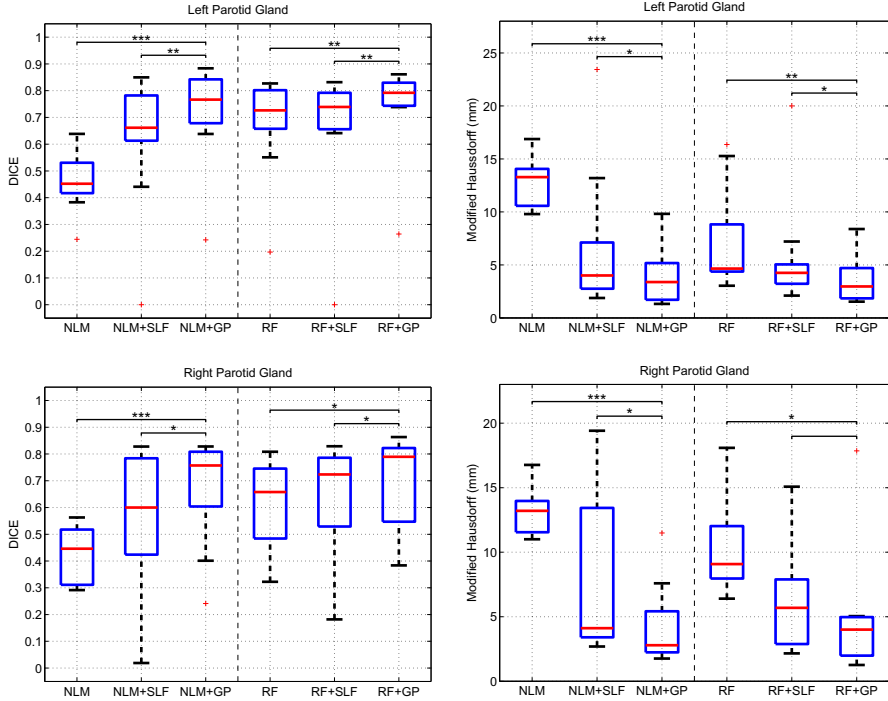


Fig. 3. Dice volume overlap and modified Hausdorff distance for left and right parotid glands. Red line indicates the median, the boxes extend to the 25th and 75th percentiles, and the whiskers reach to the most extreme values not considered outliers (red crosses). *, **, and *** indicate significance levels at 0.05, 0.01, and 0.001. For each baseline method (NLM, RF), the performance of the basic method, the variant that employs spectral label fusion (SLF) [14] and the variant based on Gaussian processes proposed here (GP) is reported.

promotes label maps that follow image structures. In this example, labels are propagated to the thin ends of the left parotid gland, which improves the segmentation in comparison to the initial labeling.

3 Experiments

We evaluate the method on 16 CT scans of patients with head and neck cancer. Each image was labeled by a trained anatomist for treatment planning. Images contain between 80 and 200 axial slices with a slice thickness of 2.5mm. The in-plane resolution is 0.9mm, slice size is 512×512 pixels. All 16 images have the left parotid gland labeled. The right parotid gland was consumed by a tumor in one patient. Experiments are performed on 10 datasets for left parotid gland and 9 datasets for right parotid gland that have not been selected for training the RF classifiers. We quantify the segmentation quality by calculating

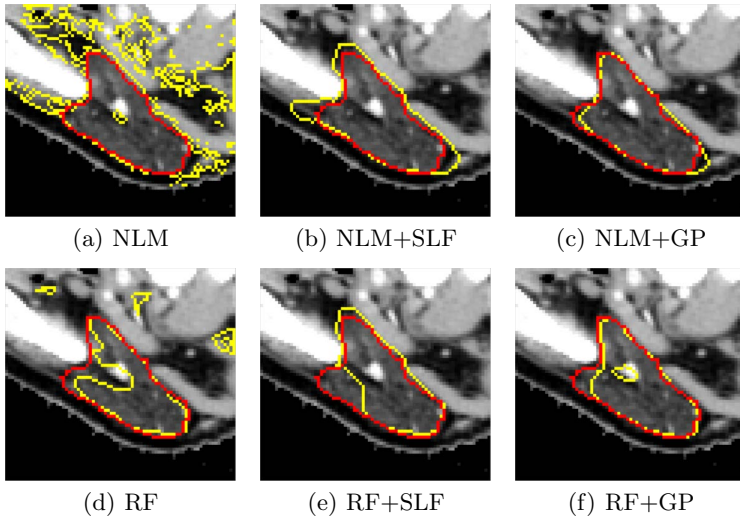


Fig. 4. Examples of automatic segmentation results for different methods are shown in yellow. Manual delineations are shown in red.

the Dice volume overlap score and modified Hausdorff distance between the automatic and manual segmentations. We compare our method to spectral label fusion (SLF) [14], which was previously demonstrated to refine segmentations based on image contours.

Fig. 3 presents the results for both parotid glands for different algorithms and $\sigma^2 = 1$. Non-local means (NLM) and random forests (RF) serve as initial label maps. The segmentation with NLM leads to many false positives, causing worse performance than RF. Applying spectral label fusion improves the segmentation results. The Gaussian process (GP) segmentation achieves the significantly best results in our experiments. A reason for the improvement of GP in comparison to SLF is that SLF votes on small image regions. If these regions are not well defined or if the baseline segmentation algorithm cannot gather enough votes in a region, this can cause large errors. The outlier of zero dice overlap for NLM+SLF of the left parotid illustrates this case. Fig. 4 shows example results for all methods.

4 Conclusion

We proposed a novel probabilistic approach for improving atlas-based segmentation. The key contribution is a contour-driven distribution over label maps that is supported by features in the image. We employ Gaussian process regression to obtain MAP estimates of label maps, on which the voting is performed. The initial label map is estimated with two different patch-based segmentation approaches, non-local means segmentation and random forest classification.

Our experiments in segmentation of the parotid glands show improved performance when the proposed method is used to refine the atlas-based label maps.

Acknowledgements. This work was funded in part by the Humboldt foundation, the National Alliance for Medical Image Computing (NIH NIBIB NAMIC U54-EB005149) and the NeuroImaging Analysis Center (NIH NCRR NAC P41-RR13218 and NIH NIBIB NAC P41-EB-015902)

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. *IEEE Trans. on Pat. Anal. Mach. Intel.* 33(5), 898–916 (2011)
2. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
3. Chen, A., Noble, J.H., Niemann, K.J., Deeley, M.A., Dawant, B.M.: Segmentation of parotid glands in head and neck CT images using a constrained active shape model with landmark uncertainty. In: *SPIE*, vol. 8314, p. 83140P (2012)
4. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
5. Han, X., Hibbard, L.S., O’connell, N.P., Willcut, V.: Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. In: *Medical Image Analysis for the Clinic: A Grand Challenge*, pp. 297–304 (2010)
6. 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)
7. Qazi, A.A., Pekar, V., Kim, J., Xie, J., Breen, S.L., Jaffray, D.A.: Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach. *Medical Physics* 38, 6160 (2011)
8. Ramus, L., Malandain, G.: Multi-atlas based segmentation: Application to the head and neck region for radiotherapy planning. In: *Medical Image Analysis for the Clinic: A Grand Challenge*, pp. 281–288 (2010)
9. Rasmussen, C., Williams, C.: *Gaussian processes for machine learning*. MIT Press (2006)
10. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, C.: Quo vadis, atlas-based segmentation? In: *Handbook of Biomedical Image Analysis*, pp. 435–486 (2005)
11. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. *IEEE Trans. Med. Imaging* 30(10), 1852–1862 (2011)
12. Song, Z., Tustison, N., Avants, B., Gee, J.C.: Integrated graph cuts for brain MRI segmentation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 831–838. Springer, Heidelberg (2006)
13. Sudderth, E.B., Jordan, M.I.: Shared segmentation of natural scenes using dependent Pitman-Yor processes. In: *NIPS*, pp. 1585–1592 (2008)
14. Wachinger, C., Golland, P.: Spectral label fusion. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III*. LNCS, vol. 7512, pp. 410–417. Springer, Heidelberg (2012)
15. Wang, H., Suh, J.W., Das, S., Pluta, J., Altinay, M., Yushkevich, P.: Regression-based label fusion for multi-atlas segmentation. In: *CVPR*, pp. 1113–1120 (2011)

Discriminative Parameter Estimation for Random Walks Segmentation

Pierre-Yves Baudin^{1,2,3,4,5,6}, Danny Goodman^{1,2,3}, Puneet Kumar^{1,2,3},
Noura Azzabou^{4,5,6}, Pierre G. Carlier^{4,5,6},
Nikos Paragios^{1,2,3}, and M. Pawan Kumar^{1,2,3}

¹ Center for Visual Computing, École Centrale Paris, FR

² Université Paris-Est, LIGM (UMR CNRS), École des Ponts ParisTech, FR

³ Équipe Galen, INRIA Saclay, FR

⁴ Institute of Myology, Paris, FR

⁵ CEA, I² BM, MIRCen, IdM NMR Laboratory, Paris, FR

⁶ UPMC University Paris 06, Paris, FR

Abstract. The Random Walks (RW) algorithm is one of the most efficient and easy-to-use probabilistic segmentation methods. By combining contrast terms with prior terms, it provides accurate segmentations of medical images in a fully automated manner. However, one of the main drawbacks of using the RW algorithm is that its parameters have to be hand-tuned. We propose a novel discriminative learning framework that estimates the parameters using a training dataset. The main challenge we face is that the training samples are not fully supervised. Specifically, they provide a hard segmentation of the images, instead of a probabilistic segmentation. We overcome this challenge by treating the optimal probabilistic segmentation that is compatible with the given hard segmentation as a latent variable. This allows us to employ the latent support vector machine formulation for parameter estimation. We show that our approach significantly outperforms the baseline methods on a challenging dataset consisting of real clinical 3D MRI volumes of skeletal muscles.

1 Introduction¹

The Random Walks (RW) algorithm is one of the most popular techniques for segmentation in medical imaging [5]. Although it was initially proposed for interactive settings, recent years have witnessed the development of fully automated extensions. In addition to the contrast information employed in the original formulation [5], the automated extensions incorporate prior information based on appearance [4] and shape [1].

It has been empirically observed that the accuracy of the RW algorithm relies heavily on the relative weighting between the various contrast and prior terms. Henceforth, we refer to the relative weights of the various terms in the RW objective function as parameters. At present, researchers either rely on a user to

¹ Supplementary materials at: <http://hal.inria.fr/hal-00830564>

hand-tune the parameters or on exhaustive cross-validation [1,4]. However, both these approaches quickly become infeasible as the number of terms in the RW objective function increase.

In contrast to the RW literature, the problem of parameter estimation has received considerable attention in the case of discrete models such as CRFs [9]. Recent years have witnessed the emergence of structured-output support vector machine (Structured SVM) as one of the most effective discriminative frameworks for supervised parameter estimation [10,11]. Given a training dataset that consists of pairs of input and their ground-truth output, structured SVM minimizes the empirical risk of the inferred output with respect to the ground-truth output. The risk is defined by a user-specified loss function that measures the difference in quality between two given outputs.

We would like to discriminatively learn the parameters of the RW formulation. To this end, a straightforward application of structured SVM would require a training dataset that consists of pairs of inputs as well as their ground-truth outputs—in our case, the *optimal* probabilistic segmentation. In other words, we require a human to provide us with the output of the RW algorithm for the best set of parameters. This is an unreasonable demand since the knowledge of the optimal probabilistic segmentation is as difficult to acquire as it is to hand-tune the parameters itself. Thus we cannot directly use structured SVM to estimate the desired parameters.

In order to handle the above difficulty, we propose a novel formulation for discriminative parameter estimation in the RW framework. Specifically, we learn the parameters using a weakly supervised dataset that consists of pairs of medical acquisitions and their hard segmentations. Unlike probabilistic segmentations, hard segmentations can be obtained easily from human annotators. We treat the optimal probabilistic segmentation that is *compatible* with the hard segmentation as a latent variable. Here, compatibility refers to the fact that the probability of the ground-truth label (as specified by the hard segmentation) should be greater than the probability of all other labels for each pixel/voxel. The resulting representation allows us to learn the parameters using the latent SVM formulation [3,8,12].

While latent SVM does not result in a convex optimization problem, its local optimum solution can be obtained using the iterative concave-convex procedure (CCCP) [13]. The CCCP involves solving a structured SVM problem, which lends itself to efficient optimization. In order to make the overall algorithm computationally feasible, we propose a novel efficient approach for ACI based on dual decomposition [2,7]. We demonstrate the benefit of our learning framework over a baseline structured SVM using a challenging dataset of real 3D MRI volumes.

2 Preliminaries

We will assume that the input \mathbf{x} is a 3D volume. We denote the i -th voxel of \mathbf{x} as $\mathbf{x}(i)$, and the set of all voxels as \mathcal{V} . In a hard segmentation, each voxel is assigned a label $s \in \mathcal{S}$ (for example, the index of a muscle). We will use \mathbf{z} to

represent the human annotation (that is, the class labels of the voxels in \mathbf{x}) in binary form:

$$\mathbf{z}(i, s) = \begin{cases} 1 & \text{if voxel } i \in \mathcal{V} \text{ is of class } s \in \mathcal{S}, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

In other words, the binary form \mathbf{z} of the annotation specifies delta distribution over the putative labels for each voxel. Our training dataset is a collection of training images \mathbf{x} and hard segmentations \mathbf{z} : $\mathcal{D} = \{(\mathbf{x}_k, \mathbf{z}_k)\}_k$. Note that we use subscript k to denote the input index within a dataset, and parenthetical i to denote a voxel within a particular input.

2.1 Random Walks Segmentation

The RW algorithm provides a probabilistic—or soft—segmentation of an input \mathbf{x} , which we denote by \mathbf{y} , that is,

$$\mathbf{y}(i, s) = \Pr[\text{voxel } i \text{ is of class } s], \forall i \in \mathcal{V}, s \in \mathcal{S}. \quad (2)$$

When using one contrast term and one prior model, the RW algorithm amounts to minimizing the following convex quadratic objective functional:

$$E(\mathbf{y}, \mathbf{x}) = \mathbf{y}^\top L(\mathbf{x}) \mathbf{y} + w^{\text{prior}} \|\mathbf{y} - \mathbf{y}_0\|_{\Omega_0(\mathbf{x})}^2, \quad (3)$$

$$= \mathbf{y}^\top L(\mathbf{x}) \mathbf{y} + E^{\text{prior}}(\mathbf{y}, \mathbf{x}). \quad (4)$$

Here, \mathbf{y}_0 is a reference prior probabilistic segmentation dependent on appearance [4] or shape [1], and $\Omega_0(\mathbf{x})$ is a diagonal matrix that specifies a voxel-wise weighting scheme for \mathbf{x} . The term $L(\mathbf{x})$ refers to a combinatorial Laplacian matrix defined on a neighborhood system \mathcal{N} based on the adjacency of the voxels. It is a block diagonal matrix—one block per label—with all identical blocks, where the entries of the block $L^b(\mathbf{x})$ use the typical Gaussian kernel formulation (see [5]). The relative weight w^{prior} is the parameter for the above RW framework. The above problem is convex, and can be optimized efficiently by solving a sparse linear system of equations. We refer the reader to [1,5] for further details.

2.2 Parameters and Feature Vectors

In the above description of the RW algorithm, we restricted ourselves to a single Laplacian and a single prior. However, our goal is to enable the use of numerous Laplacians and priors. To this end, let $\{L_\alpha\}_\alpha$ denote a known family of Laplacian matrices and $\{E_\beta(\cdot)\}_\beta$ denote a known family of prior energy functionals. In section 4, we will specify the family of Laplacians and priors used in our experiments. We denote the general form of a linear combination of Laplacians and prior terms as:

$$L(\mathbf{x}; \mathbf{w}) = \sum_\alpha w_\alpha L_\alpha(\mathbf{x}), E^{\text{prior}}(\cdot, \mathbf{x}; \mathbf{w}) = \sum_\beta w_\beta E_\beta(\cdot, \mathbf{x}), \mathbf{w} \geq 0. \quad (5)$$

Each term $E_\beta(\cdot, \mathbf{x})$ is of the form:

$$E_\beta(\mathbf{y}, \mathbf{x}) = \|\mathbf{y} - \mathbf{y}_\beta\|_{\Omega_\beta(\mathbf{x})}^2, \quad (6)$$

where \mathbf{y}_β is the β -th reference segmentation and $\Omega_\beta(\mathbf{x})$ is the corresponding voxel-wise weighting matrix (which are both known). We denote the set of all parameters as $\mathbf{w} = \{w_\alpha, w_\beta\}_{\alpha, \beta}$. Clearly, the RW energy (4) is linear in \mathbf{w} , and can therefore be formulated as:

$$E(\mathbf{y}, \mathbf{x}; \mathbf{w}) = \mathbf{y}^T L(\mathbf{x}; \mathbf{w}) \mathbf{y} + E_{\text{prior}}(\mathbf{y}, \mathbf{x}; \mathbf{w}), \quad (7)$$

$$= \mathbf{w}^T \psi(\mathbf{x}, \mathbf{y}), \quad (8)$$

where $\psi(\mathbf{x}, \mathbf{y})$ is known as the joint feature vector of \mathbf{x} and \mathbf{y} . Note that by restricting the parameters to be non-negative (that is, $\mathbf{w} \geq \mathbf{0}$), we ensure that the energy functional $E(\cdot, \mathbf{x}; \mathbf{w})$ remains convex.

2.3 Loss Function

As mentioned earlier, we would like to estimate the parameters \mathbf{w} by minimizing the empirical risk over the training samples. The risk is specified using a loss function that measures the difference between two segmentations. In this work, we define the loss function as the number of incorrectly labeled voxels. Formally, let $\hat{\mathbf{y}}$ denote the underlying hard segmentation of the soft segmentation \mathbf{y} , that is, $\hat{\mathbf{y}}(i, s) = \delta(s = \operatorname{argmax}_{s \in \mathcal{S}} \mathbf{y}(i, s))$, where δ is the Kronecker function. The loss function is defined as

$$\Delta(\mathbf{z}, \mathbf{y}) = 1 - \frac{1}{|\mathcal{V}|} \hat{\mathbf{y}}^T \mathbf{z}, \quad (9)$$

where \mathcal{V} is the set of all voxels, and $|\cdot|$ denotes the cardinality of a set.

3 Parameter Estimation Using Latent SVM

Given a dataset $\mathcal{D} = \{(\mathbf{x}_k, \mathbf{z}_k), k = 1, \dots, N\}$, which consists of inputs \mathbf{x}_k and their hard segmentation \mathbf{z}_k , we would like to estimate parameters \mathbf{w} such that the resulting inferred segmentations are accurate. Here, the accuracy is measured using the loss function $\Delta(\cdot, \cdot)$. Formally, let $\mathbf{y}_k(\mathbf{w})$ denote the soft segmentation obtained by minimizing the energy functional $E(\cdot, \mathbf{x}_k; \mathbf{w})$ for the k -th training sample, that is,

$$\mathbf{y}_k(\mathbf{w}) = \operatorname{argmin}_{\mathbf{y}} \mathbf{w}^T \psi(\mathbf{x}_k, \mathbf{y}). \quad (10)$$

We would like to learn the parameters \mathbf{w} such that the empirical risk is minimized. In other words, we would like to estimate the parameters \mathbf{w}^* such that

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w}} \frac{1}{N} \sum_k \Delta(\mathbf{z}_k, \mathbf{y}_k(\mathbf{w})). \quad (11)$$

The above objective function is highly non-convex in \mathbf{w} , which makes it prone to bad local minimum solutions. To alleviate this deficiency, it can be shown that the following latent SVM formulation minimizes a regularized upper bound on the risk for a set of samples $\{(\mathbf{x}_k, \mathbf{z}_k), k = 1, \dots, N\}$:

$$\min_{\mathbf{w} \geq \mathbf{0}} \lambda \|\mathbf{w}\|^2 + \lambda' \|\mathbf{w} - \mathbf{w}_0\|^2 + \frac{1}{N} \sum_k \xi_k, \quad (12)$$

$$\text{s.t.} \quad \min_{\mathbf{y}_k, \Delta(\mathbf{z}_k, \mathbf{y}_k)=0} \mathbf{w}^T \psi(\mathbf{x}_k, \mathbf{y}_k) \leq \mathbf{w}^T \psi(\mathbf{x}_k, \bar{\mathbf{y}}_k) - \Delta(\mathbf{z}_k, \bar{\mathbf{y}}_k) + \xi_k, \forall \bar{\mathbf{y}}_k, \forall k,$$

where the slack variable ξ_k represents the upper bound of the risk for the k -th training sample. Note that we have added two regularization terms for the parameters \mathbf{w} . The first term $\|\mathbf{w}\|^2$, weighted by hyperparameter λ , ensures that we do not overfit to the training samples. The second term $\|\mathbf{w} - \mathbf{w}_0\|^2$, weighted by hyperparameter λ' , ensures that we do not obtain a solution that is very far away from our initial estimate \mathbf{w}_0 . The reason for including this term is that our upper bound to the empirical risk may not be sufficiently tight. Thus, if we do not encourage our solution to lie close to the initial estimate, it may drift towards an inaccurate set of parameters. In section 4, we show the empirical effect of the hyperparameters λ and λ' on the accuracy of the parameters.

While the upper bound of the empirical risk derived above is not convex, it was shown to be a difference of two convex functions in [12]. This observation allows us to obtain a local minimum or saddle point solution using the CCCP algorithm [12,13], outlined in Algorithm 1, which iteratively improves the parameters starting with an initial estimate \mathbf{w}_0 . It consists of two main steps at each iteration: (i) step 3, which involves estimating a compatible soft segmentation for each training sample—known as annotation consistent inference (ACI); and (ii) step 4, which involves updating the parameters by solving problem (13). In the following subsections, we provide efficient algorithms for both the steps.

Algorithm 1. The CCCP method for parameter estimation using latent SVM.

Input: Dataset \mathcal{D} , λ , λ' , \mathbf{w}_0 , ε

1: Set $t = 0$. Initialize $\mathbf{w}_t = \mathbf{w}_0$.

2: **repeat**

3: Compute $\mathbf{y}_k^* = \operatorname{argmin}_{\mathbf{y}_k, \Delta(\mathbf{z}_k, \mathbf{y}_k)=0} \mathbf{w}_t^\top \psi(\mathbf{x}_k, \mathbf{y}_k), \forall k$.

4: Update the parameters by solving the following problem

$$\begin{aligned} \mathbf{w}_{t+1} = \operatorname{argmin}_{\mathbf{w} \geq \mathbf{0}} & \lambda \|\mathbf{w}\|^2 + \lambda' \|\mathbf{w} - \mathbf{w}_0\|^2 + \frac{1}{N} \sum_k \xi_k, \\ \text{s.t. } & \mathbf{w}^\top \psi(\mathbf{x}_k, \mathbf{y}_k^*) \leq \mathbf{w}^\top \psi(\mathbf{x}_k, \bar{\mathbf{y}}_k) - \Delta(\mathbf{z}_k, \bar{\mathbf{y}}_k) + \xi_k, \forall \bar{\mathbf{y}}_k, \forall k, \end{aligned} \quad (13)$$

5: $t = t + 1$

6: **until** The objective function of problem (12) does not decrease below tolerance ε .

3.1 Annotation Consistent Inference

Given an input \mathbf{x} and its hard segmentation \mathbf{z} , ACI requires us to find the soft segmentation \mathbf{y} with the minimum energy, under the constraint that it should be compatible with \mathbf{z} (see step 3 of Algorithm 1). We denote the ground truth label of a voxel i by s_i , that is, $s_i = \operatorname{argmax}_s \mathbf{z}(i, s)$, and the set of all voxels by \mathcal{V} . Using our notation, ACI can be formally specified as

$$\min_{\mathbf{y} \in \mathcal{C}(\mathcal{V})} \mathbf{y}^\top L(\mathbf{x}; \mathbf{w}) \mathbf{y} + E^{\text{prior}}(\mathbf{y}, \mathbf{x}; \mathbf{w}). \quad (14)$$

Here, $\mathcal{C}(\mathcal{V})$ is the set of all compatible probabilistic segmentations, that is,

$$\mathbf{y}(i, s) \geq 0, \forall i \in \mathcal{V}, \forall s \in \mathcal{S}, \quad (15)$$

$$\sum_{s \in \mathcal{S}} \mathbf{y}(i, s) = 1, \forall i \in \mathcal{V}, \quad (16)$$

$$\mathbf{y}(i, s_i) \geq \mathbf{y}(i, s), \forall i \in \mathcal{V}, \forall s \in \mathcal{S}. \quad (17)$$

Constraints (15) and (16) ensure that \mathbf{y} is a valid probabilistic segmentation. The last set of constraints (17) ensure that \mathbf{y} is compatible with \mathbf{z} . Note that in the absence of constraints (17), the above problem can be solved efficiently using the RW algorithm. However, since the ACI problem requires the additional set of compatibility constraints, we need to develop a novel efficient algorithm to solve the above convex optimization problem. To this end, we exploit the powerful dual decomposition framework [2,7]. Briefly, we divide the above problem into a set of smaller subproblems defined using overlapping subsets of variables. Each subproblem can be solved efficiently using a standard convex optimization package. In order to obtain the globally optimal solution of the original subproblem, we pass *messages* between subproblems until they agree on the value of all the shared variables. For details on the ACI algorithm, please refer to the supplementary materials of this paper.

3.2 Parameter Update

Having generated a compatible soft segmentation, the parameters can now be efficiently updated by solving problem (13) for a fixed set of soft segmentations \mathbf{y}_k^* . This problem can be solved efficiently using the popular cutting plane method (for details on this algorithm, please refer to [6]). Briefly, the method starts by specifying no constraints for any of the training samples. At each iteration, it finds the most violated constraint for each sample, and updates the parameters until the increase in the objective function is less than a small epsilon.

In this work, due to the fact that our loss function is not concave, we approximate the most violated constraint as the predicted segmentation, that is,

$$\bar{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmin}} \mathbf{w}^\top \psi(\mathbf{x}, \mathbf{y}). \quad (18)$$

The above problem is solved efficiently using the RW algorithm.

4 Experiments

Dataset. The dataset consists of 30 MRI volumes of the thigh region of dimensions $224 \times 224 \times 100$. The various segments correspond to 4 different muscle groups together with the background class. We randomly split the dataset into 80% for training and 20% for testing. In order to reduce the training time for both our method and the baselines, we divide each volume into 100/2 volumes of dimension $224 \times 224 \times 2$.

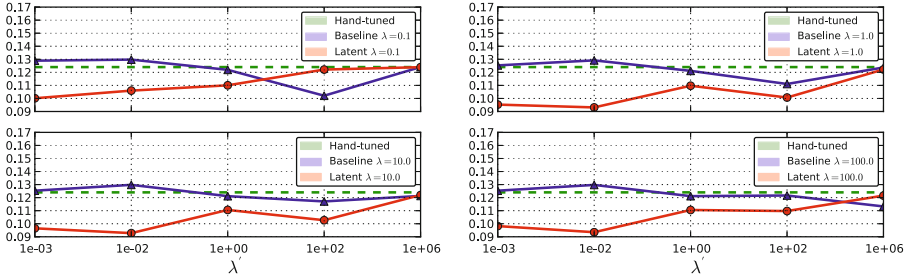


Fig. 1. Estimated risk $\Delta(\mathbf{y}_k^*, \mathbf{y}_k(\mathbf{w}))$ for three different methods

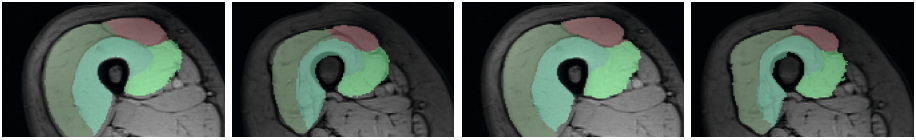


Fig. 2. Method comparison: (columns 1 & 2) segmentations using \mathbf{w}_0 ; (columns 3 & 4) segmentations using learned \mathbf{w} using latent structured SVM. The latter are closer to expert segmentation.

Laplacians and Prior Terms. We use 4 different Laplacians (generated with different weighting functions). Furthermore, we use two shape priors based on [1] and one appearance prior based on [4]. This results in a total of 7 parameters to be estimated.

Methods. The main hypothesis of our work is that it is important to represent the unknown optimal soft segmentation using latent variables. Thus we compare our method with a baseline structured SVM that replaces the latent variables with the given hard segmentations. In other words, our baseline estimates the parameters by solving problem (13), where the imputed soft segmentations \mathbf{y}_k^* are replaced by the hard segmentations \mathbf{z}_k . During our experiments, we found that replacing the hard segmentation with a pseudo soft segmentation based on the distance transform systematically decreased the loss of the output. Thus the method referred to as "Baseline" uses a structured SVM with distance-transform "softened" segmentations.

Results. Fig. 1 shows the test loss for three different methods: (i) the initial hand-tuned parameters \mathbf{w}_0 ; (ii) the baseline structured SVM with distance transforms; and (iii) our proposed approach using latent SVM. As can be seen from Fig. 1, latent SVM provides significantly better results than the baselines—even when using the distance transform. For the 4×5 hyperparameter settings that we report (that is, four different values of λ and 5 different values of λ'), latent SVM is significantly better than SVM in 15 cases, and significantly worse in only 2 cases. Note that latent SVM provides the best results for very small values of λ' , which indicates that the upper bound on the empirical risk is tight. As expected, for sufficiently large values of λ' , all the methods provide similar results.

For the best settings of the corresponding hyperparameters, the percentage of incorrectly labeled voxels as follows: (i) for \mathbf{w}_0 , 13.5%; (ii) for structured SVM, 10.0%; and (iii) for latent SVM, 9.2%. Fig. 2 shows some example segmentations for the various methods.

5 Discussion

We proposed a novel discriminative learning framework to estimate the parameters for the probabilistic RW segmentation algorithm. We represented the optimal soft segmentation that is compatible with the hard segmentation of each training sample as a latent variable. This allowed us to formulate the problem of parameter estimation using latent SVM, which upper bounds the empirical risk of prediction with a difference of convex optimization program. Using a challenging clinical dataset of MRI volumes, we demonstrated the efficacy of our approach over the baseline method that replaces the latent variables with the given hard segmentations. The latent SVM framework can be used to estimate parameters with partial hard segmentations. Such an approach would allow us to scale the size of the training dataset by orders of magnitude.

References

1. Baudin, P.-Y., Azzabou, N., Carlier, P.G., Paragios, N.: Prior knowledge, random walks and human skeletal muscle segmentation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 569–576. Springer, Heidelberg (2012)
2. Bertsekas, D.: *Nonlinear Programming*. Athena Scientific (1999)
3. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multi-scale, deformable part model. In: CVPR (2008)
4. Grady, L.: Multilabel random walker image segmentation using prior models. In: CVPR (2005)
5. Grady, L.: Random walks for image segmentation. PAMI (2006)
6. Joachims, T., Finley, T., Yu, C.N.: Cutting-plane training of structural SVMs. *Machine Learning* (2009)
7. Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: ICCV (2007)
8. Smola, A., Vishwanathan, S., Hoffman, T.: Kernel methods for missing variables. In: AISTATS (2005)
9. Szummer, M., Kohli, P., Hoiem, D.: Learning cRFs using graph cuts. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 582–595. Springer, Heidelberg (2008)
10. Taskar, B., Guestrin, C., Koller, D.: Max-margin Markov networks. In: NIPS (2003)
11. Tsochantaridis, I., Hofmann, T., Joachims, T., Altun, Y.: Support vector machine learning for interdependent and structured output spaces. In: ICML (2004)
12. Yu, C.N., Joachims, T.: Learning structural SVMs with latent variables. In: ICML (2009)
13. Yuille, A., Rangarajan, A.: The concave-convex procedure (CCCP). *Neural Computation* (2003)

Fully Automatic X-Ray Image Segmentation via Joint Estimation of Image Displacements

Cheng Chen¹, Weiguo Xie¹, Jochen Franke², Paul A. Grützner²,
Lutz-P. Nolte¹, and Guoyan Zheng¹

¹ Institute for Surgical Technologies and Biomechanics, Universität Bern, Switzerland

² BG Clinic Ludwigshafen, Ludwigshafen, Germany

Abstract. We propose a new method for fully-automatic landmark detection and shape segmentation in X-ray images. Our algorithm works by estimating the displacements from image patches to the (unknown) landmark positions and then integrating them via voting. The fundamental contribution is that, we jointly estimate the displacements from all patches to multiple landmarks together, by considering not only the training data but also geometric constraints on the test image. The various constraints constitute a convex objective function that can be solved efficiently. Validated on three challenging datasets, our method achieves high accuracy in landmark detection, and, combined with statistical shape model, gives a better performance in shape segmentation compared to the state-of-the-art methods.

1 Introduction

Segmenting anatomical regions such as the femur and the pelvis in the clinical X-ray images provides invaluable information for computer aided diagnosis [1,2], surgery planning and image-guided intervention [3], and three-dimensional (3D) model reconstruction [4,5]. Manual landmarking and segmentation are both time-consuming and error-prone. Therefore, automatic landmark detection and shape segmentation techniques have been an active research topic in medical image analysis [5]-[10].

Landmarking and segmentation X-Rays has to deal with many challenges such as poor image illumination, unknown image projection and unexpected appearance caused by trauma or implants. Many algorithms have been proposed, such as methods based on local image features [1,3] and model-based methods [5,6]. A large body of methods rely on machine learning techniques, such as the shape regression machine proposed in [7] and the marginal space learning method in [8]. Perhaps the most popular method up to now is the voting scheme based on random forest (RF) regression [9,10], which estimates the displacements from a set of sampled image patches to the landmark by random forest (RF) regression, and then the landmark position is estimated by aggregating the votes made by all the patches. In [2], Lindner et al. combine this method with a constrained local model (CLM) for automatic segmentation of proximal femur.

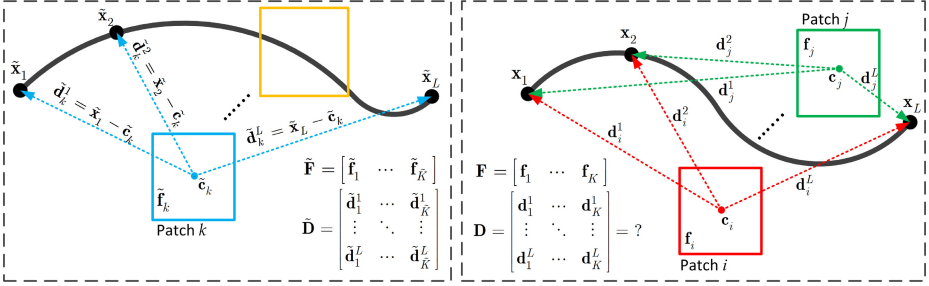


Fig. 1. Schematic illustration of the semi-supervised joint estimation of image displacements. Left: training data. Right: test data.

There are two key components behind the success of RF voting based method. The first is the strategy of positioning landmarks by estimating its relative displacements with regard to other image parts (e.g. patches). This is reasonable since medical image is highly structured. The second is the discriminative power of the RF model. In this paper, we focus on improving the estimation of the displacements from image patches to landmarks. In previous methods based on RF regression, the displacement from each patch to each landmark is estimated independently by the pre-trained RF model, i.e. the displacements are determined solely by the training data. Our method is fundamentally different, as we estimate the displacements from all patches to multiple landmarks all together. This joint estimation scheme allows us to exploit the mutual interactions among the displacements that are being estimated by considering the geometric constraints. In this way, our joint displacement estimation method achieves better accuracy. We tested our method on three large and challenging datasets: CompleteFemur, ProximalFemur and Pelvis. The experimental results show that our method achieves better performance compared to the state-of-the-art methods.

2 Landmark Detection by Joint Displacement Estimation

2.1 Problem Formulation

Training Data. Assume that we are interested in L landmarks, and the ground-truth position of these landmarks are known in a set of training images. As shown in Fig. 1 (left), $\tilde{x}_l \in \mathbb{R}^2$ is the position of the l th landmark. Furthermore, we randomly sample a number of square patches around all the landmarks. For the k th patch, we denote $\tilde{c}_k \in \mathbb{R}^2$ as its center position, $\tilde{f}_k \in \mathbb{R}^{d_f}$ as its visual feature, and $\tilde{d}_k^l = \tilde{x}_l - \tilde{c}_k \in \mathbb{R}^2$ is the displacement from the k th patch center to the l th landmark. We totally sample \tilde{K} patches over all the training images, and we denote $\tilde{F} = [\tilde{f}_1, \dots, \tilde{f}_{\tilde{K}}] \in \mathbb{R}^{d_f \times \tilde{K}}$ as the matrix of features of all training patches, and $\tilde{D} \in \mathbb{R}^{2L \times \tilde{K}}$, whose element $\tilde{D}_{ij} = \tilde{d}_j^i$, as the matrix of displacements.

Test Data. During test, we have a new image, on which we want to estimate the positions of the L landmarks, as shown in Fig. 1 (right). We randomly sample

K patches, where $\mathbf{c}_k \in \mathbb{R}^2$ and $\mathbf{f}_k \in \mathbb{R}^{d_f}$ are the center position and the visual feature of the k th patch that we sampled. We denote $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_k] \in \mathbb{R}^{d_f \times K}$ as the matrix of features of all test patches.

Strategy. To estimate the position of the L landmarks on the test image, we first want to estimate $\{\mathbf{d}_k^l\}_{k=1\dots K, l=1\dots L}$, which is the displacement from each patch to each landmark. Then, $\{\mathbf{c}_k + \mathbf{d}_k^l\}_{k=1\dots K}$ will be the set of votes of the l th landmark's position from all the test patches, from which we can compute the response image by a voting scheme (details on response image will be presented in Section 2.4). Therefore, if we denote $\mathbf{D} \in \mathbb{R}^{2L \times K}$, whose element $\mathbf{D}_{ij} = \mathbf{d}_j^i$, as the matrix of displacements in the test image, our goal is to estimate \mathbf{D} .

2.2 Objective Function

First, we construct a compound displacement matrix:

$$\hat{\mathbf{D}} = [\tilde{\mathbf{D}} \mathbf{D}] = \begin{bmatrix} \tilde{\mathbf{d}}_1^1 & \cdots & \tilde{\mathbf{d}}_{\tilde{K}}^1 & \mathbf{d}_1^1 & \cdots & \mathbf{d}_K^1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{d}}_1^L & \cdots & \tilde{\mathbf{d}}_{\tilde{K}}^L & \mathbf{d}_1^L & \cdots & \mathbf{d}_K^L \end{bmatrix} \in \mathbb{R}^{2L \times (\tilde{K} + K)} \quad (1)$$

The left part (the first \tilde{K} columns) of $\hat{\mathbf{D}}$ contains the displacements in the training images, and the right part (the last K columns) is the displacements in the test image. Note that we have $\mathbf{D} = \hat{\mathbf{D}}\mathbf{Q}$ if we define $\mathbf{Q} = \begin{bmatrix} \mathbf{0}_{\tilde{K} \times K} \\ \mathbf{I}_K \end{bmatrix}$, where \mathbf{I}_n is the $n \times n$ identity matrix, and $\mathbf{0}_{m \times n}$ is the $m \times n$ zero matrix.

Treating $\hat{\mathbf{D}}$ as a variable, our problem can be converted to finding the optimal $\hat{\mathbf{D}}$. Then, the optimal \mathbf{D} can be computed by $\mathbf{D} = \hat{\mathbf{D}}\mathbf{Q}$. To this end, we design an objective with regard to $\hat{\mathbf{D}}$:

$$E(\hat{\mathbf{D}}) = E_g(\hat{\mathbf{D}}) + \alpha E_s(\hat{\mathbf{D}}) + \beta E_p(\hat{\mathbf{D}}) + \gamma E_l(\hat{\mathbf{D}}) \quad (2)$$

Ground-Truth Penalty $E_g(\hat{\mathbf{D}})$. The left part of $\hat{\mathbf{D}}$ should be close to the ground-truth displacements in the training data, which is encoded in $\tilde{\mathbf{D}}$. Therefore, we want to minimize the Ground-truth Penalty:

$$E_g(\hat{\mathbf{D}}) = \frac{1}{2L\tilde{K}} \left\| \hat{\mathbf{D}} \begin{bmatrix} \mathbf{I}_{\tilde{K}} \\ \mathbf{0}_{K \times \tilde{K}} \end{bmatrix} - \tilde{\mathbf{D}} \right\|_F^2 = \frac{1}{2L\tilde{K}} \left\| \hat{\mathbf{D}}\mathbf{P} - \tilde{\mathbf{D}} \right\|_F^2 \quad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm, and $\mathbf{P} = \begin{bmatrix} \mathbf{I}_{\tilde{K}} \\ \mathbf{0}_{K \times \tilde{K}} \end{bmatrix}$.

Smooth Mapping Penalty $E_s(\hat{\mathbf{D}})$. First, we construct a compound feature matrix $\hat{\mathbf{F}} = [\tilde{\mathbf{f}} \mathbf{f}] \in \mathbb{R}^{d_f \times (\tilde{K} + K)}$. Now, each column of $\hat{\mathbf{F}}$ is the feature of a patch, and each column of $\hat{\mathbf{D}}$ is the displacement vector (to all landmarks) of a patch. Then, $\|\text{col}_i(\hat{\mathbf{F}}) - \text{col}_j(\hat{\mathbf{F}})\|_{L_2}$ is the L_2 feature distance of a pair of patches

(i, j) , where $\text{col}_i(\hat{\mathbf{F}})$ denotes the i th column of $\hat{\mathbf{F}}$. From all pairwise distances, we construct a binary affinity matrix $\mathbf{S} \in \{0, 1\}^{(\bar{K}+K)(\bar{K}+K)}$, where $\mathbf{S}_{ij} = 1$ if and only if the i th and the j th patches are mutually $k(k = 20)$ nearest neighbors in the feature space. Note that the edges in the affinity matrix might link two training patches, two test patches, or a training patch and a test patch.

The mapping from the feature space to the displacement space should be smooth. That is, for every pair of patches (i, j) , if they are similar in the feature space, their displacements to landmarks should also be similar. We define the Smooth Mapping Penalty $E_s(\hat{\mathbf{D}})$ as the violation from this assumption:

$$E_s(\hat{\mathbf{D}}) = \frac{1}{2L \sum_{i \neq j} \mathbf{S}_{ij}} \sum_{i \neq j} \mathbf{S}_{ij} \left\| \text{col}_i(\hat{\mathbf{D}}) - \text{col}_j(\hat{\mathbf{D}}) \right\|_{L_2}^2 \quad (4)$$

For each pair of patches, E_s introduces a high penalty if the two patches are similar in the feature space (i.e. $\mathbf{S}_{ij} = 1$) but their displacements are very different (i.e. $\left\| \text{col}_i(\hat{\mathbf{D}}) - \text{col}_j(\hat{\mathbf{D}}) \right\|_{L_2}$ is large). If we construct \mathbf{M} as the (trace normalized) laplacian matrix [11] of \mathbf{S} , E_s can be compactly written as:

$$E_s(\hat{\mathbf{D}}) = \frac{1}{L} \text{Tr} \left(\hat{\mathbf{D}} \mathbf{M} \hat{\mathbf{D}}^\top \right) \quad (5)$$

Patch Offset Penalty $E_p(\hat{\mathbf{D}})$. Each column of \mathbf{D} is the displacements from a single patch in the test image to all the landmarks. If we take the subtraction of two columns $\text{col}_{i-j}(\mathbf{D}) = \text{col}_i(\mathbf{D}) - \text{col}_j(\mathbf{D})$, it can be written as: $\text{col}_{i-j} \mathbf{D} =$

$$\mathbf{D}(\mathbf{e}_i^K - \mathbf{e}_j^K) = \begin{bmatrix} \mathbf{d}_i^1 - \mathbf{d}_j^1 \\ \dots \\ \mathbf{d}_i^L - \mathbf{d}_j^L \end{bmatrix}, \text{ where } \mathbf{e}_i^K \text{ is a } K \text{ dimensional column vector whose}$$

i th element is 1 and all other elements are 0s. From Fig. 1 (right), we can see that $\mathbf{d}_i^1 - \mathbf{d}_j^1 = \dots = \mathbf{d}_i^L - \mathbf{d}_j^L = \mathbf{c}_j - \mathbf{c}_i$, because $(\mathbf{d}_i^1, \mathbf{d}_j^1), \dots, (\mathbf{d}_i^L, \mathbf{d}_j^L)$ form triangles with the same edge $\mathbf{c}_j - \mathbf{c}_i$. Therefore, we impose a penalty $E_p^{i-j}(\mathbf{D}) = \left\| \mathbf{D}(\mathbf{e}_i^K - \mathbf{e}_j^K) - \bar{\mathbf{c}}_{j-i} \right\|_F^2$, where $\bar{\mathbf{c}}_{j-i}$ is the L times vertical replicate of $\mathbf{c}_j - \mathbf{c}_i$. We can include a penalty for each pair (i, j) of columns. For efficiency reasons, we eliminate redundancies and use $K - 1$ pairs to define the Patch Offset Penalty:

$$E_p(\hat{\mathbf{D}}) = \frac{1}{2LK} \sum_{i=1}^{L-1} E_p^{i-(i+1)}(\mathbf{D}) = \frac{1}{2LK} \left\| \hat{\mathbf{D}} \mathbf{Q} \mathbf{U} - \bar{\mathbf{C}} \right\|_F^2 \quad (6)$$

where $\mathbf{U} = [\mathbf{e}_1^K - \mathbf{e}_2^K, \dots, \mathbf{e}_{K-1}^K - \mathbf{e}_K^K]$ and $\bar{\mathbf{C}} = [\bar{\mathbf{c}}_{2-1} \dots \bar{\mathbf{c}}_{K-(K-1)}]$.

Landmark Offset Penalty $E_l(\hat{\mathbf{D}})$. Now we investigate the subtraction of rows in \mathbf{D} . For each pair of rows (i, j) , we can write $\text{row}_{i-j}(\mathbf{D}) = \text{row}_i(\mathbf{D}) - \text{row}_j(\mathbf{D}) = (\mathbf{e}_i^{2L} - \mathbf{e}_j^{2L})^\top \mathbf{D}$. Note that \mathbf{D} has $2L$ rows with interleaving coordinates in two axis as each \mathbf{d}_j^i is two dimensional. For any even number i between 1 and $2L$,

we have $\begin{bmatrix} \text{row}_{(i-1)-(i+1)}(\mathbf{D}) \\ \text{row}_{(i)-(i+2)}(\mathbf{D}) \end{bmatrix} = [\mathbf{d}_1^{i/2} - \mathbf{d}_1^{i/2+1}, \dots, \mathbf{d}_K^{i/2} - \mathbf{d}_K^{i/2+1}]$. Note that,

$(\mathbf{d}_1^{i/2}, \mathbf{d}_1^{i/2+1}), \dots, (\mathbf{d}_K^{i/2}, \mathbf{d}_K^{i/2+1})$ all form triangles with edge $\mathbf{x}_{i/2+1} - \mathbf{x}_{i/2}$ (Fig. 1 (right)). Therefore, although the value $\mathbf{x}_{i/2+1} - \mathbf{x}_{i/2}$ is not known, we however know that elements of $\text{row}_{i-(i+2)}(\mathbf{D})$ should be identical, i.e. with zero variance:

$$\text{Var}(\text{row}_{i-(i+2)}(\mathbf{D})) = (\mathbf{e}_i^{2L} - \mathbf{e}_{i+2}^{2L})^\top \mathbf{D} \mathbf{H}_k \mathbf{H}_k^\top \mathbf{D}^\top (\mathbf{e}_i^{2L} - \mathbf{e}_{i+2}^{2L}) = 0 \quad (7)$$

where \mathbf{H}_k is the $K \times K$ centering matrix $\mathbf{H}_k = \mathbf{I}_k - \frac{1}{K} \mathbf{1}_{K \times K}$, where $\mathbf{1}_{K \times K}$ is the $K \times K$ matrix whose elements are all 1s. Eq. (7) holds for every pair, and since variance is always non-negative, this is equivalent to say that the summation of all variances is also zero, and we define Landmark Offset Penalty:

$$E_l(\hat{\mathbf{D}}) = \frac{1}{2LK} \sum_{i=1}^{2L-2} \text{Var}(\text{row}_{i-(i+2)}(\mathbf{D})) = \frac{1}{2LK} \text{Tr}(\mathbf{V} \hat{\mathbf{D}} \mathbf{Q} \mathbf{H}_k \mathbf{H}_k^\top \mathbf{Q}^\top \hat{\mathbf{D}}^\top \mathbf{V}^\top) \quad (8)$$

where $\mathbf{V} = [\mathbf{e}_1^{2L} - \mathbf{e}_3^{2L}, \mathbf{e}_2^{2L} - \mathbf{e}_4^{2L}, \dots, \mathbf{e}_{2L-2}^{2L} - \mathbf{e}_{2L}^{2L}]^\top$.

2.3 Optimization

Substituting Eqs. (3),(5),(6) and (8) into Eq. (2), we get the final objective function. We can prove that Eq. (2) is convex, and the derivative is:

$$\partial E(\hat{\mathbf{D}}) / \partial \hat{\mathbf{D}} = \hat{\mathbf{D}} \mathbf{A} + \mathbf{B} \hat{\mathbf{D}} \mathbf{C} + \mathcal{G} \quad (9)$$

where $\mathbf{A} = \frac{1}{LK} \mathbf{P} \mathbf{P}^\top + \frac{2\alpha}{L} \mathbf{M} + \frac{\beta}{LK} \mathbf{Q} \mathbf{U} \mathbf{U}^\top \mathbf{Q}^\top$, $\mathbf{B} = \frac{\gamma}{LK} \mathbf{V}^\top \mathbf{V}$, $\mathbf{C} = \mathbf{Q} \mathbf{H}_k \mathbf{H}_k^\top \mathbf{Q}^\top$, $\mathcal{G} = -\frac{\tilde{\mathbf{D}} \mathbf{P}^\top}{LK} - \frac{\beta \tilde{\mathbf{C}} \mathbf{U}^\top \mathbf{Q}^\top}{LK}$. We use gradient descend for the global optimum $\hat{\mathbf{D}}$.

Note that in our optimization process we jointly optimize the training data $\tilde{\mathbf{D}}$ and the test data \mathbf{D} . We could also take another way where we optimize only \mathbf{D} . However, the advantage of our approach is that we can exploit the complex relations between each training and test patch which is reflected in the matrix \mathbf{S} in the term E_s .

2.4 Constructing Response Image

After we find the optimum $\hat{\mathbf{D}}$, we have $\mathbf{D} = \hat{\mathbf{D}} \mathbf{Q}$, and $\{\mathbf{c}_k + \mathbf{d}_k^l\}_{k=1 \dots K}$ will be the set of votes for the position of the l th landmark. From these votes we perform kernel density estimation, which gives us the probability of landmark at each pixel location of the image (this is called the *response image* of the landmark). The response images of all landmarks will be used later for shape segmentation.

3 Shape Segmentation via Statistical Shape Model

Our shape segmentation algorithm works by combining the landmark detection with a statistical shape model. First, we detect a small set of global landmarks by exhaustively search in the image in different scales and rotations.

According to the position of these global landmarks, the image is scaled, translated and rotated to compensate for the global transformation. Then, the landmarks defining the shape are detected. During this step, we divide the whole shape into multiple subshapes consisting of several nearby successive landmarks. The method introduced in Section 2 is performed on each subshape. Then, shape segmentation is performed using response images of landmarks along with a statistical shape model. Instead of using the classical Active Shape Model [13] via PCA, we adopt the shape model based on sparse representation technique introduced in [14].

4 Experiments

Data: We tested our method on three tasks: segmentation of complete femur, proximal femur and pelvis. The X-ray data come from our clinical partner:

- CompleteFemur: 80 training images, 109 test images.
- ProximalFemur: 100 training images, 188 test images.
- Pelvis: 100 training images, 163 test images.

Note that a considerable part of the images are post-operative x-ray radiographs after trauma or joint replacement surgery, which significantly increases the challenge due to large variation of femur/pelvis appearance and the presence of implants (as can be seen in Fig. 2). As a indication, we made a manual counting, which shows that 32% of the test images contain implants.

Implementation Details: Each shape is divided into subshapes of 4 successive landmarks (i.e. $L=4$ in Section 2). The landmark detection algorithm in Section 2 is performed on each subshape, and then the segmentation of the whole shape is derived as in Section 3. In Section 2, for the patch visual feature, we use multi-level HoG (Histogram of Oriented Gradient [12]) feature with block sizes 1×1 and 2×2 . Each block is divided into 2×2 cells and for each cell an 18 dimensional HoG feature is extracted by histogramming the gradient direction of each pixel. Therefore, feature dimension $d_f = 360$. For each subshape, we sample $\tilde{K} = 2000$ training patches and $K = 2000$ test patches. For the objective function, we use $\alpha = 0.1, \beta = 0.1, \gamma = 0.1$.

4.1 Results

Fig. 2 shows the qualitative result for all the three segmentation tasks. We can see that despite the challenges such as significant variation of appearance, poor image contrast and implants, our method achieves good results. For quantitative evaluation, we manually annotate the contour on all the test images, and the segmentation error is then calculated by the average point-to-curve distance between the landmarks of the segmented shape and the annotated contour. The error is then converted from pixel unit to physical unit (mm) as the pixel spacing of our radiograph is known. Table 1 shows the result. Note that to calculate the success rate, the failure cases are manually identified.

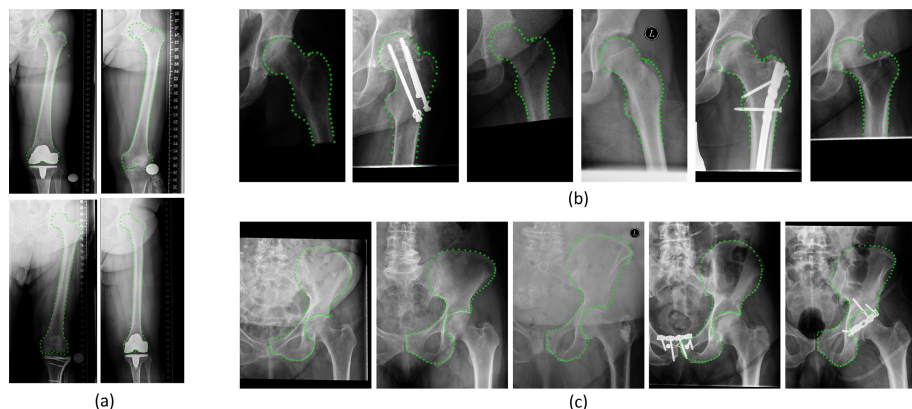


Fig. 2. Segmentation result on (a): CompleteFemur; (b): ProximalFemur; (c): Pelvis

Table 1. Quantitative evaluation. Numbers are in unit mm.

Anatomy	Success rate	Median	Minimum	Maximum	Mean	Standard deviation.
CompleteFemur	100%	1.2	0.7	3.4	1.3	0.4
ProximalFemur	98.4%	1.3	0.6	3.8	1.4	0.6
Pelvis	98.8%	1.9	1.0	4.3	2.2	0.8

Table 2. Comparison of our method with RF method. Numbers are in unit mm.

	CompleteFemur		ProximalFemur		Pelvis	
	Mean	Std	Mean	Std	Mean	Std
Our method	1.3	0.42	1.4	0.62	2.2	0.80
RF method	1.8	0.55	1.3	0.54	2.5	0.78
p-value	0.00005		0.31		0.05	

Our method processes one image of resolution 3000×3000 in about 2 minutes, with an unoptimized Matlab implementation on a computer with 3.0GHz CPU.

We compare our method with the random forest (RF) regression method. We use the same parameter for both methods when applicable (e.g. the same patch feature, the same number of training/test patches). For RF, we use 5 trees per forest¹. The results are shown in Table 2, which shows the average error and standard deviation (in mm) of the two methods as well as the p-value. From the table we see that our method have comparable performance in proximal femur segmentation, and outperforms RF in complete femur and pelvis segmentation.

5 Conclusions

We have proposed a new method for fully-automatic landmark detection and shape segmentation in X-ray images. Our method works by jointly estimating

¹ We also tried more than 5 trees and no notable improvement was found.

the displacements from test patches to landmarks by considering both training data and geometric constraints. Experiments show that our method improves the landmark detection accuracy, and, combined with statistical shape model, can accurately segment shapes on the challenging datasets. In the future, we would like to integrate the shape model into our joint regression framework.

Acknowledgement. This work is supported by SNSF Project 51NF40-144610.

References

1. Chen, Y., Ee, X., Leow, W.-K., Howe, T.S.: Automatic extraction of femur contours from hip X-ray images. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 200–209. Springer, Heidelberg (2005)
2. Lindner, C., Thiagarajah, S., Wilkinson, J.M., Wallis, G.A., Cootes, T.F.: Accurate fully automatic femur segmentation in pelvic radiographs using regression voting. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 353–360. Springer, Heidelberg (2012)
3. Gottschling, H., Roth, M., Schweikard, A., Burgkart, R.: Intraoperative, fluoroscopy-based planning for complex osteotomies of the proximal femur. *Int. J. Med. Robot.* 1(3), 67–73 (2005)
4. Baka, N., Kaptein, B.L., Bruijne, M., van Walsum, T., Giphart, J.E., Niessen, W.J., Lelieveldt, B.P.: 2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape model. *Med. Image Anal.* 15(6), 840–850 (2001)
5. Dong, X., Zheng, G.: Automatic extraction of proximal femur contours from calibrated X-ray images using 3D statistical models: an in vitro study. *Int. J. Med. Robot.* 5(2), 213–222 (2009)
6. Cristinacce, D., Cootes, T.: Automatic feature localization with constrained local models. *Pattern Recognition* 41(19), 3054–3067 (2008)
7. Zhou, S.K., Comaniciu, D.: Shape regression machine. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 13–25. Springer, Heidelberg (2007)
8. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation of 3-D cardiac CT volumes using marginal space learning and steerable features. *IEEE T. Med. Imaging* 27(11), 1668–1681 (2008)
9. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast multiple organ detection and localization in whole-body MR Dixon sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
10. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011)
11. Kokiopoulou, E., Chen, J., Saad, Y.: Trace optimization and eigenproblems in dimension reduction methods. *Numerical Linear Algebra with Applications* 18(3), 565–602 (2011)
12. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
13. Cootes, T.F., Taylor, C.J.: Active shape models—‘smart snakes’. In: BMVC (1992)
14. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Sparse shape composition: a new framework for shape prior modeling. In: CVPR (2011)

IntellEditS: Intelligent Learning-Based Editor of Segmentations

Adam P. Harrison^{1,2}, Neil Birkbeck¹, and Michal Sofka¹

¹ Siemens Corporation, Corporate Technology, Princeton, NJ, USA

² University of Alberta, Edmonton, Canada

Abstract. Automatic segmentation techniques, despite demonstrating excellent overall accuracy, can often produce inaccuracies in local regions. As a result, correcting segmentations remains an important task that is often laborious, especially when done manually for 3D datasets. This work presents a powerful tool called Intelligent Learning-Based Editor of Segmentations (IntellEditS) that minimizes user effort and further improves segmentation accuracy. The tool partners interactive learning with an energy-minimization approach to editing. Based on interactive user input, a discriminative classifier is trained and applied to the edited 3D region to produce soft voxel labeling. The labels are integrated into a novel energy functional along with the existing segmentation and image data. Unlike the state of the art, IntellEditS is designed to correct segmentation results represented not only as masks but also as meshes. In addition, IntellEditS accepts intuitive boundary-based user interactions. The versatility and performance of IntellEditS are demonstrated on both MRI and CT datasets consisting of varied anatomical structures and resolutions.

1 Introduction

Interactive approaches to segmentation have a proven track record [9]. However, the important task of interactively editing a pre-existing but imperfect segmentation, or *pre-segmentation*, has not enjoyed much attention [6,11,8]. This is unfortunate, as manually correcting segmentations in 3D medical imaging applications can be a time-consuming but necessary task. Local corrections are often needed for fully-automatic segmentation techniques [6,8], which are powerful and time-saving but still have not matched the performance of interactive tools in key applications [9]. To use all available information, an editor should consider: (1) user interactions, (2) the presegmentation, and (3) the underlying volume or data. These sources should guide an editing algorithm that strives to (1) minimize number of edits to achieve desired result, (2) minimize segmentation error after each edit, and (3) respect user guidance.

Apart from 2D-focused tools [10,15], many 3D editors function by propagating user corrections from an interaction plane to the larger 3D volume [6,3,8]. Propagation can be achieved by minimizing an energy term constrained by user input, the presegmentation, and the volumetric data [6]. Such approaches face several key challenges. First, volumetric data is incredibly rich and varied, making it non-trivial in how to best employ it. Ideally, an editor would use volume data in a way that can generalize to different

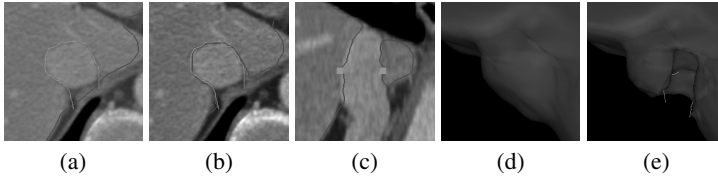


Fig. 1. Removing the vena cava from a liver mesh (red) using IntellEditS. (a) User chooses an interaction plane and draws a corrected boundary (green); (b) Mesh is updated in the plane; (c) As shown in a perpendicular plane, mesh boundary is accurately propagated in 3D; (d)-(e) The 3D propagation can also be viewed by comparing before and after surfaces in (d) and (e), respectively. As IntellEditS is executed within an MPR viewer, pre- and post-edit segmentation accuracy can be quickly assessed.

modalities, imaging qualities, and anatomical structures. Second, many presegmentations are mesh-based, meaning they must be reconciled with the voxel-based volume data. Third, user input should be as intuitive and user-friendly as possible.

This work presents an interactive editor of 3D presegmentations that simultaneously addresses all the above challenges. First, the system allows users to precisely draw new boundaries using lines within the interaction plane. This is called *splice-based interactions*. Second, a discriminative classifier uses foreground and background regions defined by the splice to model voxels inside and outside the object being edited. Classification results of unknown voxels are then incorporated within an energy functional that locally propagates user interactions to the 3D volume. Finally, when editing mesh presegmentations, a soft voxel-based representation is used, reconciling it with the volume data while still retaining a highly accurate boundary. These features culminate in a system entitled Intelligent Learning-Based Editor of Segmentations (IntellEditS) that provides a sophisticated and flexible means to rapidly edit presegmentations. Fig. 1 visually depicts the steps involved in editing.

IntellEditS advances the state of the art of *data-driven* editing through several means. For instance, many editing techniques use interactions, such as brush-based tools [6,3], that are nonintuitive for clinicians [8] as they do not allow precise correction of boundaries. In addition, state of the art data-driven editors only consider mask-based presegmentations [6,3,8], meaning that IntellEditS is the first to simultaneously perform mesh-based and data-driven editing. Finally, the above works only use local differences in volume intensity and do not attempt to model foreground and background voxels. This state of data-driven editing is in contrast to many interactive segmentation techniques, which successfully leverage learning algorithms as they build up a segmentation from scratch [16,17,13,18].

With these advances, IntellEditS fills an important gap in the state of the art of data-driven editing. The algorithmic details of IntellEditS are expounded further in the methodology section (§2). As demonstrated in the results section (§3), using identical parameters IntellEditS can perform effectively on volumetric data coming from different anatomical structures, modalities, imaging conditions, resolutions, and anisotropic characteristics. This work is concluded by a discussion of results (§4).

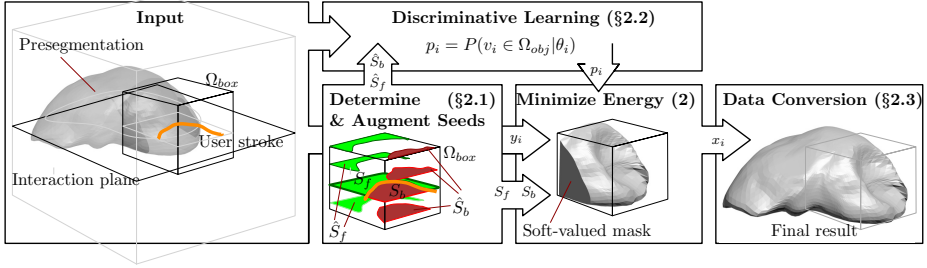


Fig. 2. Algorithm steps of IntellEditS

2 Interactive Learning-Based Editing

IntellEditS partners interactive discriminative classification with energy-based minimization in order to edit mesh- or mask-based representations of anatomical structures. Fig. 2 provides a high-level view of IntellEditS' algorithm steps.

First, a user corrects a presegmentation on a 2D interaction plane using splices. Foreground and background seed points, S_f and S_b respectively, are calculated based on user input and the presegmentation. If the presegmentation is a mesh, it is converted to a floating-point distance map, y_i , where i indexes individual voxel locations within the volume. IntellEditS only operates on a cropped region of the volume, Ω_{box} , which is specified using a bounding box around the splice line and the region of the presegmentation contained within said splice.

An augmented set of seed points, \hat{S}_f and \hat{S}_b , serve as inputs to IntellEditS' classifier, which learns how to discriminate between foreground and background regions based on a pool of features for each voxel. Features are denoted using an ordered vector of values θ_i . In formal terms, define a variable $p_i \in [0, 1]$, which describes the probability that voxel v_i is in the foreground. Based on the trained model, the classifier calculates the following posterior probability for each non-seed voxel:

$$p_i = P(v_i \in \Omega_{obj} | \theta_i), \quad (1)$$

where Ω_{obj} denotes the true boundary of the anatomical object being edited.

If the p_i values were completely accurate, the editing task would be finished. However, as complete accuracy cannot be guaranteed, p_i values are employed as part of an energy minimization formulation that incorporates the presegmentation, user seeds, and a regularizer. Formally, the energy formulation can be expressed as:

$$E(\mathbf{x}) = \sum_{e_{ij}} w_{ij} (x_i - x_j)^2 + \gamma_i \sum_{i \in \Omega_{box}} (y_i - x_i)^2 + \alpha \sum_{i \in \Omega_{box}} (p_i - x_i)^2 \quad (2)$$

$$\text{s.t.}, \begin{cases} x_i = 1, v_i \in S_f \\ x_i = 0, v_i \in S_b \end{cases},$$

where x_i are soft output potential values and w_{ij} denotes the weights assigned to the graph edges e_{ij} connecting each vertex v_i or voxel. The first summation acts as a regularizer, ensuring coherent output potentials. The second summation incorporates the

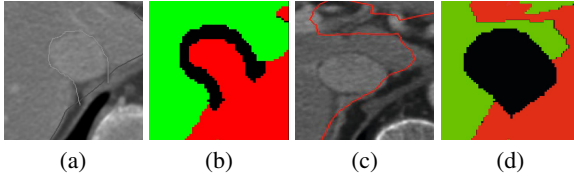


Fig. 3. Calculating Seed Points. (a) an edit and the presegmentation; (b) corresponding foreground (S_f) and background (S_b) seed points in green and red, respectively; (c) volume and presegmentation above the editing plane; (d) corresponding foreground (\hat{S}_f) and background (\hat{S}_b) augmented seed points in green and red, respectively.

presegmentation, where γ_i is a local parameter controlling its influence. IntellEditS uses the same γ_i and w_{ij} values as Grady and Funka-Lea [6]. However, floating-point presegmentation values are used instead of binary values. Finally, the third summation incorporates the per-voxel probabilities of being in the foreground. The influence of the p_i values are controlled by α , which has a context sensitive value explained in Sect. 2.2. IntellEditS minimizes (2) using the random walker algorithm [5].

Apart from using splice-based interactions and accommodating mesh-based presegmentations, the third summation in (2) represents one of the most important departures from the state of the art. While combining learning with energy minimization has proven successful in interactive segmentation, it has not been used to locally *edit* presegmentations. Other energy-minimization-based editors only employ volume data to calculate w_{ij} [6,3,10,15]. As a result, these methods neglect the highly informative ensemble of *local* foreground and background voxel features that users implicitly specify during edits. Thus, the strength of (2) rests on its use of all available sources of information—foreground/background features, presegmentation, and user seeds.

In order to apply (2) in a concrete implementation, the generation of seed points must be specified (§2.1). The seed points define the fixed regions of the segmentation and are also used to train the discriminative classifier (§2.2). After solving (2), the voxel-based output potentials are converted to a mesh-based representation (§2.3) or thresholded for mask-based presegmentations.

2.1 Determining and Augmenting Seeds

When a user draws a splice intersecting a presegmentation, she is drawing a new 2D boundary. As Fig. 3(a) and (b) demonstrate, corresponding seed points, S_f and S_b , can be inferred by flood filling the new boundary. These are only extracted from the 2D interaction plane. A buffer ensures any voxels on or close to the new boundary are not chosen as seed points, allowing IntellEditS to settle on a precise iso-contour instead.

S_f and S_b provide reliable training samples. However, since IntellEditS must propagate an edit away from the interaction plane, training samples should be extracted from the larger volume. As well, any learning algorithm’s performance partly hinges on the number of training samples used. For this reason, seed points are augmented into two larger sets, \hat{S}_f and \hat{S}_b . These are calculated by projecting locations of out-of-editing-plane voxels onto the editing plane itself. If their 2D projected location is far enough

from the user’s splice then they are included into \hat{S}_f (\hat{S}_b) if they are in the presegmentation foreground (background). One such example is provided by Fig. 3(c) and (d). While there is a possibility of incorrectly labelled samples entering the training set, experience indicates the impact, if any, to be minimal.

2.2 Discriminative Learning

Both generative [16,17] and discriminative [13,18] models have been used to interactively segment visual data. While powerful, generative-model performance hinges on selecting an appropriate model and the correct feature(s) to examine. This is a challenge when faced with different modalities, anatomical structures, image qualities, and editing contexts. Since IntellEditS’ goal is to operate effectively even on dataset types not encountered during development, it employs discriminative classification to directly model posterior probability. Classification features are drawn from a large pool of 3D Haar wavelets, where their relative influence varies based on the dataset.

This work uses a Probabilistic Boosting Tree (PBT) [14], whose nodes are composed of AdaBoost classifiers. Unlike much of previous work using PBTs, classification must execute at interactive speeds. A recent work by Birkbeck *et al.* documented a GPU PBT implementation [4], which also included fast calculation of 3D Haar features. However, since classifiers were still trained offline using multiple volumes, the algorithm was not meant to be interactive and did not address speeding up training. In contrast, IntellEditS requires *both* fast training and detection using a *single* volume. As a result, IntellEditS extends Birkbeck *et al.*’s work by implementing an interactive-speed PBT (I-PBT) for single-volume classification. Training is GPU-implemented using CUDA.

Based on feature values of \hat{S}_f and \hat{S}_b , IntellEditS trains the I-PBT to discriminate between foreground and background voxels. After training, the detection accuracy, $\rho \in [0, 1]$, of the I-PBT in classifying the training samples is calculated. The weight parameter, α , in (2) is then set to 0.5ρ , providing an automatic and performance-based tuning of the editor.

2.3 Data Conversion

Unlike previous works, IntellEditS’ goal is to edit meshes in addition to masks. Nonetheless, outside of pre- and post-processing steps to convert data, IntellEditS works within a voxel-based domain. This enables parallel execution of many of the per-voxel tasks in training and detection. As well, resampling of the volume and its features, commonly needed in simplex-based representations of voxel data, is avoided.

To retain the high resolution of mesh-based presegmentations, IntellEditS converts them to soft-valued voxel representations based on distance-map calculations. For the mesh-to-distance-map direction, IntellEditS uses the fast pseudo-normal method [2]. Converting the distance map back to a mesh is accomplished using the marching cubes algorithm [12]. The authors’ experience indicates that this conversion does not affect visual quality of the boundary. Since (2) works within a $[0, 1]$ range, distance-map values must be mapped to this range using a scale and offset. The reverse mapping must also be executed before converting output potentials to a mesh.

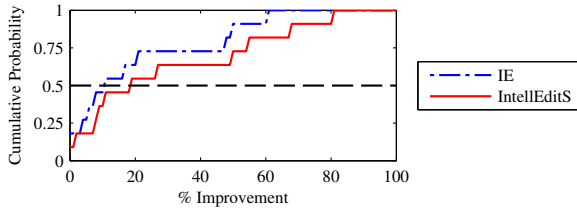


Fig. 4. Cumulative probability distribution of improvements in SPS error after IE and IntellEditS corrections. The median improvements of the IE and IntellEditS were 11 and 19% respectively and third-quartile improvements were 48 and 55%, respectively.

3 Results

Experiments tested IntellEditS’ performance on 11 CT and MRI datasets of different anatomical structures, resolutions, and anisotropies. Using the *same* single splice, experiments compared IntellEditS’ performance against an intensity-based editor (IE) that only uses the first two summations in (2). As such, IE does not model foreground and background voxels and is similar to Grady and Funka-Lea’s approach [6], except it has been significantly modified to use splice interactions and edit meshes. Each tool’s performance was gauged using symmetrical point-to-surface (SPS) error against a manually-annotated ground truth in the Ω_{box} region. These results were then compared against the original SPS error of the presegmentation.

All experiments used the same parameter values, detailed in §2. For all datasets, the I-PBT was configured to have a depth of 3 with 10 weak classifiers at each node. On average, editing time consumed 3 seconds on an 8 core machine with an NVIDIA GeForce 9800 GT video card. As Fig. 4 illustrates, IntellEditS is able to more effectively reduce segmentation errors after the same user interaction. Fig. 5 visually demonstrates the effectiveness of IntellEditS by depicting representative qualitative results drawn from the quantitative experiment.

4 Discussion and Conclusion

This work presented an interactive editing tool called IntellEditS that represents a novel and powerful way to correct presegmentations in a 3D context. Unlike previous data-driven editors, IntellEditS can edit meshes. Users employ intuitive splice-based interactions to correct presegmentations on a 2D interaction plane, which IntellEditS propagates in 3D through a combination of discriminative learning coupled with energy minimization. The practical benefits and versatility of IntellEditS were demonstrated in quantitative and qualitative experiments composed of challenging datasets of various modalities, anatomical structures, resolutions, and anisotropy. Comparing against an intensity-based editor and using the same user input, these experiments demonstrated that IntellEditS can output a mesh significantly closer to ground truth, saving valuable end-user time and effort. Interesting directions of future work include combining IntellEditS with tools that can help choose suspicious interaction planes [1], providing the

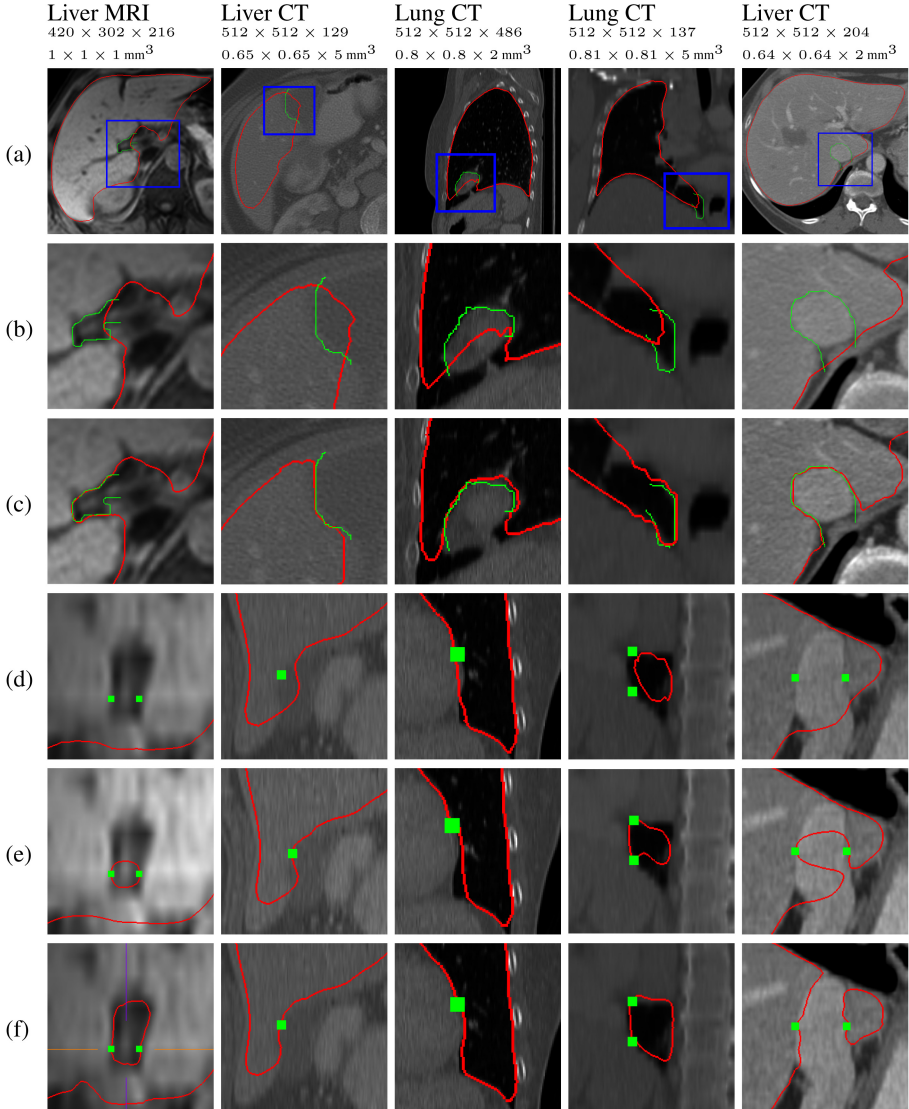


Fig. 5. Qualitative comparison of IntellEditS vs IE using the same *single* spline. (a) overall view of presegmentation and editing spline; (b) zoomed-in view of (a); (c) corrected mesh; (d) cross-section view of presegmentation; (e) and (f) same view as (d) but after the correction produced by IE and IntellEditS, respectively. Dataset dimensions and resolutions are also provided.

option to use live-wire techniques [7], and incorporating online learning into the editing process to incrementally learn a more robust model from consecutive interactions.

Acknowledgments. We thank S. Kevin Zhou, Noha El-Zehiry, and Enrico Kuhn for valuable discussions, feedback, and code that contributed to this paper.

References

1. Top, A., Hamarneh, G., Abugarbieh, R.: Active Learning for Interactive 3D Image Segmentation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 603–610. Springer, Heidelberg (2011)
2. Baerentzen, J.A., Aanaes, H.: Generating Signed Distance Fields From Triangle Meshes. Tech. rep., Informat. and Math. Mod., Tech. Univ. of Denmark (2002)
3. Beichel, R., Bauer, C., Bornik, A., Sorantin, E., Bischof, H.: Liver segmentation in CT Data: A Segmentation Refinement Approach. In: Proceedings of 3D Segmentation in the Clinic: A Grand Challenge, pp. 235–245 (2007)
4. Birkbeck, N., Sofka, M., Zhou, S.: Fast Boosting Trees for Classification, Pose Detection, and Boundary Detection on a GPU. In: 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, pp. 36–41 (2011)
5. Grady, L.: Random Walks for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(11), 1768–1783 (2006)
6. Grady, L., Funka-Lea, G.: An Energy Minimization Approach to the Data Driven Editing of Presegmented Images/Volumes. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 888–895. Springer, Heidelberg (2006)
7. Hamarneh, G., Yang, J., McIntosh, C., Langille, M.: 3D live-wire-based semi-automatic segmentation of medical images. In: Proceedings of SPIE Medical Imaging: Image Processing, vol. 5747, pp. 1597–1603 (2005)
8. Heckel, F., Moltz, J.H., Bornemann, L., Dicken, V., Bauknecht, H.C., Fabel, M., Hittinger, M., Kießling, A., Meier, S., Püsken, M., Peitgen, H.O.: 3D contour based local manual correction of tumor segmentations in CT scans. In: Proceedings of SPIE Medical Imaging: Image Processing, vol. 7259, pp. 1–9
9. Heimann, T., van Ginneken, B., et al.: Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. *IEEE Trans. Med. Imaging* 28(8), 1251–1265 (2009)
10. Jagadeesh, V., Manjunath, B.: Interactive graph cut segmentation of touching neuronal structures from electron micrographs. In: 2010 17th IEEE International Conference on Image Processing, ICIP, pp. 3625–3628 (2010)
11. Kang, Y., Engelke, K., Kalender, W.A.: Interactive 3D editing tools for image segmentation. *Medical Image Analysis* 8(1), 35–46 (2004)
12. Lorensen, W.E., Cline, H.E.: Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Computer Graphics* 21(4), 163–169 (1987)
13. Santner, J., Unger, M., Pock, T., Leistner, C., Saffari, A., Bischof, H.: Interactive Texture Segmentation using Random Forests and Total Variation. In: British Machine Vision Conference, BMVC (2009)
14. Tu, Z.: Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1589–1596 (2005)
15. Yang, H.-F., Choe, Y.: An Interactive Editing Framework for Electron Microscopy Image Segmentation. In: Bebis, G., et al. (eds.) ISVC 2011, Part I. LNCS, vol. 6938, pp. 400–409. Springer, Heidelberg (2011)
16. Yang, Q., Tang, X., Wang, C., Ye, Z., Chen, M.: Progressive Cut: An Image Cutout Algorithm that Models User Intentions. *IEEE MultiMedia* 14(3), 56–66 (2007)
17. Yang, W., Cai, J., Zheng, J., Luo, J.: User-friendly Interactive Image Segmentation Through Unified Combinatorial User Inputs. *IEEE Trans. Image Process* 19(9), 2470–2479 (2010)
18. Zhao, Y., Zhu, S.C., Luo, S.: CO3 for Ultra-fast and Accurate Interactive Segmentation. In: Proceedings of the International Conference on Multimedia, pp. 93–102. ACM (2010)

Automatic Nuchal Translucency Measurement from Ultrasonography

JinHyeong Park¹, Michal Sofka¹, SunMi Lee²,
DaeYoung Kim², and S. Kevin Zhou¹

¹ ICV TF, Siemens Corporation, Corporate Technology, Princeton, NJ 08540 USA

² H CP US PLM, Siemens Limited Seoul, Bundang Seongnam, Gyeonggi, Korea

Abstract. This paper proposes a fully automatic approach for computing Nuchal Translucency (NT) measurement in an ultrasound scans of the mid-sagittal plane of a fetal head. This is an improvement upon current NT measurement methods which require manual placement of NT measurement points or user-guidance in semi-automatic segmentation of the NT region. The algorithm starts by finding the pose of the fetal head using discriminative learning-based detectors. The fetal head serves as a robust anchoring structure and the NT region is estimated from the statistical relationship between the fetal head and the NT region. Next, the pose of the NT region is locally refined and its inner and outer edge approximately determined via Dijkstra's shortest path applied on the edge-enhanced image. Finally, these two region edges are used to define foreground and background seeds for accurate graph cut segmentation. The NT measurement is computed from the segmented region. Experiments show that the algorithm efficiently and effectively detects the NT region and provides accurate NT measurement which suggests suitability for clinical use.

1 Introduction

Nuchal Translucency (NT) refers to the fluid-filled region under the skin of posterior neck of a fetus. Increased NT in early gestation period is correlated to the high risk of major cardiac defects and chromosomal deflection including Down Syndrome [1]. The NT screening is performed in the first trimester of pregnancy using ultrasound scans. A sonographer first needs to navigate to the mid-sagittal plane containing echogenic nasal tip along with nasal bone, and translucent diencephalon in the center (Fig. 1). The plane is then stored and the measurement computed from manually placed marks. The accuracy requirement on the measurement is very high since even a slight deviation can completely change the diagnosis [1]. This makes the manual measurement and the design of automated tools difficult.

There have been semi-automatic approaches [2,3] to detect the NT measurement in 2D scans of mid-sagittal plane. In these approaches, the user is required to manually specify a region-of-interest (ROI) surrounding the NT measurement location. The ROI is used as input to the segmentation algorithm for finding upper and lower edges of the NT region. The final NT thickness is computed from

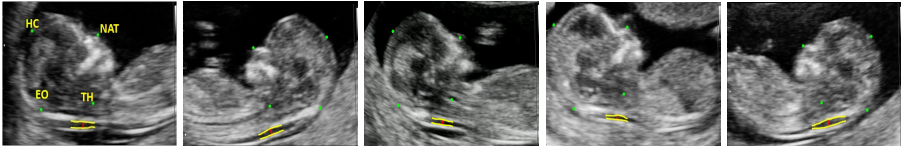


Fig. 1. Fetal ultrasound scans for measuring Nuchal Translucency (NT) in the mid-sagittal plane. NT (yellow lines) is a small fluid-filled region behind fetus neck. The NT detection is constrained by automatically found anchoring structure, fetal head, defined by stable landmarks (green): head crown (HC), throat (TH), nasal tip (NAT), and external occipital region (EO). The NT measurement is shown in red.

the traced edges as the maximum distance between them. Similarly, measuring the NT in 3D images is also performed on the mid-sagittal plane, but the plane must be first found from the acquired ultrasound volume [4].

This paper proposes a fully automatic solution for NT measurement in 2D ultrasound scans of a mid-sagittal plane. Fig. 1 depicts typical examples of NT images used for NT measurements. The algorithm first detects the fetal head by implicitly relying on the appearance of stable structures that characterize the mid-sagittal plane: echogenic nasal tip, nasal bone, and translucent diencephalon. The NT region is predicted based on a statistical model obtained from the relative poses of the fetal head and the NT region. Approximate edges of the NT region are then found using Dijkstra’s shortest path algorithm with graph weights extracted from the intensity image [5]. The edges are used to define foreground and background seed points for accurate Graph Cut segmentation algorithm [6]. Finally, the NT measurement value is computed from the segmentation result at the location of maximum thickness. To the best of our knowledge, this work is the first to propose the fully automatic NT measurement.

The paper is organized as follows. Background literature is briefly reviewed in Section 2. Section 3 formulates the problem of estimating the NT region using fetal head as an anchoring structure. The segmentation algorithm for accurately finding the NT edges is discussed in Section 4. Experimental results are presented in Section 5. The paper is concluded in Section 6.

2 Background

As mentioned in Section 1, the previously proposed methods for NT measurement in 2D ultrasound scans require manual steps [2,3]. The early techniques for automatically detecting other fetal structures relied on filtering, morphological operators, and Hough transform [7]. These techniques tend to be slow and are typically designed for a specific anatomy which makes them difficult to generalize to new structures. Chalana et al. [8] describe a method for detecting the biparietal diameter and head circumference based on active contour model. The algorithm does not use the image appearance which is necessary to increase the robustness and accuracy. Carneiro et al. [9] proposed a system for detecting and

measuring several anatomies in 2D ultrasound images using the same underlying algorithm. The method learns to discriminate between the structures of interest and background via a Probabilistic Boosting Tree classifier [10]. The appearance variations and imaging artifacts are captured by a large annotated database of images. An efficient search technique makes the system run in under half second. However, the detected structures are large (e.g. fetal limbs, abdominal circumference, crown rump length) and constraints between the structures are not exploited. Deng et al. [11] proposed an algorithm to locate NT bounding box using hierarchical model constructed from three SVM-based detectors of NT, Head, and Body. The algorithm has several limitations. It uses very coarse scale range of only 7 levels and does not provide NT orientations. Therefore, it does not provide automatic NT measurements, but only axis-aligned NT bounding boxes.

3 Automatic Estimation of the NT Region

In this section, we formally discuss the overall approach to automatically find the NT region. Let θ denote the parameter of NT pose represented by its position (x, y) , width w , height h , and orientation α , and \mathcal{I}_θ denote the observed image patch parametrized by θ . Similarly, let us define φ and \mathcal{I}_φ for fetal head. The goal is to find the parameter of the best pose $\hat{\theta}$ as:

$$\hat{\theta} = \arg \max_{\theta} P(\theta|\mathcal{I}) = \arg \max_{\theta} \int_{\varphi} P(\theta, \varphi|\mathcal{I}_\theta, \mathcal{I}_\varphi) d\varphi. \quad (1)$$

The term $P(\theta, \varphi|\mathcal{I}_\theta, \mathcal{I}_\varphi)$ inside the integral in Eq. (1) can be reformulated as follows by applying the Bayesian rule:

$$P(\theta, \varphi|\mathcal{I}_\theta, \mathcal{I}_\varphi) = \frac{P(\mathcal{I}_\theta|\theta, \varphi, \mathcal{I}_\varphi)P(\theta, \varphi|\mathcal{I}_\varphi)}{P(\mathcal{I}_\theta|\mathcal{I}_\varphi)} = \frac{P(\mathcal{I}_\theta|\theta)P(\theta, \varphi|\mathcal{I}_\varphi)}{P(\mathcal{I}_\theta|\mathcal{I}_\varphi)}, \quad (2)$$

where $P(\mathcal{I}_\theta|\theta, \varphi, \mathcal{I}_\varphi)$ can be simplified to $P(\mathcal{I}_\theta|\theta)$ because \mathcal{I}_θ is only depended by θ . By substituting Eq. (2) into Eq. (1), we obtain the following objective function:

$$\begin{aligned} \hat{\theta} &= \arg \max_{\theta} \int_{\varphi} \frac{P(\mathcal{I}_\theta|\theta)P(\theta, \varphi|\mathcal{I}_\varphi)}{P(\mathcal{I}_\theta|\mathcal{I}_\varphi)} d\varphi \\ &\approx \arg \max_{\theta} P(\mathcal{I}_\theta|\theta) \int_{\varphi} P(\theta|\varphi, \mathcal{I}_\varphi)P(\varphi|\mathcal{I}_\varphi) d\varphi \\ &= \arg \max_{\theta} P(\mathcal{I}_\theta|\theta) \int_{\varphi} P(\theta|\varphi)P(\mathcal{I}_\varphi|\varphi)P(\varphi) d\varphi. \end{aligned} \quad (3)$$

The first term in Eq. (3), $P(\mathcal{I}_\theta|\theta)$, is the likelihood of \mathcal{I}_θ given θ , which is obtained from a discriminative classifier trained to detect presence or absence of the NT region. Similarly, the likelihood $P(\mathcal{I}_\varphi|\varphi)$ is obtained from a discriminative classifier for the fetal head. The term $P(\theta|\varphi)$ represents the transition probability

of θ from φ , which is learned in a nonparametric manner to capture the geometric relationship between the fetal head and NT regions. The term $P(\varphi)$ is the prior probability for fetal head that is set as uniform.

3.1 Detection of Fetal Head and Nuchal Translucency

The algorithm employs discriminative learning [10,12,13,14] to compute the likelihoods of observing the fetal head and NT region Eq. (3). Robustness of the system is achieved by exploiting spatial relationships and multi-resolution hierarchy using Hierarchical Detection Network (HDN) [15]. In our case, the pose of the NT region is predicted using the most reliable candidates (samples) of the fetal head pose detected by the fetal head detector which is trained using three resolution levels (0.4 mm, 0.2 mm and 0.1 mm). The NT region detector uses two resolution levels (0.2 mm and 0.1 mm). The final HDN network is shown in Fig. 2-(a).

The multi-resolution hierarchy, has the advantage that the complexity of the training and detection (and therefore the complexity of the classifiers) is reduced by distributing it into each level. Global structures are best detected at the coarse resolution and provide constraints for local detection at finer resolutions. This way, the search space for the local detection is reduced which increases robustness since many image regions are never considered. In addition, computational speed is increased due to the reduced search space.

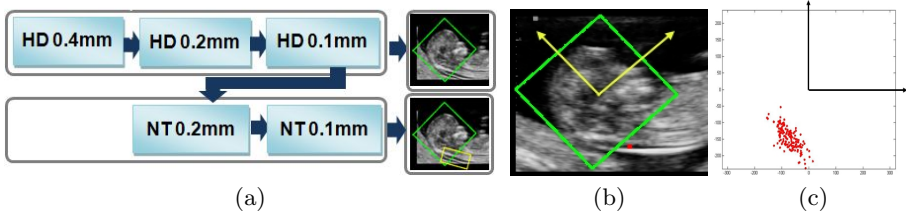


Fig. 2. (a) The hierarchy of the detectors. HD represents the fetal head detector and NT represents the NT region detector. 0.4 mm, 0.2 mm and 0.1 mm indicate the three image resolutions. (b) Fetal head bounding box (green) and NT location (red point). The two yellow axes define the local coordinate system based on the center and orientation of the fetal head. (c) The distribution of the NT location (red points) in the local coordinate system of the fetal head defined in (b). The black coordinate axes correspond to the yellow axes in (b).

It is clear that the successful detection of the NT region hinges on the modeling of the statistical relationship between the fetal head and the NT region. In this modeling, we rely on the strong anatomical prior present in the domain of fetal head scans. Fig. 2-(b) and (c) illustrate the statistical relationship between the NT position predicted from the pose of the head. For each image, the NT location is computed based on the local coordinate system defined by the head as

illustrated in Fig. 2-(b). The distribution of NT location in the local coordinate system is shown in Fig. 2-(d) where the origin is the location of the fetal head center. This distribution corresponds to $P(\theta|\varphi)$ in Eq. (3).

4 Nuchal Translucency Measurement

The NT measurement is computed from the segmentation of the NT region. The pose of the detected NT region is passed into the segmentation algorithm and the upper edge and the lower edge are initially segmented based on the directional gradient of the edge. The Dijkstras algorithm computes each path separately using two inversed gradient magnitude images which are generated by applying oriented gradient filters (rotated 180 degrees w.r.t. each other). Segmentation seeds are placed between the initial paths for foreground and outside of them for background. The seeds are used in a Graph Cut segmentation algorithm to find an accurate segmentation of the NT region. The NT measurement value is computed from as the maximum thickness of the NT region segmentation.

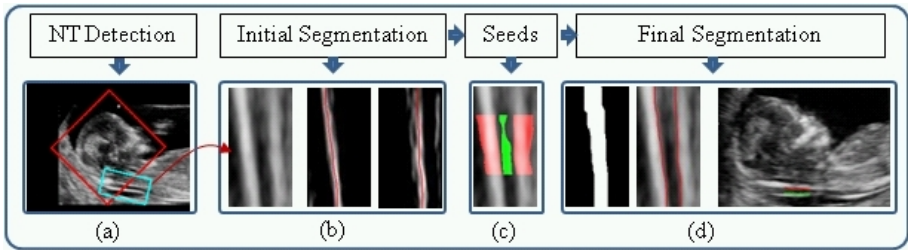


Fig. 3. Algorithm flow of NT segmentation using the NT region detection result. (a) The pose of the detected NT region is used for defining the segmentation region. (b) The initial NT region edges are obtained from Dijkstra’s shortest path on the inverse gradient magnitude as the cost. (c) The NT edges are used to define foreground (green) and background (red) seeds. (d) The final segmentation is obtained by running Graph Cut segmentation.

5 Experiments

Our dataset consists of 196 DICOM scans from the first trimester screening exams. Experts annotated the fetal head and the NT region as shown in Fig. 1. The fetal head is annotated using 4 landmarks shown as green points in Fig. 1 and they are used to define the pose of the fetal head as follows. The two points on head crown and on throat define the orientation and the height of the bounding box, and the other two points on nasal tip and external occipital region define the width of the bounding box. The two yellow lines represent the NT edges and the red line across the two edges is the one for actual measurement of NT thickness. If there is insufficient and ambiguous information in an image for annotation, experts were encouraged to choose the best guess using the context of the image.

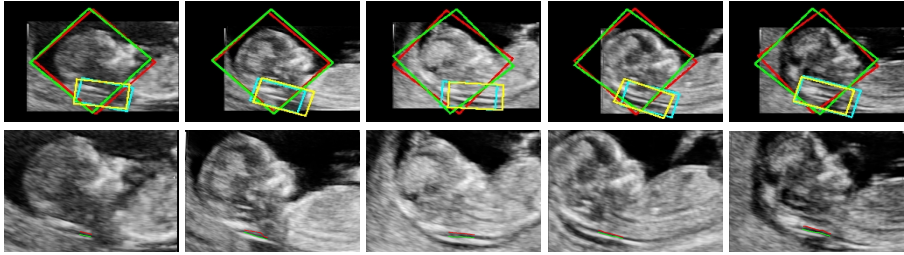
We normalized the orientation of fetal head such that the fetus always faces upwards and the head is towards the left side of the image. Standardizing the head pose helps to train more robust NT detector. During detection, the the head orientation is determined by using both the original input image and its left-to-right flipped image. The head orientation is decided based on the score of the two detection results. To achieve this, the detector is trained by generating images where the head is oriented towards the right side of the image and adding them to negative sample pool. Based on our experimental results, the algorithm can discriminate the head orientation in 100% of the cases. The images were resampled to 0.1 mm resolution. For performance evaluation, the collected data set was divided into two sets: randomly selected 80% for training and the remaining 20% for testing.

Our first experiment focuses on the robustness and accuracy of the fetal head and NT region detection. We use the following error metrics: angle difference (degree), center-to-center distance, and size difference, computed between the results from expert’s ground truth and the algorithm. Table 1 shows the error between the ground truth and the algorithm detection of all the test data. The error of the NT location is slightly higher due to the inherent ambiguity in the accurate localization along the NT region. Overall, the detected NT region can be reliably used for the segmentation.

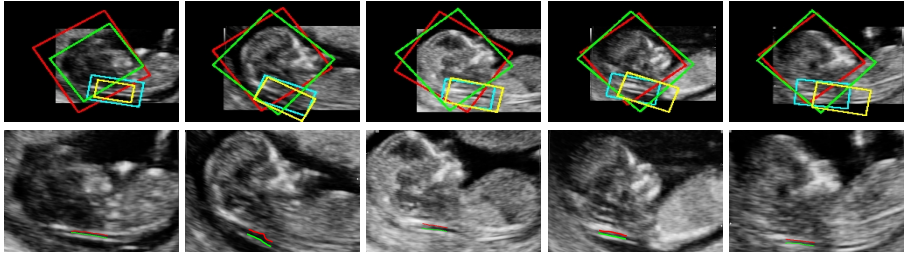
Table 1. Performance evaluation of the testing data. The angle difference is in degrees. The center and size errors are in millimeters.

Test Data (39)								
	Fetal Head				NT			
	<i>mean</i>	<i>std</i>	<i>med</i>	<i>max</i>	<i>mean</i>	<i>std</i>	<i>med</i>	<i>max</i>
Angle [deg]	6.13	4.24	5.94	17.30	3.45	2.62	2.91	9.22
Center [mm]	1.16	0.56	1.10	2.31	2.63	1.93	2.59	8.81
Size [mm]	2.63	1.56	2.36	6.72	1.28	0.94	1.21	3.33

Qualitative results and associated final NT measurements are in Fig. 4. Due to the space limitation, we only show 5 most accurate and 5 least accurate detection results along with the final NT region segmentations from 39 test results. The green and yellow boxes represent the ground truth poses of the fetal head and NT region, respectively. The red and cyan boxes correspond to the detected results. The following observations can be made from these results. As expected, the fetal head detection is very reliable because of its consistent appearance. Furthermore, the NT detection results are accurate for most cases, even for the 5 least accurate results. For most cases, the detected NT pose has lower accuracy in the horizontal direction along the NT edge. This is because of the ambiguity of the NT localization in this direction. However, the slight inaccuracy in the horizontal direction is less important for the segmentation and final NT measurement computation. The segmentation algorithm yields slightly better results of actual NT measurements for the 5 worst NT detection cases whose average error is 0.24 mm than for the 5 best NT detection results whose



(a) Best detection examples of test data and their NT segmentation results.



(b) Worst detection examples of test data and their NT segmentation results.

Fig. 4. Fetal head and NT detection results. The green and yellow boxes represent the ground truth poses of the fetal head and the NT region, respectively. The red and cyan box represent the fetal head and NT region detection results. The red and green lines highlight the upper and lower edge of the NT.

average error is 0.29 mm. It is evident that even when the NT detection box is less accurate, the NT measurement error is low and reliable.

6 Conclusion

This paper proposes a fully automatic approach for NT measurement. The algorithm starts by accurate detection of the fetal head pose. The fetal head serves as an anchoring structure and predicts the approximate pose of the NT region. The pose of the NT region is then locally refined and used to define a segmentation region. The initial edges of the NT region found by Dijkstra's shortest path are used to define seeds for accurate Graph Cut segmentation. Experimental results show that the proposed algorithm is robust and accurate. A slight inaccuracy in the detection of the NT location is caused by the ambiguity of the NT region localization along the edge of the neck but does not negatively impact the final measurement value.

The proposed algorithm can be extended to 3D ultrasound scans. To achieve this, mid-sagittal plane must be first found reliably. This can be done by relying on salient features inside the plane, but also by modeling fetal face profile after accurately detecting the fetal face [16]. This is an exciting direction of our future research and an important step towards automating routine measurements in obstetrics sonography.

References

1. Souka, A.P., Krampfl, E., Bakalis, S., Heath, V., Nicolaidis, K.H.: Outcome of pregnancy in chromosomally normal fetuses with increased nuchal translucency in the first trimester. *Ultrasound in Obstetrics and Gynecology* 18(1), 9–17 (2001)
2. Moratalla, J., Pintoffl, K., Minekawa, R., Lachmann, R., Wright, D., Nicolaidis, K.H.: Semi-automated system for measurement of nuchal translucency thickness. *Ultrasound in Obstetrics and Gynecology* 36(4), 412–416 (2010)
3. Nirmala, S., Palanisamy, V.: Measurement of nuchal translucency thickness for detection of chromosomal abnormalities using first trimester ultrasound fetal images. *Int'l J. of Computer Science and Information Security* 6(3), 101–106 (2009)
4. Won, H.S., Hyun, M.K., Lee, H.: The clinical usefulness of volume NT using three-dimensional (3D) ultrasound. Samsung-Medison Article No. WP201012-VNT (2010)
5. Mortensen, E., Barrett, W.: Intelligent scissors for image composition. In: *ACM Computer Graphics (SIGGRAPH)*, pp. 191–198 (1995)
6. Boykov, Y., Zabih, O.V., Fast, R.: approximate energy minimization via graph cuts. *IEEE Trans. PAMI* 23(11), 1222–1239 (2001)
7. Lu, W., Tan, J., Floyd, R.: Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform. *Ultrasound in Medicine & Biology* 31(7), 929–936 (2005)
8. Chalana, V., Kim, Y.: A methodology for evaluation of boundary detection algorithms on medical images. *IEEE Trans. Med. Imag.* 16(5), 642–652 (1997)
9. Carneiro, G., Georgescu, B., Good, S., Comaniciu, D.: Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Trans. Med. Imag.* 27(9), 1342–1355 (2008)
10. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition and clustering. In: *Proc. of ICCV*, pp. 1589–1596 (2005)
11. Deng, Y., Wang, Y., Chen, P., Yu, J.: A hierarchical model for automatic nuchal translucency detection. *Computers in Biology and Medicine* 42(6) (2012)
12. Zheng, Y., Barbu, A., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Trans. Med. Imag.* 27(11), 1668–1681 (2008)
13. Zhang, J., Zhou, S., Comaniciu, D.: Joint real-time object detection and pose estimation using probabilistic boosting network. In: *Proc. of CVPR* (2007)
14. Sofka, M., Ralovich, K., Birkbeck, N., Zhang, J., Zhou, S.: Integrated detection network (idn) for pose and boundary estimation in medical images. In: *ISBI* (2011)
15. Sofka, M., Zhang, J., Zhou, S.K., Comaniciu, D.: Multiple object detection by sequential Monte Carlo and Hierarchical Detection Network. In: *Proc. of CVPR* (2010)
16. Feng, S., Zhou, S., Good, S., Comaniciu, D.: Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information. In: *Proc. of CVPR*, pp. 2488–2495 (2009)

Automated Segmentation of CBCT Image Using Spiral CT Atlases and Convex Optimization

Li Wang¹, Ken Chung Chen², Feng Shi¹, Shu Liao¹, Gang Li¹, Yaozong Gao¹,
Steve GF Shen³, Jin Yan³, Philip K.M. Lee⁴, Ben Chow⁴,
Nancy X. Liu⁵, James J. Xia², and Dinggang Shen¹

¹ Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC, USA

² The Methodist Hospital Research Institute, Houston, Texas, United States

³ Shanghai Jiao Tong University Ninth Hospital, Shanghai, China

⁴ Hong Kong Dental Implant & Maxillofacial Centre, Hong Kong, China

⁵ Peking University School and Hospital of Stomatology, Beijing, China

Abstract. Cone-beam computed tomography (CBCT) is an increasingly utilized imaging modality for the diagnosis and treatment planning of the patients with craniomaxillofacial (CMF) deformities. CBCT scans have relatively low cost and low radiation dose in comparison to conventional spiral CT scans. However, a major limitation of CBCT scans is the widespread image artifacts such as noise, beam hardening and inhomogeneity, causing great difficulties for accurate segmentation of bony structures from soft tissues, as well as separating mandible from maxilla. In this paper, we presented a novel fully automated method for CBCT image segmentation. In this method, we first estimated a patient-specific atlas using a sparse label fusion strategy from predefined spiral CT atlases. This patient-specific atlas was then integrated into a convex segmentation framework based on *maximum a posteriori probability* for accurate segmentation. Finally, the performance of our method was validated via comparisons with manual ground-truth segmentations.

1 Introduction

Segmentation of the cone-beam computed topographic (CBCT) image is an essential step of generating 3D models in diagnosis and treatment planning of patients with craniomaxillofacial (CMF) deformities. It requires segmenting bony structures from soft tissues, as well as separating mandible from maxilla. Unlike expensive conventional spiral CT scanners, CBCT scanners, usually cost around \$200K, are getting popularly used clinically, even in private practice settings. However, CBCT image quality is significantly inferior to the spiral CT. Besides extremely low signal-to-noise ratio, CBCT scans have severe image artifacts, including noise, beam hardening, inhomogeneity, and truncation, thus affects image contrast and the accuracy of subsequent segmentation [1]. Furthermore, in order to better quantify the deformity, CBCT scans are usually acquired when the maxillary (upper) and mandibular (lower) teeth are in maximal intercuspation, which brings even more challenges to separate the

mandible from the maxilla [2]. To date, in order to use CBCT clinically for diagnosis and treatment planning, the segmentation must be completed manually by experts.

Manual segmentation is a tedious, time-consuming and user-dependent. Previous automated segmentation methods are mainly based on threshold and morphological operations [3], which are sensitive to the presence of the artifacts. Recently, shape information has been utilized for robust segmentation [4-6], e.g., Zhang et al. [6] proposed a deformable segmentation via sparse shape representation. However, these approaches are only applicable to objects with relatively regular shapes (e.g., mandible), but not the objects with complex shapes (e.g., maxilla). Interactive segmentation approaches [2, 7] were also provided to take advantage of both manual and automatic segmentation. To our best knowledge, there is no existing method that is able to automatically and simultaneously segment both maxilla and mandible from CBCT.

In this paper, we propose a fully automated CBCT segmentation method to 1) segment bony structures from the soft tissues, and 2) further separate the mandible from the maxilla. Specifically, we first employ a sparse label propagation strategy to estimate a patient-specific atlas from the spiral CT atlases. There are two reasons that why the spiral CT subjects are employed as the atlases: 1) although image formation process between spiral CT and CBCT is different, they share the same patterns of anatomical structures, which are captured in a patch fashion in our method to estimate the probability; 2) the images from CT scanners have better signal contrast and less noise than those from CBCT scanners, therefore less time is needed to construct atlases of manual segmentations from spiral CT images (i.e., ~30 minutes per image) than from CBCT images (i.e., ~12 hours per image) by an experienced operator. Finally, the patient-specific atlas is then integrated into a convex segmentation framework based on *maximum a posteriori probability* (MAP) for accurate segmentation.

2 Method

In this study, we aim to segment a CBCT scan into three structures/regions: mandible, maxilla (the skull without the mandible), and background. We consider image segmentation as the task of partitioning the image domain Ω into a set of three disjoint regions, such that $\Omega = \bigcup_{i=1}^3 \Omega_i$.

2.1 Subjects

We used 30 (15 males/15 females) spiral CT images with manual segmentations as atlases, and 13 (4 males/9 females) CBCT images as testing images. The 30 spiral CT images of the subjects with normal facial appearance scanned at maximal intercuspation were randomly selected from our HIPAA de-identified CT database. Their ages were 22 ± 2.6 years (range: 18-27 years). The CT matrix was 512×512 (resolution: $0.488 \times 0.488 \times 1.25$ mm³). 13 CBCT images were collected from the patients with non-syndromic dentofacial deformities and scheduled for double-jaw orthognathic surgery for their treatment. Their age were 24 ± 10 years (range: 10-49 years). The CBCT matrix was 400×400 (resolution $0.4 \times 0.4 \times 0.4$ mm³). These 30 CTs

and 13 CBCTs were labeled by two CMF surgeons who are experienced in segmentation using Mimics 10.01 software (Materialise NV, Leuven, Belgium).

2.2 Estimating Patient-Specific Atlas from Spiral CT Atlases

Atlas-based segmentation has demonstrated its robustness and effectiveness in many medical image segmentation problems [8]. Conventionally, a population-based atlas is constructed from multiple training images (e.g., manually segmented CBCTs). However, a population-based atlas often fails to provide useful guidance especially in the regions with high inter-subject anatomical variability, and leads to unsatisfactory segmentation results. One way to overcome this problem is to integrate the patient-specific information in the atlas construction. To this end, we propose to construct a patient-specific atlas by combining both population and patient information as follows.

We propose to estimate the prior, i.e., the patient-specific atlas, using a patch-based representation technique [9, 10]. The rationale is that an image patch generally provides richer information, e.g., anatomical pattern, than a single voxel. Specifically, $N = 30$ spiral CT images I_j and their corresponding segmentation maps S_j ($j = 1, \dots, N$) are nonlinearly aligned onto the space of the testing image I using ELASIX [11]. Then, for each voxel x in the testing image I , its corresponding intensity patch with size $w \times w \times w$ can be represented as a column vector $\mathbf{Q}(x)$. An initial codebook $\mathbf{B}(x)$ can be constructed with respect to all these patches $\{\mathbf{Q}_j(x) | j = 1, \dots, N\}$ from all aligned templates at the same location, i.e., $\mathbf{B}(x) = [\mathbf{Q}_1(x), \mathbf{Q}_2(x), \dots, \mathbf{Q}_N(x)]$. To alleviate the effect of registration error, the initial codebook can be extended to include more patches from the neighboring search window $\mathbf{W}(x)$, i.e., $\mathbf{B}(x) = \{\mathbf{Q}_j(y), \forall j = 1 \dots, N, y \in \mathbf{W}(x)\}$, where each patch is represented by a column vector and normalized to have the unit ℓ_2 norm. To represent the patch $\mathbf{Q}(x)$ by the codebook $\mathbf{B}(x)$, its coding vector \mathbf{c} could be estimated by many coding schemes, such as vector quantization, locality-constrained linear coding [12], and sparse coding [13]. Here, we utilize the sparse coding scheme [13] to estimate the coding vector \mathbf{c} by minimizing a non-negative Elastic-Net problem [14],

$$\min_{\mathbf{c} \geq 0} \frac{1}{2} \|\mathbf{Q}(x) - \mathbf{B}(x)\mathbf{c}\|_2^2 + \lambda_1 \|\mathbf{c}\|_1 + \frac{\lambda_2}{2} \|\mathbf{c}\|_2^2 \quad (1)$$

where the first term is the least square fitting term, the second term is the ℓ_1 regularization term which is used to enforce the sparsity constraint on the reconstruction vector \mathbf{c} , and the last term is the ℓ_2 smoothness term to enforce the similarity of coding coefficients for the similar patches. Each element of the coding vector \mathbf{c} , i.e., $c_j(y)$, reflects the similarity between the target patch $\mathbf{Q}(x)$ and the patch $\mathbf{Q}_j(y)$ in the codebook. Based on the assumption that the similar patches should share similar labels, we use the sparse coding \mathbf{c} to estimate the prior probability of the voxel x belonging to the i -th structure/region, i.e., $p_i(x) = \sum_j \sum_{y \in \mathbf{W}(x)} c_j(y) \delta_i(S_j(y))$, where $\delta_i(S_j(y)) = 1$ if the label $S_j(y) = i$; otherwise, $\delta_i(S_j(y)) = 0$. Finally, $p_i(x)$ is normalized to ensure $\sum_i p_i(x) = 1$.

2.3 Convex Segmentation Based on MAP

After obtaining the prior probability from the patient-specific atlas, we integrate it with the intensity distribution of the testing image itself for better segmentation. In the testing image, to accurately label each voxel x in the image domain Ω , we jointly consider its neighboring voxels $y \in \mathcal{O}(x)$, where $\mathcal{O}(x)$ is the neighborhood of voxel x . The regions $\{\Omega_i\}$ produce a partition of the neighborhood $\mathcal{O}(x)$, i.e., $\{\Omega_i \cap \mathcal{O}(x)\}_{i=1}^3$. We first consider the segmentation of $\mathcal{O}(x)$ based on *maximum a posteriori probability* (MAP). According to the Bayes rule:

$$p(y \in \Omega_i \cap \mathcal{O}(x) | I(y)) = \frac{p(I(y) | y \in \Omega_i \cap \mathcal{O}(x)) p(y \in \Omega_i) p(y \in \mathcal{O}(x))}{p(I(y))} \quad (2)$$

where $p(I(y) | y \in \Omega_i \cap \mathcal{O}(x))$, denoted by $p_{i,x}(I(y))$, is the structure probability density in region $\Omega_i \cap \mathcal{O}(x)$. Note that $p(y \in \Omega_i)$, i.e., $p_i(y)$, is the a priori probability of y belonging to the region Ω_i , which has been estimated in section 2.2. $p(y \in \mathcal{O}(x)) = \mathbf{1}_{\mathcal{O}(x)}(y)$ is the indicator function and $p(I(y))$ is independent of the choice of the region and can therefore be neglected. Thus, Eq. (2) can be simplified as $p(y \in \Omega_i \cap \mathcal{O}(x) | I(y)) = p_{i,x}(I(y)) p_i(y)$. Assuming that the voxels within each region are independent, the MAP will be achieved only if the product of $p_{i,x}(I(y)) p_i(y)$ across the region $\mathcal{O}(x)$ is maximized: $\prod_{i=1}^3 \prod_{\Omega_i \cap \mathcal{O}(x)} p_{i,x}(I(y)) p_i(y)$. In fact, we can use a Gaussian kernel K_ρ with scale ρ to indicate the neighborhood $\mathcal{O}(x)$ [15]. Taking a logarithm transformation and integrating all the voxels x , the maximization can be converted to the minimization of the following energy,

$$\mathcal{E}^D = - \sum_{i=1}^3 \int_{\Omega} \int_{\Omega_i} K_\rho(x-y) \log p_{i,x}(I(y)) p_i(y) dy dx \quad (3)$$

Based on [16], we can use multiple variables which take values between 0 and 1 to derive a convex formulation for Eq. (3). Since in our project, there are only 3 different regions of interest: mandible, maxilla, and background, we need only 2 segmentation variables $u_1 \in [0, 1]$ and $u_2 \in [0, 1]$ to represent the partitions $\{\Omega_i\}$: $M_1 = u_1$, $M_2 = u_2$, $M_3 = (1 - u_1)(1 - u_2)$. Therefore, Eq. (3) can be converted as follows,

$$\mathcal{E}^D = - \sum_{i=1}^3 \int_{\Omega} \int_{\Omega} K_\rho(x-y) \log p_{i,x}(I(y)) p_i(y) M_i(y) dy dx \quad (4)$$

There are many options to estimate the $p_{i,x}(I(y))$. In this paper, we utilize a Gaussian distribution model with the local mean $\mu_i(x)$ and the variance $\sigma_i^2(x)$ [17] to estimate it: $p_{i,x}(I(y)) = \exp\left(-(\mu_i(x) - I(y))^2 / 2\sigma_i^2(x)\right) / (\sqrt{2\pi}\sigma_i(x))$.

Based on the assumption that there should be no overlap between mandible and maxilla, we propose the following penalty constraint term,

$$\mathcal{E}^P = \int u_1(x) u_2(x) dx \quad (5)$$

In addition, the length regularization term [18] is defined as weighted total variation of functions u_1 and u_2 ,

$$\mathcal{E}^{\mathcal{R}} = \int g(I(x))(|\nabla u_1(x)| + |\nabla u_2(x)|) dx \quad (6)$$

where g is a non-edge indicator function that vanishes at object boundaries [18].

Finally, we define the entire energy functional below, which consists of the data fitting term $\mathcal{E}^{\mathcal{D}}$, the overlap penalty term $\mathcal{E}^{\mathcal{P}}$ and the length regularization term $\mathcal{E}^{\mathcal{R}}$:

$$\min_{u_1, u_2 \in [0, 1]} \{\mathcal{E}(u_1, u_2) = \mathcal{E}^{\mathcal{D}} + \alpha \mathcal{E}^{\mathcal{P}} + \beta \mathcal{E}^{\mathcal{R}}\} \quad (7)$$

where α and β are the positive coefficients. Based on [16], the energy functional (Eq. (7)) can be easily minimized in a fast way with respect to u_1 and u_2 .

3 Experimental Results

The parameters used in this paper were determined experimentally via cross validation. In fact, the results are relatively insensitive to the variation of the parameters. We finally chose the following parameters for all the experiments: the weight for ℓ_1 -norm term $\lambda_1 = 0.1$, the weight for ℓ_2 -norm term $\lambda_2 = 0.05$, the patch size $9 \times 9 \times 9$, the search window size $5 \times 5 \times 5$, $\rho = 3$ for the Gaussian Kernel K_ρ , and the weights $\alpha = 10$ for the overlap penalty term $\mathcal{E}^{\mathcal{P}}$ and $\beta = 10$ for the length regularization term $\mathcal{E}^{\mathcal{R}}$.

Fig. 1 demonstrates the segmentation results of different methods for one typical subject. In the first row, the volume rendering of the original intensity image, and surfaces rendering of the segmentation result obtained by majority voting (MV), patch-based fusion method [10], the proposed method by directly using the maximum class probability from step 1 (prior estimation) as the segmentation result, the proposed method with both step 1 (prior estimation) and step 2 (convex segmentation), and the manual segmentation, are shown from left to right. The following rows show the corresponding results on slices and zoomed views for better visualization, from which the artifacts can be clearly observed. Due to possible errors during image registration, the surface by MV is far from accurate, which incorrectly labels some upper teeth as lower teeth, by referring to the manual segmentation. Due to the closed-bite position and large intensity variations, the patch-based fusion method [10] cannot accurately separate the mandible from maxilla, thus mislabeling the upper teeth and lower teeth. Note that, to be a fair comparison, we have performed the similar searching scheme to derive the optimal parameters for the patch-based fusion method [10], by obtaining the final patch size of $9 \times 9 \times 9$ and the final search window size of $5 \times 5 \times 5$. Overall, the proposed method produces much more reasonable results.

We then quantitatively evaluate the performance of different methods on 13 subjects by employing Dice ratio. The average Dice ratios of different methods are shown in Fig. 2. As also shown in Table 1, with the convex segmentation integrating the estimated prior, the proposed method achieves the highest Dice ratios. To further validate the proposed method, we also evaluate its accuracy by measuring the average

surface distance error, as plotted in Fig. 3. Additionally, the Hausdorff distance is also used to measure the maximal surface-distance errors of each of 13 subjects. The average Hausdorff distance on all 13 subjects are shown in Table 1, which again demonstrates the advantage of our proposed method.

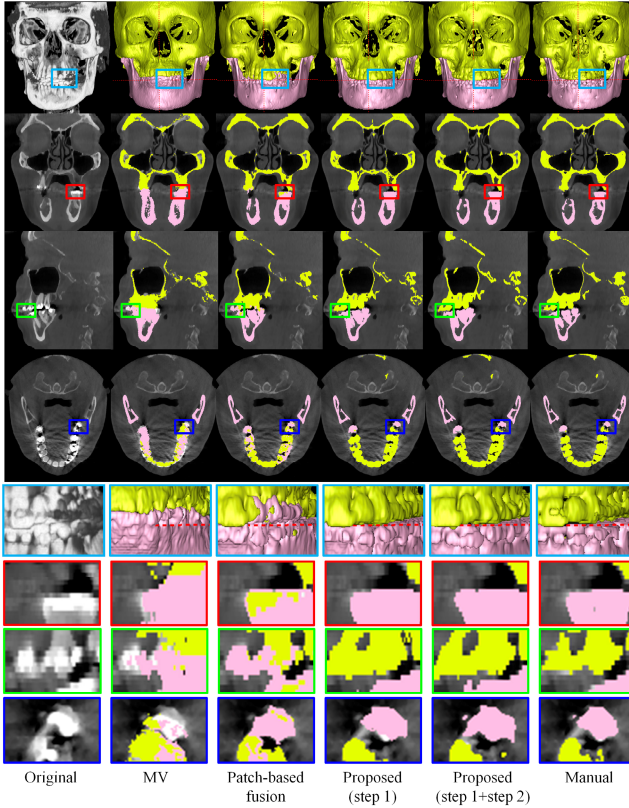


Fig. 1. Comparisons of segmentation results of 4 different methods on a typical CBCT image

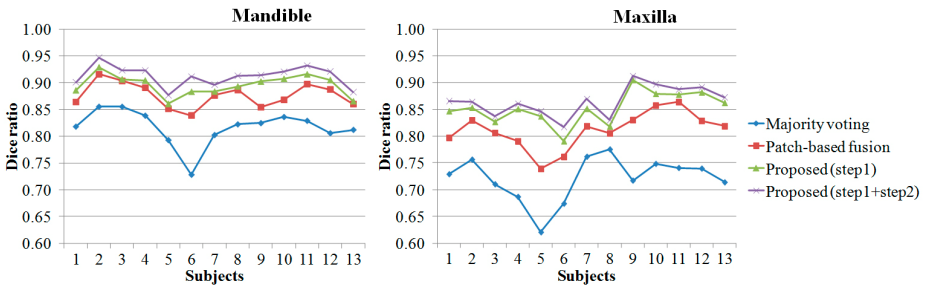


Fig. 2. Dice ratios on mandible and maxilla by 4 different methods

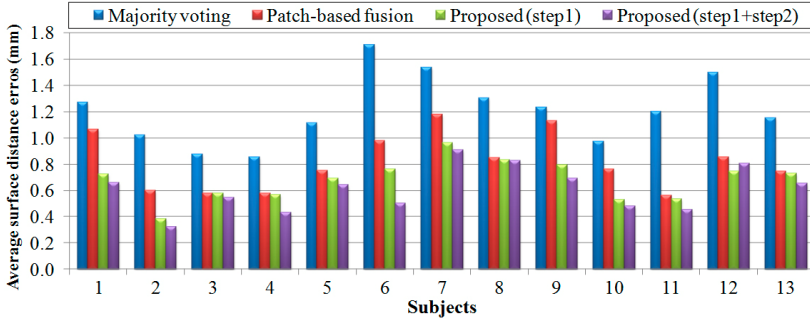


Fig. 3. Average surface distances from the surfaces obtained by 4 different methods to the ground-truth surfaces on 13 subjects

Table 1. Average Dice ratio and surface distance error (in mm) on 13 subjects

		Majority voting	Patch-based fusion [10]	Proposed (step 1)	Proposed (step 1+2)
Dice ratio	Mandible	0.82±0.03	0.88±0.02	0.89±0.02	0.91±0.02
	Maxilla	0.72±0.04	0.81±0.03	0.85±0.03	0.87±0.02
Average distance error	Mandible	1.21±0.25	0.81±0.21	0.67±0.15	0.61±0.17
Hausdorff distance error	Mandible	3.61±1.53	2.22±1.1	1.14±0.54	0.92±0.47

4 Discussion and Conclusion

We have successfully developed and validated a novel fully automated method for CBCT segmentation. We first estimated a patient-specific atlas from spiral CT atlases using a sparse label fusion strategy. Then, we integrate it into a convex segmentation framework based on MAP for segmentation. Comparing to the state-of-the-art label-fusion methods, our method achieved the best results.

The success of applying normal CT atlases to the CBCT subject with CMF deformity can be mainly attributed to the following two factors: 1) The deformation between the subject with the CMF deformity and the normal subject is first alleviated by nonlinear image registration; 2) After registration, the structure probabilities for the CBCT subject with CMF deformity are then robustly estimated by the proposed patch-based sparse technique in Section 2.2.

In our experiment, we found that increasing the number of atlases would generally improve the segmentation accuracy. For example, the combination of CT and CBCT atlases achieves slightly improved Dice ratios: 0.005 higher for mandible and 0.0046 higher for maxilla. However, more atlases also bring in larger computational cost.

In our future work, we will validate the proposed method on more dataset and further compare with other methods such as [19]. We will also improve the robustness of the proposed method by increasing the variability of the atlases such as including more datasets with different CMF deformities.

References

1. Loubele, M., Maes, F., Schutyser, F., Marchal, G., et al.: Assessment of bone segmentation quality of cone-beam CT versus multislice spiral CT: a pilot study. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* 102, 225–234 (2006)
2. Le, B.H., Deng, Z., Xia, J., Chang, Y.-B., Zhou, X.: An Interactive Geometric Technique for Upper and Lower Teeth Segmentation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 968–975. Springer, Heidelberg (2009)
3. Hassan, B.A.: *Applications of Cone Beam Computed Tomography in Orthodontics and Endodontics* (2010)
4. Kainmueller, D., Lamecker, H., Seim, H., Zinser, M., Zachow, S.: Automatic Extraction of Man-dibular Nerve and Bone from Cone-Beam CT Data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 76–83. Springer, Heidelberg (2009)
5. Gollmer, S.T., Buzug, T.M.: Fully automatic shape constrained mandible segmentation from cone-beam CT data. In: *ISBI*, pp. 1272–1275 (2012)
6. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Deformable segmentation via sparse shape representation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part II. LNCS*, vol. 6892, pp. 451–458. Springer, Heidelberg (2011)
7. Suebnukarn, S., Haddawy, P., Dailey, M., Cao, D.: Interactive Segmentation and Three-Dimension Reconstruction for Cone-Beam Computed-Tomography Images. *NECTEC Technical Journal* 8, 154–161 (2008)
8. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., et al.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46, 726–738 (2009)
9. Rousseau, F., Habas, P.A., Studholme, C.: A Supervised Patch-Based Approach for Human Brain Labeling. *TMI* 30, 1852–1862 (2011)
10. Coupé, P., Manjón, J., Fonov, V., Pruessner, J., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54, 940–954 (2011)
11. Klein, S., Staring, M., Murphy, K., Viergever, M.A., et al.: elastix: A Toolbox for Intensity-Based Medical Image Registration. *TMI* 29, 196–205 (2010)
12. Wang, J., Yang, J., Yu, K., Lv, F., et al.: Locality-constrained Linear Coding for image classification. In: *CVPR*, pp. 3360–3367 (2010)
13. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., et al.: Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227 (2009)
14. Zou, H., Hastie, T.: Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B* 67, 301–320 (2005)
15. Li, C.M., Kao, C.Y., Gore, J.C., Ding, Z.H.: Minimization of Region-Scalable Fitting Energy for Image Segmentation. *TIP* 17, 1940–1949 (2008)
16. Goldstein, T., Bresson, X., Osher, S.: *Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction*. CAM Report, UCLA (2009)
17. Brox, T., Cremers, D.: On Local Region Models and a Statistical Interpretation of the Piecewise Smooth Mumford-Shah Functional. *IJCV* 84, 184–193 (2009)
18. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *IJCV* 22, 61–79 (1997)
19. Liao, S., Gao, Y., Shen, D.: Sparse patch based prostate segmentation in CT images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS*, vol. 7512, pp. 385–392. Springer, Heidelberg (2012)

Automatic Analysis of Pediatric Renal Ultrasound Using Shape, Anatomical and Image Acquisition Priors

Carlos S. Mendoza^{1,2}, Xin Kang¹, Nabile Safdar¹,
Emmarie Myers¹, Aaron D. Martin^{1,3}, Enrico Grisan⁴,
Craig A. Peters^{1,3}, and Marius George Linguraru¹

¹ Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington DC, USA

carlos.sanchez.mendoza@gmail.com

² Signal Processing Department, University of Sevilla, Spain

³ Division of Urology, Children's National Medical Center, Washington DC, USA

⁴ Department of Information Engineering, University of Padova, Italy

Abstract. In this paper we present a segmentation method for ultrasound (US) images of the pediatric kidney, a difficult and barely studied problem. Our method segments the kidney on 2D sagittal US images and relies on minimal user intervention and a combination of improvements made to the Active Shape Model (ASM) framework. Our contributions include particle swarm initialization and profile training with rotation correction. We also introduce our methodology for segmentation of the kidney's collecting system (CS), based on graph-cuts (GC) with intensity and positional priors. Our intensity model corrects for intensity bias by comparison with other biased versions of the most similar kidneys in the training set. We prove significant improvements ($p < 0.001$) with respect to classic ASM and GC for kidney and CS segmentation, respectively. We use our semi-automatic method to compute the hydronephrosis index (HI) with an average error of 2.67 ± 5.22 percentage points similar to the error of manual HI between different operators of 2.31 ± 4.54 percentage points.

1 Introduction

Kidney segmentation in US images is a topic of limited presence in the image processing literature [1, 2]. Several limitations of US images make segmentation a particularly daunting task: poor signal-to-noise ratio, signal drop-out, missing boundaries, misplaced boundaries and reconstruction errors [3]. US segmentation methods have been previously classified according to the prior knowledge employed to improve the accuracy of results. These constraints include image-derived priors (intensity, intensity derivatives, local phase, texture), and application-derived priors (shape and motion) [3].

Two-dimensional ultrasound (US) is the widely preferred image modality for in vivo assessment of renal conditions, particularly for pediatric applications, mainly for its safety and cost-effectiveness, both prenatally and postnatally.

For the diagnosis of hydronephrosis (dilatation of the renal pelvis and calyces, often caused by obstruction of the free flow of urine from the kidney [4]) in young children, Shapiro et al. [5] have proposed a hydronephrosis index (HI), a quantitative measure in which $HI \text{ (percentage)} = 100 \times (\text{renal area} - \text{renal pelvis/calices area}) / (\text{renal area})$. Computation of HI requires manual delineations of the kidney and the pelvis/calices on a single 2D sagittal US image in which the kidney achieves its maximal longitudinal dimensions. It could be discussed whether using a single slice allows for robust shape model creation and good estimation of HI. We show in Fig. 3 in Section 3 some examples of longitudinal kidney sections and in Fig. 1 the good behaviour of the obtained shape model can be appreciated. Although the estimation of HI would certainly be improved by 3D quantification, it is not our purpose to go beyond the accuracy of the current standard clinical procedure, which is performed on properly chosen 2D slices due to higher availability of 2D acquisition equipment [5].

The purpose of the work presented herein is to increase the availability and reduce the variability of HI computation by automating the segmentation of the renal parenchyma and the renal collecting system (pelvis and calyces). This task presents several challenges. For the segmentation of renal structures there is limited availability of robust priors, especially from noisy 2D US images acquired at variable angles that give variable renal shapes. Furthermore, in US the interior of the kidney exhibits heterogeneous structures of different shapes, sizes and intensities, and many of the boundaries are lost due to the density similarity to surrounding structures and shadowing [1].

To the best of our knowledge, the automated segmentation of the renal collecting system (CS) has not been tackled by previous work. CS appears as a hypoechoic (dark) region inside the kidney, and expands into the parenchyma from the ureteropelvic junction (UPJ). Difficulties for segmentation of CS in sagittal US sections arise due to intensity offset, signal drop-out (shadowing) and loss of the 3D connectivity of the CS region when studied on 2D views. Furthermore, in very young children the renal pyramids appear more hypoechoic than in other subjects, and can be confused with CS even by the experienced eye.

In the following Sections we present our methodology for segmentation of renal structures relevant to hydronephrosis. For segmenting the kidney we introduce a modification of the classic Active Shape Models (ASM) [6], with shape and radial intensity priors based on statistical models that incorporate the relative orientation with respect to the US probe. We also propose an automatic robust initialization strategy, important for ASM under large shape uncertainty, based on particle swarm optimization. For segmentation of CS we introduce a methodology based on graph-cuts (GC) [7] with intensity and positional priors. To correct for intensity bias we developed a correction strategy based on probability density function (PDF) cross-correlations with similar cases identified in the training set. As output of the GC algorithm we obtain a set of candidate regions that are ranked according to anatomical and clinical priors.

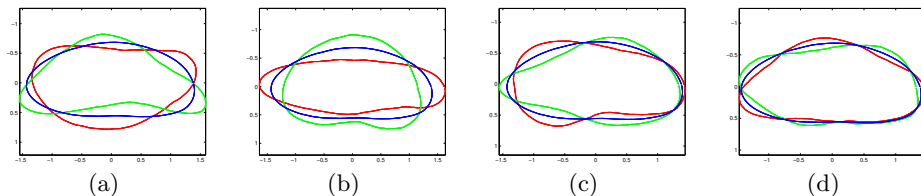


Fig. 1. Three principal modes of shape variations for the kidney shape model. (a)-(d) First to fourth mode. Blue: mean shape. Green: $\text{mean} + 3\sigma$. Red: $\text{mean} - 3\sigma$.

Ultimately, the proposed technique allows the semi-automatic analysis of the kidney and its CS from difficult US image data with low error.

2 Method

2.1 Kidney Segmentation: ASM with Image Acquisition Priors

In the classic ASM methodology [6], shape is represented by a large number of landmarks distributed along the object's contour. ASM performs by adjusting the position of the landmarks to fit pre-trained radial intensity derivatives along rays traced normally to the contour's landmarks. The new set of landmarks is then filtered through the pre-trained shape model. This procedure is re-iterated until convergence or for a given number of iterations. The shape model does not incorporate any information on the pose, particularly relevant in US images, but is instead constructed from pose-corrected instances. Our pose normalization was performed by finding the centroid of the region enclosed by the contour, by correcting orientation using the main axis of inertia, and by scaling the contour according to the average distance from the computed centroid. Then landmarks were obtained by even arc-length sampling of the contour. Furthermore, in ASM the range of search on the radial profiles is usually small. Therefore, a good initial estimation of the pose (location, orientation, scale) of the object is necessary for good convergence. In our method, the user provides two mouse-clicks along the main longitudinal axis of the kidney so that the line connecting the two clicks provides an initial approximation of centroid location (midpoint), orientation and scale.

Particle Swarm Optimization (PSO) [8] is adopted to optimize the initializing shape for the ASM (the pose, and also the shape by allowing variations along the main modes of variation). PSO was chosen because it makes few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions. In PSO the optimization is achieved by having a swarm of particles (candidate solutions) moving around in the search space. Each particle's position and velocity are initialized as uniformly distributed random vectors and each particle's position and velocity are updated according to the best solutions found in previous iterations. In our implementation the pose is obtained from the

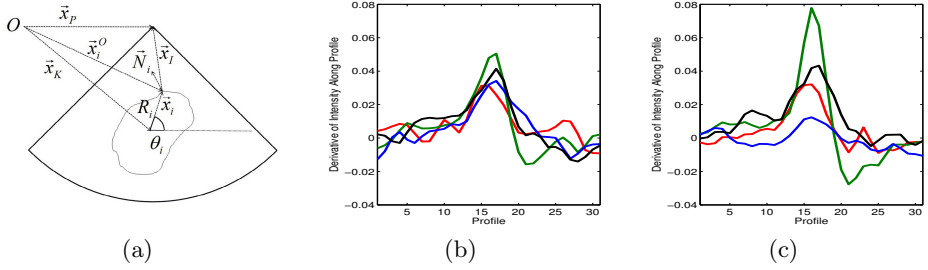


Fig. 2. (a) Definitions for computation of the incidence angle. Edge profiles without (b) and with (c) rotation correction averaged across all kidneys for landmarks spaced around the kidney contours ($\theta_i = (0 \text{ (red)}, \pi/2 \text{ (green)}, \pi \text{ (blue)}, 3\pi/2 \text{ (black)})$).

two mouse-clicks as described above, allowing variations of ± 20 pixels in location, ± 10 degrees in orientation, and ± 10 in scale. The result is a pose/shape-adjusted instance of the shape model that best fits the image features. All parameters in this and following algorithms were tuned heuristically by systematic trial and error, and further parameter space exploration was left for future work.

Both the initialization and ASM procedures rely on the sum of Mahalanobis distances between the radial intensity derivative profiles at landmark locations for the test image and training data. In our adaptation to US, in order to account for the fading of edges that are tangent to the US propagation direction, profile training must be addressed not by anatomical landmark correspondence, but instead in accordance to the angle of incidence at landmark i , located at x_i . According to the conventions in Fig. 2(a), referred to image origin O , if the position of the probe (x_p) and the kidney centroid (x_k) are considered approximately constant for all images, the incidence direction only depends on x_i . Then if we approximate the contour of the kidney by a circumference of radius R , both the normal $N_i \approx \frac{x_i}{|x_i|}$ and $x_i \approx (R \cos \theta_i, R \sin \theta_i)$ (and therefore x_l) are only a function of the angular coordinate $\alpha_i \approx f(\theta_i) \approx f(i - i_0)$, where landmark i_0 has angular coordinate θ_0 . Then, the right incidence-angle correspondence can be obtained if the set of landmarks is circulated around the contour according to the orientation θ_0 of every shape instance. As seen in Fig. 1(b-c), the kidney radial profiles become more distinct when rotation correction is applied, since the averaging over landmarks with different incidence angles is avoided.

2.2 CS Segmentation: Intensity Correction and Anatomical Priors

In renal US the structures of the CS appear as hypoechoic structures proximal to the UPJ (Fig. 3). To incorporate priors of intensity and position an appropriate framework is GC segmentation [7]. GC techniques operate on a graph analogue of an image and exhibit good performance and guaranteed convergence of energies of the following form: $E(f) = \sum_{p \in P} (D_p(f_p)) + \lambda \sum_{(p,q) \in N} (V_{p,q}(f_p, f_q))$, where $D_p(\cdot)$ is the pixel-wise regional term and $V_{p,q}(\cdot, \cdot)$ is the boundary term for

measuring the interaction potential over neighborhood N , for a labeling scheme f . Often, the regional term depends on the Bayesian likelihood of a certain pixel feature according to a pre-trained model. We propose nonparametric estimation of the likelihoods for intensity and also position via kernel density estimation (KDE) from training images.

Intensity Model. The main characteristic of the CS in US images is its hypoechoic appearance. However, US images can present significant intensity bias due to user-controlled depth gain settings and beam attenuation within the body. Other authors proposed intensity correction for US images by estimating the bias field according to complex models [9]. We make a less restrictive assumption: that the intensity distribution of the kidney suffers from a constant bias that can be determined by comparison with biased versions training kidneys. The advantage is that no assumptions are made on specific bias field models.

For each of the test images we estimate the PDF of the pixels inside the kidney according to the automatic segmentations performed as in Section 2.1. The intensity distributions of the training images for the pixels inside the kidney, parenchyma and CS are obtained according to manual segmentations from our ground truth. If f_i is the PDF of kidney intensities in training image i we define the bias b_i as

$$b_i = \underset{|n| \leq Q}{\operatorname{argmax}}(f_i \star g)[n] , \quad (1)$$

where g is the kidney PDF in the test image, and \star denotes cross-correlation. Notice that the cross-correlation is maximized over bounded values of n . If $\{C_j^g\}_{j=1\dots K}$ are the K (we used $K = 10$) training images with larger $(f_i \star g)[b_i]$ among the T training images $\{C_i\}_{i=1\dots T}$, then our intensity models $\hat{f}_g^{P/CS}$ for segmenting image g into parenchyma or collecting system (P/CS) are obtained as

$$\hat{f}_g^{P/CS} = \operatorname{KDE} \left(\left\{ \left\{ I_{k,C_j^g}^{P/CS} + b_{C_j^g} \right\}_{k=1\dots M_{C_j^g}^{P/CS}} \right\}_{j=1\dots K} \right) , \quad (2)$$

where $I_{k,C_j^g}^{P/CS}$ is the k -th intensity sample over P/CS for training image C_j^g , from the set of all $M_{C_j^g}^{P/CS}$ samples. Intuitively, we find a subset of the most similar training kidneys and for which bias this highest similarity occurs for each one of them. Then the intensity model is obtained from kernel density estimation over the bias-corrected training kidneys in the obtained subset.

Position Model. Another available prior for CS segmentation is its anatomical location with respect to the UPJ. The location of UPJ for each kidney is obtained by exploiting landmark correspondence in the ASM. To build the positional model from kidneys of different sizes we normalize the distances from each pixel to the UPJ by the distance from the kidney centroid to UPJ, in each of the T training segmentations (ground truth). The normalized positional models

$\hat{p}^{P/CS}$ for P/CS segmentation are estimated from the set of all normalized pixel locations $J_{k,i}^{P/CS}$ of each training image i as

$$\hat{p}^{P/CS} = \text{KDE}\left(\left\{\left\{J_{k,i}^{P/CS}\right\}_{k=1\dots M_i^{P/CS}}\right\}_{i=1\dots T}\right). \quad (3)$$

Graph Energy. The trained models for intensity and position can be incorporated into the region term of the GC energy:

$$D_p(P/CS) = \hat{f}_g^{P/CS}(I_p) + \alpha \cdot \hat{p}^{P/CS}(J_p). \quad (4)$$

The boundary term depends on the difference of likelihood for intensities across the n-link to being CS:

$$V_{p,q}(f_p, f_q) = \begin{cases} \beta \cdot e^{\gamma |\hat{f}_g^{CS}(I_p) - \hat{f}_g^{CS}(I_q)|} & \text{if } f_p \neq f_q, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

α , β and γ are tuning constants (we used $\alpha = 0.5$, $\beta = 5$ and $\gamma = 100$) that weight the contributions of the different terms to the GC energy.

Candidate Selection. The output of the GC segmentation can consist of several regions. However, CS is a single connected structure inside the kidney, so we post-process the results by connected-component analysis. For each of the L output connected components $\{CC_l\}_{l=1\dots L}$ we compute a score $S(CC_l)$ as follows:

$$S(CC_l) = \left(1 - \left|\frac{M^{CC_l}}{M^K}\right|\right) + \delta \left(\frac{\overline{\text{Mean}(\{I_k^{CC_l}\}_{k=1\dots M^{CC_l}})}}{\overline{\text{Mean}(\{I_k^K\}_{k=1\dots M^K})}}\right) + \epsilon \cdot \overline{\text{Mean}(\{J_k^{CC_l}\}_{k=1\dots M^{CC_l}})}, \quad (6)$$

where M^A is the number of pixels in region A , and $K = P \cup CS$ (i.e. the kidney). The bar indicates the normalization of the mean and standard deviation of the terms to the $[0, 1]$ range over all connected components $\{CC_l\}_{l=1\dots L}$. We used $\delta = 0.9$ and $\epsilon = 0.3$. Intuitively, the score S is a measure of likelihood for a region to be selected as final single connected segmentation among the regions obtained from GC. It integrates priors on region size (should be large), region relative hypoechoic quality (should be high) and location relative to UPJ (should be close).

3 Results

We collected 34 US image studies of hydronephrotic kidneys, from one Philips IU22 and three General Electrics Logiq E9 units. Two trained operators (GT1 and GT2) manually delineated the kidneys and their CS, under the supervision

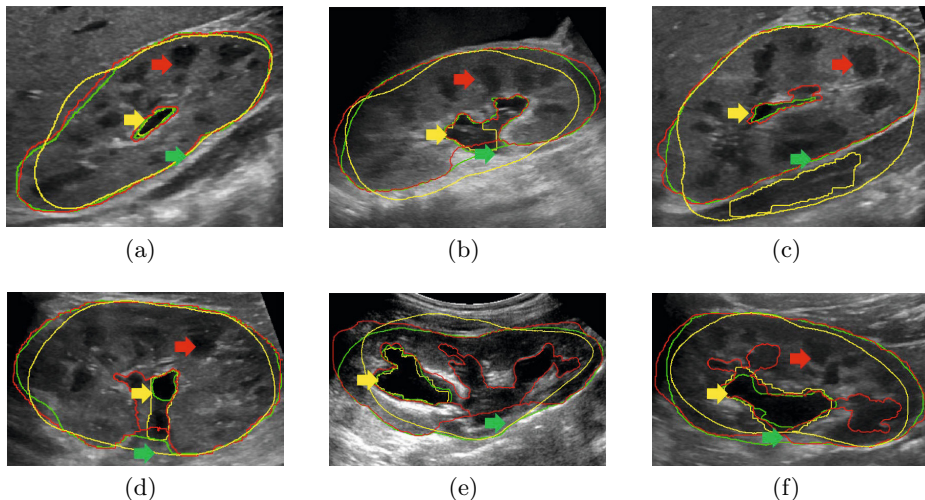


Fig. 3. Results for (a) best, (b) median and (c) worst case. (d-f) Some more results. Green: GT1. Red: GT2. Yellow: Result. Green arrow: UPJ. Red arrow: Renal pyramid. Yellow arrow: CS.

Table 1. DICE, RAD, and MPD (pixels) with respect to GT1. IO K: Kidney delineation by GT2. cKS: Kidney segmentation with classic ASM. KS: Kidney segmentation with proposed method. IO CS: CS delineations by GT2. cCS: CS segmentation with classic GC. CS: CS segmentation with proposed method.

Measure	IO K	cKS	KS	IO CS	cCS	CS
Dice	0.95 ± 0.03	0.83 ± 0.04	0.87 ± 0.05	0.76 ± 0.18	0.32 ± 0.22	0.62 ± 0.32
RAD	0.04 ± 0.04	0.16 ± 0.11	0.13 ± 0.11	0.79 ± 1.48	1.42 ± 1.54	0.85 ± 1.21
MPD	6.47 ± 4.14	21.2 ± 6.36	17.3 ± 7.50	11.7 ± 14.1	28.4 ± 41.4	25.3 ± 35.3

of a radiologist and a urologist, on a single sagittal 2D US image. All images and GT1 delineations were used for training and testing in a leave-one-out configuration. The average time for each test was 45 ± 39 s. (ASM) and 4 ± 2 s. (GC).

We compared results for kidney and CS segmentation from our methods (M) with GT1, and also compared GT2 with GT1. We obtained Dice's coefficients (DICE), relative area differences (RAD), and mean perimeter distances (MPD). Classic ASM (mean shape initialization, no rotation correction) and classic GC (no bias correction, positional prior or candidate selection) implementations were also included in the comparison. Results are shown in Table 1. Significant improvements ($p < 0.001$) were noted using our method over a classic ASM (for kidney segmentation) and classic GC (for CS segmentation). We computed HI for all test cases as in [5] using GT1, GT2 and M. The average error for M vs. GT1 was 2.67 ± 5.22 percentage points, and 2.31 ± 4.54 for GT2 vs. GT1. HIs from M and GT1 are significantly correlated ($r = 0.99$, $p < 0.001$). Fig. 3 shows segmentation results.

4 Conclusion

We have presented a segmentation framework for renal US. Our kidney segmentation method is based on an improvement of ASM via rotation correction and PSO initialization. Our framework also includes a GC segmentation algorithm for structures of the renal collecting system (pelvis and calyces). The GC energy depends on intensity and positional priors estimated from training images via KDE. Intensity bias is corrected by optimizing cross-correlations of the test image PDF with the most similar kidneys in the training set. Results show that our approach performs well in difficult images, and that the HI can be obtained with error similar to the disagreement between manual operators, suggesting the potential for computer-aided diagnosis of (pediatric) renal conditions, related to the size and geometry of the organ and its substructures, from US imaging data.

Acknowledgment. This project was supported by a philanthropic gift from the Government of Abu Dhabi to Children's National Medical Center. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the donor.

References

1. Martín-Fernández, M., Alberola-Lopez, C.: An approach for contour detection of human kidneys from ultrasound images using Markov random fields and active contours. *Med. Image. Anal.* 9(1), 1–23 (2005)
2. Xie, J., Jiang, Y., Tsui, H.: Segmentation of kidney from ultrasound images based on texture and shape priors. *IEEE Trans. Med. Imag.* 24(1), 45–57 (2005)
3. Noble, J.: Ultrasound image segmentation and tissue characterization. *Proc. Inst. Mech. Eng. H J. Eng. Med.* 224(2), 307–316 (2010)
4. Peters, C., Chevalier, R.: Congenital urinary obstruction: Pathophysiology and clinical evaluation. In: Wein, A., Kavoussi, L., Novick, A., Partin, A., Peters, C. (eds.) *Campbell-Walsh Textbook of Urology*. Elsevier Inc., Philadelphia (2012)
5. Shapiro, S., Wahl, E., Silberstein, M., Steinhardt, G.: Hydronephrosis index: A new method to track patients with hydronephrosis quantitatively. *Urology* 72(3), 536–538 (2008)
6. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) *ECCV 1998*. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
7. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(11), 1222–1239 (2001)
8. Clerc, M., Kennedy, J.: The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* 6(1), 58–73 (2002)
9. Xiao, G., Brady, M., Noble, J., Zhang, Y.: Segmentation of ultrasound B-mode images with intensity inhomogeneity correction. *IEEE Trans. Med. Imag.* 21(1), 48–57 (2002)

Joint Model-Pixel Segmentation with Pose-Invariant Deformable Graph-Priors

Bo Xiang^{1,2}, Jean-Francois Deux³, Alain Rahmouni³, and Nikos Paragios^{1,2}

¹ Center for Visual Computing, Ecole Centrale de Paris, France

² Equipe GALEN, INRIA Saclay, Île-de-France, France

³ Radiology Department, Henri Mondor Hospital, Créteil, France

Abstract. This paper proposes a novel framework for image segmentation through a unified model-based and pixel-driven integrated graphical model. Prior knowledge is expressed through the deformation of a discrete model that consists of decomposing the shape of interest into a set of higher order cliques (triplets). Such decomposition allows the introduction of region-driven image statistics as well as pose-invariant (*i.e.* translation, rotation and scale) constraints whose accumulation introduces global deformation constraints on the model. Regional triangles are associated with pixels labeling which aims to create consistency between the model and the image space. The proposed formulation is pose-invariant, can integrate regional statistics in a natural and efficient manner while being able to produce solutions unobserved during training. The challenging problem of tagged cardiac MR image segmentation is used to demonstrate the performance potentials of the method.

1 Introduction

Segmentation is one of the most well studied topics in computer vision. Model-free methods are often based on clustering, aiming at grouping together pixels with consistent intensity properties. Knowledge-driven methods, on the other hand, aim to find a solution that is a compromise between the one produced from the observations and the one expressed from the model space.

Popular examples of model-free segmentation refer to the mean-shift method [4], variational formulations such as the Mumford-Shah framework [12], or graph-based methods including normalized cuts [9], graph-cuts [2]. Due to the lack of assumptions on the geometric form of the object of interest, these methods are rather flexible in terms of spread of admissible solutions while it can also lead to erroneous results due to intensity variability, occlusions, noise presence.

Knowledge-based methods are either manifold constrained or manifold enhanced. The former class of methods models geometric variation of the object of interest and then seeks an instance of this space in the image. Active shape/appearance models [6,5] and atlas-based methods [7] are popular examples. Manifold enhanced methods aim to minimize the distance of the solution from the learned manifold, *e.g.* active contours/surface models [10,11]. Both classes of methods inherit a severe limitation with respect to pose, due to the

fact that the current solution should be brought to the same referential as the ones used in learning.

Recently, [14] proposed a pose invariance model through a higher order graph-based formulation with promising results. Due to the discrete shape model, it might produce significant segmentation errors on the boundary. Meanwhile, combined image-model segmentation has been investigated as well. The approach of [13] considered extremely simple shape priors. [3] addressed the problem within an alternating minimization approach involving both discrete and continuous optimization process, where no guarantees on the optimality properties of the obtained solution could be satisfied.

In this paper, we propose a novel pose-invariant segmentation approach that simultaneously solves the problem in both model space and image space. It is achieved by the definition of an objective function aiming to: (i) assign labels to image pixels in order to maximize the image likelihood [2], (ii) deform a point distribution model in order to maximize the geometric likelihood of the model as well as the model-to-image likelihood [14], (iii) impose consistency between the two label spaces. The resulting higher order graphical model formulation is solved by using a state of the art message passing algorithm [8]. Promising results on a challenging clinical setting demonstrate the potentials of our method.

The remainder of the paper proceeds as follows. We present the shape model in section 2 and the segmentation energy is defined in section 3. Experimental validation are shown in section 4 while section 5 concludes the paper.

2 Shape Representation

We adopt the pose invariant shape model in [14]. It consists of: (1) a set of control points distributed on the boundary (Fig.1 left); (2) a set \mathcal{C} of cliques including all possible combinations of three points; (3) a series of probability distributions p_c of each triplet $c \in \mathcal{C}$ learned from a training set.

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ denote a shape instance defined by n points, where \mathbf{x}_i denotes the coordinates of point i . Given a training set of K samples $\{\mathbf{X}^t\}_{t=1}^K$, we assume that point correspondences exist within the training set, but no need to align all the samples in a common reference space. For a triplet clique $c = (i, j, k) \in \mathcal{C}$, the co-occurrence probability of the three points $\mathbf{x}_c = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ can be represented by their inner angles $\{\alpha_c = \angle \mathbf{x}_j \mathbf{x}_i \mathbf{x}_k, \beta_c = \angle \mathbf{x}_i \mathbf{x}_j \mathbf{x}_k\}$.

$$p_c(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = p_c(\alpha_c, \beta_c) \quad (1)$$

This angle representation of triplet is invariant to global pose (*i.e.* translation, rotation, scale) of the shape of the object. With K training samples, the probability density distributions $p_c(\alpha_c, \beta_c)$ of triplet c are learned using a standard probabilistic model (*e.g.* Gaussian distribution model). Assuming that the local constraints on triplets are independent, the global shape is constrained through the accumulation of all the local constraints.

$$p(\mathbf{X}) \propto \prod_{c \in \mathcal{C}} p_c(\alpha_c, \beta_c) \quad (2)$$

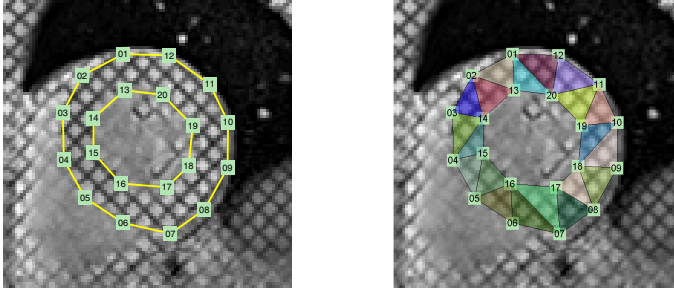


Fig. 1. Shape model. Left: point distribution model. Right: model triangulation.

The shape model inherits pose invariance from the local representation. As a result, no shape alignments to the same referential are needed for both training samples and testing shapes. Moreover, it can be easily encoded with a MRF inference due to the local interactions defined by *prior related* clique set \mathcal{C} .

In addition to prior related concerns, we introduce a *data related* clique set \mathcal{A} which decomposes the object region into triangles (Fig.1 right). Such a triangulation can be applied to any shape (heart, liver etc.) represented as a polygonal area and it should meet the conditions: (1) each triplet is a subset of the model and its corresponding triangle region should be included only in the object; (2) these triangle regions should not overlap; (3) the union of these triangle regions recovers the whole object domain. Using model triangulation facilitates to factorize the regional-driven energy as well as to introduce pixel and model interactions which will be shown in the following segmentation framework.

3 Combined Model-Pixel Based Segmentation

In this section, we propose a framework to combine both model-based and pixel-based segmentation. The aim is to simultaneously deform the shape model to an observing image and to label the image pixels as object/background within an interconnected graphical model.

3.1 MRF Formulation

Now we address the segmentation problem within a higher order Markov Random Field (MRF) formulation. The proposed graph model G consists of two sub-graphs: (1) The model-based G_m consists of a set $\mathcal{V}_m = \{1, \dots, n\}$ of *model* nodes (associated with n points in shape model) and a set of cliques \mathcal{C}_m used in model-based segmentation independently; (2) The pixel-based G_p consists of a set $\mathcal{V}_p = \{1, \dots, k\}$ of *pixel* nodes (associated with k pixels in the image) and a set of cliques \mathcal{C}_p introduced by pixel-based segmentation. Moreover, the two sub-graphs are connected with a clique set \mathcal{C}_{int} . We illustrate the graph structure in Fig.2, where the yellow upper part represents model-based G_m and the green lower part represents pixel-based G_p .

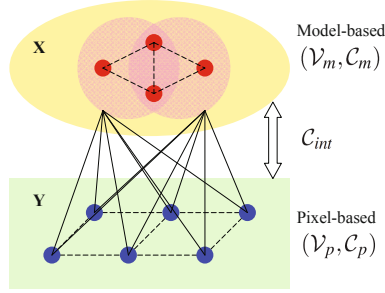


Fig. 2. MRF graphical model coupling the model space and the label space

Let $X_{i \in \mathcal{V}_m}$ (*i.e.* point coordinates) and $Y_{i \in \mathcal{V}_p}$ (*i.e.* pixel label) denote the latent random variables for model nodes and pixel nodes respectively. The variable X_i can take a configuration \mathbf{x}_i from its candidate space \mathcal{U}_i , while the variable Y_i can take a value y_i from label space \mathcal{L} . We define the pixel label space¹ $\mathcal{L} = \{0, \dots, m\}$, where m is the number of triangle parts produced by the clique set \mathcal{A} as defined in the last section. Given an image \mathbf{I} , the segmentation problem is formulated as an estimation of an optimal configuration ($\mathbf{X} = (\mathbf{x}_i)_{i \in \mathcal{V}_m}$, $\mathbf{Y} = (y_i)_{i \in \mathcal{V}_p}$) of all the nodes over model space $\mathcal{U} = \prod_{i \in \mathcal{V}_m} \mathcal{U}_i$ and labeling space \mathcal{L}^k .

$$\begin{aligned}
 (\mathbf{X}, \mathbf{Y})^{\text{opt}} &= \arg \min_{\mathbf{X} \in \mathcal{U}, \mathbf{Y} \in \mathcal{L}^k} E(\mathbf{X}, \mathbf{Y}, \mathbf{I}) \\
 E(\mathbf{X}, \mathbf{Y}, \mathbf{I}) &= E^{(1)}(\mathbf{X}, \mathbf{I}) + E^{(2)}(\mathbf{Y}, \mathbf{I}) + E^{(3)}(\mathbf{X}, \mathbf{Y})
 \end{aligned} \tag{3}$$

where the MRF energy $E(\mathbf{X}, \mathbf{Y}, \mathbf{I})$ contains model-based energy $E^{(1)}$, pixel-based energy $E^{(2)}$ and interaction-based energy $E^{(3)}$.

3.2 Model-Based Energy

This energy is composed by data-related term and prior term [14]. The data term encodes the image likelihood given a model configuration. The prior term encodes spatial constraints of a model configuration with respect to the shape prior manifold. We formulate the model-based energy $E^{(1)}(\mathbf{X}, \mathbf{I})$ as follows.

$$E^{(1)}(\mathbf{X}, \mathbf{I}) = \lambda_1 \cdot \sum_{a \in \mathcal{A}} \Phi^{(1)}(\mathbf{x}_a) + \lambda_2 \cdot \sum_{c \in \mathcal{C}} \Psi^{(1)}(\mathbf{x}_c) \tag{4}$$

where λ_1, λ_2 are the weights of data term and prior term respectively.

The data term captures the homogeneity properties of the object region inside the model. Based on the model triangulation, the regional term can be factorized into higher order potentials $\Phi^{(1)}$ on data triplet set \mathcal{A} introduced in section 2.

$$\Phi^{(1)}(\mathbf{x}_a) = \iint_{\mathbf{x}_a} \mathbf{L} ds, \quad a \in \mathcal{A} \tag{5}$$

¹ When a pixel takes a non-zero value, it is labeled as a triangle part of the object, otherwise it is labeled as background.

where data potential $\Phi^{(1)}$ encodes the integral of image likelihood function \mathbf{L} over the triangle area \mathbf{x}_a . We denote likelihood² $\mathbf{L}_i = \log \frac{p_{\text{bck}}(\mathbf{I}_i)}{p_{\text{obj}}(\mathbf{I}_i)}$.

The prior term is formulated with the prior probability $p(\mathbf{X})$ defined in Eq.(2). It is factorized into potentials $\Psi^{(1)}$ defined on prior clique set \mathcal{C} .

$$\Psi^{(1)}(\mathbf{x}_c) = -\log p_c(\mathbf{x}_c), \quad c \in \mathcal{C} \quad (6)$$

where the distribution probabilities p_c are learned from training. We generalize the model-based interactions $\mathcal{C}_m = \{\mathcal{A}, \mathcal{C}\}$, where both subsets are triplet cliques.

3.3 Pixel-Based Energy

The energy $E^{(2)}$ also consists of a data term and a prior term as in [2].

$$E^{(2)}(\mathbf{Y}, \mathbf{I}) = \lambda_3 \cdot \sum_{i \in \mathcal{V}_p} \Phi^{(2)}(y_i) + \lambda_4 \cdot \sum_{(i,j) \in \mathcal{C}_p} \Psi^{(2)}(y_i, y_j) \quad (7)$$

where λ_3, λ_4 are the weights. The data term encodes the image likelihood over the pixel assignments. The unary potential $\Phi^{(2)}$ encodes the individual penalties for labeling pixel i as object or background.

$$\Phi^{(2)}(y_i) = \begin{cases} -\log p_{\text{bck}}(\mathbf{I}_i) & \text{if } y_i = 0 \\ -\log p_{\text{obj}}(\mathbf{I}_i) & \text{otherwise} \end{cases} \quad (8)$$

where label $y_i = 0$ assigns the pixel i as background, otherwise non-zero value assigns the pixel i as object. We denote p_{bck} and p_{obj} in footnote 2. The prior term penalizes the inconsistency of the pixel labels within a neighborhood system (*e.g.* 8-connected) which is defined by pairwise clique set \mathcal{C}_p .

$$\Psi^{(2)}(y_i, y_j) = \begin{cases} 0 & \text{if } y_i = y_j \\ W & \text{otherwise} \end{cases} \quad (9)$$

where pairwise potential $\Psi^{(2)}$ constraints the neighboring pixel i and pixel j to have the same label, and W is a penalizing parameter.

3.4 Interaction-Based Energy

The interaction energy is the key of producing consistency between model space and labeling space. This consistency is held when given a shape configuration \mathbf{X} , pixels inside/close to the shape boundary should be more likely labeled as object. Due to its dependency on the global shape, it is difficult to be encoded in the framework where both model solution and pixel labeling are sought for at the same time. Using the model triangulation, the model-pixel interaction can be factorized into constraints between pixel and model parts (triplets).

$$E^{(3)}(\mathbf{X}, \mathbf{Y}) = \sum_{(i,a) \in \mathcal{C}_{int}} \Phi^{(3)}(y_i, \mathbf{x}_a), \quad \Phi^{(3)}(y_i, \mathbf{x}_a) = -\log p(y_i | \mathbf{x}_a) \quad (10)$$

² $p_{\text{obj}}, p_{\text{bck}}$ are the appearance distribution models of object and background obtained from a training set.

The potential $\Phi^{(3)}$ is a third-order term, encoding the dependency of a pixel label and a regional triplet. The interaction clique set $\mathcal{C}_{int} = \{(i, a) | i \in \mathcal{V}_p, a \in \mathcal{A}\}$ connects every pixel with every regional triangle. The distribution $p(y_i | \mathbf{x}_a)$ of the pixel label conditioned on a triplet, uses a softmax function.

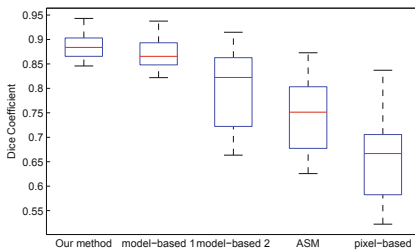
$$\begin{cases} P(y_i = l(a) | \mathbf{x}_a) = [1 + \exp(\text{dist}(i, \mathbf{x}_a))]^{-1} \\ P(y_i \neq l(a) | \mathbf{x}_a) = [1 - p(y_i = l(a) | \mathbf{x}_a)] / m \end{cases} \quad (11)$$

where label value $l(a) \in \{1, \dots, m\}$ equals to the index of triplet a ; m is the number of the regional triplets in clique \mathcal{A} . We denote $\text{dist}(i, \mathbf{x}_a)$ a signed Chamfer distance of the pixel i to the triangle \mathbf{x}_a boundary (*i.e.* negative/positive when pixel is inside/outside the triangle). Hence, all energy terms are defined in MRF formulation and we can use TRW-S algorithms [8] for MAP-MRF inference. To search for an optimal model configuration, we use the same strategy as in [1].

4 Experimental Validation

A dataset of 40 2D tagged cardiac MR images (100*100 pixels) is used to validate our method. The ground truth provided by experts is used for both training and validation. Gabor features are used as the image representation to deal with tagged MR images. We performed a leave-one-out cross validation on the whole dataset. The experiments were run on a 2.8GHz, 12GB Ram computer and our segmentation took a couple of seconds per image.

Some visual results of two test images are presented in Fig. 4. The first column is our results in both model space \mathbf{X} (yellow contours) and label space \mathbf{Y} . The second column shows the results from model-based module (using only energy $E^{(1)}$) [14] (blue contours) and pixel-based module (using only energy $E^{(2)}$) respectively. The third column provides the ground truth (green contours) and the pixel-wise difference image between our labeling result and the ground truth, where the gray pixels are correct labeled, the white/black pixels are wrongly labeled as object/background. The fourth column zooms in the area inside the red box (shown in the third column) with our model results (yellow contours) and model results by [14] (blue contours). We can see that only pixel-based method is sensitive to the complicated background and noise. The only model-based results are globally correct, but do not give accurate segmentation around the boundary locally (*e.g.* see the zoom in effects in the fourth column). The results from our



Method	mean	std
our method	0.8882	0.0263
model-based 1	0.8709	0.0298
model-based 2	0.7966	0.0761
ASM	0.7439	0.0710
pixel-based	0.6560	0.0781

Fig. 3. Comparisons on dice coefficients. Left: box plot. Right: statistic figures.

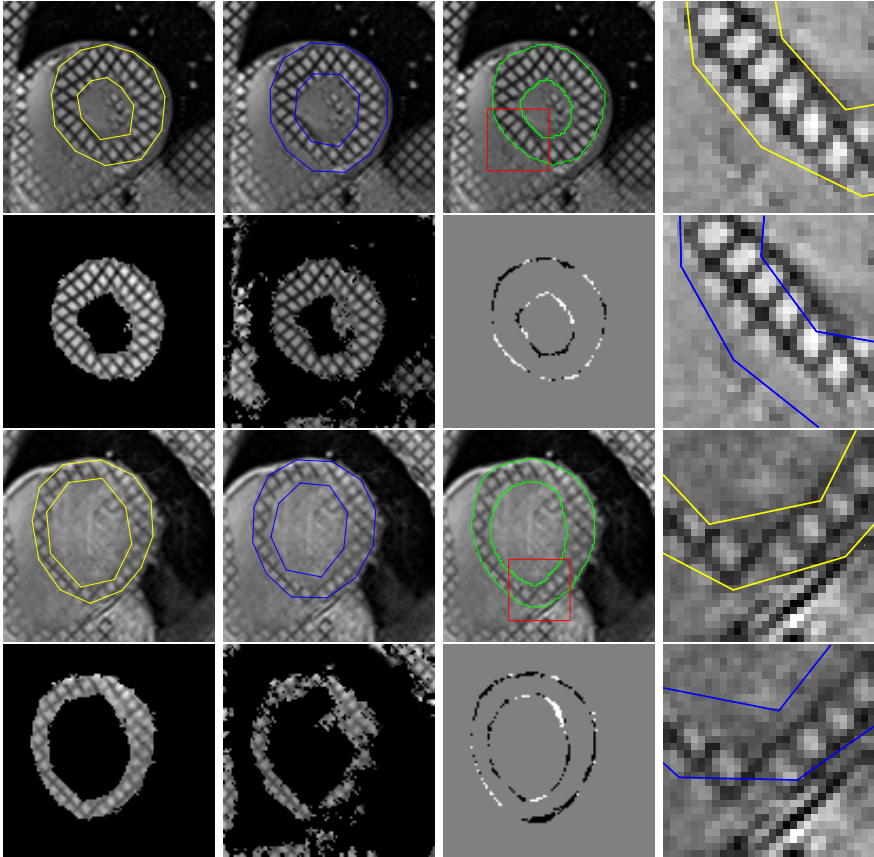


Fig. 4. Segmentation results of 2 test images. The columns from left to right are our results, only model/pixel based results, ground truth and comparison, zoom effects.

integrated framework can overcome this defect, showing flexibility to get local deviations as well as producing pixel-wise labeling result at the same time. Our method also deals well with the varying scales of the object (*e.g.* the scale of the inner contour shows large variability) thanks to the pose-invariant shape prior.

For both quantitative evaluation and comparison purposes, we present in Fig.3 the Dice coefficient distributions obtained respectively by (1) our hybrid method, (2) model-based method 1 [14], (3) model-based method 2 [1], (4) standard ASM method and (5) pixel-base method. Noted that a higher Dice coefficient implies a better segmentation result, Fig. 3 highlights the better performance of our method compared with the previous methods.

5 Conclusions

In this paper we propose a novel approach to address jointly model/image-based segmentation using a higher order graphical model. The proposed formulation

can easily encode regional support, meanwhile being able to account for shape variability unseen during training. Furthermore, it produces states of the art results in particular when exact boundary delineation is of interest through the combined model-pixel graph. To the best of our knowledge, this is the first method that recovers a consistent solution between the model and the image space in a single shot optimization framework, while being pose-invariant.

The formulation involves more weight coefficients of different energy terms that need to be tuned to optimize the result. This defect can be addressed by estimating the parameters through a MRF training algorithm which can deal with a larger number of parameters. It is also necessary to extend the method in 3D cases where numerous challenging segmentation problems do exist in particular in medical image analysis. Compact modeling of shape variability is critical in terms of complexity (number of higher order cliques). Last but not least, understanding the varying importance of these cliques with respect to modeling of shape variations is also important. Recent progress on MRF learning could be a natural path towards adjusting the local contributions of the model.

References

1. Besbes, A., Komodakis, N., Langs, G., Paragios, N.: Shape priors and discrete mrfs for knowledge-based segmentation. In: CVPR, pp. 1295–1302 (2009)
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. *IJCV* 70, 109–131 (2006)
3. Bray, M., Kohli, P., Torr, P.: POSECut: Simultaneous segmentation and 3D pose estimation of humans using dynamic graph-cuts. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 642–655. Springer, Heidelberg (2006)
4. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. *PAMI* 24, 603–619 (2002)
5. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. *PAMI* 23(6), 681–685 (2001)
6. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models-their training and application. *CVIU* 61, 38–59 (1995)
7. Glocker, B., Komodakis, N., Paragios, N., Glaser, C., Tziritas, G., Navab, N.: Primal/dual linear programming and statistical atlases for cartilage segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 536–543. Springer, Heidelberg (2007)
8. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. *PAMI* 28(10), 1568–1583 (2006)
9. Shi, J., Malik, J.: Normalized cuts and image segmentation. *PAMI* 22 (2000)
10. Staib, L., Duncan, J.: Boundary finding with parametrically deformable models. *PAMI* 14(11), 1061–1075 (1992)
11. Taron, M., Paragios, N., Jolly, M.: Registration with uncertainties and statistical modeling of shapes with variable metric kernels. *PAMI* 31(1), 99–113 (2009)
12. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the mumford and shah model. *IJCV* 50, 271–293 (2002)
13. Wang, C., de La Gorce, M., Paragios, N.: Segmentation, ordering and multi-object tracking using graphical models. In: ICCV, pp. 747–754 (2009)
14. Xiang, B., Wang, C., Deux, J., Rahmouni, A., Paragios, N.: 3d cardiac segmentation with pose-invariant higher-order mrfs. In: ISBI, pp. 1425–1428 (2012)

Abdominal Multi-organ CT Segmentation Using Organ Correlation Graph and Prediction-Based Shape and Location Priors

Toshiyuki Okada¹, Marius George Linguraru², Masatoshi Hori¹,
Ronald M. Summers³, Noriyuki Tomiyama¹, and Yoshinobu Sato¹

¹ Department of Radiology, Graduate School of Medicine Osaka University, 2-2
Yamadaoka, Suita, Osaka 565-0871, Japan

toshi@image.med.osaka-u.ac.jp

² Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National
Medical Center, Washington DC, USA

³ National Institutes of Health, Clinical Center, Radiology and Imaging Sciences, 10
Center Drive Bethesda, MD 20892, USA

Abstract. The paper addresses the automated segmentation of multiple organs in upper abdominal CT data. We propose a framework of multi-organ segmentation which is adaptable to any imaging conditions without using intensity information in manually traced training data. The features of the framework are as follows: (1) the organ correlation graph (OCG) is introduced, which encodes the spatial correlations among organs inherent in human anatomy; (2) the patient-specific organ shape and location priors obtained using OCG enable the estimation of intensity priors from only target data and optionally a number of untraced CT data of the same imaging condition as the target data. The proposed methods were evaluated through segmentation of eight abdominal organs (liver, spleen, left and right kidney, pancreas, gallbladder, aorta, and inferior vena cava) from 86 CT data obtained by four imaging conditions at two hospitals. The performance was comparable to the state-of-the-art method using intensity priors constructed from manually traced data.

Keywords: shape prediction, intensity model, partial least squares.

1 Introduction

Several general frameworks for abdominal multi-organ CT segmentation have been proposed. Recent approaches utilize a number of pairs of CT data and their manual traces, called atlases, as training data. Wolz et al. proposed a novel hierarchical method based on nonrigid registration and weighted label fusion using a large dataset of atlases [1]. The method uses sum of squared differences between target data and CT data in atlases, which means that the target data will be required to have similar contrast patterns as CT data in the atlases. Therefore, applicability to different imaging conditions (ICs) may be limited. In addition, nonrigid registration involves high computational cost and difficulty

in dealing with discontinuous deformations among abdominal organs. Machine-learning based approaches have also been studied [2,3]. Although these methods were validated for the various organs, they were not evaluated for organs having large inter-subject variations on shape and location, such as the pancreas and gallbladder. Since the features of the classifiers include the original CT value, the performance for unknown ICs is not clear.

Some multi-organ segmentation methods utilize prior models constructed from atlases [4,5]. Linguraru et al. used 4D graph incorporating multi-organ relations and two phases of intensity priors [4]. Okada et al. used patient-specific statistical atlases based on shape prediction from pre-segmented organ regions [5], which is also regarded as a generalization of some methods specialized to pancreas segmentation [6]. These methods used intensity priors based on intensity information obtained from traced CT data (hereafter we call “supervised intensity information”), and thus again the application to different ICs will be limited. Finally, Freiman et al. proposed a kidney segmentation method [7], in which the estimation of shape and intensity priors was performed by EM algorithm. Although intensity-based non-rigid registration between target and CT data in atlases was needed for initializing shape and intensity priors, only target data was used to estimate intensity prior without the atlases.

In this paper, we propose a framework for multi-organ segmentation, which is adaptable to any ICs without supervised intensity information. The features of the framework are as follows: (1) the organ correlation graph (OCG) is introduced, which encodes the spatial correlations among organs inherent in human anatomy. The concept of OCG is similar to [5], however, we provide fully automated methods for the construction and utilization of OCG. Then (2) the patient-specific organ shape and location priors obtained using OCG enable the estimation of intensity priors without using supervised intensity information and nonrigid registration. We evaluated the proposed method using 86 abdominal CT datasets obtained by four ICs at two hospitals.

2 Methods

We analyze eight organs, that is, the liver, spleen, left and right kidneys, gallbladder, aorta, inferior vena cava (IVC), and pancreas. The fundamental idea is to incorporate inter-organ spatial correlations to attain stable and accurate segmentation of the target organs. To do so, we first perform the segmentation of relatively stable organ in their position and shape (hereafter called “predictor” organs), and then segment other less stable organs whose positions and shapes are expected to be well-predicted by pre-segmented organs. In order to realize the above concept, we begin with a single organ segmentation method described in 2.1. In 2.2, we have a basic segmentation module to analyze multiple organs by using shape and location priors incorporating prediction from predictor organs. In 2.3, we obtain intensity priors from untraced CT data using the prediction-based priors in 2.2. Finally, we introduce OCG to assemble the single segmentation modules to establish a multi-organ segmentation system.

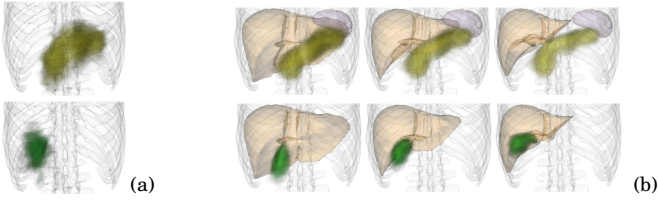


Fig. 1. Conventional and proposed probabilistic atlas (PA). (a) Conventional PA. (b) Proposed prediction-based PA. Upper: Pancreas. Lower: Gallbladder.

2.1 Single Organ Segmentation Method

As priors of the target organs, we utilize a statistical shape model (SSM) and a probabilistic atlas (PA) as shape and location priors [8], and an intensity model (IM) represented as Gaussian mixture model of the intensity distribution as intensity prior. After CT data is spatially transformed to the normalized space defined by the liver dome top and the rectangle circumscribing the bone tissue regions, voxel-wise MAP segmentation using PA and IM is performed, followed by SSM fitting with IM [8]. Finally graph-cut refinement is performed [9].

2.2 Prediction-Based Shape and Location Priors

We assume that $n - 1$ predictor organs are available for the target organ. Let a set of predictor organ shapes be \mathbf{s} and the target organ shape be \mathbf{v} , and let S and V be training data of the predictor and target organ shapes, respectively. We denote the function of PLSR (partial least squares regression) trained using S and V by $PLSR(\mathbf{s}; S, V)$ [10]. Given the pre-segmented predictor organ shapes \mathbf{s}^* , the prediction equation is given by $\mathbf{v} = \mathbf{v}^* + \mathbf{r}$, where $\mathbf{v}^* = PLSR(\mathbf{s}^*; S, V)$, \mathbf{v} is the true shape, and \mathbf{r} denotes the residual after the prediction. We represent \mathbf{r} using PA and SSM [5].

To obtain PA and SSM of \mathbf{r} , we use the average shape of V as reference shape \mathbf{v}'_0 . The true shape \mathbf{v} is transformed using the 3D deformation field generated by correspondences from the predicted shape \mathbf{v}^* to reference \mathbf{v}'_0 , and then the transformed true shape \mathbf{v}' is obtained. Now, we have $\mathbf{r}' = \mathbf{v}' - \mathbf{v}'_0$ of all training data, which are the residuals in the reference space, and then PA and SSM are generated. When we use PA and SSM for segmentation, the inverse of the 3D deformation field is used to bring them to the patient space. Figure 1 shows the prediction-based PA for the pancreas and gallbladder (and the comparison with the conventional PA). Segmentation is performed using the method described in 2.1 by replacing the conventional PA and SSM by the above-described prediction-based ones.

2.3 Estimating Intensity Model from Untraced CT Data

Two types of intensity models are constructed: (1) target-data specific intensity model (TD-IM) where only one CT dataset, which is the segmentation target is

used; (2) imaging-condition specific intensity model (IC-IM), where a sufficient number of CT data acquired using the same IC as the target are used.

The TD-IM is constructed as follows. Let I_0 be the target CT data. First, PA is binarized using the threshold value determined by the 1-percentile in the probability distribution of PA. The binarized region R is the initial volume of interest (VOI). The histogram H_0 is calculated from the intensity distribution of I_0 within R . IM is calculated by fitting a Gaussian mixture model (GMM) to H_0 . MAP segmentation is applied to I_0 using PA and IM, and R is updated by the extracted region. The processes calculating IM with R and updating R by MAP are repeated sufficient times.

The IC-IM is constructed as follows. Let $I = \{I_i\}(i = 1, \dots, N)$ be N optional CT data acquired using the same IC as I_0 . Note that manual traces are unavailable for these CT data. TD-IMs are calculated for each $I_i(i = 0, 1, \dots, N)$. Let $H = \{H_i\}(i = 0, 1, \dots, N)$ be the calculated histograms. The average histogram $\bar{H} = \frac{1}{N+1} \sum_{i=0}^N H_i$ is calculated. IM is calculated by fitting GMM to \bar{H} .

Segmentation is performed using the method described in 2.1 by replacing conventional IM by the above-described TD-IM or IC-IM.

2.4 Multi-organ Segmentation Based on Organ Correlation Graph

The OCG is defined as a set of nodes and directed edges. Each node corresponds to an organ, and each directed edge denotes the correlation from one organ to the other. We consider a simple OCG defined by three nodes and two edges (Fig. 2 (a)). Two nodes having out-edges are regarded as the predictor organs, whose segmentation is assumed to be stable enough as predictors, and the other having in-edge as the target organ or response.

Among the eight organs, liver segmentation has been studied intensively, and is now sufficiently accurate and stable by itself [1,4,5,8]. The spleen and kidneys have also been studied by a sufficient number of works, which showed relatively good accuracy [1,4,5]. Finally, the segmentation of other organs has not been well-studied nor shown to be so accurate. Based on the above observations, we classified the eight organs into the following three types.

Type 1, $V_{C1} = \{Liver\}$: Only out-edges are defined (no in-edges). Segmentation is performed unconditionally.

Type 2, $V_{C2} = \{Spleen, Right\ kidney, Left\ kidney\}$: Both in- and out-edges are defined. Segmentation is performed under the condition that the segmentation is completed for at least one node connected by an in-edge. In addition, the organ can be a predictor for a node connected by an out-edge.

Type 3, $V_{C3} = \{Pancreas, Gallbladder, Aorta, IVC\}$: Only in-edges are defined (no out-edges). Segmentation is performed conditionally like for V_{C2} . This type cannot be a predictor.

Given the constraints on the above three types of nodes, the edges connections representing organ correlations are automatically defined based on shape predictability by PLSR [10]. Let $V_p = V_{C1} \cup V_{C2}$ be a set of nodes for prediction and $V_r = V_{C2} \cup V_{C3}$ be a set of nodes for response. Let V_{all} be all possible

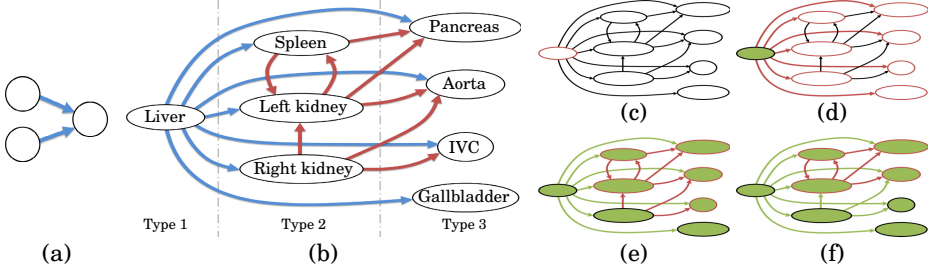


Fig. 2. Multi-organ segmentation based on organ correlation graph (OCG). (a, b) OCG. Blue and red edges indicate the directed edges from a node in Types 1 and 2, respectively. (c-f) Sequential segmentation steps based on OCG. Red border indicates the nodes to be segmented. Green node indicates segmented nodes.

combinations of the nodes in V_p . For each set in V_{all} , the prediction of node $v_r \in V_r$ is performed by applying PLSR and the prediction error is calculated [10]. Then a set of nodes $V_{min}(v_r)$, which has minimum prediction error, is selected. Let $E(v_r) = \{\{v, v_r\} | v \in V_{min}(v_r)\}$ be a set of directed edges to v_r . OCG $G = \langle V, E \rangle$ is defined as $V = V_{C1} \cup V_{C2} \cup V_{C3}$ and $E = \bigcup_{v_r \in V_r} E(v_r)$. Figure 2 (b) shows OCG constructed from the eight organs.

Let v be a node to be segmented, $V_{out}(v)$ be a set of nodes having the edge directed from v , and $V_{in}(v)$ be a set of nodes having the edge directed to v . v is segmented using prediction-based priors constructed from the nodes in $V_{in}(v)$. After the segmentation of v , all nodes in $V_{out}(v)$ become ready for segmentation at the next stage. Let V_{target}^j be a set of nodes ready for segmentation at Stage j , and $V_{extracted}$ a set of segmented nodes. Let $Segmentation(v; V_{predictor})$ denote the segmentation of v based on prediction from $V_{predictor}$. The process of multi-organ segmentation based on OCG is formulated as follows:

1. $j \leftarrow 0$, $V_{target}^j \leftarrow V_{C1}$, $V_{extracted} \leftarrow \emptyset$.
2. For each $v \in V_{target}^j$, apply $Segmentation(v; V_{in}(v) \cap V_{extracted})$.
3. $V_{target}^{j+1} \leftarrow \bigcup_{v \in V_{target}^j} V_{out}(v)$, $V_{extracted} \leftarrow V_{extracted} \cup V_{target}^j$, $j \leftarrow j + 1$.
4. Repeat steps 2 and 3 a fixed number of times.

Fig. 2 (c) to (f) show the process of multi-organ segmentation based on OCG.

3 Experimental Results

We tested the proposed methods using 86 abdominal CT data obtained by four different ICs at two hospitals: Osaka University Hospital and National Institutes of Health (NIH). Table 1 shows the details of the four conditions. Figure 3 shows examples of CT data. In all data, the eight organs were manually segmented. No intensity information combined with manual traces was used in the proposed TD-IM and IC-IM.

Table 1. Details of CT scans used in experiment

Dataset	Institution	Phase	# of cases	Slice thickness [mm]
A	Osaka Univ. Hospital	late arterial phase	10	2.5
B	Osaka Univ. Hospital	late arterial phase	39	0.625
C	NIH	portal venous phase	25	0.68 - 1.25
D	NIH	non-contrast	12	1.0

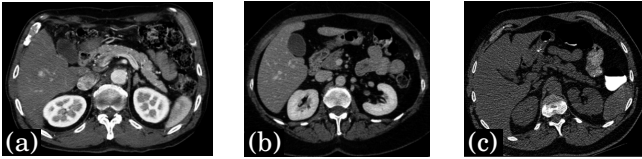


Fig. 3. Examples of CT cross-sectional images of our data. (a) Late arterial phase. (b) Portal venous phase. (c) Non (intra-venous) contrast but artifact due to oral contrast.

The segmentation accuracy using the proposed TD-IM and IC-IM was compared with the intensity model constructed from manual traces (hereafter called “supervised IC-IM”). Leave-one-out cross validation was performed for the evaluation of segmentation accuracy. The supervised IC-IM was constructed by leave-one-out method for each IC. Jaccard index (JI) and Dice coefficient (DC) were used for evaluation. The multi-organ segmentation based on OCG was performed until Stage 4. For liver segmentation, only TD-IM was used.

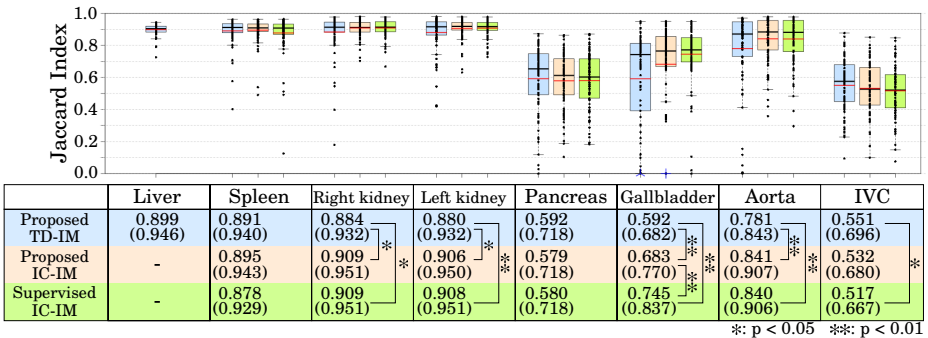


Fig. 4. Summary of accuracy evaluation for three intensity models. Average JIs of 86 cases using proposed TD-IM (blue box plot), proposed IC-IM (orange box plot), and supervised IC-IM (green box plot) for each of eight organs were plotted. Red and black lines in box indicate average and median JIs, respectively. In the liver, only the result using TD-IM was shown. Average JI and DC (shown in parenthesis) values and statistical significance are also shown below the plots.

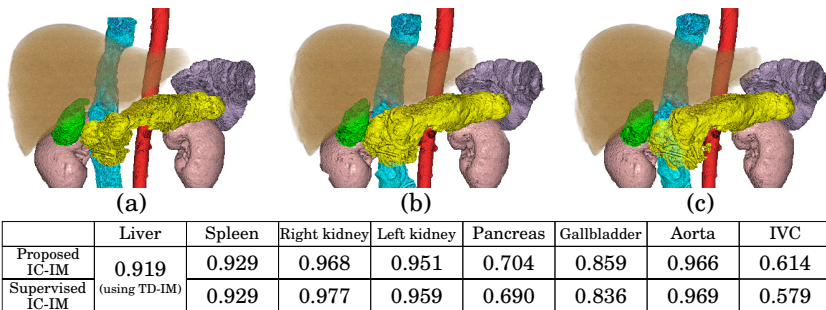


Fig. 5. Typical segmentation result. (a) Manual segmentation. (b) Proposed method using IC-IM. (c) Result using supervised IC-IM. Orange, purple, pink, yellow, green, red, and cyan volumes indicate the liver, spleen, kidneys, pancreas, gallbladder, aorta, and IVC, respectively. Jaccard index of each segmented organ is also shown. In the liver, JI using the proposed TD-IM is shown.

Figure 4 shows the segmentation accuracy for the three intensity models. Figure 5 shows typical segmentation results. Average JIs of the liver, spleen, right and left kidneys using the proposed intensity model were around 0.90 for all methods. These performances were comparable to state-of-the-art segmentation methods based on the availability of large training sets with manual tracings. In the pancreas and IVC, TD-IM achieved the best performance with average JI of 0.592 and 0.551, respectively. In the gallbladder (GB), supervised IC-IM achieved the best performance with average JI of 0.745. No significant accuracy improvement was observed between the proposed IC-IM and supervised IC-IM except for the GB.

The similarity between the proposed IC-IM and supervised IC-IM was calculated using normalized cross correlation (NCC). All NCCs were higher than 0.9 (average was 0.98 ± 0.03), except for GB where NCC was 0.64.

4 Discussion and Conclusion

We have presented a framework for multi-organ segmentation from abdominal CT which is adaptable to any contrast ICs without using supervised intensity information. The performance was comparable to the supervised intensity prior by using additional (untraced) CT data for intensity prior modeling. Results demonstrate that initialization of organ location and shape becomes accurate enough to estimate the intensity model in an unsupervised manner by using OCG and prediction-based priors.

Among the eight analyzed organs, GB was the only organ for which the supervised intensity information was significantly useful. Even for GB, however, if CT cholangiography data are used, in which the gallbladder contrast is enhanced, our method may be adaptable while the conventional method needs a number of manual traces to obtain additional supervised intensity information.

In the proposed method, failures of liver segmentation affect the subsequent prediction and segmentation of other organs. In non-contrast data, the prediction and segmentation of GB failed when the segmented liver region was leaked into GB. The low NCC of IM on GB was due to this leakage problem in addition to the small size of GB. Although this problem also occurred in IVC, its segmentation could be recovered using prediction-based priors.

Future work will include adding other abdominal organs, such as gastrointestinal tract, and the application to diseased organ segmentation.

Acknowledgments. This work is partly supported by KAKENHI No. 21103003.

References

1. Wolz, R., Chu, C., Misawa, K., Mori, K., Rueckert, D.: Multi-organ Abdominal CT Segmentation Using Hierarchically Weighted Subject-Specific Atlases. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 10–17. Springer, Heidelberg (2012)
2. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Pluim, J.P.W., Dawant, B.M. (eds.) Medical Imaging 2009: Image Proceedings, SPIE, vol. 7259, p. 725902 (2009)
3. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of CT images. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
4. Linguraru, M.G., Pura, J.A., Pamulapati, V., Summers, R.M.: Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. *Med. Image Anal.* 16(4), 904–914 (2012)
5. Okada, T., Linguraru, M.G., Hori, M., Suzuki, Y., Summers, R.M., Tomiyama, N., Sato, Y.: Multi-organ segmentation in abdominal CT images. In: 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, pp. 3986–3989 (2012)
6. Shimizu, A., Kimoto, T., Kobatake, H., Nawano, S., Shinozaki, K.: Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. *Int. J. Comput. Assist. Radiol. Surg.* 5(1), 85–98 (2010)
7. Freiman, M., Kronman, A., Esses, S.J., Joskowicz, L., Sosna, J.: Non-parametric iterative model constraint graph min-cut for automatic kidney segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 73–80. Springer, Heidelberg (2010)
8. Okada, T., Shimada, R., Hori, M., Nakamoto, M., Chen, Y.W., Nakamura, H., Sato, Y.: Automated segmentation of the liver from 3D CT images using probabilistic atlas and multi-level statistical shape model. *Acad. Radiol.* 15(11), 1390–1403 (2008)
9. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Eighth IEEE International Conference on Computer Vision, pp. 105–112 (2001)
10. Rao, A., Aljabar, P., Rueckert, D.: Hierarchical statistical shape analysis and prediction of sub-cortical brain structures. *Med. Image Anal.* 12(1), 55–68 (2008)

Multimodal Image Driven Patient Specific Tumor Growth Modeling

Yixun Liu¹, Samira M. Sadowski², Allison B. Weisbrod², Electron Kebebew²,
Ronald M. Summers¹, and Jianhua Yao¹

¹Radiology and Imaging Sciences, NIH

²Endocrine Oncology Branch, National Cancer Institute, NIH

Personalized tumor growth model using clinical imaging data is valuable in tumor staging and therapy planning. In this paper, we build a patient specific tumor growth model based on longitudinal dual phase CT and FDG-PET. We propose a reaction-advection-diffusion model integrating cancerous cell proliferation, infiltration, metabolic rate and extracellular matrix biomechanical response. We then develop a scheme to bridge our model with multimodal radiologic images through intracellular volume fraction (ICVF) and Standardized Uptake Value (SUV). The model was evaluated by comparing the predicted tumors with the observed tumors in terms of average surface distance (ASD), root mean square difference (RMSD) of the ICVF map, the average ICVF difference (AICVFD) of tumor surface and the tumor relative volume difference (RVD) on six patients with pathologically confirmed pancreatic neuroendocrine tumors. The ASD between the predicted tumor and the reference tumor was 2.5 ± 0.7 mm, the RMSD was $4.3 \pm 0.6\%$, the AICVFD was $2.6 \pm 0.8\%$, and the RVD was $7.7 \pm 1.9\%$.

1 Introduction

Quantitatively characterizing the tumor spatial-temporal progression is valuable in staging tumor and designing optimal treatment strategies. Tumor growth not only relies on the properties of cancer cells, but also depends on dynamic interactions among cancer cells, and between cells and their constantly changing microenvironment. The complexity of the cancer system motivates the study of the tumor growth using mathematical models [1] [2] [3]. Swanson et al. [1] assumed an infiltrative growth of the tumor cells, while considering differences in cell diffusion in white and gray matter. Clatz et al. [2] modeled locally anisotropic migration patterns by integrating information from diffusion tensor images (DTI). Hogeia et al. [3] included the mechanical properties of the lesion on surrounding structures to model mass effect.

All these works use a reaction-diffusion model, originally introduced by Turing over 60 years ago, to study cell proliferation and infiltration. In this work, we will extend the reaction-diffusion model to incorporate cell metabolic rate based on the energy conservation law. The travelling front of the reaction-diffusion model offers the benefit that the model prediction (front) can be directly connected with the anatomical CT and MRI via the identifiable tumor boundary in the image. However, tumor boundary only provides limited tumor physiological information. In this paper,

we extract underlying tumor cell fraction in each image voxel using dual phase CT image and incorporate it into the growth model.

In this paper, we target image driven patient specific tumor modeling using routine clinical CT and FDG-PET data. To bridge the gap between the model and the imaging data, we introduce energy conservation law into the modeling and developed a reaction-advection-diffusion model to incorporate cell proliferation, infiltration, metabolic rate and mass effect. We further incorporate intracellular volume fraction (ICVF) derived from dual phase CT data and glucose metabolic rate from FDG-PET (2-[18F] Fluoro-2-deoxyglucose positron emission tomography) to measure model physiological parameters.

We evaluate the proposed model on pancreatic neuroendocrine tumor. The only work on the pancreatic tumor modeling that we are aware of is [4], in which the authors used a compartment model to divide the cell population into three subpopulations: primary tumor cells, metastasis-enabled cells and metastasized cells. The migration rate between subpopulations and the growth rate and death rate within each subpopulation were estimated based on autopsy data. In this paper, we focus on the way to combine routine clinical multimodal images to study the growth of the primary solid tumor.

2 Method

In this section, we first derive a reaction-advection-diffusion model incorporated with cell metabolic rate via energy conservation law, and then describe how to adapt the model to associate it with routine dual phase CT and FDG-PET.

2.1 Tumor Growth Model

According to the tumor logistical growth law presented in [1], the number of the newly created cells can be described by,

$$\frac{dN}{dt} = \rho N \left(1 - \frac{N}{K}\right) \quad (1)$$

where N is the number of cells, a function of time t . ρ is spatial-temporal invariant proliferation rate. This law describes that the tumor grows exponentially at the beginning and then gradually slows down as approaching the carrying capacity K ($N < K$).

As a tumor progresses, the parts with sufficient nutrients and oxygen grow faster, and those suffering vascular inefficiencies will develop into necrosis, suggesting a heterogeneous or spatial-temporal varying proliferation function $\rho(\mathbf{x}, t)$. The metabolic energy conservation law presented by West et al. [5] quantitatively describes the relationship between the metabolic energy and the ontogenetic growth, providing the theoretical foundation to explore the heterogeneity of the proliferation rate. The energy conservation law states that the incoming energy $B(t)$ required for tumor growth is allocated to two parts,

$$B = NB_c + E_c \frac{dN}{dt} \quad (2)$$

where the first term represents the energy to maintain the existing cells and the second term represents the energy to create new cells. B_c is the metabolic rate of a single cell, and E_c is the energy required to create a cell. Both B_c and E_c are assumed constant during tumor growth. Replace $\frac{dN}{dt}$ in equation (2) with $\rho N(1 - \frac{N}{K})$,

$$B = NB_c + E_c \rho N \left(1 - \frac{N}{K}\right) \Rightarrow \rho = \frac{KB - KB_c N}{E_c N(K - N)} \quad (3)$$

The proliferation rate ρ in equation (3) is a function of time t . However, in clinical practice, ρ is only available at specific time points when B and N are measurable. Thus, we approximate ρ at time t between 0 and T through linear interpolation,

$$\begin{aligned} \rho(t) &= \rho(0) + \frac{t}{T}(\rho(T) - \rho(0)) = \\ & \frac{KB_0 - KB_c N_0}{E_c N_0(K - N_0)} + \frac{t}{T} \left(\frac{KB_T - KB_c N_T}{E_c N_T(K - N_T)} - \frac{KB_0 - KB_c N_0}{E_c N_0(K - N_0)} \right) \end{aligned} \quad (4)$$

where B_0, B_T, N_0 and N_T are the measured metabolic rate and cell numbers at time 0 and T , respectively. Apply model (1) to each voxel (millions of cells within $1mm^3$) at position \mathbf{x} , and add a diffusion term as that in the reaction-diffusion model [1] to account for cancerous cell infiltration into surrounding tissues, leading to a reaction-diffusion model,

$$\frac{\partial N}{\partial t} = D \nabla^2 N + \rho N \left(1 - \frac{N}{K}\right) \quad (5)$$

where the first term is the diffusion term, and the second term is the reaction (proliferation) term. D is the diffusivity or infiltration rate. Equation (5) describes that the rate of cell number change equals the sum of the net dispersal of cancerous cells and the net proliferation of cancerous cells. Note that both N and ρ are a function of position \mathbf{x} and time t . The cell number at position \mathbf{x} is not only affected by the proliferation and diffusion (Brownian movement) but also affected by the underlying mechanical deformation (so-called mass effect), which is caused by the growing cells impacting on the extracellular matrix. An advection term is added into model (5) to account for the tumor cells being displaced as a consequence of the underlying mechanical deformation [3],

$$\begin{aligned} \frac{\partial N}{\partial t} &= D \nabla^2 N - \nabla \cdot (N \mathbf{v}) + \rho N \left(1 - \frac{N}{K}\right) && \text{reaction - advection - diffusion} \\ \nabla \cdot \boldsymbol{\sigma} + \mathbf{b} &= \mathbf{0} && \text{momentum} \\ \boldsymbol{\sigma} &= (\lambda \nabla \cdot \mathbf{u}) + \mu (\nabla \mathbf{u} + \nabla \mathbf{u}^T) && \text{constitutive} \\ \mathbf{v} &= \frac{\partial \mathbf{u}}{\partial t} && \text{kinematics} \end{aligned} \quad (6)$$

where the tumor cell drift velocity \mathbf{v} depends on the displacement vector \mathbf{u} induced by the balance between Cauchy stress tensor $\boldsymbol{\sigma}$ and body force \mathbf{b} . λ and μ are unknown Lamé's coefficients in linear elasticity. The body force \mathbf{b} originated from the growing cells is proportional to the local gradient of the tumor cell density,

$$\mathbf{b} = -p \nabla N \quad (7)$$

where p is an unknown positive constant.

Model (6) allows the incorporation of cell metabolic rate and accounts for cell proliferation, infiltration, metabolism and mass effect, but not directly connected to clinical imaging data. We will further develop model (6) in the following two sections in order to associate the model parameters with measurements from CT and FDG-PET. The proposed model accounts for the cell metabolic rate and directly connects to routine CT and FDG-PET, which makes our model fundamentally novel and different from the model in [3].

2.2 Associate the Model with Dual Phase CT

Tissues within a voxel are considered to be made of three well known compartments: (1) a vascular space through which the blood flows; (2) an extravascular extracellular space (EES) which provides the supporting structure of the tissues; and (3) the cellular space. The extracellular volume fraction (ECVF) of the studied voxel at position \mathbf{x} and time t can be estimated by equation (8) using dual phase CT [6],

$$ECVF(\mathbf{x}, t) = \frac{HU_{post}(\mathbf{x}, t) - HU_{pre}(\mathbf{x}, t)}{(HU_{post_bloodpool} - HU_{pre_bloodpool}) / (1.0 - Hct_{blood})} \tag{8}$$

where the numerator is the HU enhancement brought by the contrast distributed in the extracellular space (EES plus vascular space) of the studied voxel, and the denominator is the HU enhancement brought by the contrast distributed in the whole space (EES plus vascular space plus cellular space) of the reference blood pool voxel. The ratio of the HU enhancement is a measure of the fraction of the extracellular space, i.e., ECVF because the HU enhancement is proportional to the volume of the space, in which the contrast reaches equilibrium. HU_{post} and HU_{pre} are HU of the post-contrast CT image and the pre-contrast CT image, respectively. $HU_{post_bloodpool}$ and $HU_{pre_bloodpool}$ are average HU of the blood pool of the post-contrast CT and pre-contrast CT, respectively. The hematocrit Hct_{blood} is the volume percentage (%) of red blood cells in blood, which varies from patient to patient, but can be measured by the blood sample.

ECVF's complement ICVF can be calculated by,

$$\theta(\mathbf{x}, t) = 1.0 - ECVF(\mathbf{x}, t) \tag{9}$$

In equation (6), the cell number N is difficult to be directly measured by CT image. We adapt the reaction-advection-diffusion equation by replacing $N(\mathbf{x}, t)$ with $K\theta(\mathbf{x}, t)$ based on the assumption that all cells have similar volumes,

$$\frac{\partial \theta}{\partial t} = D\nabla^2 \theta - \nabla \cdot (\theta \mathbf{v}) + \rho \theta (1 - \theta) \tag{10}$$

where (replace N_0 and N_T in (4) with $K\theta_0$ and $K\theta_T$, respectively)

$$\rho = \frac{B_0 - a\theta_0}{b\theta_0 - b\theta_0^2} + \frac{t}{T} \left(\frac{B_T - a\theta_T}{b\theta_T - b\theta_T^2} - \frac{B_0 - a\theta_0}{b\theta_0 - b\theta_0^2} \right) \tag{11}$$

where $a = KB_c$ and $b = KE_c$. Both parameters a and b have specific biological meanings, representing the energy to maintain K cells and create K cells, respectively. Similarly, replacing N in (7) with $K\theta$ leads to $\mathbf{b} = -pK\nabla\theta = -q\nabla\theta$.

2.3 Associate the Model with FDG-PET

Normally, when the scanning time is longer than 45 min post-injection, the glucose metabolic rate MR_{glc} or B can be approximated by [7],

$$B = MR_{glc} = \frac{Glc}{100.0} \cdot PET(t) / \left(\frac{dose}{body\ weight} \right) \Rightarrow B = MR_{glc} = c \times SUV(t) \quad (12)$$

where Glc is the glucose concentration in arterial plasma. $PET(t)$ denotes the radioactive tracer FDG¹⁸ concentration in tissue at time t that is measurable from PET. k is a constant that is not dependent on the particular subject being studied. $c = \frac{1.0}{LC \cdot k / 100.0}$, a lumped unknown parameter. The numerator in (12) is widely used as standardized uptake value (SUV), which is proportional to MR_{glc} since both k and LC are constants [7].

Replace B_0 and B_T in equation (11) with $c \times SUV_0$ and $c \times SUV_T$, respectively,

$$\rho = \frac{\alpha SUV_0 - \beta \theta_0}{\theta_0 - \theta_0^2} + \frac{t}{T} \left(\frac{\alpha SUV_T - \beta \theta_T}{\theta_T - \theta_T^2} - \frac{\alpha SUV_0 - \beta \theta_0}{\theta_0 - \theta_0^2} \right) \quad (13)$$

where $\alpha = c/b, \beta = a/b$.

Equation (6) with its reaction-advection-diffusion equation replaced with equation (10) constitutes our proposed model (or state equations from an optimal control standpoint [8]). The proposed model constitutes the forward problem with unknown control parameter $\mathbf{g} = (\alpha, \beta, D, \lambda, \mu, q)$, which can be estimated by fitting the model predicted ICVF θ (a function of \mathbf{g}) with the observed ICVF θ_T ,

$$J(\mathbf{g}) = \int_{\Omega} (\theta(x) - \theta_T(x))^2 dx + \gamma (\mathbf{g} - \mathbf{g}_r) \cdot (\mathbf{g} - \mathbf{g}_r) \quad (14)$$

where the first term measures the degree of similarity, and the second term is Tikhonov regularization term to recover a locally unique solution close to a reference solution $\mathbf{g}_r = (\alpha_r, \beta_r, D_r, \lambda_r, \mu_r, q_r)$ defined as: $\alpha_r = 2.3 \times 10^{-3} g^{-1} \cdot ml \cdot day^{-1}$, $\beta_r = 1.9 \times 10^{-2} day^{-1}$, $D_r = 0.13 mm^2/day$, $\lambda_r = 9310 Pa$, $\mu_r = 1034 Pa$, $q_r = 200 Pa$. The reference solution was derived from literature and was just a rough estimation of the real solution. It is not necessary to be accurate since it is only used to define a region in which the real solution is located. γ controls the balance of these two terms, which is obtained by a trial-and-error strategy. Functional (14) along with model (state) equations constitutes a coupled PDE-constrained optimization problem, which is solved by the one-shot method presented in [8].

3 Results

To study tumor growth, we have developed a dedicated protocol spanning for several years to collect patients with pancreatic tumors. The desirable longitudinal data needs to satisfy the requirements: 1) the tumor should be big enough (volume $> 20 mm^3$) to allow us to ignore the error induced by segmentation and registration, 2) at least three time points and each time point includes both dual-phase CT and FDG-PET, and 3) without any treatments. Usually, a tumor will be surgically removed when it becomes sufficiently big. The contradictive requirements 1) and 3) lead to the difficulty to obtain desirable data.

We evaluate the proposed model by comparing the predicted ICFV and tumor with the measured ICFV and tumor at the 2nd follow-up. The predicted ICFV was produced by growing the ICFV (using model equation) from the 1st follow-up for the period between the 1st and 2nd follow-up with the parameters estimated from the longitudinal data at the baseline and the 1st follow-up. The predict tumor is an isosurface extracted from the predicted ICFV based on a threshold. Six patients with pathologically confirmed untreated pancreatic neuroendocrine tumors were enrolled in our experiment. Fig.1 shows the longitudinal post-contrast CT, fused PET/CT, estimated SUV map (with decay correction) and the ICFV maps (only the relative heterogeneous ICFV region is shown).

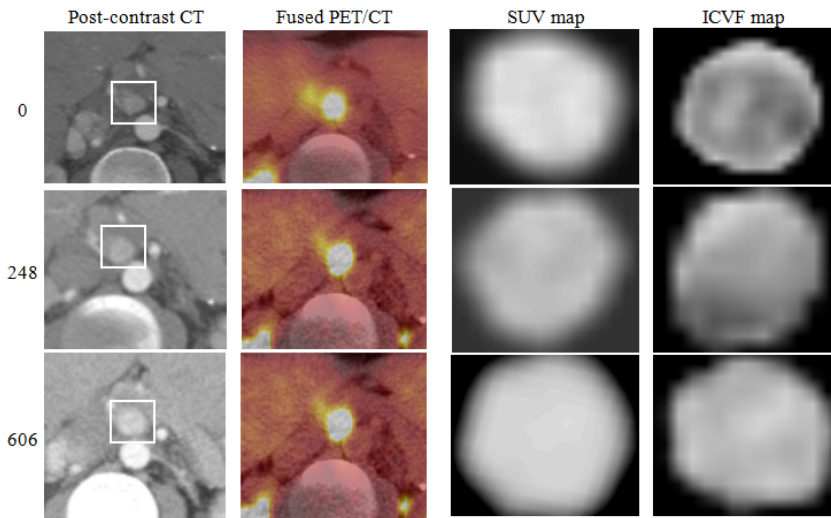


Fig. 1. Longitudinal original and intermediate results. The rows correspond with baseline ($T=0$ day), 1st follow-up ($T=248$ days), and 2nd follow-up ($T=606$ days) and the columns correspond with post-contrast CT, fused PET/CT, estimated SUV and ICFV maps. The white bounding box highlights the tumor.

Fig.2 shows the comparison between the reference results of the 2nd follow-up and the prediction results of two patients. The first row demonstrates a similar distribution of the ICFV map between the reference and the prediction results for both two patients: cell number decreases from the center to the rim of the tumor. The predicted center region (blue) is more homogeneous than the reference part, which might be caused by the exclusion of the complex heterogeneous tumor microenvironment in our model. The comparison of the isocontours of the ICFV map is shown in the second row. The inner most contour shows larger discrepancy, but the outer contours agree well with each other, suggesting a more heterogeneous cell distribution in the center region of the reference tumor, which also can be observed in the gray scale ICFV (the last column of Fig.1). In the third row, we compare the ICFV on the surface of the segmented tumor in post-contrast CT of the 2nd follow-up. We focus on this surface because it is one that can be identified in the image data with our naked eyes. Both patients show similar ICFV distribution and the second patient demonstrated a

more homogeneous ICVF distribution on the tumor surface than the first patient. We assume the average ICVF on the tumor surface to be the threshold that defines the detectable tumor boundary, which makes sense since ICVF is a main factor affecting HU of CT. We use this threshold to extract the isosurface (predicted tumor) from the predicted ICVF map to compare with the segmented tumor in terms of average surface distance and relative volume difference.

The quantitative evaluation regarding the root mean square difference (RMSD) of ICVF map, the average ICVF difference (AICVFD) of tumor surface, the average surface distance (ASD) between the predicted tumor surface and the segmented (reference) tumor surface, and the relative tumor volume difference (RVD, ratio of the volume difference and the volume of the reference volume) are listed in Table 1, in which the RMSD is $4.3 \pm 0.6\%$, the AICVFD is $2.6 \pm 0.8\%$, the ASD is 2.5 ± 0.7 mm, and the RVD is $7.7 \pm 1.9\%$. The predicted ICVF value is slightly large, but the boundary prediction (clinically relevant) is very promising with an average error around 2.5mm. In fact, there is no work to compare with our work regarding ICVF because we are the first to introduce ICVF into the modeling.

We conducted the global nonlinear non-monotonic sensitivity analysis using extended Fourier Amplitude Sensitivity Test (eFAST). Both the first order (S_i) and total order (S_{Ti}) sensitivity indexes show the biological parameters: D , α and β are consistently higher than the three biomechanical parameters: μ , λ and q , which suggests the biological parameters affect the ICVF more than the biomechanical parameters. In the biological parameters, the diffusion D is highest ($S_i=0.4206$, $S_{Ti}=0.7422$), probably disclosing the aggressive infiltration of the pancreatic tumor. In the biomechanical parameters, μ and λ dominate the explanation of the variation, which makes sense since these two parameters control the stiffness and incompressibility of the tissue.

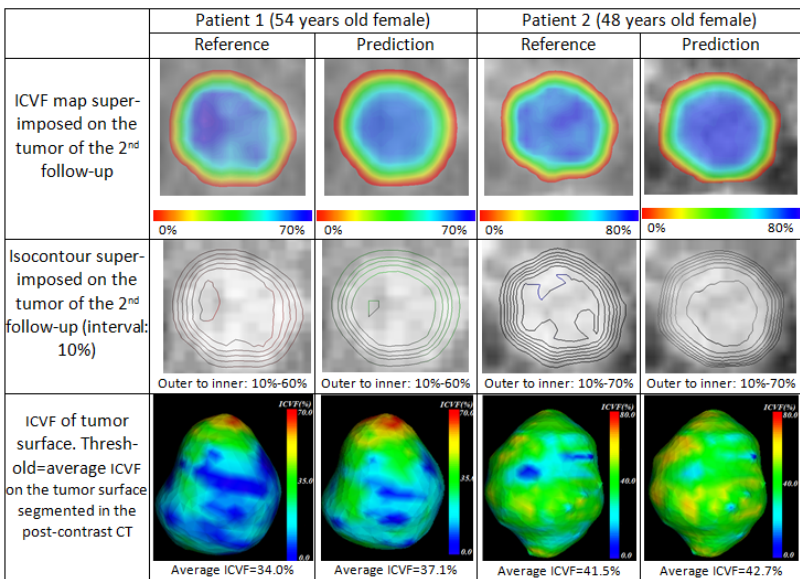


Fig. 2. Comparison between the reference (the 2nd follow-up) and the prediction of two patients regarding ICVF map, isocontour and ICVF of tumor surface

Table 1. Quantitative evaluation. ICVF ([min%, max%]), RMSD (%), AICVFD (%), ASD (mm), RVD (%). The value in the parenthesis of the ASD is the threshold to extract the predicted tumor from predicted ICVF map. $\gamma = 0.6$.

Id	ICVF	Predicted ICVF	RMSD	AICVFD	ASD	RVD
1	[0,73]	[0,81]	4.3	2.9	2.1(42.8)	8.3
2	[0,75]	[0,88]	5.2	2.4	2.5(38.1)	7.6
3	[0,69]	[0,65]	4.6	3.1	3.3(34.0)	10.4
4	[0,75]	[0,79]	3.7	1.2	3.1(41.5)	8.2
5	[0,79]	[0,77]	3.8	3.5	1.5(39.2)	4.5
6	[0,66]	[0,78]	4.1	2.2	2.4(32.2)	7.1
<i>mean ± std</i>	[0,73±5]	[0,81±4]	4.3±0.6	2.6±0.8	2.5±0.7	7.7±1.9

4 Conclusions and Future Work

In this paper, we presented a tumor growth model, which is characterized by incorporating cell metabolic rate into the reaction-diffusion model and being driven by routine clinical imaging data based on ICVF and SUV. The experiment on pancreatic neuroendocrine tumors demonstrated the promise of the proposed model. Other than the characteristics of tumor itself such as the aggressiveness measured by the metabolic rate, tumor microenvironment is also essential for the tumor growth. In the future, besides dual phase CT and FDG-PET, we will introduce DCE-MRI to measure vasculature/perfusion region and FMISO-PET to measure hypoxia region in order to capture some parts of the complex tumor microenvironment.

References

1. Swanson, K.R., Alvord, E.C., Murray, J.D.: A quantitative model for differential motility of gliomas in grey and white matter. *Cell Prolif.* 33, 317–329 (2000)
2. Clatz, O., Maxime, S., Pierre-Yves, B., Herve, D., Simon, K.W., Grole, M., Nicholas, A.: Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. *IEEE Transactions on Medical Imaging* 24, 1334–1346 (2005)
3. Hoge, C., Davatzikos, C., Biros, G.: An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects. *J. Math. Biol.* 56, 793–825 (2008)
4. Haeno, H., Mithat, G., Meghan, B.D., Joseph, M.H., Christine, A.L., Franziska, M.: Computational modeling of Pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell* 148, 362–375 (2012)
5. West, G.B., Brown, G.H., Enquist, B.J.: A general model for ontogenetic growth. *Nature* 413, 628–631 (2001)
6. Nacif, M.S., Kawel, N., Lee, J.J., Chen, X., Yao, J., Zavodni, A., Sibley, C.T., Lima, J.A., Liu, S., Bluemke, D.A.: Interstitial myocardial fibrosis assessed as extracellular volume fraction by low radiation dose cardiac CT. *Radiology* 264(3), 876–883 (2012)
7. Huang, S.: Anatomy of SUV. *Nuclear Medicine & Biology* 27, 643–646 (2000)
8. Gunzburger, M.D.: Perspectives in Flow Control and Optimization. *Advances in Design and Control*, pp. 14–21. SIAM (2003)

Patient-Specific Biomechanical Modeling of Ventricular Enlargement in Hydrocephalus from Longitudinal Magnetic Resonance Imaging

Yasheng Chen^{1,2}, Zheng Fan³, Songbai Ji⁴,
Joseph Muenzer^{5,6}, Hongyu An^{1,2}, and Weili Lin^{1,2}

¹ Biomedical Research Imaging Center, Dept. of

² Radiology, ³ Neurology, ⁵ Pediatrics and ⁶ Genetics,

University of North Carolina at Chapel Hill, NC, 27599

⁴ Thayer School of Engineering, Dartmouth College, Hanover, NH 03755

{yasheng_chen, zheng_fan, joseph_muenzer,

hongyuan, weili_lin}@med.unc.edu,

songbai.ji@dartmouth.edu

Abstract. Ogden type of hyperelastic constitutive law has recently emerged in modeling ventricular enlargement in hydrocephalic brain with finite element method, but this material property for brain tissue has not been investigated in a patient-specific setting in hydrocephalus. Consequently, the accuracy of the simulated ventricular enlargement using this hyperelastic tissue property remains unknown. In this study, we evaluated this brain material model in four patients with communicating hydrocephalus under a small trans-mantle pressure difference (TPMD) between brain ventricle and subarachnoid space (<1mmHg). Based upon changes in ventricular geometries obtained with sequential MRI, we found that this hyper-elastic model has a great flexibility and accuracy in modeling ventricular enlargement (with errors less than 1mm). Our study supports the utility of this hyperelastic constitutive law for future hydrocephalus modeling and suggests that the observed ventricular enlargement in these patients may be caused by a slight increase in TMPD.

Keywords: Brain finite element modeling, Brain mechanics, Hydrocephalus, Hyper-elastic brain modeling, Finite element analysis, Ventricular enlargement.

1 Introduction

Hydrocephalus is defined as an active enlargement of brain ventricles caused by the impairment of cerebrospinal fluid (CSF) homeostasis. CSF is mainly produced by the choroid plexus in the lateral and third ventricles. After flowing into the fourth ventricle through aqueduct of Sylvius, CSF is reabsorbed in the subarachnoid space (SAS) through arachnoid granulations in the sagittal sinus. Hydrocephalus is referred to as communicating hydrocephalus when no apparent obstruction is presented within the brain ventricular system [1].

Finite element method (FEM) based simulation of brain ventricle enlargement started with a biphasic model for brain tissue. Nagashima et al employed Biot's consolidation

theory to model brain tissue as a biphasic poroelastic medium in 2D, and demonstrated that increased hydraulic conductivity may lead to periventricular CSF edema. The transmante pressure difference between ventricle and SAS (TMPD) applied to the ventricular wall was 20mmHg [2]. Pena et al demonstrated that distribution of periventricular edema in acute hydrocephalus is highly related to not only the increased intra-ventricular pressure, but also the natural geometry of the ventricle [3]. Clatz et al combined the biphasic brain tissue model with a scalar model of CSF production-resorption cycle and used Monroe-Kellie assumption (the total volume of brain, CSF and blood is a constant due to the confinement of skull) to simulate the interaction between brain tissue and CSF production under the assumption of infinitesimal deformation [4]. As pointed out in [5], these works have limitations by using both linear elastic tissue property and linear geometrical analysis. More recently, Momjian and Bichsel introduced nonlinearity to this poroelastic model in 2D by varying the Young's modulus as a function of local parenchymal void ratio and relaxing the internal stress [6].

Different from the previously discussed biphasic tissue model, through testing porcine brain tissue, Miller et al have found that brain tissue deforms like a hyperelastic material with a linear viscoelasticity [7, 8]. Due to the large time scale considered for ventricular enlargement, this model was simplified to an Ogden type of hyperelastic constitutive law and was first applied to simulate ventricular enlargement in hydrocephalus by Taylor and Miller [9]. Based upon this hyperelastic model, they suggested that the modulus in previous biphasic modeling of hydrocephalus may be too high, and a considerably lower modulus value of approximately 584Pa should be used [9]. In addition, based upon generic brain geometry, Dutta-Roy et al demonstrated that the Biot's consolidation theory based biphasic brain tissue model is not advantageous when compared to the hyperelastic model in simulating ventricular enlargement [5]. In this work, this hyperelastic model will be further evaluated in terms of deformation capability and accuracy in modeling ventricular enlargement in a patient-specific setting.

As pointed out in [10], one important limitation in current FEM-based numerical analyses in hydrocephalus is the lack of quantitative comparison between the FEM simulation and the actual patient-specific anatomical changes caused by hydrocephalus, which provides a critical feedback for us to gain more insights into brain mechanics in hydrocephalus. To address this limitation, we performed patient-specific geometrical modeling of ventricle enlargement using FEM simulation and the accuracy of the simulated deformation was assessed by using the same patient's sequential MRI as the ground truth. In particular, we sought to determine whether FEM simulation with this hyperelastic model using the parameters as in [5] can model the ventricular enlargement accurately especially under a small pressure loading. To the best of our knowledge, our study is the first 3D nonlinear FEM analysis of ventricular enlargement in hydrocephalus in a patient-specific setting.

2 Methods

2.1 Image Acquisition and Preprocessing

This is an institution review board (IRB) approved study and written consents were obtained prior to image acquisition. We included four patients with communicating

hydrocephalus caused by lysosomal storage disorder. In these patients, ventricular enlargement is a chronic process developed in years. The T1 weighted images of these four patients were acquired 24 months, 28 months, 17 months, and 57 months apart, respectively. No treatment was administered for these patients. The T1 weighted images from the later scans were registered onto the same patient's initial scan using the linear registration in FSL toolkit [11]. Ventricular segmentation was performed with an atlas based approach and manual editing was performed when mis-segmentation occurred with the altered T1 intensity due to the disease. The brain supporting structures including falx cerebri and tentorium cerebelli were manually delineated due to their low image contrast in T1. Tetrahedral meshes were generated using the segmented brain images with a constrained Delaunay process [12].

2.2 Biomechanical Modeling

The hyperelastic material model for the brain was originally proposed by Miller et al as a linear viscoelastic material to account for the nonlinear and strain rate dependent behavior of brain tissue during deformation [7, 8]. Due to the large time scale considered for ventricular enlargement, this model was simplified to an Ogden type of hyperelastic model assuming a potential quasi-static loading with zero strain rate [5]. The hyperelastic constitutive law for brain parenchyma is given as:

$$W = 2\mu/\alpha^2(\lambda_1^\alpha + \lambda_2^\alpha + \lambda_3^\alpha) + \frac{1}{D}(J^{el} - 1)^2, \quad (1)$$

where W is the potential function; λ_i s are the principle stretches, μ is the initial shear modulus (155.77Pa) and α is a material constant (-4.7) [5, 9]. J^{el} is the elastic volume ratio and D is a material coefficient inversely related to the initial bulk modulus. Effectively, this model has only three parameters including μ , α , and D , and λ_i s and J^{el} are computed from nodal displacements. Poisson's ratio (PR) is computed as $(6/(D\mu) - 2)/(12/(D\mu) + 2)$. Previously, Dutta-Roy et al have chosen PR as 0.35, 0.49 and 0.5 representing brain tissue as compressible, nearly incompressible, and fully incompressible materials [5]. It should be noted that even though we use a very fine mesh, potential locking may exist when the PR approximates 0.5. We varied PR in a range of [0.35, 0.49] at a step size of 0.02 for all the four patients. The rigid supporting structures including falx cerebri and tentorium cerebelli were modeled as isotropic elastic material with a Young's modulus of 3.15×10^7 Pa and a PR of 0.45 [13]. It is also worth pointing out that our study is the first to consider these supporting structures in hydrocephalus modeling even though the Young's modulus of these supporting structures is approximately 105 times higher than the initial Young's modulus of the hyperelastic brain tissue model (~ 584 Pa).

We have applied both large (~ 10 mmHg) and small pressure loadings (< 1 mmHg) to the ventricular wall in FEM simulation. The large pressure loading is used to evaluate the capability of this hyperelastic constitutive law in modeling large deformation in patient-specific geometrical setting.

Even though all current FEM-based studies in hydrocephalus load the model with an assumed TMPD [2-6, 9], intriguingly, several experimental studies (including both patient and animal studies) did not detect such a pressure difference [14-16].

These results suggest that the assumed TMPDs (~ 20 mmHg) in previous FEM studies [2, 3] were too high and their FEM models may not truly represent in-vivo biological conditions. In our study, we assume that the maximal TMPD applied to the ventricular wall is with 1mmHg (133.47Pa) as that in [5], and such a small TMPD is expected to be within the experimental variations in [14-16]. We also varied TMPD in the range of 0.1–1 mmHg (with a step size of 0.1 mmHg) jointly with different PRs discussed previously.

Two types of brain-skull boundary conditions have been used in previous hydrocephalus modeling studies. The majority employed the Dirichlet boundary condition by fixing brain mesh surface nodes in all degrees-of-freedom [2-4, 6, 9]. The only exception is a nonlinear boundary condition constraining the brain surface nodes on the bottom while allowing the remainder to move within the cranial cavity as an arbitrary 3mm gap between brain and skull [17]. This boundary condition was originally designed to account for brain shift due to the loss of CSF upon craniotomy for brain tumor resection. Since all subjects in this study had an intact skull, we employed the Dirichlet boundary condition accordingly.

2.3 Evaluation of Simulation Accuracy

The distance from one node in the simulated ventricular surface to the ground truth geometry (the ventricular surface from the later scan) was defined as the minimal distance from this node to all the triangular patches consisting of the ground truth ventricle (Eq. [2]).

$$D(i) = \min \text{dist}(\vec{x}_i, Tri_j) \quad i \in M, j \in N \quad (2)$$

where $D(i)$ is the shortest distance from node i on the simulated ventricular surface to all the triangular patches (Tri) on the ground truth ventricle from the later MRI. M is the total number of nodes in the simulated ventricular surface and N represents the total number of triangular patches on the ground truth ventricle. We further compute the weighted sum of $D(i)$ from all the nodes as a measure of error of the simulation (Eq. (3)). The weight for one node was computed as the ratio of $1/3$ of the total area of all the neighboring triangular patches over the total area of the entire simulated ventricular surface (Eq. (3)). The purpose is to account for the area differences of the triangular patches after mesh is deformed.

$$w_i = \sum_{j \in N_i} 1/3 * Area_j / \sum_{j=1}^N Area_j; D_e = \sum_{i=1}^M w_i D(i) \quad (3)$$

N_i is the number of neighboring patches to node i , and $Area_j$ is the area of the triangular patch j on the surface. For reference, the same measure was also computed between the initial and the ground truth ventricles of the same patient (referred to as D_{init}).

3 Results

3.1 Deformation Capability of Hyperelastic Modeling

The registered images from the later scans to the initial scans are given in Fig. 1 for all four patients. The rendered ventricular surfaces from patient1 are given for the initial

(Fig. 2(a)) and later (Fig. 2(b), the ground truth) scans. The ventricular enlargement was especially apparent as the thickening of the frontal and occipital horns of the lateral ventricle (indicated by the arrows in Fig. 2(a) and 2(b)). Simulated ventricular surfaces with two sets of TMPD and PR are given in Fig. 2(c) and Fig. 2(d), respectively.

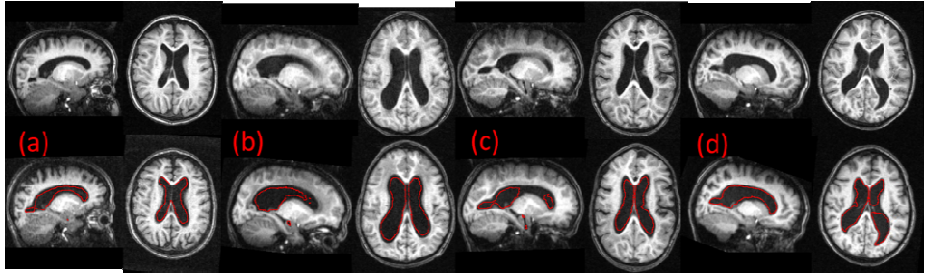


Fig. 1. T1 images from the four patients with the ventricle contour (red curves) from the initial visit (top panel) superimposed on the T1 images from the later visit (bottom panel)

In order to demonstrate the capability of this hyperelasticity constitutive law in modeling an extremely large deformation, we applied a TMPD of 10mmHg with a PR of 0.35 to determine whether the FEM analysis using this hyperelastic model is able to converge, because a TMPD of 4.0 ± 3.6 mmHg was reported in patients with hydrocephalus before [18]. The simulations converged at the TMPDs of 9.57mmHg (Fig 2 (d) for patient1), 10mmHg, 9.87mmHg and 10mmHg for these four patients, respectively. Thus, even without resorting to an adaptive numerical technique such as arbitrary Lagrangian and Eulerian FEM [19], this hyperelastic model for brain tissue allows us to simulate extremely large ventricular enlargement.

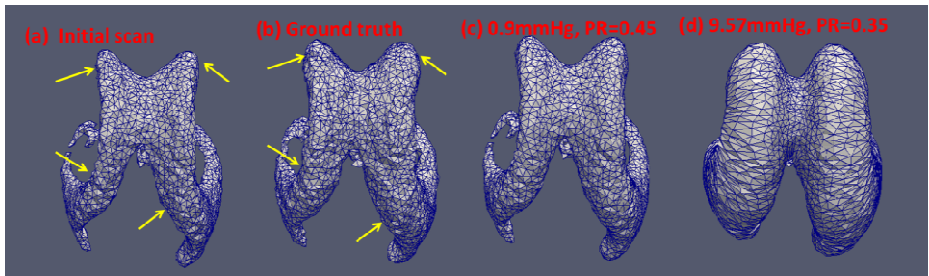


Fig. 2. The rendered ventricle shapes at the initial (a) and later (b) scans and from simulations with various conditions (c and d)

3.2 Accuracy of FEM Simulation

We varied PR and TMPD to minimize the fitting errors. In the 80 simulations for each patient performed with ABAQUS, the errors (computed via Eq. (3)), and the combinations of TMPD and PR having the least error for all patients are given in Fig. 3. The minimal errors were found at (TMPD=0.9mmHg, PR= 0.45), (TMPD=1mmHg,

PR=0.45), (TMPD=1mmHg, PR=0.45) and (TMPD=0.6mmHg, PR = 0.41) for these four patients, respectively. The minimal errors D_e for the simulated ventricular surfaces were (0.773mm, 0.983mm, 0.940mm, 0.718mm), which were (49.7%, 48.4%, 51.1%, 48.8%) of D_{init} (1.554mm, 2.031mm, 1.841mm, 1.472mm). For instance, the best approximation for patient 1 (Fig. 2(c)) bears a very high resemblance to the ground truth (Fig. 2(b)). The distance maps to the ground truth geometry from the initial scan (Fig. 4(a) and (c)) and the best simulation (Fig. 4(b) and (d)) in both dorsal and ventral views are rendered in Fig. 4. In Fig. 5, the 2D contours from the best simulations and their corresponding ground truth geometries were overlaid on the later MRI in three orthogonal slice locations for these four patients.

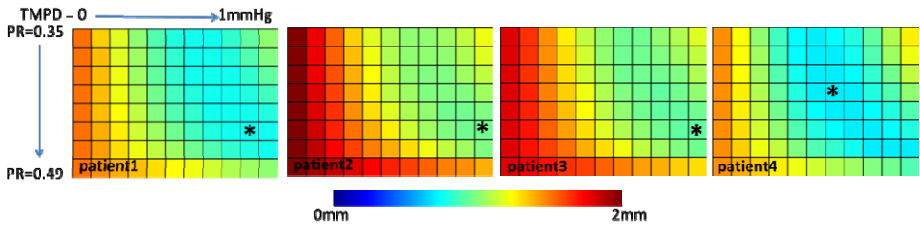


Fig. 3. The errors from the 80 simulations for the four patients with different combinations of TMPD and PR (with * indicating the parametric pair with the least error)

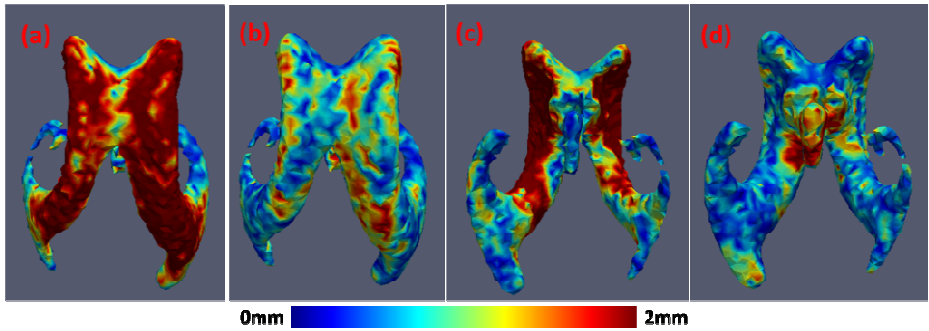


Fig. 4. The rendered maps of D_{init} (a and c) and D_e (b and d) for patient 1

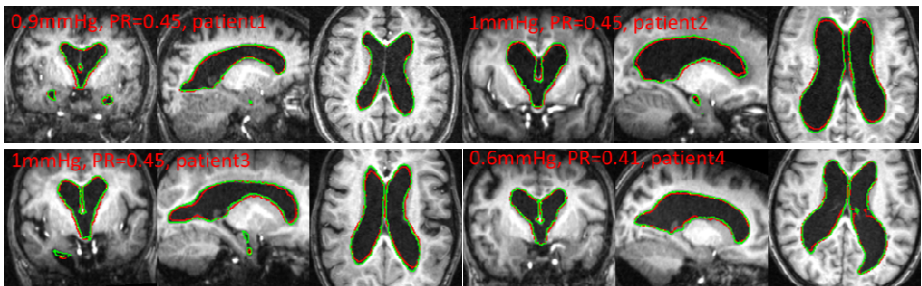


Fig. 5. The 2D contours comparing the ventricular boundaries obtained from the best simulation result (red) and the ground truth (green) super-imposed on the later MR scans

4 Conclusions

Our initial experience with the hyperelastic modeling supports its future application in investigating hydrocephalus. We have demonstrated that with the previously suggested parameters, this model is not only able to approximate the deformation observed in these four patients with high accuracy (less than 1mm in error), but also flexible enough to achieve large deformations even under a large TMPD in line with previous experimental reports (4.0 ± 3.6 mmHg [18]).

Our results suggest that the investigation of the mechanical factors in communicating hydrocephalus may warrant the investigation of the existence of a small TMPD, which may be able to explain the controversies among some of the previous hydrocephalus studies. For instance, the existence of TMPD in hydrocephalic patients with ventricular enlargement has been reported in [18, 20] but not in [14-16]. We speculate that such a small TMPD as in communicating hydrocephalus may not be readily detected in these experiments due to instrumental variations.

Limitations in our study are noted. Measurements of TMPD and tissue property in these patients were not conducted due to the invasive nature of the involved procedures. Evidence from magnetic resonance elastography suggests that hydrocephalic brain tissue may be more compliant due to the weakened mechanical structures [21, 22]. If this is the case for our patients, even though we may have underestimated the ventricular enlargement under the same TMPD, the major conclusion that a small TMPD may lead to ventricular dilation in these patients remains to be true. Due to the same limitation, we did not consider the potential temporal changes and spatial inhomogeneity of brain tissue property. We varied Poisson's ratio to account for deviation from the originally proposed model parameters. At last, in the initial scans of these patients, brain tissue may have already been strained due to ventricular enlargement and our analysis may not start from an ideal tension-free setting.

References

1. Rekte, H.L.: A contemporary definition and classification of hydrocephalus. *Semin. Periatr. Neurol.* 16, 9–15 (2009)
2. Nagashima, T., Tamaki, N., Matsumoto, S., Horwitz, B., Seguchi, Y.: Biomechanics of hydrocephalus: a new theoretical model. *Neurosurgery* 21, 898–904 (1987)
3. Pena, A., Bolton, M.D., Whitehouse, H., Pickard, J.D.: Effects of brain ventricular shape on periventricular biomechanics: a finite-element analysis. *Neurosurgery* 45, 107–116; discussion 116–108 (1999)
4. Clatz, O., Litrico, S., Delingette, H., Paquis, P., Ayache, N.: Dynamic model of communicating hydrocephalus for surgery simulation. *IEEE Trans. Biomed. Eng.* 54, 755–758 (2007)
5. Dutta-Roy, T., Wittek, A., Miller, K.: Biomechanical modeling of normal pressure hydrocephalus. *J. Biomech.* 41, 2263–2271 (2008)
6. Momjian, S., Bichsel, D.: Nonlinear poroplastic model of ventricular dilation in hydrocephalus. *J. Neurosurg.* 109, 100–107 (2008)

7. Miller, K., Chinzei, K.: Constitutive modelling of brain tissue: experiment and theory. *J. Biomech.* 30, 1115–1121 (1997)
8. Miller, K.: Constitutive model of brain tissue suitable for finite element analysis of surgical procedures. *J. Biomech.* 32, 531–537 (1999)
9. Taylor, Z., Miller, K.: Reassessment of brain elasticity for analysis of biomechanisms of hydrocephalus. *J. Biomech.* 37, 1263–1269 (2004)
10. Rekte, H.L.: The usefulness of mathematical modeling in hydrocephalus research. *Childs. Nerv. Syst.* 10, 13–18 (1994)
11. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(suppl. 1), S208–S219 (2004)
12. Si, H.: Adaptive tetrahedral mesh generation by constrained Delaunay refinement. *Int. J. Numer. Meth. Engrg.* 75, 856–880 (2008)
13. Kleiven, S.: Predictors for traumatic brain injuries evaluated through accident reconstructions. *Stapp. Car. Crash. J.* 51, 81–114 (2007)
14. Penn, R.D., Lee, M.C., Linninger, A.A., Miesel, K., Lu, S.N., Stylos, L.: Pressure gradients in the brain in an experimental model of hydrocephalus. *J. Neurosurg.* 102, 1069–1075 (2005)
15. Shapiro, K., Kohn, I.J., Takei, F., Zee, C.: Progressive ventricular enlargement in cats in the absence of transmantle pressure gradients. *J. Neurosurg.* 67, 88–92 (1987)
16. Stephensen, H., Tisell, M., Wikkelsø, C.: There is no transmantle pressure gradient in communicating or noncommunicating hydrocephalus. *Neurosurgery* 50, 763–771; discussion 771–763 (2002)
17. Wittek, A., Miller, K., Kikinis, R., Warfield, S.K.: Patient-specific model of brain deformation: application to medical image registration. *J. Biomech.* 40, 919–929 (2007)
18. Conner, E.S., Foley, L., Black, P.M.: Experimental normal-pressure hydrocephalus is accompanied by increased transmantle pressure. *J. Neurosurg.* 61, 322–327 (1984)
19. Chen, Y., Ji, S., Wu, X., An, H., Zhu, H., Shen, D., Lin, W.: Simulation of brain mass effect with an arbitrary Lagrangian and Eulerian FEM. *Med. Image. Comput. Comput. Assist. Interv.* 13, 274–281 (2010)
20. Hoff, J., Barber, R.: Transcerebral mantle pressure in normal pressure hydrocephalus. *Arch. Neurol.* 31, 101–105 (1974)
21. Streitberger, K.J., Wiener, E., Hoffmann, J., Freimann, F.B., Klatt, D., Braun, J., Lin, K., McLaughlin, J., Sprung, C., Klingebiel, R., Sack, I.: In vivo viscoelastic properties of the brain in normal pressure hydrocephalus. *NMR Biomed.* 24(4), 385–392 (2011)
22. Freimann, F.B., Streitberger, K.J., Klatt, D., Lin, K., McLaughlin, J., Braun, J., Sprung, C., Sack, I.: Alteration of brain viscoelasticity after shunt treatment in normal pressure hydrocephalus. *Neuroradiology* 54(3), 189–196 (2011, 2012)

Simulation of Lipofilling Reconstructive Surgery Using Coupled Eulerian Fluid and Deformable Solid Models

Vincent Majorczyk, Stéphane Cotin, Christian Duriez, and Jeremie Allard

Inria Nord-Lille Europe, France

Abstract. We present a method to simulate the outcome of reconstructive facial surgery based on fat-filling. Facial anatomy is complex: the fat is constrained between layers of tissues which behave as walls along the face; in addition, connective tissues that are present between these different layers also influence the fat-filling procedure. To simulate the end result, we propose a method which couples a 2.5D Eulerian fluid model for the fat and a finite element model for the soft tissues. The two models are coupled using the computation of the mechanical compliance matrix. Two contributions are presented in this paper: a solver for fluids which couples properties of solid tissues and fluid pressure, and an application of this solver to fat-filling surgery procedure simulation.

Keywords: Simulation, fluid-solid interaction, reconstructive surgery.

1 Introduction

The Parry-Romberg syndrome is a progressive hemifacial atrophy. It is due to disorders of central nervous system and is characterized by a degeneration of tissues beneath the skin. The syndrome affects generally one side of the face and distorts the nose and the mouth. Causes are still unknown, an autoimmune mechanism is suspected. The reconstructive surgery is the only way to restore the face with the adding of fat. In addition, there are less severe cases of facial dystrophy due to side effect of drugs or insulin-resistance for which drug solutions are not considered sufficient compared with reconstructive surgery [1].

The fat-filling procedure consists of injecting fat in subcutaneous areas. The main literature concerning about the fat-filling surgery are the result analyses of the operation. Surgeons test different methods of injection, or different products, and they compare measures before the operation, short-term and long-term to define the viability of their method [2]. As it is difficult to predict the results, most of the time surgeons rely on their experience to plan the operation and calculate the required volume of fat. So the surgeon must become familiar with the possible depths of injection (subdermal, intramuscular, supraperiosteal), and the amounts of fat to get the desired change [3].

The facial anatomy is complex: there are many layers of tissues (skin, Superficial Musculo-Aponeurotic System ...) with different properties [4]. Furthermore, ligaments, nerves and blood vessels connect the layers, change the stiffness of

tissues and create obstacles for injected fat. Each patient has a particular anatomy, and surgeons may need help planning difficult cases.

The long-term aim of this project is to use simulation and patient specific data to predict the behavior of injected fat under the skin. So the anatomy needs to be modeled realistically [5] to understand the role of each element.

The model presented in this paper is an 2.5D eulerian fluid. The fluid is injected between two layers of tissues, moves along a curved surface and has a thickness. The fluid solver uses the compliance properties of the solid to calculate the pressure of fluid. This pressure acts to surfaces and deforms solids.

2 Related Works

In simulation, while we can find a large amount of works about breast augmentation, it mainly concerns implants [6] and not the fat injection. The aim of these works consists to plan the surgery operation, finds the best implant and predicts the result. These works use the finite element method to simulate deformable tissues. The reconstruction of facial atrophy may use computed-assisted lipofilling [7], it only considers the volume difference of fat between the original model and the wished result.

The simulation of fluid-structure coupling is often particularly challenging. Each problem depends on the aim of the simulation, on the required precision and on the available data. The coupling method often needs to be adapted.

In gynaecology, Karry [8] describes a model with viscous fluid between soft tissues. It concerns a microbicide gel injected in the vagina to analyze the repartition of gel due to the pressure of membranes. This model is composed of a viscous fluid between two deformable solids. However, the proposed model is restricted to the simple case of an homogeneous solid.

Lagrangian particle-based methods allow to describe volume of solid and fluid as particles and generate repulsive forces among particles [9]. Solenthaler *et al.* [10] fixes the problem of incompressibility, but introduces a costly iterative scheme.

Eulerian models have inherent difficulties to resolve moving domain boundaries. Moreover, the method needs to avoid fluid loss and to correctly transfer forces between fluid and solid. Klingner [11] defines the topology of the domain at each step-time and handles accurately the behaviour of boundaries. Other methods handle moving boundaries within an Eulerian grid by modeling partly filled cells [12–14].

3 Fluid-Solid Model

The fat flows along a surface between two layers of tissues. In this model the fat is considered as a fluid and tissues are represented by deformable solids. The purpose of this section is to describe the fluid-solid coupling.

3.1 Fluid Model

The fluid model needs to be physically realistic and stable to simulate fluid between two deformable surfaces. The base of the method is described in [15]: it is an Eulerian fluid model which consists of a velocity field and a pressure field. The domain of the surface is in the form of a staggered grid representation [16]. Further, the surface is curved, the model uses a distortion matrix to adapt data (velocities and pressure) from facial topology to unitary cells [17]. We extend the model to add a fluid height field and take it into account in all equations to obtain a 2.5D formulation.

The method considers a succession of four operations: computation of external forces, the diffusion to simulate viscosity, the advection for turbulence and the projection to force fluid incompressibility. Regarding fluid-solid interactions, it is important to detail the projection mechanism.

The projection step allows the input fluxes in a cell to be equal to the output fluxes. So there is neither surplus nor loss of matter. At the beginning of the step the error between input flux and output flux is computed (eq.1). The error is named divergence D and is the sum of flux around the cell. The flux is the velocity u_{edge} multiplied by the area A_{edge} between adjacent cells.

$$D_{cell} = \sum_{edge} A_{edge} u_{edge}(t) \quad (1)$$

The pressure is a scalar field of values defined within each cell. It creates an additional flux proportional to the pressure differences between adjacent cells. For an incompressible fluid, the pressure can be computed as the solution enabling the divergence of the resulting velocity field to be zero, i.e. the divergence of the fluxes introduced by difference of pressure within a cell P_{cell} and in an adjacent cell P_{adj} should counterbalance the initial cell divergences:

$$D_{cell} = \sum_{edge} A_{edge} (P_{cell} - P_{adj}) \quad (2)$$

By gathering the equations of all cells, we obtain a matrix system of the form:

$$\mathbf{d} = \mathbf{J}\mathbf{p} \quad (3)$$

Where \mathbf{d} is a vector with the divergence values D_{cell} , \mathbf{p} is a vector which contains pressure values P , and \mathbf{J} is the assembling matrix which includes area coefficients A_{edge} .

3.2 Deformable Solid

Good visual results can be obtained with discrete approaches (like spring-mass or particles) that are very fast to compute but not physics-based. On the other hand, numerical solutions, like the finite element method, allow to integrate accurately the constitutive law (strain/stress relation) of the deformable solid. The computation cost of such methods could be highly reduced when using

corotational strain measure [18] which is valid for deformations with large transformation (displacement and rotation).

This formulation gives the possibility to compute the compliance matrix (inverse of the matrix of the dynamic system) which have a pivotal role in the interaction calculation [19]. Equation (eq.4) describes the dynamic of a non-linear deformable object with external constraints applied :

$$\mathbf{M}\dot{\mathbf{v}} = \mathbf{f}_{ext} - \mathbb{F}(q, v) + \mathbf{H}^T \lambda \quad (4)$$

where \mathbf{M} is mass matrix, q and v are the position and velocity vectors. \mathbb{F} is a function which describes the internal visco-elastic forces. \mathbf{f}_{ext} is the external forces such as the gravity. λ is a vector of applied constraints forces, multiplied by a constraint space to mechanical space correspondence matrix \mathbf{H}^T .

The constraints are given as a set of (in)equations :

$$\delta = \left[\mathbf{H} \underbrace{\left(\frac{\mathbf{M}}{h^2} + \frac{d\mathbb{F}}{hdv} + \frac{d\mathbb{F}}{dq} \right)}_{\mathbf{K}^{-1}} \mathbf{H}^T \right] \lambda \quad (5)$$

δ corresponds to the desired constraint displacements and \mathbf{K}^{-1} , the inverse of matrix \mathbf{K} which is homogeneous to stiffness matrix, is called the compliance matrix. So, (eq. 5) corresponds to the displacement of the fluid-solid interface due to forces applied on the object.

3.3 Fluid-Solid Coupling

We combine the above models to create a coupled system. The interactions between the fluid and the solid are handled by combining the fluid pressure and mechanical constraints equations.

From the point a view of the fluid, any motion of the solid which introduces height variations δ_h changes the incompressibility pressure equation (eq.3) into:

$$\mathbf{J}\mathbf{p} + \mathbf{A}\delta_h = \mathbf{d} \quad (6)$$

\mathbf{A} is the matrix containing area coefficients converting height displacements to volumes. The height variation is a modification of the thickness of the fluid per time-step due to the pressure of the fluid on the solid. This volume change corresponds to having a flow normal to the surface of the cell. This flow is added to the divergence to compute the pressure.

From the point of view of the deformable solid, the coupling is considered as a set of constraints pushing on the surface exposed to the pressure of the fluid:

$$\lambda = \mathbf{A}^T \mathbf{p} \quad (7)$$

Combining with (eq.5), we obtain:

$$\delta_h = \mathbf{H}\mathbf{K}^{-1}\mathbf{H}^T \lambda \quad (8)$$

\mathbf{H} is a matrix which adapts forces applied on the fluid-solid interface to the solid. That is to say, the fluid interface and the solid may have different topologies. Next, we merge (eq.7) and (eq.8) with the pressure equation (eq.6) :

$$\mathbf{J}\mathbf{p} + \underbrace{(\mathbf{A}\mathbf{H}\mathbf{K}^{-1}\mathbf{H}^T\mathbf{A}^T)}_{\mathbf{C}}\mathbf{p} = \mathbf{d} \quad (9)$$

\mathbf{C} is a compliance expressed in the volume/pressure space. As the fluid is a closed domain, the above equation can be difficult to solve. To avoid this problem, we use numerical regularization and reinforcement of the diagonal of the matrix by replacing \mathbf{C} with a regularization matrix \mathbf{R} whose diagonal is the sum of all contributions of volume displacement on the domain (N cells) by a pressure on the considered cell i :

$$\mathbf{R}_{i,i} = \sum_{j=1}^N \mathbf{C}_{i,j} \quad (10)$$

$$(\mathbf{J} + \mathbf{R})\mathbf{p} = \mathbf{d} \quad (11)$$

The model allows to calculate pressure using fluid data and solid data. By introducing the matrix \mathbf{R} , we have the volume changes of the fluid-film domain when applying pressure on the deformable walls. The role of this matrix is to provide some improvements to the numerical level: it essentially stabilizes the system and helps the convergence of the system. Thus, when solving this equation, we guaranty the incompressibility of the fluid. Finally, the pressures are transformed to forces to affect the motion of the deformable tissues.

4 Results

4.1 Simple Square

We first tested the behavior of our model on a square domain, where the fluid is placed between a rigid solid and a deformable solid. The simulation starts with an tunnel-like fluid volume in the center. The displacement over time of the fluid-solid surface is given in Fig.1. It shows a convergence of results with the use of different time-steps and mesh resolutions. The third graph corresponds to simulations with different stiffness. The simulations shows the coherence of the mathematical model: whether spatial, temporal, respect of fluid incompressibility or general behavior.

4.2 Facial Fat-Filling

To validate our simulation on actual patient data, we start with a geometrical mesh of part of the face around the jaw. This three-dimensional mesh is obtained by laser scan of the skin of the patient's face. The surface mesh is then extruded and three layers of different tissues (skin, subcutaneous and SMAS¹.) are created

¹ Superficial Muscular Aponeurotic System

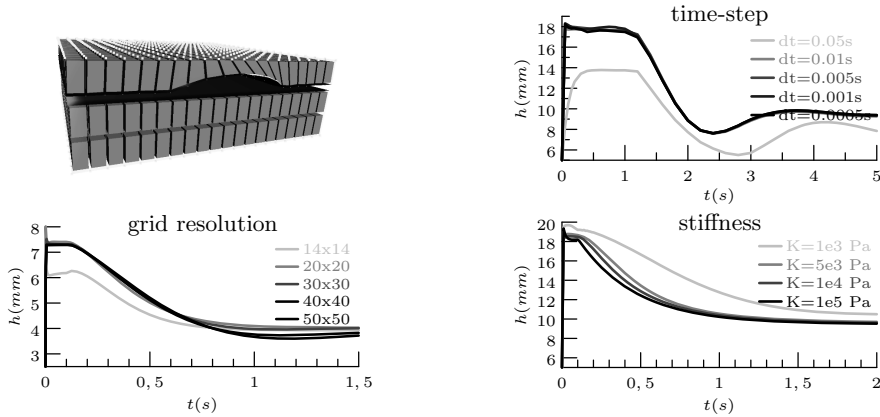


Fig. 1. Test simulation with a simple square (bottom layer: rigid ; top layer: deformable). The graphs show simulation results with different time-step, resolution and rigidity (Young's Modulus); curves represent the height of fluid over time of a point at the center of the square.

as hexahedral meshes. These meshes support a finite element computation of the tissue deformation, where the constitutive law is based on geometrically non linear elasticity (large displacements, small strains). Mechanical parameters (Young's modulus) of each layer is derived from data available in the literature: muscle 0.5 MPa [20], skin between 0.09 MPa [22] and 0.5 MPa [21]. The fat may be injected between any of the two layers. Adding fat near the skin modifies locally the shape of the cheek and adding fat near the SMAS corrects more globally the shape. Each layer is attached to one another by constraints to create zones separated by ligaments or walls.

The simulation illustrated in Fig.2 is composed of two layers of tissues (muscle and skin) and we inject fat to fill two zones: at the masseter and at the angle of inferior maxillary, corresponding to a good part of what was performed during the actual surgery. The simulated fat injection deforms the tissues based on the fluid pressure and tissue stiffness, it generates a change in the shape of the cheek. Then we compared the result of the simulation to the actual shape of the patient after surgery. Fig.2 shows the resulting deformation and the comparison. Although some differences exist between our results of the simulated fat-filling and the surgical outcome, the deformations are similar and the behaviour of model provides realistic results.

4.3 Performance Measurements

The time-step used in the above simulations is 10 ms. The computation time for each step is 0.6 seconds on average. Obtaining the compliance matrix C is currently the most expensive operation, because it involves the inverse of the matrix K whose size is proportional to the number of nodes of the deformable model (1244 hexahedrons compose each layer and 220 cells for the fluid).

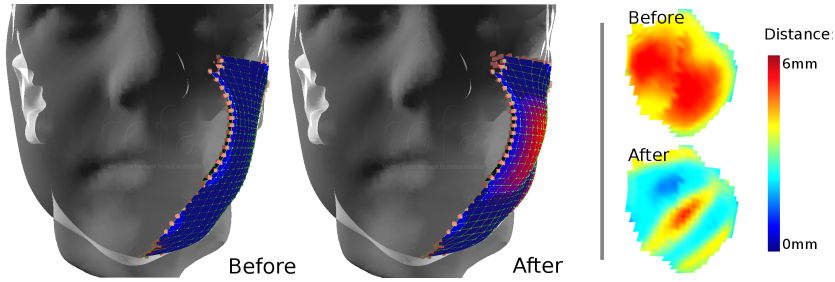


Fig. 2. Simulation with patient-specific data. The patient is suffering from Parry-Romberg syndrome. The blue zone is the simulated side. At the left, the patient before operation and the patient with simulated injection. At the right, is the Hausdorff distance with the result of surgical operation.

5 Conclusion

We presented a simulation method for fat-filling surgery procedure, based on a fluid-solid coupling interaction model. This model can take into account the variation in stiffness due to each patient anatomy.

Further works could be investigated in three areas : improving computational speed, handling large fluid viscosity, and using better ligament model and more generally a better knowledge of the anatomy. The computation speed could be significantly improved by optimizing the inversion of the mechanical stiffness matrix, or using one of several possible approximation methods. Fat is a really material in-between fluid and solid, but visco-elastic properties are not included in the fluid model. Afterwards, the model of tissue is constituted of layers which are constrained at many points to simulate obstacles. The constraints do not use properties of ligaments and should be more elastic. In addition, the tissue layers currently have a uniform thickness. The use of more accurate description of tissues should provide more realistic tissue behaviors.

Acknowledgements. The authors would like to thank Guillaume Captier, plastic surgeon at CHRU Montpellier, for providing us with the data. We would also like to thank Gerard Subsol and Benjamin Gilles, researchers at Montpellier Laboratory of Informatics, Robotics, and Microelectronics for their valuable feedback on this project.

References

1. Calmy, A., Bernasconi, E., Meier, C.A., Hirschel, B., Toutous-Trellu, L.: 10 years of lipodystrophy—and so many uncertainties. *Rev. Med. Suisse* 4(184), 2755–2757 (2008)
2. Fontdevila, J.J., et al.: Assessing the long-term viability of facial fat grafts: an objective measure using computed tomography. *Aesthetic Surgery Journal* 28(4), 380–386 (2008)

3. Coleman, S.R.: Structural Fat Grafting. *Grabb and Smith's Plastic Surgery*, 6th edn., pp. 480–485
4. Keeve, E., Girod, S., Kikinie, R., Girod, B.: Deformable Modeling of facial tissue for craniofacial surgery simulation. *Comput. Aided Surg.* 3(5), 228–238 (1998)
5. Mazza, E., Barbarino, G.G.: 3D mechanical modeling of facial soft tissue for surgery simulation. *Facial Plast. Surg. Clin. North Am.* 19(4), 523–537 (2011)
6. Lapuebla-Ferri, A.S., et al.: A patient-specific FE-based methodology to simulate prosthesis insertion during an augmentation mammoplasty. *Medical Engineering and Physics* 21, 1–9 (2011)
7. Hoehnke, C., Eder, M., Papadopoulos, N.A., Zimmermann, A.: Minimal invasive reconstruction of posttraumatic hemi facial atrophy by 3-D computer-assisted lipofilling. *Journal of Plastic, Reconstructive and Aesthetic Surgery* 60, 1138–1144 (2007)
8. Karri, S.: 2D Thin-Film Flow of a Non-Newtonian Fluid Between Elastic Boundaries. Master-Thesis of the University of Kansas (2011)
9. Muller, M.: Fast and robust tracking of fluid surfaces. In: *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA 2009*, pp. 237–245. ACM, New York (2009)
10. Solenthaler, B., Pajarola: Predictive-Corrective Incompressible SPH. *ACM Transactions on Graphics* 28(3), 401–406 (2009)
11. Klingner, B.: Fluid animation with dynamic meshes. In: *SIGGRAPH 2006*, pp. 820–825 (2006)
12. Johansen, H.: A cartesian grid embedded boundary method for poisson's equation on irregular domains. *Journal of Computational Physics* 147(1), 60–85 (1998)
13. Batty, C., Bertails, F., Bridson, R.: A fast variational framework for accurate solid-fluid coupling. *ACM Trans. Graph.* 26(3), 100 (2007)
14. Peskin, C.: The immersed boundary method. *Acta Numerica* 11, 479–517 (2002)
15. Stam, J.: Stable-fluids. In: *SIGGRAPH 1999*, pp. 121–128 (1999)
16. Fedkiw, R., et al.: Visual simulation of smoke. In: *SIGGRAPH 2001*, pp. 15–22 (2001)
17. Stam, J.: Flows on surfaces of arbitrary topology. *ACM Transactions on Graphics* 22(3), 724–731 (2003)
18. Muller, M., Gross, M.: Interactive virtual materials. In: *Graphics Interface 2004*, pp. 239–246 (2004)
19. Courtecuisse, H., Allard, J., Duriez, C., Cotin, S.: Preconditioner-based contact response and application to cataract surgery. In: *Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891*, pp. 315–322. Springer, Heidelberg (2011)
20. Duck, F.: *Physical properties of tissues: a comprehensive reference book*. Academic Press, London (1991)
21. Samani, A., et al.: A 3D contact problem finite element model for breast shape deformation derived from mri data. In: *ASB 1999 Annual Conference* (1999)
22. Schnabel, J.A., et al.: Validation of non-rigid registration using finite element methods. In: *Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082*, pp. 344–357. Springer, Heidelberg (2001)

Towards a Better Understanding of Pelvic System Disorders Using Numerical Simulation

Pauline Lecomte-Grosbras¹, Mouhamadou Nassirou Diallo^{1,2,3},
Jean-Francois Witz¹, Damien Marchal², Jeremie Dequidt², Stéphane Cotin²,
Michel Cosson³, Christian Duriez², and Matthias Brieu¹

¹ Laboratoire de Mécanique de Lille, École Centrale de Lille

² INRIA - University of Lille 1

³ CHRU Lille - University of Lille 2

Abstract. Genital prolapse is a pathologic hyper-mobility of the organs that forms the pelvic system. Although this is common condition, the pathophysiology of this disorder is not well known. In order to improve the understanding of its origins, we recreate - virtually - this biomechanical pathology using numerical simulation. The approach builds on a finite element model with parameters measured on several fresh cadavers. The meshes are created from a MRI of a healthy woman and the simulation includes the mechanical interactions between organs (contacts, ligaments, adhesion...). The model is validated through comparison of functional mobilities of the pelvic system observed on a dynamic MRI. We then propose to modify, step by step, the model and its parameters to produce a pathologic situation and have a better understanding of the process. It is not a formal proof but the numerical experiments reinforce the clinical hypothesis on the multifactorial origins of the pathology.

1 Introduction

Genital prolapse is a pathologic hyper-mobility of the pelvic system's organs. It induces a strong discomfort and incontinence. This pathology affects one third of women, whatever the age [1] and more than 60% of women older than 60 [2]. To improve quality of life, surgical treatments exist with an intervention rate of 12 % at the age of 80. Unfortunately, these interventions have a failure of 30 % independently of the surgical technic. The main reason of these failures relates to the complex and multifactorial physiopathology of the prolapse which is still not perfectly understood.

In order to better understand the physiological mobilities and to improve the surgical technics, several works have started to model the pelvic system. For instance, an important work on tissue characterization has been led, using destructive methods [3][4]. Other work address the problem of the evolution of soft connective tissue characteristics with respect to the age and/or the pathology [4] [5]. These results provide a range of tissue properties and evolution for several structures that can possibly be compared. But they can not provide a global understanding of the respective contribution of each element on the deformations

of pelvic organs. Some previous work present FEM based models of the pelvic system. For instance, FEM model is used to simulate the bladder filling [7]. In [8], bladder filling is used to get the material properties by inverse simulation. In [9] a FEM approach completed by characterization measurements provides a first FEM model dedicated to pelvic system mobilities. In [11], the work is completed by a study of the ligaments modeling. The problem of modeling the interactions between the pelvic organ is faced in [6] but the solution is only 2D.

In this paper, we build on these works to provide a complete 3D FEM simulation with a wide variety of boundary conditions (contacts with friction, role of ligaments, adhesion and fixations), in order to reproduce the healthy pelvic mobilities that are validated with a dynamic MRI. But the main contribution of this work is to use and degrade the FEM model till progressively reaching a simulation of pathological cases. The topology of the ligament support system is also modified while carefully relying on anatomical and clinical references. We then compare the mobilities obtained in simulation to the ones observed clinically. Finally, the pathologic example of cystocele illustrates how this simulation tool provides new evidence on the phenomena at the origin of the disease.

2 Bio-mechanical Model of the Pelvic System

In order to build the FEM simulation of the pelvic system mobilities, it is necessary to have a geometrical model of the anatomy, the mechanical properties of tissues and a good modeling and numerical tools for the boundary conditions (attachments, loadings, contacts, etc.). This section describes how the physics-based model of pelvic system mobilities is obtained.

Geometrical Model. The geometry of the key structures of the pelvic system is obtained from segmentation of MR images of voluntary healthy women. The images are segmented to get the geometrical contours. The image is taken while the bladder is full in order to have a good contrast on the image. For the vagina and the rectum, a gel is inserted in order to increase the contrast. From the data we get also the contour of the uterus and the bony pelvis. However, the MRI does not provide enough signal for the segmentation of the ligament system of the pelvic floor. Its geometry was reconstructed according to anatomical data, available in the literature [12].

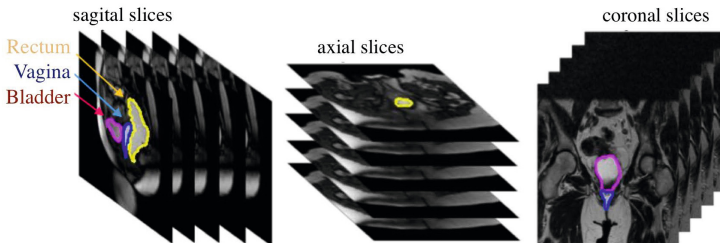


Fig. 1. Geometric model construction: contour segmentation

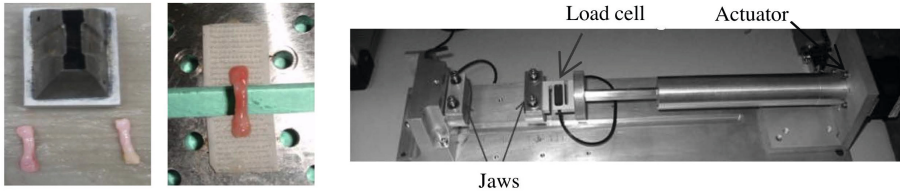


Fig. 2. Protocol and experimental set-up for the mechanical characterization of soft tissues. The samples are cut using a punch (left). Their geometry was defined in order to obtain a homogeneous deformation.

Mechanical Properties of Soft-Tissues. Experimental measurements were conducted in order to estimate the mechanical properties of the tissues that form the pelvic system. The protocol used for ex vivo characterization has been carefully built (control of hygrometry of the tissue, body temperature, rate of deformation for visco-elasticity, etc.) [9] [10]. A portable uniaxial traction machine (see Figure 2) was developed and used in order to do the experiments at the morgue, on 18 fresh cadavers (without pathology) with a mean age of 75. Several mechanical tests were realized on each cadaver (2 to 3 per organ) which makes a total of more than 300 tests. These tests are completed by measurements on pathologic tissues (see section 3) for a total of up to 2000 tests. Biaxial tension tests have been considered but would require too large samples for the considered organs (with our machine: 9cm x 9cm). Moreover, our experimental results reveal that there is no statistical evidences of anisotropy for pelvic soft tissues. In our measurements, if we sort the organs according to their stiffness, the order is always the same. The vagina is the organ made of the stiffer tissue, whereas the rectum and the bladder are softer. Even if the dispersion is important between cadavers, we can define - statistically - average mechanical properties for each organs if we define classes of population related to age and nulliparous or not. The following table provides the mechanical properties used for the tissues to build the *healthy* model:

Anatomical structure	Young Modulus	
	median (interquartile range)	Thickness/Diameter(mm)
Rectum	0,54 (0.66) MPa	3
Vagina	0,67 (2,22) MPa	4
Bladder	0,24 (1.5) MPa	3
Pelvic floor (quantitative data)	3 MPa	10
Uterosacral Ligament	0,78 (1.98) MPa	8
Round Ligament	1,32 (1.08) MPa	4
Large Ligament	2.22 (4.5) MPa	4

For the simulation, we use the Finite Element code available in SOFA [13]. The organs of the pelvic system are mostly hollow, so for modeling their deformation, we use triangular elements based on shell theory (both membrane and bending energies are modeled). Due to its high rigidity, the uterus is modeled as a rigid body. For modeling the ligaments, series of beam elements are used. The tissues are considered as almost incompressible due to their high level of water content. Thus we use a Poisson ratio of 0.45 for deformable shells.

Boundary Conditions. At this stage, the model includes the organ geometry and their mechanical properties, but the boundary conditions play a central role in the resulting mobilities. The suspension and the support systems of the organ are needed, as well as the modeling of the interactions between the organs.

For the ligaments, the literature often highlights the role of the round, large and uterosacral ligaments. This is in line with our observations during the dissection and measuring programs on cadavers. But, to obtain better results on the mobilities, we have also modeled some existing ligaments that are less documented: paravaginal and umbilical. The role of the paravaginal ligament is fundamental to limit the retroversion of the uterus and the umbilical ligament limits the mobility at the top of the bladder. The figure 3 presents two views on the models with the organs that are modeled and the ligaments.

The analysis of the displacement field of a dynamic MRI shows that there is two zones where the organs are fixed: where the bladder is in contact with the pubis, and where the rectum is in contact with the sacrum. These adhesions are modeled using fixed boundary conditions on the models. Finally, in order to model the mechanical interactions between the organs and reproduce the role played by the fascias, we use a friction contact model, as proposed in 2D in [6]. The model relies on Signorini conditions and Coulomb friction law and leads to complex non-smooth dissipative behaviors. We benefit from the numerical treatment of such unilateral contact constraints that is provided by SOFA [13].

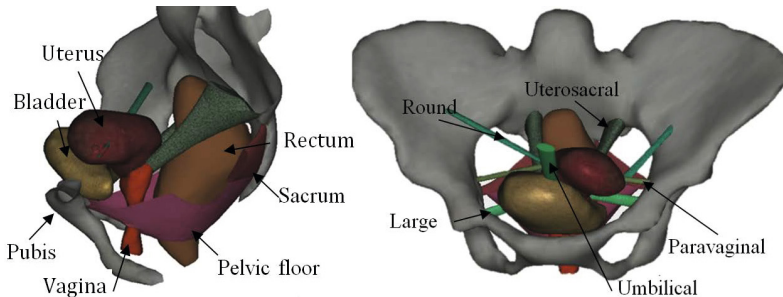


Fig. 3. Complete model of the pelvic system: the four organs, the pelvic floor, the umbilical ligament and the four symmetric pairs of ligaments: round, paravaginal, large and uterosacral

Validation for Healthy Tissues. The figure 4 shows a comparison between the displacement that are measured by dynamic MRI of a healthy woman. The displacements are obtained when simulating a push effort. The data in [12] indicates that such effort can be modeled by a pressure P , in the order of 10 MPa, oriented at a 45 degree angle. This is a strong hypothesis that is very difficult to verify, yet the difference between the measured displacements and simulated ones is less than 1 cm while the maximal displacements is more than 10 cm.

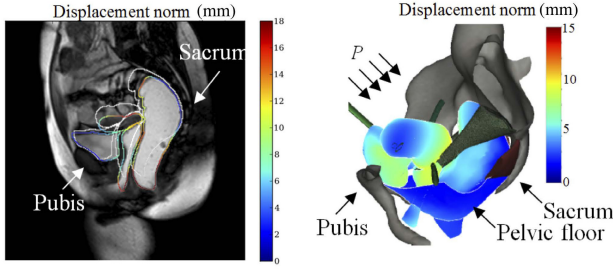


Fig. 4. The displacements of the organs are measured on a dynamic MRI to validate the displacements obtained by FEM simulation on healthy woman

3 Simulation of the Multi-factorial Origins of Pathologies

According to mechanical studies, pelvic sagging pathologies are mostly due to the elongation of the ligaments but other factors like the change of tissues properties, and the degradation of the fascias are also mentioned. However, in real life, it is impossible to evaluate separately the impact of each of these factors. Yet it is a relevant information for reparative surgery. In this work, we propose a first answer based on a set of numerical experiences, on the healthy biomechanical model of the pelvic system. These experiences illustrate the role that the different factors can play in the pathology.

The Role of Ligaments. For obtaining healthy physiological mobilities, we have placed, modeled and use several ligaments. If we refer to anatomical books, and in particular to comparative anatomy [14], it can be seen that all female mammals have round, large, paravaginal and umbilical but only biped mammal have uterosacral ligaments. In order to understand the reason, we have realized some simulation in which the direction of the gravity changes compared to the body axis. Figure 5 shows the force exerted on the ligaments compared to the direction of gravity: it highlights the role of the ligaments according to the body angle.

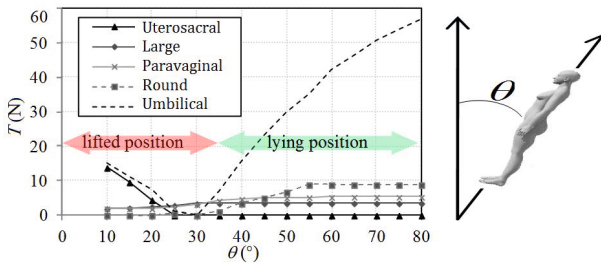


Fig. 5. Forces in the ligaments according to different position of the body. Calculated by simulation.

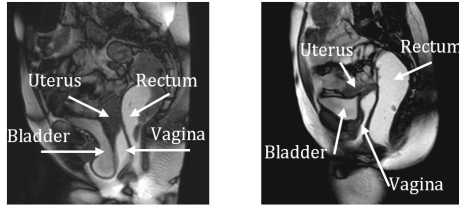


Fig. 6. Left: Dynamic MRI of a cystocele pathology, Right Dynamic MRI of healthy tissues. The motion of the bladder is completely different.

These simulations allow to confirm that in the lifted position ($\theta < 35^\circ$) the uterosacral ligaments play a predominant role, while in a lying position, the contribution of the paravaginal and umbilical ligaments is crucial. These results allow to understand why important work published in pelvic surgery studies consider empirically, that the uterosacral ligaments play a key role on the pelvic mobilities.

Modeling the Prolapse Pathology. The following simulations consist in degrading progressively the healthy physiological model in order to recreate, artificially, a frequent pathological mobility: the cystocele. The cystocele is a pathology that leads abnormally important motion of the bladder that can lead to an important sagging of the vagina and the bladder and to incontinence problems. Figure 6 (left) shows a MRI issued from a dynamic sequence in the sagittal plane in the case of a cystocele pathology. During the push effort, the bladder undergoes an important displacement and tends to move out of the body. This pathologic mobility can be compared with healthy mobilities during the same effort (Figure 6 right).

In order to identify the role of each element in the pathology, we intentionally degrade the functional healthy model in order to reproduce the pathologic motion of the bladder. Several steps were used to degrade the model. We conducted additional experimental measures of characterization using biological tissues taken during prolapse surgeries (issued from about 40 patients) and also

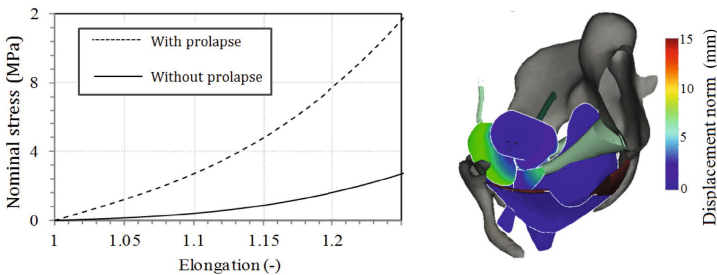


Fig. 7. Comparison of the mechanical behavior of vagina tissues of cadaver, with and without genital prolapse - Simulation obtained using pathologic constitutive law

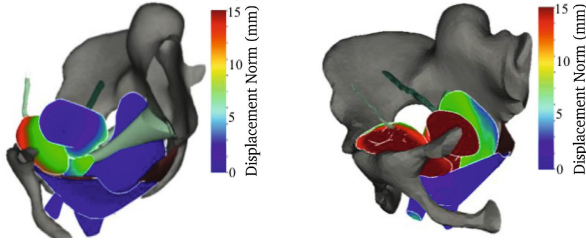


Fig. 8. Simulation of the elongation of the ligaments (left), elongation of ligaments and degradation of fascia between bladder and pubic bone

on 5 pathologic fresh cadavers (approximately the same age than the non-pathologic ones). These experiments highlight the discrepancies in the characteristics of pathological and healthy tissues. Paradoxically, contrary to what we might expect, the pathologic tissues are often stiffer than the healthy ones (see Figure 7 left). The simulation of the mobilities while using the properties of the pathologic tissues is not sufficient to reproduce a cystocele by simulation (Figure 7 right).

It is often considered that the cystocele is due to an elongation of ligaments, induced by micro damages cumulated through the years. We progressively augment the length of the uterosacral ligament in order to get, on the biomechanical model, the same level of displacement that are observed for the cervix on pathological data. We obtain an increasing of the length of about 10%. However, we notice that the mobilities are not sufficient to re-create, by simulation, a cystocele. We then applied successively the 10% augmentation of the length for all ligaments (especially the umbilical ligament in the case of the bladder) and even a complete rupture of umbilical ligament but the obtained displacements were not sufficient to be considered as pathologic.

In fact, an other factor of cystocele is the degradation of the adhesions of the bladder on the pubis, that are considered as perfect in healthy situations. The simulation presented on the right of figure 8, is obtained by combining a 10% elongation of the ligaments and the degradation of the mechanical properties of the fascias between the bladder and the public bone. The perfect adhesion is replaced by a friction contact law combined with an elastic link with a low stiffness. The obtained mobilities for the cervix and the bladder now corresponds to the ones that can be observed during a cystocele.

Discussion: The successive degradation of the biomechanical model of the pelvic system mobilities allows to switch from healthy mobilities to pathological mobilities. The simulation reinforce the hypothesis of a multi factorial origin for the pathology for cystocele: it seems to be due to a change of traumatic elongation of the ligaments and to a degradation of the adhesive fascias. Of course, the impact of these results is limited by the fact that we can not validate the simulations of the pathology by ground truth data. Indeed, these data do not exist as we have created artificially a pathology starting from a healthy set of data !

4 Conclusion and Perspective

This paper presents a simulation tool of the pelvic system that is based on characterization of the tissues, modeling the boundary conditions and providing Finite Element approach. The main contribution is to provide an assessment, thanks to numerical experiments, to the hypothesis made by the medical community on the origin of the cystocele. In future work, we plan to work on patient-specific models based on a fusion of 3D images and 2D dynamic data with our simulation. From pathologic mobilities observed in the images, we will try to identify the main factors that induce the pathologic mobilities and adapt the parameters of our simulation accordingly. Then, we could then compare the results with experimental ground truth data collected during reparative surgery and validate this numerical simulation tool. If predictive, we could then use it for individual planning of surgery.

References

1. Samuelsson, E., Victor, F., Tibblin, G., Svardsudd, K.: Signs of genital prolapse in a Swedish population of women 20 to 59 years of age and possible related factors. *Am. J. Obstet. Gynecol.* 180, 299–305 (1999)
2. Swift, S.: The distribution of pelvis organ support in a population of female subjects seen for routine gynecologic health care. *Am. J. Obstet. Gynecol.* 183, 277–285 (2000)
3. Ettema, G.J., Goh, J.T., Forwood, M.R.: A new method to measure elastic properties of plastic-viscoelastic connective tissue. *Med. Eng. Phys.* 20, 308–314 (1998)
4. Goh, J.T.: Biomechanical properties of prolapsed vaginal tissue in pre and postmenopausal women. *Int. Urogynecol. J. Pelvic. Floor Dysfunct.* 13, 76–79 (2002)
5. Gilchrist, A.S., Gupta, A., Eberhart, R.C., Zimmern, P.E.: Biomechanical properties of anterior vaginal wall prolapse tissue predict outcome of surgical repair. *J. Urol.* 183(3), 1069–1073 (2010)
6. Chen, Z.W., Joli, P., Feng, Z.-Q.: Finite element modeling of interactions between pelvic organs due to pressure. *CSMA* (2011)
7. Krywonos, J., et al.: MRI image-based FE modelling of the pelvis system and bladder filling. *Comput. Method. Biomec.* 13(6), 669–676 (2010)
8. Balakrishna, H., et al.: PelvicSim-A Computational-Experimental System for Biomechanical Evaluation of Female Pelvic Floor Organ Disorders and Associated Minimally Invasive Interventions. *Stud. Health Technol. Inform.* 119, 182 (2005)
9. Rao, G.V., Rubod, C., Brieu, M., Bhatnagar, N., Cosson, M.: Experiments and finite element modeling for the study of prolapsed in the pelvic system. *Comput. Method. Biomec.* 13(3), 349–357 (2010)
10. Rubod, C., Brieu, M., Cosson, M., Rivaux, G., Clay, J.C., Gabriel, B.: Biomechanical properties of human pelvic organs. *Urol.* 79(4), 1346–1354 (2012)
11. Vallet, A., Witz, J.F., Brieu, M., Rubod, C., Cosson, M.: Simulation of pelvic mobility: topology optimization of ligamentous system. *Comput. Method. Biomec.* 14(1), 159–163 (2011)
12. Kamina, P.: *Clinical Anatomy*, book number 4, Maloine edn. (2008)
13. Faure, F., et al.: SOFA: A Multi-Model Framework for Interactive Physical Simulation. *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, 283–321 (2012), <http://www.sofa-framework.org>
14. Baronne, R.: *Comparative anatomy of domestic mammals*, Vigot edn. (2009)

Constructive Real Time Feedback for a Temporal Bone Simulator

Yun Zhou¹, James Bailey¹, Ioanna Ioannou², Sudanthi Wijewickrema²,
Gregor Kennedy³, and Stephen O'Leary²

¹ Department of Computing and Information Systems, University of Melbourne

² Department of Otolaryngology, University of Melbourne

³ Centre for the Study of Higher Education, University of Melbourne

Abstract. As demands on surgical training efficiency increase, there is a stronger need for computer assisted surgical training systems. The ability to provide automated performance feedback and assessment is a critical aspect of such systems. The development of feedback and assessment models will allow the use of surgical simulators as self-guided training systems that act like expert trainers and guide trainees towards improved performance. This paper presents an approach based on Random Forest models to analyse data recorded during surgery using a virtual reality temporal bone simulator and generate meaningful automated real-time performance feedback. The training dataset consisted of 27 temporal bone simulation runs composed of 16 expert runs provided by 7 different experts and 11 trainee runs provided by 6 trainees. We demonstrate how Random Forest models can be used to predict surgical expertise and deliver feedback that improves trainees' surgical technique. We illustrate the potential of the approach through a feasibility study.

Keywords: real time feedback, surgical simulation, random forest.

1 Introduction

Over the past two decades, a variety of virtual reality simulations have been developed for surgical training purposes, using novel techniques such as 3D illusion, haptic feedback and augmented reality. These advanced high fidelity simulations offer many potential benefits for surgical training, but also raise new challenges [7]. One potential benefit is the ability to use surgical simulators as self-guided learning tools, thus reducing the burden of work on surgical trainers. However, to realise this benefit, simulators must possess the ability to provide timely meaningful feedback to trainees, in order to facilitate effective learning through deliberate practice [2].

In minimally invasive surgery (MIS), [3,9] have applied data mining techniques to evaluate surgical processes or identify surgical gestures. This type of surgery typically requires surgeons to manipulate a set of tools in a prescribed way, and there are identifiable “gestures” associated with correct surgical technique. On the other hand, open surgery such as temporal bone surgery often utilises a

small instrument set of surgical drills and suction devices and there are many ways to achieve a correct outcome. As such, it is difficult to identify specific gestures that represent good surgical technique. Furthermore, the time frame of the analysis is typically longer than in MIS tasks, thus increasing the complexity of identifying underlying motion patterns. Thus, evaluating performance in open surgery simulators is a challenging task.

Most existing work [8,6] on automated performance evaluation in open surgery simulators is limited to assessment of surgical outcomes. Work on the provision of online feedback is still in its infancy. One such work is [8], where users are provided with an evaluation console allowing review of their performance based on surgical motion metrics. Interactive feedback took the form of coloured voxels indicating whether the correct region was drilled. While this type of feedback provides some guidance to achieve the correct surgical outcome, it provides no assistance in improving surgical technique, which can be equally important.

We introduce a method based on Random Forests (RF) [1] to design and deliver online technique feedback within a temporal bone surgical simulator, which can be generalised to other types of open surgery. First, a RF model is built from drilled region data to predict expertise. During a simulator task, if this model predicts that a user is a trainee, a second model combining RF and nearest neighbour search is used to generate human understandable feedback where necessary. The RF model used to generate such feedback is based on surgical stroke data, such as stroke force and length. We evaluated the RF approach against a baseline and our experimental results suggest that RF is a robust technique suitable for expertise classification and feedback delivery during an ongoing temporal bone surgical simulation.

In summary, the paper makes the following contributions: 1) We present the first virtual surgery system which can provide automatic real time feedback to improve surgical technique. 2) The first use of random forest classifiers in virtual simulations as the basis for providing feedback to users.

The remainder of the paper is organised as follows. We first introduce the simulator dataset that was used to train the RF feedback models. We proceed to explain how the two RF models are used to assess expertise and provide feedback. Finally, we define two evaluation metrics to measure the quality of the feedback and present the results of our experiments.

2 Method

Simulation Metrics. Training data was collected using the University of Melbourne temporal bone simulator [5]. This simulator displays a 3D temporal bone model based on segmented micro-CT data and provides haptic feedback through a Sensable PHANToM Desktop haptic device. The simulator can be used to perform any temporal bone drilling task and it records two kinds of performance measures at a sample rate of approximately 15 Hz: outcome measures and technique measures. Outcome measures consist of a time series of drilled voxel positions. Technique measures include motion-based metrics, simulator parameters

Table 1. Technique measures derived from the Temporal Bone Simulator

Motion-based	Simulator parameters	Proximity
stroke duration	drill burr size	distance to facial nerve
stroke distance	zoom level	distance to hearing bone
stroke speed		distance to membrane
stroke acceleration		distance to dura
stroke force		distance to sigmoid
stroke straightness		distance to tendon
stroke centroid distance		distance to round window
# bone drilled by stroke		

and proximity data as shown in Table 1¹. A k-cos [4] approach is used to segment drill trajectories into a series of surgical strokes for the calculation of motion-based metrics.

We collected 16 expert and 11 trainee temporal bone simulation runs. The data was provided by 7 different experts and 6 trainees. The training data was unevenly distributed due to limitations in the availability of trainees, but this does not affect the training of RF models significantly. Each simulator run consisted of three surgical tasks: cortical mastoidectomy, followed by posterior tympanotomy and cochleostomy. Cortical mastoidectomy is the preparatory step of many ear operations while posterior tympanotomy and cochleostomy are parts of cochlear implantation surgery.

Random Forest Based Feedback. The Random Forest [1] algorithm builds a strong classifier out of an ensemble of decision trees. A single decision tree (DT) is not a stable classifier, because a small change in the training set can significantly alter the tree structure. To overcome this drawback, RF creates a set of randomly selected subsets from the training data and each subset is used to build a DT. In each tree, nodes are also split using a random subset of all features (i.e. measures) in the data. Each tree classifies each data point (in our case, as expert or trainee), and RF uses the majority vote as the final prediction.

The first step of our approach was to predict the expertise of the surgeon. This step was vital to provide feedback appropriate to the surgeon’s expertise. We built a RF model to carry out this task. From a surgical point of view, drilled bone regions (i.e. voxels) are highly related to surgeon expertise so they are appealing to use as features to train a classifier. However, the bone volume contained a large number of voxels (9090750 to be precise) and not every voxel is related to expertise. Thus, it was necessary to employ a feature selection step: mutual information was used to extract the top 10% of voxels based on their capability to distinguish expertise. The RF tree was built using these voxels as

¹ The stroke force metric refers to the force (in Newtons) being generated by the haptic device motors in response to user interaction.

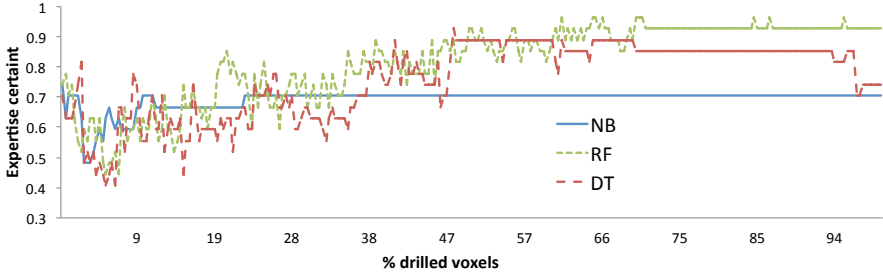


Fig. 1. Expertise classification certainty against surgical task progress. X-axis is the percentage of removed voxels. Y-axis is the expertise classification certainty.

features. We note that previous work by Sewell et al [8] used Naive Bayes(NB) to predict expertise. We chose RF over NB in our method, since the features (voxels) are not independent. If a voxel has been drilled, its neighbours are highly likely to have been drilled as well. Figure 1 illustrates expertise classification certainty using three methods. Prediction models were trained at multiple times during the surgical task, according to when a certain number of voxels were drilled. We expected that the accuracy of prediction models would increase as the number of drilled voxels increased. Figure 1 shows that we can predict the expertise of a user with increasing certainty, by using more information about the voxels drilled so far. By deploying a model that considers the locations of the first 37% (approximately) of voxels drilled, we can be around 80% certain of whether the user is an expert or trainee. In the end, RF misclassified only 1 out of 27 simulation runs. A single DT was generally better than NB, but it is unstable, since we see a large decrease in certainty, even near the end of the task. Overall, we can see that drilled voxel measures provided a very good prediction of expertise. Once we are 80% confident about the trainee’s expertise, we can start delivering real time feedback to improve their performance. We note that while drilled voxels provided good expertise classification, they could not provide useful guidance on improving surgical technique. Therefore we used the surgical technique measures shown in Table 1 to create a second RF model for the purposes of feedback generation.

Human trainers often suggest ways to improve surgical technique as part of their feedback. We attempt to mimic this interaction by making suggestions to change one technique feature at a time (e.g. “use longer strokes”). Therefore we begin by selecting the feature on which to provide feedback. A naive way to do this is to select the feature possessing the highest association with expert technique. Work in [1] proposed a way to compute feature importance by randomly shuffling the values of each feature across the dataset. Feature importance is defined as $imp(f) = err(d_f) - err(d)$, where d_f denotes the data with shuffled values for feature f , d is the original dataset, and err denotes the classification error. We computed $imp(f)$ for each feature and chose the feature of highest importance as our global feature. This feature was used as the baseline (naive)

Input: si = new stroke, eg = expert stroke group, dist = distance function

Output: fn

```

1 es = nearest(si, eg, dist);
2 F = feature vote array;
3 for each tree in forest do
4   l1=classify si; l2=classify es;
5   if l1 is trainee and l2 is expert then
6     f = feature id at which si and es go to different branches;
7     if es[f] > si[f] then F[f+];
8     else F[f-];
9   end
10 end
11 fn = maxIndex(F);

```

Algorithm 1. Random Forest feedback algorithm

feedback in our feasibility study. This naive approach can provide basic feedback to improve one feature towards expertise, but other features may be just as important at different times during the surgical task. Therefore we propose a dynamic way to deliver feedback using a joint RF model and nearest neighbour approach, outlined in Algorithm 1. The algorithm begins when we identify a user as a trainee using the first RF model introduced above. Once the user has performed a stroke, the algorithm identifies the most similar expert stroke (from a historical database) using a nearest neighbour strategy. Instead of using Euclidean distance, we use the distance function derived from RF [1]. The expert stroke serves as a reference for delivering feedback. In order to choose the specific feedback feature, the user stroke and the reference stroke are classified by each tree in the RF feedback model. In a given tree, provided both strokes have been classified correctly, we compute the first feature on which the strokes are split into different branches and this feature receives one vote. Once a feature in a given tree is chosen, we calculate the degree of change (in terms of magnitude and direction) on that feature between the user stroke and the reference expert stroke. As we iterate through the forest, we store the votes for each feature in each direction in an array F . The size of F is twice the number of features, since we count votes for increase (f^+) and decrease (f^-) separately.

Figure 2 shows a running example. This forest contains four decision trees. The dashed cyan line indicates the path of an expert stroke while the solid magenta line is the path of a trainee stroke. Expert leaves are dashed cyan buckets while trainee leaves are magenta solid buckets. In the first tree, the trainee and expert were classified correctly and the split was at the zoom feature. Experts used a lower zoom value, so the vote for $zoom^-$ increased by one. In the second and fourth trees, the split feature suggested a decrease in force, so $force^-$ received two votes. In the third tree both strokes were classified incorrectly, so this tree was ignored. The final feedback chosen would be to “decrease force”.

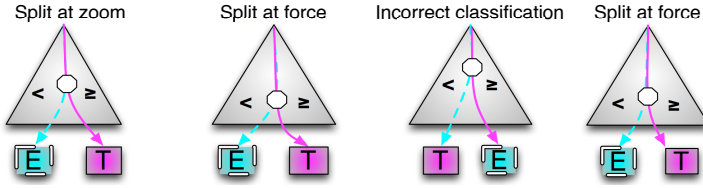


Fig. 2. Example of voting in a forest with four trees (E=expert, T=trainee)

3 Feasibility Study and Results

Comprehensively evaluating the quality of feedback is difficult and would require prospective methods, such as a randomised controlled trial. A controlled trial is beyond the scope of this paper, as the aim is to present the RF-based feedback method and evaluate its practical potential. To evaluate our methods, we conducted a feasibility study based on two metrics we designed to assess feedback quality.

The first metric is *recovery rate*. We generated a “synthetic trainee” as follows: 1) Randomly select a stroke made by an expert. 2) Randomly select a feature from the selected expert stroke. 3) Randomly change the value of this feature. 4) If the altered stroke is classified as trainee by the RF model, go to step 5, otherwise discard the stroke and go to step 2. 5) Label the altered stroke as a “synthetic trainee stroke”. Then we input this synthetic trainee stroke to the RF feedback model, and if the suggested feedback corresponded to the modification we made, the number of correct suggestions was increased by one. We repeated this process 10 times on each expert stroke in the dataset and calculated the recovery rate as $\frac{\# \text{ correct suggestions}}{\# \text{ expert strokes}}$.

Table 2 presents an example of recovery rate computation using both the baseline and RF approaches. Suppose we created 5 synthetic trainee strokes by changing the features listed in the first column. We assume that force is the global feature used as the baseline, so the baseline will only make suggestions about force. The only correct suggestion by the baseline is for the first stroke, hence the recovery rate is 1/5. On the other hand, RF feedback makes 4 correct suggestions so the recovery rate is 4/5.

The second method to assess our RF model is to take each trainee stroke, apply the suggested feedback (i.e. manipulate the stroke’s characteristics according to the feedback), and then determine whether the altered stroke is classified as ‘expert’ by the RF model. We call this metric *promotion rate* and it is equal to $\frac{\# \text{ trees which classify stroke as expert in RF}}{\# \text{ total trees in RF}}$.

To calculate the above measures for our data set, each simulator run was first divided into five stages or sub-tasks. These stages have different surgical goals and characteristics, so we created separate RF models for each stage. We set the number of trees in each RF to 500, which is large enough to tolerate the noise caused by variability in surgical performance. All experiments were

Table 2. Example of recovery rate computation

feature change	baseline	RF feedback
force ⁺	force ⁺	speed ⁺
speed ⁻	force ⁻	speed ⁻
zoom ⁻	force ⁺	zoom ⁻
zoom ⁺	force ⁺	zoom ⁺
stroke length ⁺	force ⁻	stroke length ⁺

conducted using a ten-fold cross validation scheme. In each fold we used 24 runs as the training set and the remaining 3 runs as the test set. For the recovery rate calculation, we changed the value of features by a random percentage ranging from 10% to 50%. For the promotion rate calculation we changed the value of the proposed feature by 10% to 200% in the suggested direction of the feedback. We used a range of value changes to reflect the real life situation, where a trainee is unlikely to be able to achieve the exact suggested correction.

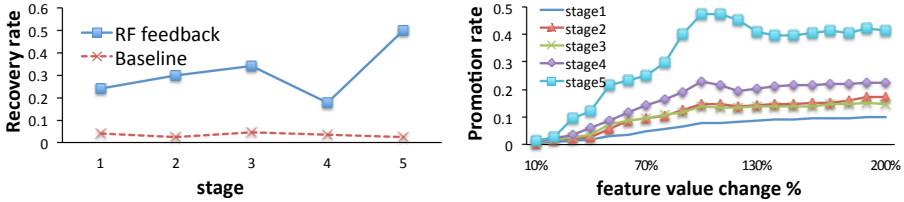


Fig. 3. Average recovery rate and promotion rate across 5 stages. Curve colour is significant.

Figure 3 shows the results. Both rates are expected to increase across the stages, since later stages involved a more restricted surgical work area with less freedom of movement. As shown in Figure 3, the later the stage, the higher the rate for both metrics. This suggests that there were more easily identifiable differences between experts and trainees in later stages. The recovery rate for stage 4 is an outlier and requires further investigation. One possible explanation is that the trainees in our dataset were more skilled in this stage.

RF feedback achieved significantly higher recovery rates than the baseline (using 95% significance t-test). The peak point at stage 5 appears at 100% change, which is consistent with our expectations since it is the exact suggested correction. RF feedback achieved 50% in stage 5 for both rates. This is a good result because it shows that stroke technique improved considerably by making just one feature correction. However, the promotion rate appears to taper off after 120% change, suggesting that change in a single feature has limited potential to improve overall technique. This is unlikely to be a serious problem, since our approach is not limited to providing feedback on only one feature. In a real simulated training situation, the model could provide a series of suggestions, gradually guiding the trainee towards expertise.

4 Discussion and Conclusion

We have presented a method to automatically deliver online constructive feedback on surgical technique within a temporal bone surgical simulation. This approach is generalisable to other types of open surgery simulation. Our evaluation showed that the RF based approach is effective at classifying expertise and outperformed the baseline in feedback quality measures. The measures of recovery rate and promotion rate demonstrated the feasibility of this approach. Further work including controlled trials is needed to evaluate the feedback system in situ. In addition, future work will also focus on automatic approaches to improve promotion rate. One possible direction might be to investigate correlations between metrics when generating and responding to feedback. In general, there remain intriguing open questions regarding automated feedback in simulation-based training, such as when to provide feedback and how to provide it (such as auditory, visual or haptic).

References

1. Breiman, L., Schapire, E.: Random forests. *Mach. Learn.* 45(1), 5–32 (2001)
2. Ericsson, K.A.: Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad. Med.* 79(10), S70–S81 (2004)
3. Forestier, G., Lalys, F., Riffaud, L., Trelhu, B., Jannin, P.: Classification of surgical processes using dynamic time warping. *J. Biomed. Inform.* 45(2), 255–264 (2012)
4. Hall, R., Rathod, H., Maiorca, M., Ioannou, I., Kazmierczak, E., O’Leary, S., Harris, P.: Towards haptic performance analysis using K-metrics. In: Pirhonen, A., Brewster, S. (eds.) HAID 2008. LNCS, vol. 5270, pp. 50–59. Springer, Heidelberg (2008)
5. Kennedy, G., Ioannou, I., Zhou, Y., Bailey, J., O’Leary, S.: Mining interactions in immersive learning environments for real-time student feedback. *Australas J. Educ. Tec.* 29(2) (2013)
6. Kerwin, T., Wiet, G., Stredney, D., Shen, H.W.: Automatic scoring of virtual mastoidectomies using expert examples. *IJCARS* 7(1), 1–11 (2012)
7. Kneebone, R.: Simulation in surgical training: educational issues and practical implications. *Med. Educ.* 37(3), 267–277 (2003)
8. Sewell, C., Morris, D., Blevins, N., Dutta, S., Agrawal, S., Barbagli, F., Salisbury, K.: Providing metrics and performance feedback in a surgical simulator. *Comput. Aided Surg.* 13(2), 63–81 (2008)
9. Stylopoulos, N., Cotin, S., Maithel, S., Ottensmeyer, M., Jackson, P., Bardsley, R., Neumann, P., Rattner, D., Dawson, S.: Computer-enhanced laparoscopic training system (CELTS): bridging the gap. *Surg. Endosc.* 18(5), 782–789 (2004)

Lattice Boltzmann Method for Fast Patient-Specific Simulation of Liver Tumor Ablation from CT Images

Chloé Audigier^{1,2}, Tommaso Mansi², Hervé Delingette¹, Saikiran Rapaka², Viorel Mihalef², Puneet Sharma², Daniel Carnegie⁴, Emad Boctor³, Michael Choti⁴, Ali Kamen², Dorin Comaniciu², and Nicholas Ayache¹

¹ INRIA Sophia-Antipolis, Asclepios Research Group, Sophia-Antipolis, France

² Siemens Corporation, Corporate Research and Technology, Imaging and Computer Vision, Princeton, NJ, USA

³ Dept. of Radiology, Johns Hopkins Medical Institutions, Baltimore, MD, USA

⁴ Dept. of Surgery, Johns Hopkins Medical Institutions, Baltimore, MD, USA

Abstract. Radio-frequency ablation (RFA), the most widely used minimally invasive ablative therapy of liver cancer, is challenged by a lack of patient-specific planning. In particular, the presence of blood vessels and time-varying thermal diffusivity makes the prediction of the extent of the ablated tissue difficult. This may result in incomplete treatments and increased risk of recurrence. We propose a new model of the physical mechanisms involved in RFA of abdominal tumors based on Lattice Boltzmann Method to predict the extent of ablation given the probe location and the biological parameters. Our method relies on patient images, from which level set representations of liver geometry, tumor shape and vessels are extracted. Then a computational model of heat diffusion, cellular necrosis and blood flow through vessels and liver is solved to estimate the extent of ablated tissue. After quantitative verifications against an analytical solution, we apply our framework to 5 patients datasets which include pre- and post-operative CT images, yielding promising correlation between predicted and actual ablation extent (mean point to mesh errors of 8.7 mm). Implemented on graphics processing units, our method may enable RFA planning in clinical settings as it leads to near real-time computation: 1 minute of ablation is simulated in 1.14 minutes, which is almost 60× faster than standard finite element method.

1 Introduction

In spite of recent advances in cancer therapy, treatment of primary and metastatic tumors of the abdomen, including the liver, remains a significant challenge. Hepatocellular carcinoma (HCC) for example is one of the most common malignancies (more than 1 million cases per year), with increasing frequency in Western countries [1]. Unfortunately, less than 25% of patients with primary or secondary liver cancer are candidates for resection or transplantation, which are considered as the most effective treatments. Consequently, ablative therapies such as radio-frequency ablation (RFA) has raised increasing interest.

RFA consists in placing a probe within the target area. Electrodes at the tip of the probe create heat, which is conducted into the surrounding tissue, causing coagulative necrosis at temperatures above 50°C . Success of the procedure depends on the complete coverage of the tumor by the generated necrosis area, which relies on optimal probe placements and the extend of heat delivery. However, the latter is challenged by the hepatic blood vessels that dissipate heat, thus potentially reducing RFA efficiency and increasing risks of recurrence.

Several studies [2,3,4] have investigated finite element method (FEM) to compute heat diffusion in liver and predict the optimal placement of the RFA probes. Heat sink as well as various cellular necrosis models [5] have been studied. However, to the best of our knowledge, none of these models rely on patient-specific data. In particular, the vascular system of the liver is neglected or simplified and blood flow circulation is not computed based on patient-specific clinical information. Moreover, FEM is computationally demanding (execution time is in the range of hours), which is not suitable for clinical purposes.

This paper presents a multi-physics model for efficient patient-specific planning of RFA based on medical images such as CT or MRI (Sec. 2). In particular, we rely on the Lattice Boltzmann Method (LBM) to compute heat diffusion in the liver tissue. The LBM offers high scalability, second order accuracy in space and the simplicity of implementation on a uniform Cartesian grid [6]. In Sec. 3, we demonstrate the validity of our algorithm against an analytical solution and its predictive power is evaluated on patient data. Sec. 4 concludes the paper.

2 Method

Fig. 1 illustrates the different steps of our method. Starting from a preoperative clinical CT image, we extract the liver geometry and the venous systems. Next, the bioheat equation is solved using LBM. The liver is highly vascularized and modeling the impact of all vessels is out of reach. Therefore, only the effect of large vessels are described explicitly with the Pennes model [7]. Small ones are represented implicitly in the parenchyma as a porous medium using the Wulff and Klinger model [8]. The bioheat equation is weakly coupled to a computational fluid dynamics (CFD) solver to accurately take into account the effect of blood circulation on the dissipated heat, while the blood flow in the porous tissue is computed by solving the Darcy's equation [9]. The heat transfer depends on the blood flow, which is not modified as the organ is heated (the effect of

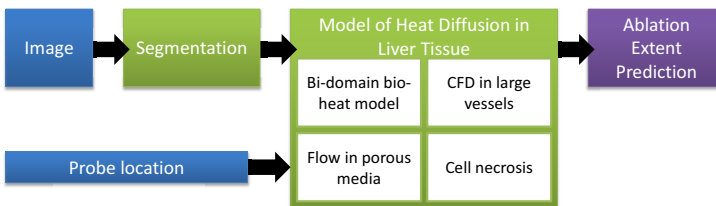


Fig. 1. Steps of the proposed method (blue: input, green: processes, purple: output)

heat on the viscosity of the flow is neglected as well as the coagulation effect). This assumption allowed us to compute CFD and porous flow only once, at the beginning of the algorithm, thus speeding up the process. LBM and CFD are calculated on a Cartesian grid while the porous solver is executed using FEM on a linear multi-domain tetrahedral mesh for increased accuracy. The resulting flow is tri-linearly rasterized on the Cartesian grid after computation. Finally, a cellular necrosis model is employed to compute cell death due to overheating [5].

2.1 Model of Patient Liver Anatomy

Preoperative images are semi-automatically segmented, yielding a detailed anatomical model of patient's liver, including parenchyma, tumors, hepatic veins, vena cava, and portal vein (Fig. 2, left panel). For each structure, we define seeds inside and outside of the area of interest. Then, the random-walker algorithm is employed [10] to automatically estimate the boundaries of the structure. The process can be refined interactively. From the segmentation, a level set representation of the liver, without tumor and vessels is computed. A multi-label mask image is also created to identify the structures of interest for the simulation. Finally, a tetrahedral multi-domain mesh is generated based on the resulting multi-label mesh (www.cgal.org) for computing the porous flow.

2.2 Model of Heat Transfer in Liver Tissue

Computing heat diffusion in biological tissues amounts to solving the coupled bioheat equations derived from the theory of porous media, where each elementary volume is assumed to comprise both tissue and blood with a certain fraction. As current imaging techniques do not allow to estimate the accurate ratio between blood and liver tissue, two main simplifications of the bioheat equations have been proposed. The *Pennes model* [7] assumes constant blood temperature, which holds close to large vessels, where blood velocity is high. The model writes:

$$(1 - \epsilon)\rho_t c_t \frac{\partial T}{\partial t} = (1 - \epsilon)Q + (1 - \epsilon)\nabla \cdot (d_t \nabla T) + H(T_{b0} - T) \quad (1)$$

For small vessels, the *Wulff-Klinger (WK) model* [8] assumes equilibrium between tissue and blood temperatures, with a blood volume fraction $\epsilon \ll 1$:

$$(1 - \epsilon)\rho_t c_t \frac{\partial T}{\partial t} = (1 - \epsilon)Q + (1 - \epsilon)\nabla \cdot (d_t \nabla T) - \epsilon\rho_b c_b \mathbf{v} \cdot \nabla T \quad (2)$$

In both equations, T , Q , \mathbf{v} and T_{b0} stand for temperature, source term, blood velocity and the mean temperature (assumed constant) of the blood in large vessels. The other parameters are listed in Table 1. In our framework, we use either the Pennes model or the WK model according to the spatial location in the anatomy. Assuming that blood vessels and surrounding tissue are isolated from each other, we compute the temperature by solving the diffusion equation:

$$\rho_t c_t \frac{\partial T}{\partial t} = Q + \nabla \cdot (d_t \nabla T) \quad (3)$$

Table 1. Values from literature [9] of the parameters used in the simulation

parameter	description	value
ρ_b, ρ_t	blood and tissue densities	$1.06 \times 10^3 \text{ kg m}^{-3}$
c_b	blood heat capacity	$4.18 \times 10^3 \text{ J(kg K)}^{-1}$
c_t	tissue heat capacity	$3.6 \times 10^3 \text{ J(kg K)}^{-1}$
c_t^*	tissue heat capacity in dead cells	$0.67 \times 10^3 \text{ J(kg K)}^{-1}$
d_b, d_t	blood and tissue heat conductivities	$0.512 \times (1 + 0.00161 \times (T - 310)) \text{ W(m K)}^{-1}$
H	convective transfer coefficient	$24.4 \times 10^4 \text{ W (m}^3 \text{ K)}^{-1}$
ϵ	blood volume fraction	0.1
κ	permeability	$4.0 \times 10^{-11} \text{ m}^2$
μ	dynamic viscosity of the blood	0.0035 Pa s
φ_{vcin}	vena cava inflow	2.0 L min^{-1}
φ_i	flow through the inlets of the hepatic veins	1.6 L min^{-1}
p_0	vena cava outlet pressure	3 mmHg
k_f	forward rate constant	$3.33 \times 10^{-3} \text{ s}^{-1}$
k_b	backward rate constant	$7.77 \times 10^{-3} \text{ s}^{-1}$
T_k	parameter of cell state model	40.5°C

everywhere in the domain, to which we add the cooling term $H(T_{b0} - T)/(1 - \epsilon)$ when a point belongs to a large vessel (Pennes model) or $-\epsilon\rho_b c_b \mathbf{v} \cdot \nabla T/(1 - \epsilon)$ when it belongs to the parenchyma (WK model).

2.3 Model of the Patient Hepatic Venous Circulation System

Blood velocity \mathbf{v} inside the parenchyma is calculated according to Darcy’s law : $\mathbf{v} = -\kappa/(\mu\epsilon^{2/3})\nabla p$ where p is the pressure. This amounts to solving the Laplace equation : $\nabla \cdot (-\kappa/(\mu\epsilon^{2/3})\nabla p) = 0$. At the border of the liver, Neumann boundary conditions are employed. Dirichlet boundary conditions are applied at the tip of the portal and hepatic veins, to define the pressure drop between them. As we cannot estimate these pressures *in-vivo*, we rely on a CFD model of the hepatic venous circulation system (Fig. 2, right panel). We used a full 3D CFD solver (incompressible Navier-Stokes equations with viscous terms, expressed in an Eulerian framework which embeds the domain boundary using a level set representation of the segmented vessels [11]). The blood is modeled as a Newtonian fluid with pre-specified density ρ_b and viscosity μ . Let φ_p be the portal

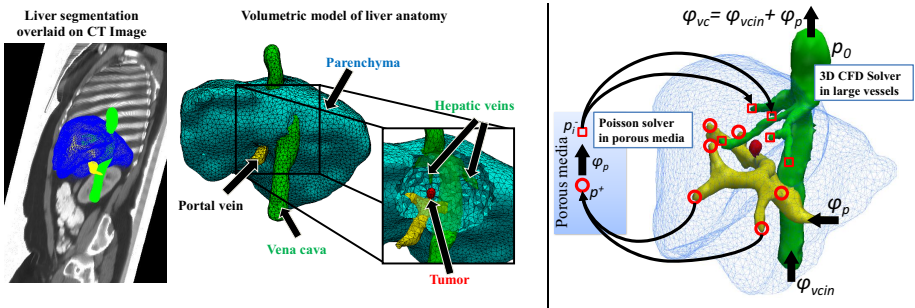


Fig. 2. Left: A detailed anatomical model of the liver is estimated from a standard clinical CT image. Right: Model of the hepatic circulatory system. Arrows denote blood flow. Circles and squares denote portal and hepatic vessel tips. See text for details.

vein inflow and $\varphi_{vc} = \varphi_{vc_{in}} + \varphi_p$ the vena cava outflow (conservation of mass, the hepatic artery is neglected). We also fix the vena cava outlet pressure in the range of physiological values of healthy patients. The values are listed in Table 1.

Blood flow and pressure distribution within the vena cava and hepatic veins: A plug profile velocity field is applied at the inlets (squares in Fig. 2), computed from the outflow φ_p and each inlet cross-sectional area. The CFD calculation give the downstream pressures p_i^- for each inlet of the hepatic vein.

Blood flow and pressure distribution within the portal vein: To estimate the upstream pressure p^+ , assumed constant, of the portal vein outlets (circles in Fig. 2), we solve the Darcy’s law and optimize over p^+ such that the computed perfused flow through the hepatic vein inlets matches the one computed using 3D CFD. Once p^+ is estimated, we compute the blood flow using the CFD solver.

2.4 Model of Cellular Necrosis

Tissue necrosis is computed based on the simulated temperatures using a 3-state model [5]. The model computes the variation of concentration of alive (A), vulnerable (V) and dead (D) cells over time according to the state equation :

$$A \begin{array}{c} \xrightarrow{k_f(T)} \\ \xleftarrow{k_b} \end{array} V \xrightarrow{k_f(T)} D \tag{4}$$

where $k_f(T) = \bar{k}_f e^{T/T_k} (1 - A)$ and k_b are the rates of cell damage and recovery respectively. This equation results in 3 coupled ODEs solved with a first order explicit scheme at each vertex of the Cartesian grid, yielding a spatially-varying cell state field used in the bioheat solver to update the heat capacity during the computation (Table 1). The initial conditions are chosen as in [5].

2.5 Lattice Boltzmann Formulation of the Bioheat Equations

The bioheat model is solved on an isotropic Cartesian grid using LBM with 7-connectivity topology and Neumann boundary conditions. For stability reason, we use a Multiple-Relaxation-Time model. The boundaries are treated according to the level set representation using linear interpolation without requiring advanced meshing techniques. The governing equation at position $\mathbf{x} = (x, y, z)$ for the edge \mathbf{e}_i is given by (5). $\mathbf{f}(\mathbf{x}) = \{f_i(\mathbf{x})\}_{i=1..7}$ is the vector of distribution function with $f_i(\mathbf{x})$, the probability of finding a particle travelling along the edge \mathbf{e}_i of the node \mathbf{x} at a given time, $c = \Delta x/\Delta t$, $c_s^2 = 1/4$, Δx is the spacing.

$$\mathbf{f}(\mathbf{x} + \mathbf{e}_i \Delta x, t + \Delta t) = \mathbf{f}(\mathbf{x}, t) + \mathbf{A}[\mathbf{f}^{eq}(\mathbf{x}, t) - \mathbf{f}(\mathbf{x}, t)] + \boldsymbol{\omega} \Delta t H(T_{b0} - T(\mathbf{x}, t)) \tag{5}$$

$f_i^{eq}(\mathbf{x}, t) = \omega_i T(\mathbf{x}, t) [1 + \frac{\mathbf{e}_i \cdot \mathbf{V}}{cc_s^2}]$ and $\boldsymbol{\omega} = \{\omega_i\}_{i=1..7}$ the vector of weighting factors [6]. The temperature is computed as $T(\mathbf{x}, t) = \sum_{i=1}^7 f_i(\mathbf{x}, t)$ and is updated at every node of the grid for every timestep. Finally, we model the heat source term through a Dirichlet boundary condition at the location of the probe.

3 Experiments and Results

All experiments were executed on a Windows 7 desktop machine (Intel Xeon, 2.80 GHz, 45GB RAM, 24 CPUs) with a Nvidia Quadro 6000 1.7 GB (448 CUDA cores).

3.1 Quantitative Verification against Analytical Solution

To evaluate our model, we compared its behavior on a regular cuboid domain with the 3D analytical solution of a source of mass M released at \mathbf{x}_0 at time t_0 :

$$T(\mathbf{x}, t) = \frac{M}{[4\pi(t - t_0)D]^{3/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_0 - (t - t_0)\mathbf{v}\|^2}{4D(t - t_0)}\right)$$

of the advection-diffusion equation : $\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \nabla \cdot (D\nabla T)$. Parameters were chosen to get heat diffusion in physiological range: $D = 0.1 \text{ mm}^2/\text{s}$, $\mathbf{v} = (2, 0, 0) \text{ mm}/\text{s}$, $M = 35000 \text{ }^\circ\text{C}/\text{mm}^3$, $t_0 = -50 \text{ s}$, yielding a Gaussian-shape source term of 70°C at the center at time $t = 0$. In our LBM solver, we initialized the temperature values at each point with the analytical solution at time $t = 0$. The temperature at several points of the domain was reported. Our implementation was qualitatively close to the analytical solution (Fig. 3). For a given resolution, an upper and lower bound for the time-step were provided by the simulated physics and the Courant-Friedrichs-Lewy conditions. As expected, the smaller the spatial resolution, the more accurate the solution. A time-step of 75 ms and a resolution of $1 - 2 \text{ mm}$ appeared to be a good compromise between accuracy and computational cost. From a computational point of view, experiments showed a speed-up of 11 with parallel optimization (OpenMP) and 45 with graphical processing units (GPU) implemented on CUDA with respect to a single-core implementation of LBM. After a quantitative verification of the FEM simulation against the analytical solution, experiments showed that a $60\times$ speed-up was obtained with respect to FEM for a similar accuracy.

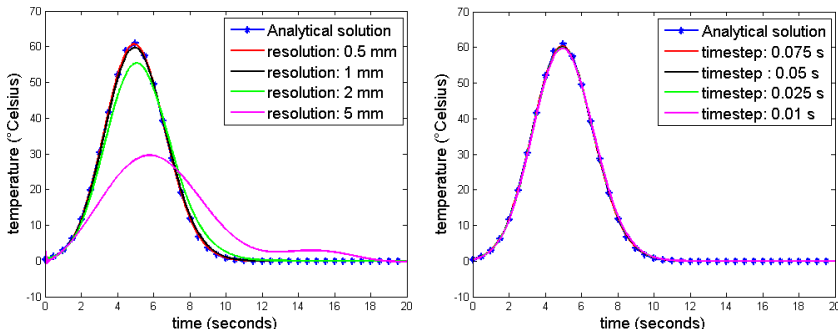


Fig. 3. *Left:* Spatial convergence analysis for a fixed time-step of 0.01 s. *Right:* Time convergence analysis for a resolution of 1 mm compared to the analytical solution. As one can see, the proposed framework quickly converges to the right solution.

3.2 Evaluation on Patient Data

We evaluated our model on 5 patients, with 7 ablations (patient 5 had 3 tumors ablated) for whom pre- and post-operative CT images were available. For all patients, nominal tissue parameters were employed. Clinical RFA protocol requires that the probe is deployed within the tumor with a diameter defined pre-operatively according to the size of the tumor, and then maintained for 7 minutes after the target temperature of 105°C was reached, as measured by the probe thermistors. For large tumors, the process was iterated with sequentially increasing diameters. After anatomical model extraction, we emulated the RFA protocol by placing the virtual probe at the center of the tumor. Cells around the probe tip within the probe diameter sphere were heated at 105°C during 7 minutes or 2 times 7 minutes. The simulation continued for 3 more minutes without the probe so that each cell reach a steady state. Qualitatively, computed ablation followed closely the boundaries of the vessels, due to the heat sink effects of the blood. The shape of the ablated area also depended on the heat advection due to the small arteries (Fig. 4). Cell death area computed using the model compared qualitatively well with the observed postoperative necrosis zone (Fig. 4, the lesion was manually segmented by an expert and rigidly registered to preoperative image). Quantitatively, average point-to-mesh errors (Table 2) were within clinical variability as they were lower than the different size configurations of the probes. More importantly, the simulation predicted that the selected protocols covered the entire lesion, which is the clinical criterion for ablation planning. One minute is computed in 1.14 minutes, in comparison, using FEM it takes 1.14 hours. To the best of our knowledge, this is the first time that near real-time simulations of RFA ablation could be achieved. All cases presented a larger necrosis area compared to the ground truth, the diffusion coefficient used from the literature was too high to get a perfect match.

4 Discussion and Conclusion

We have presented a first patient-specific model of liver tumor ablation allowing near real-time computation. As we rely on LBM, our framework does not require advanced meshing techniques to solve bioheat equations and the level set representations of the structures are directly computed from images. We focused on

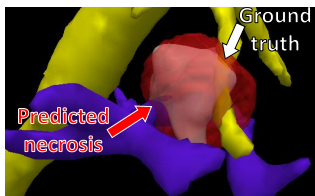


Fig. 4. Predicted necrosis compared qualitatively well with ground truth (patient 2)

Table 2. Evaluation on Patient Data

patient	point-to-mesh error	probe diameter
1	9.5 ± 5.9 mm	4 cm then 5 cm
2	4.6 ± 3.3 mm	3 cm
3	10.4 ± 6.5 mm	4 cm then 5 cm
4	11.7 ± 6.3 mm	4 cm then 5 cm
5-1	7.3 ± 5.1 mm	4 cm then 5 cm
5-2	8.2 ± 6.2 mm	4 cm then 5 cm
5-3	9.0 ± 5.5 mm	4 cm then 5 cm

modeling heat propagation and cell death based on a patient image taking into account the heat sink effect of blood vessels and porous circulation in the liver. Despite possible biases in establishing correspondences between the post- to the pre-operative images, and the use of nominal biological parameters, which are not patient-specific, our model provided promising results, opening new opportunities in the RFA planning and guidance, even if the target is to go beyond real-time, as we need 1-2 minutes of computation for clinical use. We have not considered the effects of the arterial flow, but it would be straightforward to include it for improved accuracy, provided the hepatic artery is visible in the image. It is worth noting that the veins account for more than 70 % of the blood flow [12]. Future works include validation on larger cohorts of patient, and a full coupling of blood flow and heat transfer models for more accurate predictions.

Acknowledgments. We thank Marta Mesa Gonzalez for segmenting all the cases.

References

1. El-Serag, H.B., Davila, J.A., Petersen, N.J., McGlynn, K.A.: The continuing increase in the incidence of hepatocellular carcinoma in the united states: An update. *Ann. Intern. Med.* 139, 817–823 (2003)
2. Chen, X., Saidel, G.M.: Mathematical modeling of thermal ablation in tissue surrounding a large vessel. *J. Biomech.* 131 (2009)
3. Jiang, Y., Mulier, S., Chong, W., Diel Rambo, M., Chen, F., Marchal, G., Ni, Y.: Formulation of 3D finite elements for hepatic radiofrequency ablation. *IJMIC* 9, 225–235 (2010)
4. Kröger, T., Pätz, T., Altrogge, I., Schenk, A., Lehmann, K., Frericks, B., Ritz, J., Peitgen, H., Preusser, T.: Fast estimation of the vascular cooling in RFA based on numerical simulation. *Open Biomed. Eng. J.* 4, 16–26 (2010)
5. O’Neill, D., Peng, T., Stiegler, P., Mayrhauser, U., Koestenbauer, S., Tscheliessnigg, K., Payne, S.: A three-state mathematical model of hyperthermic cell death. *Ann. Biomed. Eng.* 39, 570–579 (2011)
6. Rapaka, S., Mansi, T., Georgescu, B., Pop, M., Wright, G.A., Kamen, A., Comaniciu, D.: LBM-EP: Lattice-boltzmann method for fast cardiac electrophysiology simulation from 3D images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 33–40. Springer, Heidelberg (2012)
7. Pennes, H.H.: Analysis of tissue and arterial blood temperatures in the resting human forearm. *J. Appl. Physiol.* 85, 5–34 (1998)
8. Klinger, H.: Heat transfer in perfused biological tissue I: General theory. *B. Math. Biol.* 36, 403–415 (1974)
9. Payne, S., Flanagan, R., Pollari, M., Alhonnoro, T., Bost, C., O’Neill, D., Peng, T., Stiegler, P.: Image-based multi-scale modelling and validation of radio-frequency ablation in liver tumours. *Philos. T. Roy. Soc. A.* 369, 4233–4254 (2011)
10. Grady, L.: Random walks for image segmentation. *IEEE T. Pattern Anal. Mach. Intell.* 28, 1768–1783 (2006)
11. Ralovich, K., et al.: Hemodynamic assessment of pre- and post-operative aortic coarctation from MRI. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 486–493. Springer, Heidelberg (2012)
12. Schenk Jr., W.G., McDonald, J.C., McDonald, K., Drapanas, T.: Direct measurement of hepatic blood flow in surgical patients: with related observations on hepatic flow dynamics in experimental animals. *Ann. Surg.* 156, 463–469 (1962)

Registration of a Validated Mechanical Atlas of Middle Ear for Surgical Simulation

Guillaume Kazmitcheff^{1,2,*}, Christian Duriez¹, Mathieu Miroir²,
Yann Nguyen^{2,3}, Olivier Sterkers^{2,3},
Alexis Bozorg Grayeli^{2,3,4}, and Stéphane Cotin¹

¹ Shacra, Inria, Université Lille 1, Villeneuve d'Ascq, France

² UMR-S 867, Inserm, Université Paris 7 Denis Diderot, Paris, France

³ AP-HP, Hôpital Beaujon, Service ORL, Clichy, France

⁴ Present Address : Université de Médecine, Service ORL, Dijon, France

Abstract. This paper is centered on the development of a new training and rehearsal simulation system for middle ear surgery. First, we have developed and validated a mechanical atlas based on finite element method of the human middle ear. The atlas is based on a microMRI. Its mechanical behavior computed in real-time has been successfully validated. In addition, we propose a method for the registration of the mechanical atlas on patient imagery. The simulation can be used for a rehearsal surgery with the geometrical anatomy of a given patient and with mechanical data that are validated. Moreover, this process does not necessitate a complete re-built of the model.

Keywords: Simulation & training systems, Atlases, Head and neck.

1 Introduction

Learning surgery requires a large amount of practice to acquire experience, especially for complex tasks that involve knowledge or delicate gestures achievement. In this context, training simulations are crucial to preserve the patient safety during the learning process. However, to gain interesting experience from simulation, the simulated environment must be highly realistic. In this paper, we present a new approach based on a validated mechanical atlas that is adapted with patient data. We applied this approach for a simulator of middle ear surgery.

Middle ear surgery is a microsurgery of hearing rehabilitation for conductive loss. This surgery is particularly complex due to the high susceptibility to trauma and the sub-millimetric size of the anatomical structures, such as the ossicular chain (Fig. 1). Training on cadavers is the most realistic and appreciate solution but it is expensive and there is a risk of infection [1]. Synthetic models are mainly used for economical and safety reasons [2]. Some virtual simulators exist for ear surgery, like the Visible Ear [3] or the Voxel-Man (Voxel-Man Group, Hamburg, Germany), which are mostly centered on drilling tasks. The main limitations of

* The authors would like to thank Collin Ltd. (Bagneux, France) for financial support.

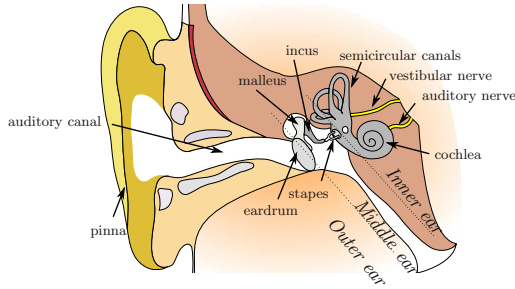


Fig. 1. Anatomy of the human ear. The tympanic membrane and the ossicular chain with the malleus, the incus, and the stapes compose the middle ear.

the simulation are the low number of pathological scenarios. Yet, the anatomy of the middle ear is dissimilar between patients and can lead to different surgical approaches or even to a surgical contraindication.

The objective of this study is the design of a mechanical atlas for the ossicular chain dedicated to the simulation of the middle ear surgery. The mechanical behavior of our atlas is confronted to measurements on human temporal bone specimens. As the atlas is developed for an interactive surgical simulator, the computation efficiency is also investigated. Finally, a registration method of the mechanical atlas to fit patient imagery is also described.

2 Validated Mechanical Atlas Dedicated to Real-Time

This section presents a biomechanical model of the middle ear behavior. The approach is built on finite element methods (FEM) with parameters that were collected from literature. The model is validated for both dynamic small deformations and static large transformations. We then optimize the computation to use the model, without changes, in an interactive real-time simulation context.

Design and Specifications: The three-dimensional geometric model obtained from a micro-Magnetic Resonance Imaging of the middle ear [4] is used to create our atlas. The missing ligaments are manually reconstructed according to the anatomical data using Blender. A volumetric model of the ossicular chain is generated using the Computational Geometry Algorithms Library (CGAL) that allows to create a coherent mesh while defining several domains. We then assign adequate parameters to these different domains (as illustrated in Figures 2(a) and 3(a) with red and blue meshes for respectively, high and low stiffnesses in the domain). The eardrum and the ossicular chain (meshed respectively by 2492 and 4361 tetrahedral deformable elements) are implemented in SOFA, a real-time medical simulation project [5]. Both models use a finite element deformable model (large displacements and rotations) derived from [6], with respectively, membrane and volume elastic energy. The mass is lumped at the

Table 1. Mechanical parameters used in our atlas

Anatomical structures		Density(kg/m ³)	Young modulus (N/m ²)
Eardrum	Pars tensa	1.20 x 10 ³ [7]	3.34 x 10 ⁷ [7]
	Pars flaccida	1.20 x 10 ³ [7]	1.10 x 10 ⁷ [8]
Malleus	Handle	3.70 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
	Head	2.55 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
	Neck	4.53 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
Incus	Body	2.36 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
	Short process	2.26 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
	Long process	5.08 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
Stapes	-	2.20 x 10 ³ [8], [9]	1.41 x 10 ¹⁰ [8]
Ligaments	Superior	2.50 x 10 ³ [7]	4.90 x 10 ⁴ [8],[10]
	Incudostapedial	2.50 x 10 ³ [7]	6.00 x 10 ⁵ [8]
	Incudomalleolar	2.50 x 10 ³ [7]	1.41 x 10 ¹⁰ [8]
	Malleus anterior	2.50 x 10 ³ [7]	2.10 x 10 ⁷ [7]
	Malleus lateral	2.50 x 10 ³ [7]	6.70 x 10 ⁴ [10]
	Incus posterior	2.50 x 10 ³ [7]	6.50 x 10 ⁵ [7]
	Tendons	Tensor tympani	2.50 x 10 ³ [7]
	Stapedial	2.50 x 10 ³ [7]	5.20 x 10 ⁵ [7]

nodes. The tympanic annulus, the stapedial annular ligament, the cochlea, the ligaments and the tendons of the middle ear are used as boundaries of our FEM. The density and the Young modulus parameters of our atlas are chosen according to published data on middle ear mechanics [7], [8], [9], [10], as reported in table 1.

Validation of the atlas: The dynamics of the atlas model is evaluated using the transfer function analysis (TFA) of the ossicular chain. The TFA consists in the measurement of the stapes footplate velocity when a sinusoid pressure is applied to the malleus (Fig. 2(a)). The pressure is set to 0.632 Pa equivalent to 90 dB (SPL) and the frequency ranged from 250 to 8000 Hz like in a physiological condition of hearing. This test cannot be performed in real-time due to the very small time steps required to simulate vibrations at high frequency. The TFA result provides the stapes footplate velocity transfer function (STF) which is increasing for frequency below 1000 Hz and decreasing for higher frequency (Fig. 2(b)). This characteristic shape is also observed by other FEM simulation [8] (but non targeted towards real-time application) and by reported studies on human temporal bone [11]. The large repartition of the TFA reported by Rosowski et al. is explained by the individual anatomical variability of the middle ear components and by the storage condition of the specimens. This test assesses the dynamics of our model at different frequencies (which is linked to the mass and stiffness values) but the deformations are limited to small displacements.

Consequently, we have also conducted the analysis of the umbo displacement under static pressure load (UDSP) to assess the model with large deformation

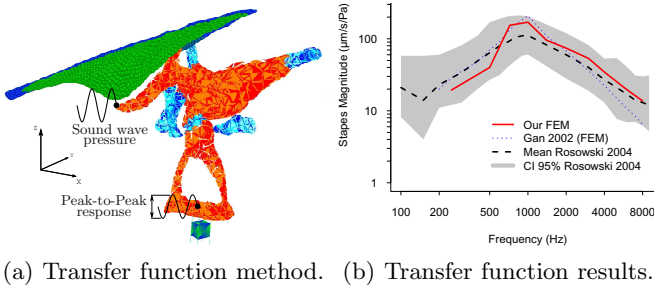


Fig. 2. Evaluation of the atlas in physiological condition using the transfer function analysis. The motion of the stapes footplate is analyzed when a sinusoid pressure is applied to the malleus.

transformation at equilibrium. The UDSP corresponds to the observation of the umbo motion when a static pressure is applied to the eardrum (Fig. 3(a)). We use pressures ranged from -4000 to 4000 Pa which are comparable to pressures involved in a conventional tympanometry diagnostic. As shown in figure 3(b) the results of the simulation of our atlas are close to published observations on human temporal bones [12] or previously published FEM model [13]. Moreover, the ratio of the umbo displacement between negative and positive force is around 1.76 ± 0.31 ($n=8$) for our atlas, which is similar to the ratio between 1.8 - 1.9 obtained on temporal bones. The analysis of the UDSP allows to assess the realism of the middle ear model when the displacements induced by the deformation are significant compared to the size of the structure (more than 15%). This range of deformation with large displacements is similar to those encountered when simulating a surgery.

Interactive surgical simulation: The goal of this work is to build a training simulation system of a new generation, in which the physics-based models are not over-simplified for real-time performance. Real-time is a needed condition to

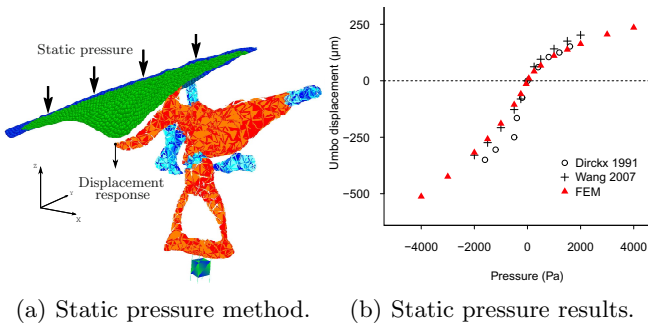


Fig. 3. Evaluation of the atlas in surgery condition using the static pressure. The motion of the umbo is analyzed when a static pressure is applied to the eardrum.

allow for the interactivity required for a surgical simulator. This constraint often leads to the choice of plausible models rather than validated physics-based approaches. In this work, we have deliberately chosen to base the real-time simulation on the atlas model that is validated using TFA and UDSP tests. As previously mentioned, the number of elements used in the FEM models is not excessive. We assessed the convergence of the FEM at several resolutions, between the presented model and a model composed of 210689 elements the mean square error using TFA is $2.8 \mu\text{m/s/Pa}$, less than 1 dB. Due to the relatively high stiffness/mass ratio, the use of implicit integration is necessary. Still, due to the strong heterogeneities in the model, we had a bad condition number on the system of equations generated by the FEM model. Consequently, iterative solvers, like the conjugate gradient algorithm had difficulties to reach the convergence. Moreover, the factorization cost induced by direct solvers was too heavy for real-time performance. Thus, we have opted for the solution presented in [14], in which a preconditionner is factorized asynchronously in a separate parallel thread (i.e. without additional computation cost in the simulation process). This solution strongly improves the convergence of the conjugate gradient algorithm used in our simulation. The time step of the backward Euler scheme that we use in the simulation is set at 0.04 s. An average of 40 frames per second is observed with no interaction and collision. In the worst case (collision response with more than 15 friction contact points), the number of frame rates falls down to 20Hz. The framework SOFA allows for haptic rendering with Phantom Omni devices (Sensable, Wilmington, MA). The collision response and the haptic algorithm is based on a multithreaded approach [15]. It allows for updating the force feedback at higher rates than the simulation. Using these components of SOFA, we are able to use the atlas model without any changes in the parameters, for a real-time simulation and contact response with surgical instruments.

3 Patient Specific Model

The atlas is developed for training and for rehearsal simulation of the middle ear surgery. Therefore, we implement a registration method to fit our validated mechanical model to the patient image data. The goal is to perform the registration of the atlas (mechanical model, parameters, boundary conditions...) to avoid time-consuming work of segmentation and the re-parametrization. First, a rigid semi-automatic approach is used to place approximately the atlas in the same area and orientation as the patient anatomy. This method is followed by a deformable registration that uses the FEM of the atlas and the implementation of additional spring forces that depends on the intensity of the pixel.

The rigid registration is based on the manual selection of 3 points: The umbo (\tilde{p}_a), the incudomalleolar joint (\tilde{p}_b) and the incudostapedial joint (\tilde{p}_c). These points correspond to distinct landmarks on the ossicular chain that are easily identifiable on the images. These landmarks are also placed on our atlas model, noted p_x (Fig. 4(a)) to allow for a simple rigid registration.

In practice, after the rigid registration, the middle ear components on the images are close to the surface of our atlas. But to enhance the precision of the

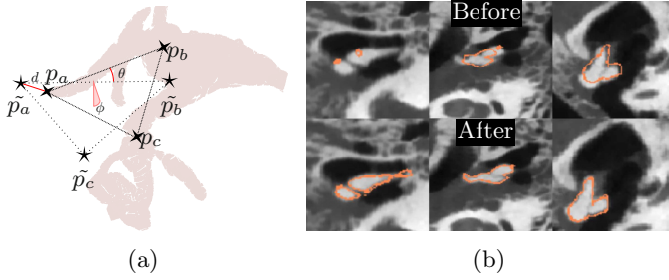
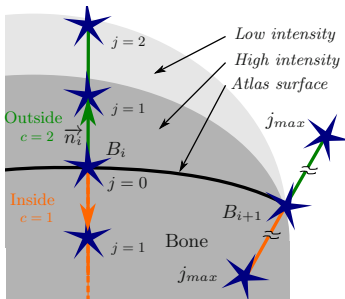


Fig. 4. Registration of our atlas. (a) presents the method of the rigid registration, and (b) the results of the deformable approach, with a representation of the surface of the atlas in orange before (top) and after (down) the deformable registration process.

registration, we perform a deformation based on the variation of the imaging’s pixel intensity around the atlas and elastic energy from FEM model. We start from the barycenter B_i , of each triangle i of the atlas surface mesh and we explore the neighborhood along the normal vector n_i inside the mesh ($c = 1$) and outside ($c = 2$) using the equation 1. Where X_i^c is the coordinate of the exploration point along the normal vector, $step$ corresponds to the smallest thickness of the imaging slices, and j to the exploration iteration parameter (Fig. 5). The equation 2 computes a pixel intensity ratio between two successive points around the atlas in order to detect an intensity shift meaning that we are at a bone boundary on the patient image. Thus, a border point is detected and associated to each points that compose the triangle i . As a point belongs to several triangles, each point of the surface have several associated border points. Thus, we compute the barycenter ω_k of these associated border points for each point k of the surface mesh. Then a spring is implemented between ω_k and k to deform the mechanical FEM of the atlas. We perform several iterations of the simulation with an update of the border points at each step until an equilibrium position is reached.

Imagery acquisition of four left ears is obtained using a NewTom 5G Cone Beam Computed Tomography (QR SRL, Verona, Italy). The figure 4(b) repre-



$$X_i^c(j) = B_i + (-1)^c \cdot n_i \cdot j \cdot step \quad (1)$$

$$R_i^c(j) = \frac{Intensity(X_i^c(j))}{Intensity(X_i^c(j-1))} \quad (2)$$

In figure 5 :

$$R_i^2(j = 2) < 1 - threshold$$

$$R_i^1(j = 2) \approx 1$$

Fig. 5. Detection of the intensity shift around the atlas surface

sents the surface of our mechanical atlas over the patient images before and after the deformable process. The results presented are obtained after 300 iterations performed in 110 seconds on a conventional computer. The registration accuracy is compared to manual segmentations using MeshDev [16], a mean error of 0.201 mm is found for our registration algorithm and 0.187 mm between manual segmentations. This error is within the imaging resolution, 0.26 mm, corresponding to the voxel diagonal.

4 Discussion

The objective of the atlas is to supply models for a new surgical tool for training and rehearsal purpose. The simulator is planned to simulate the otosclerosis surgery including stapes footplate drilling or placing the prosthesis to restore the sound transmission. Our simulation, based on a mechanical atlas, provides realistic and validated results for the behavior of the middle ear model. Indeed, results show that our simulation is realistic in regard to experimental observation on human ear. Moreover, the computational efficiency of our approach allows real-time interactions, making it suitable for use in a training simulator.

In middle ear surgery, the anatomy of the patient has a strong influence on the choice of surgical procedure or on a contraindication statement. To account for rehearsal simulation, information from the patient should be taken into consideration. A registration approach of our model to fit the anatomical data is developed and tested. The main advantage of this registration approach is that the process is performed on a tested and validated mechanical model. Indeed, all the mechanical parameters such as the meshing, the boundaries conditions, etc, are still implemented into our atlas after the registration. However, only the geometry of the atlas is adapted to patient data, the mechanical results provided by the registered models are similar to the atlas model.

To build a predictive patient specific model, we should also use the mechanical parameters (such as the Young moduli of each structures) that correspond to each individual patient. Indeed, these parameters differ from a patient to another and the search of parameters from the patient data is still future work. Using such a predictive model, the simulation could find new applications. Namely, it could be used as a planning pre-operative tool to anticipate the choices that are currently done during the surgery: selection of ossicular bones to repair, type of implantable prosthesis, etc.

5 Conclusion

A mechanical atlas of the human middle ear was implemented for surgical training and rehearsal simulator. The realism of the atlas behavior was evaluated in physiological and in surgery condition. The computation efficiency of the developed atlas allows for real time computation simulation and for haptic interactions. To our best knowledge, this is the first paper that presents such a real-time simulation. An interactive method is reported to adapt the proposed mechanical

atlas to the individual anatomy of the patient in few minutes. A patient specific model is obtained directly from the evaluated mechanical model avoiding the time-consuming work of a manual segmentation, mechanical parametrization and evaluation. The presented study is the basis of a microsurgical simulator design for training, practice and planning for otological surgery.

References

1. George, A.P., De, R.: Review of temporal bone dissection teaching: how it was, is and will be. *J. Laryngol. Otol.* 124, 119–125 (2010)
2. Bakhos, D., Velut, S., Robier, A., Alzahrani, M., Lescanne, E.: Three-dimensional modeling of the temporal bone for surgical training. *Otol. Neurotol.* 2, 328–334 (2010)
3. Trier, P., Noe, K.Ø., Sørensen, M.S., Mosegaard, J.: The visible ear surgery simulator. *Stud. Health Technol. Inform.* 132, 523–525 (2008)
4. Nicholson, D.T., Chalk, C., Funnell, W.R.J., Daniel, S.J.: A randomized controlled study of a computer-generated three-dimensional model for teaching ear anatomy. *Med. Educ.* 40, 1081–1087 (2006)
5. Faure, F., Duriez, C., Delingette, H., Allard, J., Gilles, B., Marchesseau, S., Talbot, H., Courtecuisse, H., Bousquet, G., Peterlik, I.: SOFA: A Multi-Model Framework for Interactive Physical Simulation. *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery* 11, 283–321 (2012)
6. Felippa, C.A.: A systematic approach to the element independent corotational dynamics of finite elements. Report of Center for Aerospace Structures (2000)
7. Koike, T., Wada, H., Kobayashi, T.: Modeling of the human middle ear using the finite-element method. *J. Acoust. Soc. Am.* 111, 1306–1317 (2002)
8. Gan, R.Z., Sun, Q.: Finite element modeling of human ear with external ear canal and middle ear cavity. In: *Proc. 2nd Joint IEEE EMBS/BMES Conf.*, vol. 1, pp. 264–265 (2002)
9. Kirikae, I.: *The Structure and Function of the Middle Ear*. Tokyo Press (1960)
10. Beer, H.J., Bornitz, M., Drescher, J., Schmidt, R., Hardtke, H.J.: Finite element modeling of the human eardrum and applications. In: Hüttenbrink, K.B. (ed.) *Proc. Middle Ear Mechanics in Research and Otosurgery*, pp. 40–47. University of Technology, Dresden (1996)
11. Rosowski, J.J., Huber, A.M., Ravicz, M.E., Goode, R.L.: Are temporal bones useful models of human middle ear mechanics? In: *27th Meeting of the Association for Research in Otolaryngology*, p. 275 (2004)
12. Dirckx, J.J.J., Decraemer, W.F.: Human tympanic membrane deformation under static pressure. *Hearing Res.* 51, 93–105 (1991)
13. Wang, X., Cheng, T., Gan, R.Z.: Finite-element analysis of middle-ear pressure effects on static and dynamic behavior of human ear. *J. Acoust. Soc. Am.* 122, 906–917 (2007)
14. Courtecuisse, H., Allard, J., Duriez, C., Cotin, S.: Asynchronous Preconditioners for Efficient Solving of Non-linear Deformations. In: *Proc. VRIPHYS* (2010)
15. Saupin, G., Duriez, C., Cotin, S.: Contact Model for Haptic Medical Simulations. In: *Proc. 4th ISBM*, pp. 157–165 (2008)
16. Roy, M., Fofou, S., Truchetet, F.: Generic Attribute Deviation Metric for Assessing Mesh Simplification Algorithm Quality. In: *Proc. IEEE ICIP*, pp. 817–820 (2002)

Surgical Gesture Segmentation and Recognition

Lingling Tao, Luca Zappella, Gregory D. Hager, and René Vidal

Johns Hopkins University, 3400 North Charles Street, Baltimore, MD, 21218, USA

Abstract. Automatic surgical gesture segmentation and recognition can provide useful feedback for surgical training in robotic surgery. Most prior work in this field relies on the robot's kinematic data. Although recent work [1,2] shows that the robot's video data can be equally effective for surgical gesture recognition, the segmentation of the video into gestures is assumed to be known. In this paper, we propose a framework for joint segmentation and recognition of surgical gestures from kinematic and video data. Unlike prior work that relies on either frame-level kinematic cues, or segment-level kinematic or video cues, our approach exploits both cues by using a combined Markov/semi-Markov conditional random field (MsM-CRF) model. Our experiments show that the proposed model improves over a Markov or semi-Markov CRF when using video data alone, gives results that are comparable to state-of-the-art methods on kinematic data alone, and improves over state-of-the-art methods when combining kinematic and video data.

Keywords: surgical gesture segmentation, surgical gesture recognition, time series analysis, conditional random fields, structured output learning.

1 Introduction

Robotic minimally invasive surgery (RMIS) offers several advantages over traditional surgery, including better precision, smaller incisions, and quicker recovery time. However, reductions in the amount of one-on-one teaching [3], together with the steep learning curve of RMIS [4], advocate for the development of a new teaching paradigm where surgical skill is automatically assessed and timely feedback is automatically provided.

Advances in machine learning, computer vision, and speech processing can be exploited for this purpose. For instance, segmentation and recognition methods can be used to decompose a surgical task (e.g., suturing) into a sequence of gestures (e.g., grab needle, position needle, insert needle), and perform skill assessment based on how well a sequence of gestures is executed. As shown in [5], gesture recognition can also be used to accomplish shared or cooperative control tasks that are triggered based on what the user is doing. Hence, we could envisage robotic gesture recognition being used, for example, to trigger appropriate information displays. All these applications require segmentation and recognition of surgical gestures, which is the main focus of this paper.

Most of the prior work on automatic segmentation and recognition of surgical gestures relies on the kinematic observations of the robot's motion. The temporal evolution of such observations is typically modeled using a Hidden Markov Model (HMM), where each gesture corresponds to one or more states of the HMM and the transitions among consecutive gestures are modeled by the HMM transition probabilities. Different papers [6,7,8,9,10,11,12] use different models for the observations associated with

each gesture, including discrete HMMs, Gaussian HMMs, factor analyzed HMMs and Sparse HMMs. While generally successful, these methods rely mostly on *local cues* from a few frames, thus failing to capture *global cues* about the whole execution of a gesture. To address this issue, [11] uses Switched Linear Dynamical Systems (SLDSs), which model the dynamics of the whole execution of a gesture with an LDS. However, this comes at a steep computational cost because inference for SLDSs is intractable.

More recently, video-based solutions have drawn the attention of the research community. Most of the prior work focuses on the detection of the instruments used during surgery or in the operating room [13,14,15,16,17,18,19,20] using techniques such as dynamic time warping, support vector machines and HMMs. However, these techniques use only frame-level features, such as color, texture and shape-based cues. Moreover, the desired solution is a high-level recognition based on the presence of some tools, rather than fine-grained gesture recognition based on motion data.

To the best of our knowledge, the only existing works that use state-of-the-art computer vision algorithms for surgical gesture recognition from video data are [1,2]. These works show that video data can be as discriminative as kinematic data when appropriate features and algorithms are used. Specifically, [1,2] obtain very high gesture recognition rates using *global* bag-of-spatio-temporal features (BoSTF) and LDSs to model the whole execution of a gesture. However, this is possible only because the temporal segmentation of the video into gestures is assumed to be known. Recent work in computer vision addresses the joint segmentation and recognition of generic actions in videos using conditional Random fields (CRFs), where the sequence of gestures is obtained by minimizing the energy of the CRF. The work of [21] adopts a Markov CRF model whose energy depends on which objects are present in each frame and their interactions, while the work of [22] adopts a semi-Markov CRF model based on global features extracted from many frames. However, these methods have not been combined for joint gesture segmentation and recognition, nor have they been applied to surgical gestures.

In this paper, inspired by the use of graphical models in speech processing [23], we propose a combined Markov/semi-Markov conditional random field (MsM-CRF) model for joint segmentation and recognition of surgical gestures from kinematic and video data. Our MsM-CRF model captures both local cues (thanks to the Markov component) and global cues (thanks to the semi-Markov component). Moreover, we exploit the high accuracy of the BoSTF approach from [1,2] to model the unary potentials of the CRFs using two kinds of spatio-temporal features. Our experiments on a typical surgical training setup show that our model improves over a Markov or semi-Markov CRF model on video data alone, gives state-of-the-art results on kinematic data alone, and improves over state-of-the-art methods by combining kinematic and video data.

2 Joint Segmentation and Recognition of Surgical Gestures

2.1 MsM-CRF Model

Let $\mathcal{V} = \{I_t\}_{t=1}^T$ be a sequence of observations, where I_t represents the observations at frame t . Let $\mathcal{C} = \{1, \dots, C\}$ be the set of possible gesture labels. Our goal is to find the sequence of frame-level gestures $\mathcal{Y} = \{Y_t^F\}_{t=1}^T$, where $Y_t^F \in \mathcal{C}$ denotes the gesture label at frame t . Since each gesture may span a few consecutive frames, we can divide

the time interval $[1, T]$ into M segments, where the i -th segment is $[t_{i-1}, t_i]$, such that the frame gesture label within a segment does not change. Here t_i is the last frame of i , $t_0 = 1$, and $t_M = T$. We can then define the sequence of segment-level gestures as $\{\mathbf{Y}_i^S\}_{i=1}^M$, where $\mathbf{Y}_i^S = Y_t$ for all $t \in i$. In what follows, we will use the superscripts F and S to denote variables at the frame and segment levels, respectively.

We represent \mathcal{V} with a graphical model $\mathcal{G} = (\mathcal{N}^F, \mathcal{E}^F, \mathcal{N}^S, \mathcal{E}^S)$. In the case of CRFs, each node $N_t^F \in \mathcal{N}^F$ denotes a frame t , hence $|\mathcal{N}^F| = T$. In the case of semi-CRFs, each node $N_i^S \in \mathcal{N}^S$ denotes the collection of frames in segment i , hence $|\mathcal{N}^S| = M$. In both cases, an edge $e_j \in \mathcal{E}$ denotes the connection between consecutive nodes N_j and N_{j+1} . We model the conditional distribution of the sequence of labels \mathcal{V} given \mathcal{V} , with a Gibbs distribution: $p(\mathcal{Y}|\mathcal{V}) \propto \exp(E(\mathcal{Y}, \mathcal{V}))$, where $E(\mathcal{Y}, \mathcal{V})$ is an energy function defined on the cliques of \mathcal{G} . In our MsM-CRF model, we have:

$$E(\mathcal{Y}, \mathcal{V}) = \lambda^{FU} \sum_{t=1}^T \psi_t^{FU}(Y_t^F; \mathcal{V}) + \lambda^{FP} \sum_{t=1}^{T-1} \psi_{t,t+1}^{FP}(Y_t^F, Y_{t+1}^F; \mathcal{V}) + \lambda^{SU} \sum_{i=1}^M \psi_i^{SU}(\mathbf{Y}_i^S; \mathcal{V}) + \lambda^{SP} \sum_{i=1}^{M-1} \psi_{i,i+1}^{SP}(\mathbf{Y}_i^S, \mathbf{Y}_{i+1}^S; \mathcal{V}), \quad (1)$$

where ψ^{FU} and ψ^{FP} are the CRF unary and pairwise potentials, while ψ^{SU} and ψ^{SP} are the semi-CRF unary and pairwise potentials, each one weighted by its own λ factor.

CRF Unary. This potential gives the score of assigning a gesture label to a single frame. For kinematic data, the score is computed from the output of an SVM classifier, with an RBF kernel, trained for each possible gesture on the raw data of each frame. For video data, this score is obtained from the output of a classifier applied to a histogram of features extracted from a neighborhood of the frame. Specifically, during training the spatio-temporal video features are clustered by K -means to form a dictionary of visual words. Each frame is then represented with a histogram of words and these histograms are used to train an SVM classifier with a χ^2 -RBF kernel. In both cases, the logarithm of the probability returned by regression of the SVM output is used as a unary score.

Semi-CRF Unary. This potential gives the score of assigning a gesture label to a segment, thereby capturing global features related to the overall gesture. For kinematic data, we train an SVM classifier with RBF kernel for each gesture on the average of the raw data within each segment. For video data, we represent each segment by the histogram of words accumulated over all the frames that correspond to the segment using the same dictionary of visual words described before. These histograms are then used to train an SVM classifier with χ^2 -RBF kernel for each gesture. Hence, we use the logarithm of the probability returned by regression of the SVM output as our unary term. This way of computing the most likely label for each segment corresponds exactly to the approach followed in [1,2].

Spatio-temporal Features. We use two kinds of spatio-temporal features. The first one is a concatenation of histograms of oriented gradients (HOG) and histograms of optical flows (HOF) extracted from a cuboid centered around each STIP point [24]. Since STIP points tend to be sparse in space, we also use the dense features presented in [25], which

consist of HOG, HOF, and histograms of motion boundaries and velocities (in term of x and y coordinates) computed around dense trajectories.

CRF Pairwise. This potential captures the probability of switching from gesture label g_k to g_j when moving from one frame to the next. Since a gesture is composed of many frames, this potential encourages the frame labels to be temporally coherent. We capture the relationship between adjacent frames using the transition probability

$$P_{g_k, g_j}^F = \frac{\# \text{ frames switching from } g_k \text{ to } g_j}{\# \text{ frames with label } g_k} \quad (2)$$

computed from the training set. We then set $\psi_{t, t+1}^{FP}(Y_t^F, Y_{t+1}^F; \mathcal{V}) = \log(P_{Y_t^F, Y_{t+1}^F}^F)$.

Semi-CRF Pairwise. This potential captures the probability of switching from gesture label g_k to g_j when moving from one segment to the next. Since each segment represents a single instance of a gesture, two consecutive segments should not have the same label. Thus, this potential encourages a switch from one gesture label to a different one. We capture the relationship between adjacent segments using the transition probability

$$P_{g_k, g_j}^S = \frac{\# \text{ segments switching from } g_k \text{ to } g_j}{\# \text{ segments with label } g_k} \quad (3)$$

computed from the training set. We then set $\psi_{i, i+1}^{SP}(\mathbf{Y}_i^S, \mathbf{Y}_{i+1}^S; \mathcal{V}) = \log(P_{\mathbf{Y}_i^S, \mathbf{Y}_{i+1}^S}^S)$.

2.2 Efficient Inference and Learning

Inference. The energy in (1) can be re-written as $E(\mathcal{Y}, \mathcal{V}) = \mathbf{w}^\top \Psi(\mathcal{Y}; \mathcal{V})$, where

$$\mathbf{w} = \begin{pmatrix} \lambda_{FU} \\ \lambda_{FP} \\ \lambda_{SU} \\ \lambda_{SP} \end{pmatrix} \quad \text{and} \quad \Psi(\mathcal{Y}; \mathcal{V}) = \begin{pmatrix} \sum_{t=1}^T \psi_t^{FU}(Y_t^F; \mathcal{V}) \\ \sum_{t=1}^{T-1} \psi_{t, t+1}^{FP}(Y_t^F, Y_{t+1}^F; \mathcal{V}) \\ \sum_{i=1}^M \psi_i^{SU}(\mathbf{Y}_i^S; \mathcal{V}) \\ \sum_{i=1}^{M-1} \psi_{i, i+1}^{SP}(\mathbf{Y}_i^S, \mathbf{Y}_{i+1}^S; \mathcal{V}) \end{pmatrix}. \quad (4)$$

Given the MsM-CRF model parameters in \mathbf{w} and a test video \mathcal{V} , we can perform joint gesture segmentation and recognition by solving the inference problem $\mathcal{Y}^* = \arg\max_{\mathcal{Y}} E(\mathcal{Y}, \mathcal{V})$. One can show that the energy in (1) is equivalent to an energy that depends only on the segment labels $\{\mathbf{Y}_i^S\}$. The maximization of the resulting energy can be done by a Viterbi-like dynamic programming algorithm, as described in [22].

Learning. Given B training videos $\{\mathcal{V}_i\}_{i=1}^B$ and their corresponding labelings $\{\bar{\mathcal{Y}}_i\}_{i=1}^B$, we learn the parameters \mathbf{w} using a method based on structural SVM [26]. Specifically, let us refer to any labeling of \mathcal{V}_i that is different from $\bar{\mathcal{Y}}_i$ as a negative example, and denote the set of negative examples for a video \mathcal{V}_i as \mathcal{Y}_i^- . Given $\mu > 0$, we learn the parameters \mathbf{w} by solving the following optimization problem:

$$\{\mathbf{w}^*, \{\xi_i^*\}_{i=1}^B\} = \arg \min_{\mathbf{w}, \{\xi_i\}_{i=1}^B} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{\mu}{B} \sum_{i=1}^B \xi_i, \quad \text{subject to} \quad (5)$$

$$\begin{aligned} \text{(a) } & \forall i = 1, \dots, B, \forall \mathcal{Y} \in \mathcal{Y}_i^-, \mathbf{w}^\top (\Psi(\bar{\mathcal{Y}}_i; \mathcal{V}_i) - \Psi(\mathcal{Y}; \mathcal{V}_i)) \geq \ell(\bar{\mathcal{Y}}_i, \mathcal{V}) - \xi_i \\ \text{(b) } & \forall i = 1, \dots, B, \xi_i \geq 0 \quad \text{and} \quad \text{(c) } \mathbf{w} \geq \mathbf{0}. \end{aligned} \quad (6)$$

The intuition behind the first inequality is that we want the energy at the ground truth labeling $\mathbf{w}^\top \Psi(\bar{\mathcal{Y}}_i; \mathcal{V}_i)$ to be greater than the energy of any wrong labeling $\mathbf{w}^\top \Psi(\mathcal{Y}; \mathcal{V}_i)$ by the loss $\ell(\bar{\mathcal{Y}}_i, \mathcal{Y})$ while allowing some slack ξ_i . The loss function $\ell(\bar{\mathcal{Y}}_i, \mathcal{Y})$ measures the error in the labeling \mathcal{Y} as the fraction of misclassified frames. Since the number of constraints is exponentially large, we use the cutting plane algorithm [27] to find \mathbf{w} .

3 Experiments

Data. We evaluate our approach using the dataset in [28], which contains three different surgical tasks, suturing (SU), needle passing (NP) and knot tying (KT), each performed by 8 surgeons with three different skill levels (expert, intermediate and novice). Each surgeon performs around 3 to 5 trials for each task, which gives 39 trials for SU, 26 trials for NP and 36 trials for KT. Each trial lasts, on average, 2 minutes, and both kinematic and video data are recorded at a rate of 30 frames per second. Kinematic data consists of 78 motion variables (positions, rotation angles, and velocities of the master/patient side manipulators), whereas video data is taken from the first person view point of the robot and consists of jpeg images of size 320×240 . The data is manually labelled with 15 surgical gestures (e.g., reach needle, position needle, etc.) described in Figure 1.

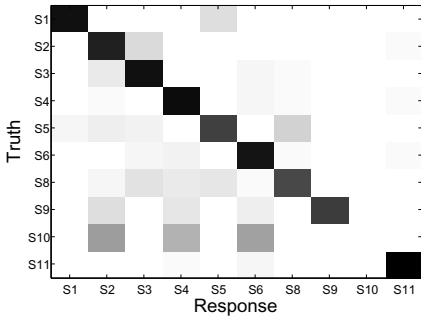
Setup. Following [1,2], we use two different test setups. The first one is the *leave-one-super-trial-out* (LOSO), where super trial i (i.e., trial i from each user) is held out for testing. The second, and more challenging setup, is the *leave-one-user-out* (LOUO), where all the trials from user i are held-out for testing. We compute the video CRF unary terms using a neighborhood of 25 frames, and the CRF and semi-CRF unaries using a dictionary of 300 words. To speed up the computation of the Viterbi-like algorithm, we perform inference every 10 frames, and assume that the maximum length of a segment is 400 frames. We compare the performance of the proposed MsM-CRF model with that of a CRF and a semi-CRF model. Note that the semi-CRF model with STIP features can be seen as an extension of the method presented in [1,2] to the case where the segmentation is unknown. The average percentage of correctly classified frames is shown in Table 1. For each technique, the type of features used is indicated within parenthesis. For MsM-CRF, the first feature refers to the CRF unaries and the second one to the semi-CRF unaries. For instance, MsM-CRF(kin-STIP) means that kinematic data are used for the CRF unary, while STIP features are used for the semi-CRF unary.

Results on Video Data. Notice that the combination of local and global features from video data used by the MsM-CRF model always improves over the CRF and semi-CRF models, for all tasks and test setups. Notice also that using dense features leads to better results than using of STIP features. This should be related to the fact that the features of [25] include motion boundaries and velocities in addition to HOG and HOF. Finally, notice that the results from [1] are 2-12% better, but they assume known segmentation.

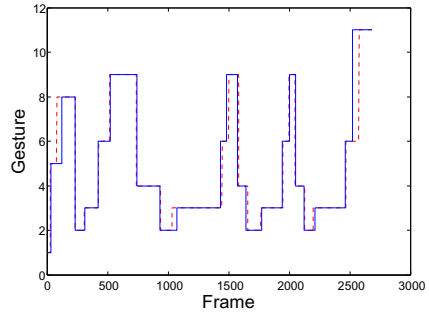
Results on Kinematic Data. In this case, the CRF model outperforms the semi-CRF model. Arguably, this is because the feature used in the unary term of the semi-CRF model (average of the data in a temporal window) is too simple. This is also reflected on the MsM-CRF results, which are similar to those of the CRF. The results of sparse-HMMs [12] are better than those of MsM-CRF(dense-dense) in the LOSO setup. However, in the more challenging LOUO setup MsM-CRF(dense-dense) seems to generalize

Table 1. Average percentage of correctly classified frames on the dataset in [28]. Best results for methods that do not assume known segmentation are highlighted in boldface.

	Method	LOSO			LOUO		
		SU	KT	NP	SU	KT	NP
Video	1) CRF(STIP)	70.86%	68.33%	55.12%	61.12%	62.63%	52.58%
	2) semi-CRF(STIP)	67.91%	67.07%	56.49%	51.71%	40.04%	45.03%
	3) MsM-CRF(STIP-STIP)	73.32%	70.95%	63.31%	66.28%	66.53%	58.85%
	4) CRF(dense)	76.51%	69.16%	62.23%	68.80%	60.17%	54.52%
	5) semi-CRF(dense)	65.83%	44.82%	56.22%	59.41%	41.46%	46.89%
	6) MsM-CRF(dense-dense)	79.04%	72.04%	68.81%	71.76%	66.94%	60.39%
	7) BoSTF(STIP,known segment.) [1]	84.87%	84.03%	72.16%	75.72%	79.05%	61.73%
Kinematic	8) CRF(kin)	81.62%	81.06%	74.56%	68.65%	67.38%	46.44%
	9) semi-CRF(kin)	63.20%	39.20%	54.15%	62.24%	44.28%	38.36%
	10) MsM-CRF(kin-kin)	80.99%	79.39%	74.85%	69.03%	64.28%	52.39%
	11) sparse-HMM [12]	81.1%	82.6%	76.1%	67.8%	65.7%	59.3 %
Mix	12) MsM-CRF(kin-STIP)	82.49%	80.50%	76.41%	70.09%	68.43%	54.41%
	13) MsM-CRF(kin-dense)	82.81%	81.10%	76.82%	72.60%	68.83%	57.08%



(a) Confusion matrix, darker colors mean higher percentage.



(b) Ground-truth “- -” and prediction “-” for one suturing trial.

Fig. 1. Results for the MsM-CRF(kin-dense) algorithm on the Suturing-LOSO setup. S1 reaching for needle with right hand. S2 positioning needle. S3 pushing needle through tissue. S4 transferring needle from left to right. S5 moving to center with needle in grip. S6 pulling suture with left hand. S7 pulling suture with right hand. S8 orienting needle. S9 using right hand to help tighten suture. S10 loosening more suture. S11 dropping suture at end and moving to end points. S12 reaching for needle with left hand. S13 making ‘C’ loop around right hand. S14 right hand reaches for suture. S15 both hands pull.

better. Overall, our method is able to achieve state-of-the-art results on kinematic data, and its performance on video data is comparable to that on kinematic data.

Results on Video and Kinematic Data. We also exploit the flexibility of the MsM-CRF framework to combine both kinematic data (in the CRF-unaries) and video data (in the semi-CRF unaries). The resulting MsM-CRF(kin-dense) model performs almost always better than any other technique that uses only kinematic or video data.

Fig. 1(a) shows the confusion matrix for the MsM-CRF(kin-dense) method on the Suturing-LOSO setup, while Fig. 1(b) shows the predicted and ground-truth sequence of gestures for one suturing trial. From these figures it is possible to appreciate the quality of the results produced. Note that gesture 10 is never classified correctly. This is due to the fact that this gesture appears very rarely, hence, it is not possible to learn a robust classifier. Fig. 1(b) shows that most of the errors appear around the switching times. Typically, the prediction switches either too early or with some delay. This might be due to the fact that for speed convenience we perform inference every 10 frames, hence, the results could be slightly improved by sacrificing some computational time.

Computing Time. A final note on the computational complexity. For completing one held-one-out experiment, in which usually around 30 trials are used for training and 8 for testing, the training stage took around 4 hours with a Matlab implementation on a x86@3.33GHz processor, and testing for one trial usually required around 1 minute.

4 Conclusions

We have proposed a combined Markov/semi-Markov CRF model for temporal gesture segmentation and recognition. Our model can capture local features and interactions between frames, as well as global characteristics of each gesture and interactions between gestures. We have shown on a typical surgical dataset that the MsM-CRF model always improves with respect to the CRF or semi-CRF frameworks used alone. We have also observed that the MsM-CRF model based on dense features is more robust in the LOUO setup. Moreover, thanks to the flexibility of the MsM-CRF, we were able to present a hybrid solution where kinematic and video data were both used. Such a hybrid solution achieved results that are similar or superior to those of state-of-the-art algorithms.

Acknowledgments. Work funded by Sloan Foundation, and NSF 0931805 and 0941362.

References

1. Béjar Haro, B., Zappella, L., Vidal, R.: Surgical gesture classification from video data. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 34–41. Springer, Heidelberg (2012)
2. Zappella, L., Béjar, B., Hager, G., Vidal, R.: Surgical gesture classification from video and kinematic data. *Medical Image Analysis* (2013)
3. Barden, C., Specht, M., McCarter, M., Daly, J., Fahey, T.: Effects of limited work hours on surgical training. *Obstetrical & Gynecological Survey* 58(4), 244–245 (2003)
4. Lenihan, J., Kovanda, C., Seshadri-Kreaden, U.: What is the learning curve for robotic assisted gynecologic surgery? *J. of Minimally Invasive Gynecology* 15(5), 589–594 (2008)
5. Padoy, N., Hager, G.D.: Human-machine collaborative surgery using learned models. In: IEEE Conference on Robotics and Automation, pp. 5285–5292 (2011)
6. Rosen, J., Hannaford, B., Richards, C., Sinanan, M.: Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills. *IEEE Trans. Biomedical Eng.* 48(5), 579–591 (2001)
7. Reiley, C.E., 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, Part I. LNCS, vol. 5761, pp. 435–442. Springer, Heidelberg (2009)
8. Loukas, C., Georgiou, E.: Surgical workflow analysis with Gaussian mixture multivariate autoregressive (GMMAR) models: a simulation study. *Computer Aided Surgery* (2013)

9. Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M.: Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. *Computer Aided Surgery* 7(1), 49–61 (2002)
10. Leong, J.J.H., Nicolaou, M., Atallah, L., Mylonas, G.P., Darzi, A., Yang, G.Z.: 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)
11. Varadarajan, B.: Learning and inference algorithms for dynamical system models of dextrous motion. PhD thesis, Johns Hopkins University (2011)
12. Tao, L., Elhamifar, E., Khudanpur, S., Hager, G.D., Vidal, R.: Sparse hidden Markov models for surgical gesture classification and skill evaluation. In: Abolmaesumi, P., Joskowicz, L., Navab, N., Jannin, P. (eds.) *IPCAI 2012. LNCS*, vol. 7330, pp. 167–177. Springer, Heidelberg (2012)
13. Wolf, R., Duchateau, J., Cinquin, P., Voros, S.: 3D tracking of laparoscopic instruments using statistical and geometric modeling. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part I. LNCS*, vol. 6891, pp. 203–210. Springer, Heidelberg (2011)
14. Mung, J., Vignon, F., Jain, A.: A non-disruptive technology for robust 3D tool tracking for ultrasound-guided interventions. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part I. LNCS*, vol. 6891, pp. 153–160. Springer, Heidelberg (2011)
15. Richa, R., Bó, A.P.L., Poignet, P.: Robust 3D visual tracking for robotic-assisted cardiac interventions. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part I. LNCS*, vol. 6361, pp. 267–274. Springer, Heidelberg (2010)
16. Blum, T., Feußner, H., Navab, N.: Modeling and segmentation of surgical workflow from laparoscopic video. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part III. LNCS*, vol. 6363, pp. 400–407. Springer, Heidelberg (2010)
17. Lalys, F., Riffaud, L., Morandi, X., Jannin, P.: Automatic phases recognition in pituitary surgeries by microscope images classification. In: Navab, N., Jannin, P. (eds.) *IPCAI 2010. LNCS*, vol. 6135, pp. 34–44. Springer, Heidelberg (2010)
18. Lalys, F., Riffaud, L., Morandi, X., Jannin, P.: Surgical phases detection from microscope videos by combining SVM and HMM. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) *MICCAI 2010. LNCS*, vol. 6533, pp. 54–62. Springer, Heidelberg (2011)
19. Lalys, F., Riffaud, L., Bouget, D., Jannin, P.: An application-dependent framework for the recognition of high-level surgical tasks in the OR. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part I. LNCS*, vol. 6891, pp. 331–338. Springer, Heidelberg (2011)
20. Lalys, F., Riffaud, L., Bouget, D., Jannin, P.: A framework for the recognition of high-level surgical tasks from video images for cataract surgeries. *IEEE Transactions on Biomedical Engineering* 59(4), 966–976 (2012)
21. Fathi, A., Farhadi, A., Rehg, J.M.: Understanding egocentric activities. In: *IEEE International Conference on Computer Vision*, pp. 407–414 (2011)
22. Shi, Q., Cheng, L., Wang, L., Smola, A.J.: Human action segmentation and recognition using discriminative semi-markov models. *Int. Journal of Computer Vision* 93(1), 22–32 (2011)
23. Andrew, G.: A hybrid markov/semi-markov conditional random field for sequence segmentation. In: *Conf. on Empirical Methods in Natural Language Processing*, pp. 465–472 (2006)
24. Laptev, I.: On space-time interest points. *International Journal of Computer Vision* 64(2-3), 107–123 (2005)
25. Wang, Y., Tran, D., Liao, Z.: Learning hierarchical poselets for human parsing. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1705–1712 (2011)
26. Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y.: Large margin methods for structured and interdependent output variables. *J. of Machine Learning Research* 6, 1453–1484 (2005)
27. Joachims, T., Finley, T., Yu, C.N.J.: Cutting-plane training of structural svms. *Machine Learning* 77(1), 27–59 (2009)
28. Reiley, C.E., Lin, H.C., Varadarajan, B., Vagolgyi, B., Khudanpur, S., Yuh, D.D., Hager, G.D.: Automatic recognition of surgical motions using statistical modeling for capturing variability. In: *Medicine Meets Virtual Reality*, pp. 396–401 (2008)

Towards Intra-operative OCT Guidance for Automatic Head Surgery: First Experimental Results

Jesús Díaz Díaz¹, Dennis Kundrat¹, Kim-Fat Goh¹,
Omid Majdani², and Tobias Ortmaier¹

¹ Institute of Mechatronic Systems, Leibniz Universität Hannover,
Appelstr. 11a, D-30167 Hannover, Germany
{jesus.diazdiaz,dennis.kundrat,kimfat.goh,
tobias.ortmaier}@imes.uni-hannover.de
<http://www.imes.uni-hannover.de>

² Clinic for Laryngology, Rhinology and Otology, Hannover Medical School,
Carl-Neuberg-Str. 1, D-30625 Hannover, Germany
majdani.omid@mh-hannover.de

Abstract. In recent years, optical coherence tomography (OCT) has gained increasing attention not only as an imaging device, but also as a guidance system for surgical interventions. In this contribution, we propose OCT as an external high-accuracy guidance system, and present an experimental setup of an OCT combined with a cutting laser. This setup enables not only in situ monitoring, but also automatic, high-accuracy, three-dimensional navigation and processing. Its applicability is evaluated simulating a robotic assisted surgical intervention, including planning, navigation, and processing. First results demonstrate that OCT is suitable as a guidance system, fulfilling accuracy demands of interventions such as the cochlear implant surgery.

Keywords: optical coherence tomography, laser, navigation system, guidance system, cochlear implant surgery.

1 Introduction

The cochlear implant (CI) surgery is a surgical procedure during which an electrode is inserted into the cochlea in order to electrically stimulate the auditory nerve. Current research investigates the realization of a single-channel using *Robot Assisted Surgery* (RAS) and the direct insertion from the outer lateral skull to the cochlea. This surgical intervention demands an accuracy of 0.5 mm. In this contribution, we focus on the use of OCT as intra-operative monitoring and guidance system for this purpose.

OCT was established in 1991. Its working principle is based on the interference of back-reflected laser light from a sample with reference laser light in a Michelson interferometer. OCT typically has a resolution in the micron-scale and is highly sensitive. It is contact-free and, therefore, a nondestructive imaging device, capable not only of scanning the surface, but also of obtaining three-dimensional

tissue information. OCT is used in wide range of applications and usually as qualitative imaging system. With an increased utilization in the field of medical engineering, quantitative applications gain importance. The idea to use OCT as a guidance system may be as old as OCT itself. The visualization using OCT during a surgical procedure and its feedback, enables the user to control the tissue processing at the micron-scale. Boppart et al. [1] used and proposed OCT for surgical guidance by manually imaging a region of interest. Recently, more sophisticated approaches combining OCT with other tools have been developed and are used in a guided manner. In the field of ophthalmology, the integration of OCT in an microsurgical instrument enables the surgeon to perform OCT guided retinal microsurgery by visualization of internal structures of the eye [7]. Zhao et al. [6] combined OCT with MRI for neurosurgery guidance. During the insertion and while navigating with MRI, Zhao et al. use real-time 2D OCT to image adjacent structures and navigate the surgeon. These examples have in common, that the guidance is based on forward-imaging. The adjacent target region of the instrument is imaged with OCT, which is used as internal guidance system without a direct feedback to the planning.

In this contribution, we propose a novel setup of combined OCT and cutting laser as a monitoring, navigation and processing system for RAS in hard tissue. Moreover, we introduce OCT as an external navigation system for laser ablation. Since OCT is used as stand-alone external guidance system, intra-operative OCT data has to be matched to (pre-operative) planning data. In order to demonstrate OCT's suitability with regard to the stated accuracies, experiments are performed by simulating a surgical intervention, including planning, navigation and processing. The experimental setup, methods, and workflow are introduced in section 2. The results are presented in section 3 and discussed in section 4.

2 Setup and Methods

The following hardware components are part of the proposed system:

- tool for processing: cutting laser.
- tracking system: high-accurate OCT.
- tracking landmarks: spherical artificial fiducial landmarks.
- positioning system: high-accurate parallel robot.
- sample: imaging and navigation phantom.

Recent approaches for navigated material removal involve state-of-the-art optical tracking systems in an *eye-to-hand* configuration. When using OCT, an *eye-in-hand* configuration is more appropriate due to the limitation of the OCT's working distance. In clinical applications, a suitable robot iteratively positions the combined laser and OCT. The ablation procedure starts after reaching the target pose with respect to the patient. In the present paper, however, and only for the experimental setup, an *eye-to-hand* configuration is used to position the sample and not the tool. Methodology and relative motion with respect to the phantom nevertheless remain the same.

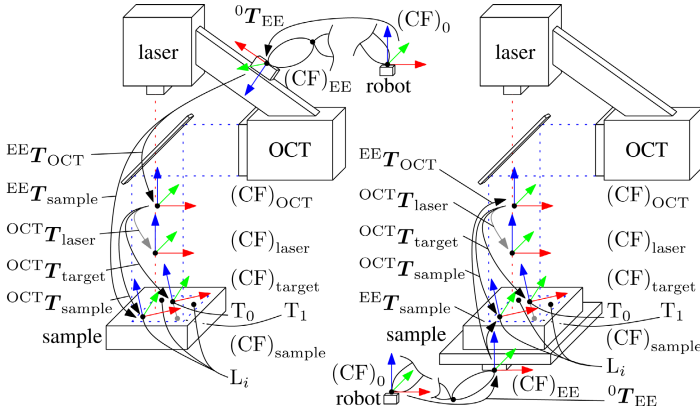


Fig. 1. *Eye-in-hand* (left) and *eye-to-hand* (right) of the experimental system with its components, coordinate frames and transformation matrices

2.1 Experimental Setup

The experimental system is sketched out in figure 1 and shown in figure 2 (left).

The cutting laser is the erbium-doped yttrium aluminium garnet (Er:YAG) laser of Pantec Biosolutions AG (model DPM-15). It is a pulsed solid-state laser with a wavelength of $\lambda_{\text{laser}} = 2940$ nm. The functionality of the laser is expanded with scan components. Therefore, the entity of laser and scanner is converted into a three axis laser system, defining a coordinate system $(\text{CF})_{\text{laser}}$. The working space of this entity, in the following just laser, has a working space of 10 mm in each dimension.

The OCT used in the optical setup is the system GANYMEDE of Thorlabs, Inc. The OCT has a center wavelength of $\lambda_{\text{OCT}} = 930$ nm. The maximum field of view has been enlarged to image approximately $20 \text{ mm} \times 20 \text{ mm} \times 2.7 \text{ mm}$, defining a coordinate system $(\text{CF})_{\text{OCT}}$. Furthermore, a geometric calibration [2] of this OCT has been performed in order to reduce the imaging error.

The optical paths of OCT and laser are combined by a dichroic mirror for an approximate co-axial propagation of the beams and a spatial overlap of the working spaces, keeping the relative configuration constant. This important feature enables an in situ imaging and control of the ablation process. We refer to [4] for further information and first results.

Using OCT as tracking device requires to adapt the tracking landmarks to this technology. In this contribution, we focus on artificial fiducial landmarks. Due to previous results, we choose spheres of titan with a diameter of 1 mm. The sample used for the evaluation of the navigation accuracy, i.e., the phantom, is composed of two parts (see figure 2 (right)). The first part is the carrier for the fiducial landmarks and the target area. This second part is a cuboid made of wood. The artificial fiducial landmarks are positioned not only on the front, but also on the flip side of the phantom for evaluation purposes, defining an upper

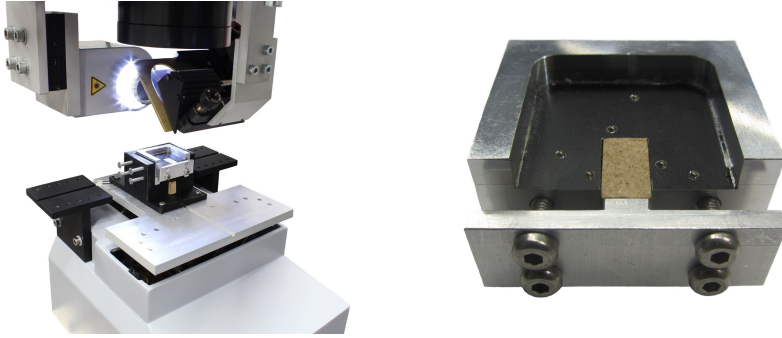


Fig. 2. Experimental setup (left) with OCT (upper left), laser (upper middle), phantom (middle), and robot (lower middle). Phantom (right) including fiducial landmarks and wooden target.

and lower fiducial landmark plane, respectively, and defining a sample coordinate frame $(CF)_{\text{sample}}$. The configuration of all fiducial landmarks has been measured with the coordinate measurement machine Zeiss ZMC 550. Both parts have a relevant depth of approximately 12 mm.

The phantom is positioned using a high accuracy parallel robot, the F-206.S HexAlign™ 6 Axis-Hexapod of Physik Instrumente (PI) GmbH & Co. KG.

2.2 System Calibration

The aim of this subsection is the description of the methods used to determine the transformation between rigid components. First, coordinate frames $(CF)_{\text{laser}}$ and $(CF)_{\text{OCT}}$ have to be registered, i.e., the homogenous transformation matrix ${}^{\text{OCT}}T_{\text{laser}}$ has to be determined. An arbitrary material is positioned in the common working space. We perform the ablation by a limited number of single pulses removing a small part of material. After appropriate filtering for noise reduction, the surface including the ablation spot is segmented using snakes [5], i.e., by choosing the curve in the image that minimizes an energy functional composed of internal and external energy. For the external part, an energy map based on the diffusion of the gradient vectors [9] have been used. Using the segmented curves, the volume centroid is calculated. Ablation and image processing is repeated while positioning the material in several different depths. The laser is described by a point and a direction. We use the data of ablation spots to calculate the point of origin by evaluating the spot size as a function of depth, and the direction by calculating the line of best fit in terms of least squares.

Second, the homogenous transformation matrix ${}^{\text{EE}}T_{\text{sample}}$ between the robot's end effector (EE) and the sample has to be determined. The basic idea is to select different poses of the robot, such that the sample's region of interest is in the OCT imaging volume. After positioning the EE, OCT images are acquired of the sample including the fiducial landmarks. The OCT data is processed automatically. The centroids of the fiducial landmarks are calculated with a template

matching algorithm using cross correlation. On the one hand, the localized center points of the fiducial landmarks define the sample coordinate frame $(CF)_{\text{sample}}$ with respect to the OCT coordinate frame $(CF)_{\text{OCT}}$, i.e., ${}^{\text{OCT}}\mathbf{T}_{\text{sample}}$. On the other hand, the transformation ${}^0\mathbf{T}_{\text{EE}}$ is defined through the relative position of robot EE with respect to robot basis. The workflow is repeated for different poses of the EE, acquiring pair of matrices ${}^0\mathbf{T}_{\text{EE}}^{(m)}$ and ${}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(m)}$ for the m -th repetition. Due to the hand-to-eye configuration, and in order to calculate the unknown ${}^{\text{EE}}\mathbf{T}_{\text{sample}}$, the following set of algebraic equations

$$A \cdot {}^{\text{EE}}\mathbf{T}_{\text{sample}} = {}^{\text{EE}}\mathbf{T}_{\text{sample}} \cdot B, \quad (1)$$

$$A = \left({}^0\mathbf{T}_{\text{EE}}^{(m)}\right)^{-1} \cdot {}^0\mathbf{T}_{\text{EE}}^{(n)}, \quad B = \left({}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(m)}\right)^{-1} \cdot {}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(n)} \quad (2)$$

is solved for a pair or set of matrices using methods introduced by Tsai et al. [8].

2.3 Navigation

The conventional workflow starts with an appropriate pre-operative imaging of the patient in order to acquire data the planning is going to be based on. This imaging, generally using CT, is omitted, since a ground-truth of the sample, being the patients replacement, is well-known. We realize the planning by localizing artificial tracking landmarks L_k ($k = 1, 2, \dots$), relative to which we define an entry T_0 and an exit target point T_1 , i.e., a target transformation matrix ${}^{\text{sample}}\mathbf{T}_{\text{target}}$. This enables us to define a target pose of the sample

$${}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(\text{target})} = {}^{\text{OCT}}\mathbf{T}_{\text{laser}} \cdot {}^{\text{laser}}\mathbf{T}_{\text{target}} \cdot \left({}^{\text{sample}}\mathbf{T}_{\text{target}}\right)^{-1}, \quad {}^{\text{laser}}\mathbf{T}_{\text{target}} = \mathbf{I}. \quad (3)$$

Intra-operatively, the iterative process is the alternation of tracking landmarks and comparing actual to target sample pose. The residual error of these two data sets is minimized performing a singular value decomposition of the weighted mean fiducial covariance matrix. Tracking is realized through calculating the centroid of the fiducial landmarks L_k ($k = 1, 2, \dots$) with a template matching algorithm using cross correlation. This results in the sample's pose in the i -th iteration ${}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(i)}$. The difference transformation between actual and target sample pose

$$\Delta\mathbf{T}_{\text{sample}}^{(i)} = \left({}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(i)}\right)^{-1} \cdot {}^{\text{OCT}}\mathbf{T}_{\text{sample}}^{(\text{target})}, \quad (4)$$

enables us to calculate the positioning of the robot's EE for iteration step $i + 1$

$${}^0\mathbf{T}_{\text{EE}}^{(i+1)} = {}^0\mathbf{T}_{\text{EE}}^{(i)} \cdot {}^{\text{EE}}\mathbf{T}_{\text{sample}} \cdot \Delta\mathbf{T}_{\text{sample}}^{(i)} \cdot \left({}^{\text{EE}}\mathbf{T}_{\text{sample}}\right)^{-1}. \quad (5)$$

The *fiducial registration error* (FRE) [3] of the two set of fiducials, target and actual, as well as the estimated *target registration error* (TRE) [3] of the entry target point are used as quality criterion. If these values fall below a certain threshold, we stop the navigation and start with the ablation.

The evaluation is carried out by imaging the upper as well as the bottom part of the sample after ablation and comparing actual and planned target points.

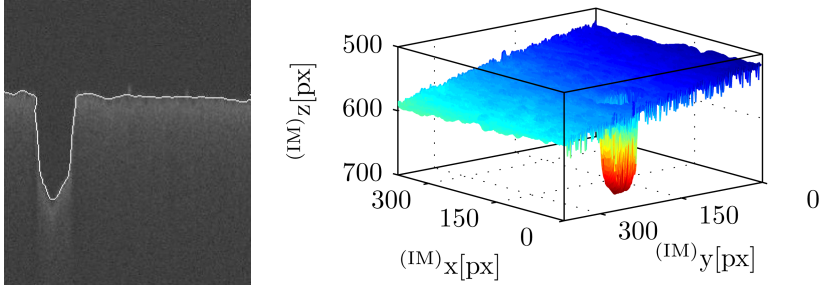


Fig. 3. 2D OCT image (B-scan) of ablation spot with segmented surface (left) and segmented 3D surface (right)

The errors err_{in} and err_{out} at the target entry and exit point, respectively, as well as the angle error err_{α} between trajectories are used for evaluation. The error $err_{35\text{ mm}}$ is extrapolated at a depth of 35 mm from the surface, which is approximately the distance from the outer lateral skull to the cochlea, using the intercept theorem and the entry and exit point.

3 Results

The registration of the OCT and laser has been carried out by positioning and cutting a sample of wood at 9 different axial positions. The (laser) parameters for current, puls duration, puls frequency, scanner coordinate have been chosen to $I = 220$ A, $\Delta t = 180$ μs , $f = 200$ Hz, and $\mathbf{x}_{laser} = (0, 0, 5)^T$ mm, respectively. Dense and calibrated volume OCT scans of the ablation spots with a spatial resolution of $8.2 \mu\text{m} \times 8.2 \mu\text{m} \times 2.6 \mu\text{m}$ for a scan region of $3.0 \text{ mm} \times 3.0 \text{ mm} \times 2.7 \text{ mm}$ have been acquired. An example of the image processing for calculating the center of the ablation spot in the OCT image data is given in figure 3 (left), showing an original 2D OCT image of an ablation spot. The segmented surface, i.e., the snake is superimposed. The segmentation of the surface of the complete 3D volume is presented in figure 3 (right). The real configuration of both system components is unknown, so the results can only be evaluated in terms of precision. The mean distance between localized ablation spots and line of best fit is $3.5 \mu\text{m}$ having a standard deviation of $1.5 \mu\text{m}$ (see figure 4 (left)).

The registration of the EE and sample has been carried out performing a hand-eye calibration positioning the EE, and, therefore, the sample in ten poses. The different poses have a translational width of maximal ± 2 mm and rotational width of maximal $\pm 5^\circ$. With the calibrated OCT measurement system, dense volume scans of the sample have been acquired. A spatial resolution of $15.0 \mu\text{m} \times 15.0 \mu\text{m} \times 2.6 \mu\text{m}$ for a scan region of $15.0 \text{ mm} \times 15.0 \text{ mm} \times 2.7 \text{ mm}$ has been chosen. Figure 4 shows for the i -th step the translational (middle) and rotational (right) part of the matrix ${}^{EE}\mathbf{T}_{sample}^{(i)} \cdot ({}^{EE}\mathbf{T}_{sample}^{(10)})^{-1}$, being ${}^{EE}\mathbf{T}_{sample}^{(i)}$ the result an hand-eye calibration with the the first i poses. Both registrations show high convergence and small residual errors.

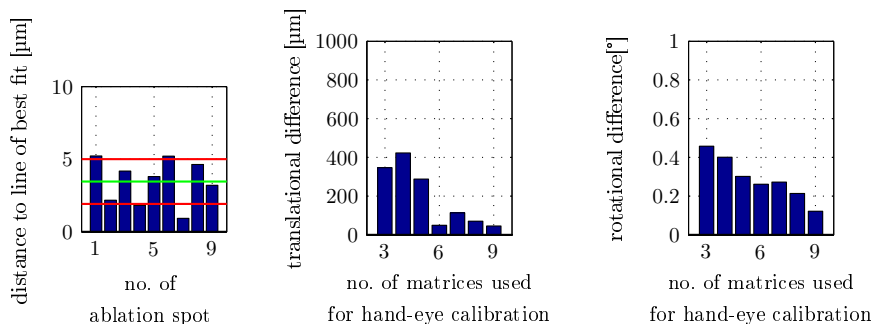


Fig. 4. Distance of localized ablation spots to line of best fit (left). Translational (middle) and rotational (right) error with respect to converged result $({}^{EE}\mathbf{T}_{\text{sample}}^{(10)})^{-1}$. The rotational error is the angle of the axis-angle representation of the difference matrix.

The key experiment including navigation and cutting has been carried out by pre-positioning the (evaluation) sample laterally in the center and axially approximately at the focal distance of OCT and cutting laser, respectively. With the calibrated OCT measurement system, dense volume scans have been acquired choosing the same parameters as for the hand-eye calibration. The (laser) parameters for current, puls duration, and puls frequency remain unchanged avoiding a possible "pointing" of the laser. The ablation is carried out performing the scanner a truncated cone geometry with a diameter of 3000 μm at the upper and 200 μm at the lower end of the cutting geometry. The height is of 10000 μm . The entry point T_0 is planned to be the center point of the three landmark fiducials on the upper side of the evaluation sample. The target exit point T_1 is defined through the intersection of the normal of the upper fiducial landmark plane and the lower fiducial landmark plane at an approximate distance from T_0 of 12 mm. The navigation and cutting is performed ten times. Generally, the iterations of the navigation, i.e., the repositioning of the robot, stop, when the FRE and TRE fall below a threshold of 10 μm . Then, we start with the ablation. The navigation errors at the last iteration step and the ablation errors are as follows:

	exp. 1	exp. 2	exp. 3	exp. 4	exp. 5	exp. 6	exp. 7	exp. 8	exp. 9	exp. 10
FRE [μm]	4.0	4.1	15.1	7.5	9.3	7.9	9.7	7.1	4.3	10.1
TRE [μm]	1.2	3.1	10.3	3.7	1.9	1.5	2.0	2.4	1.3	7.6
err _{in} [μm]	32.2	46.6	51.2	53.3	49.1	58.2	39.9	31.3	21.3	22.5
err _{out} [μm]	98.0	121.7	83.1	66.0	33.8	58.2	105.0	166.3	156.36	116.2
err _{35mm} [μm]	226.1	336.9	341.3	237.1	133.0	179.5	379.2	473.7	462.1	312.5
err _{α} [°]	0.3	0.5	0.64	0.44	0.28	0.27	0.68	0.76	0.76	0.49

The trials have been carried out in a series of two (exp. 1-2), four (exp. 3-6), and four (exp. 7-10) experiments. Each of these series has been performed with slight variations, e.g., with a different laser OCT registration. The three series

show consistent results, with all experiments fulfilling the necessary accuracy for CI surgery.

4 Conclusion

This contribution reports ten trials of OCT guided laser ablation, all of which consistently resulted in an error of less than 0.5 mm. Although the number of repetitions is not sufficient to assume statistical significance, the results fulfill the accuracy demands of interventions such as CI surgery and, thus, lend preliminary support to the assumption that OCT may be used as an external high-accuracy guidance system. Simulating a robotic assisted surgical intervention, we demonstrated the feasibility and potential of the combined setup of laser and OCT for navigation and processing.

Acknowledgments. The research reported in this paper was supported by the DFG (Deutsche Forschungsgemeinschaft) grants HE 2445/23-1, RE 1488/15-1, and MA 4038/3-1. We want to thank Christian Seiffert and Dipl.-Ing. Moritz Krauß from the Institute of Measurement and Automatic Control of the Leibniz Universität Hannover for the measurements they carried out, facilitating this contribution.

References

1. Boppart, S., Herrmann, J., Pitris, C., Bouma, B., Tearney, G.: Interventional optical coherence tomography for surgical guidance. In: Conf. Lasers and Electro-Optics 1998, pp. 123–124 (1998)
2. Díaz Díaz, J., Rahlves, M., Majdani, O., Reithmeier, E., Ortmaier, T.: A one step vs. a multi step geometric calibration of an optical coherence tomography. In: Proc. SPIE Photonics WEST/BiOS 2013. SPIE (2013)
3. Fitzpatrick, J., West, J., Maurer Jr., C.R.: Predicting error in rigid-body point-based registration. *IEEE Tran. Medical Imaging* 17(5), 694–702 (1998)
4. Fuchs, A., Schultz, M., Krüger, A., Kundrat, D., Díaz Díaz, J., Ortmaier, T.: Online measurement and evaluation of the er:yag laser ablation process using an integrated oct system. In: Proc. DGBMT Jahrestagung, pp. 434–437 (2012)
5. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *Int. J. Computer Vision*, 321–331 (1988)
6. Liang, C.P., Kim, I.K., Makris, G., Desai, J., Gullapalli, R.L., Chen, Y.: Concurrent multi-scale imaging combining optical coherence tomography with MRI for neurosurgery guidance. In: Proc. SPIE Photonics WEST/BiOS. SPIE (2013)
7. Liu, X., Balicki, M., Taylor, R.H., Kang, J.U.: Automatic online spectral calibration of fourier-domain oct for robotic surgery. In: Mahadevan-Jansen, A., Vo-Dinh, T., Grundfest, W.S. (eds.) Proc. SPIE Photonics WEST/BiOS, vol. 7890, p. 78900X. SPIE (2011)
8. Tsai, R., Lenz, R.: A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Tran. Robotics and Automation* 5(3), 345–358 (1989)
9. Xu, C., Prince, J.: Gradient vector flow: a new external force for snakes. In: IEEE Conf. Computer Vision and Pattern Recognition, pp. 66–71. IEEE (1997)

Configurable Automatic Detection and Registration of Fiducial Frames for Device-to-Image Registration in MRI-Guided Prostate Interventions

Junichi Tokuda, Sang-Eun Song, Kemal Tuncali,
Clare Tempany, and Nobuhiko Hata

Department of Radiology, Brigham and Women's Hospital and
Harvard Medical School, Boston, MA, USA
{tokuda,sam,ctempany,hata}@bwh.harvard.edu, ktuncali@partners.org

Abstract. We propose a novel automatic fiducial frame detection and registration method for device-to-image registration in MRI-guided prostate interventions. The proposed method does not require any manual selection of markers, and can be applied to a variety of fiducial frames, which consist of multiple cylindrical MR-visible markers placed in different orientations. The key idea is that automatic extraction of linear features using a line filter is more robust than that of bright spots by thresholding; by applying a line set registration algorithm to the detected markers, the frame can be registered to the MRI. The method was capable of registering the fiducial frame to the MRI with an accuracy of 1.00 ± 0.73 mm and 1.41 ± 1.06 degrees in a phantom study, and was sufficiently robust to detect the fiducial frame in 98% of images acquired in clinical cases despite the existence of anatomical structures in the field of view.

1 Introduction

Magnetic Resonance Imaging (MRI) is an advantageous option as an intra-operative imaging modality for image-guided prostate interventions. While transrectal ultrasound (TRUS) is the most commonly used imaging modality to guide core needle prostate biopsy in the United States, the limited negative predictive value of the TRUS-guided systematic biopsy has been argued [1]. To take advantage of MRI's excellent soft tissue contrast, researchers have been investigating the clinical utility of MRI for guiding targeted biopsies [2]. MRI-guided prostate biopsies are often assisted by needle guide devices [3, 4] or MRI-compatible manipulators [5–7]. These devices allow the radiologist to insert a biopsy needle accurately into the target defined within the MRI coordinate space.

Registering needle guide devices to the MRI coordinate system is essential for accurate needle placement [3–5]. These devices are often equipped with MR-visible passive markers to be localized in the MRI coordinate system. Because the locations of those markers in the device's own coordinate system are known,

one can register the device's coordinate system to the MRI coordinate system by detecting the markers on an MR image. However, the detection and registration of markers on an MR image are not always simple to achieve, because simple thresholding does not always provide robust automatic detection due to noise from other sources such as the patient's anatomy. Even if the markers are successfully detected, associating them with the individual markers is another hurdle for device-to-image registration. Existing methods rely on specific designs of fiducial frames or MR sequences [3, 4], restricting the device design.

In this paper, we propose a novel method for robust automatic fiducial frame detection and registration that can be applied to a variety of fiducial frame designs. The only requirement for the frame design is that the frame has at least three cylindrical MR-visible markers asymmetrically arranged. The key idea behind the method is that extraction of 3D linear features from cylindrical markers using a line filter is more robust than that of bright spots on the image by thresholding; by matching the cylindrical shapes detected on an MR image and a model of the fiducial frame, one can register the frame to the MRI coordinate system. We conducted phantom and clinical studies to evaluate the accuracy and the detection rate of the method using an existing MRI-visible fiducial frame [4].

2 Methods

2.1 Requirements for a Fiducial Frame

Our method is designed to detect a fiducial frame consisting of multiple MR-visible cylindrical markers on an MR image and register a model of the fiducial frame to the detected markers. The cylindrical markers can be an MR skin marker product, or sealed tubes filled with liquid that produces MR signal. The frame must be rotationally asymmetric to obtain a unique solution in marker registration. The configuration of the fiducial frame is modeled as a model line set $\{l_1^M, \dots, l_{NM}^M\}$ in our registration algorithm. Each line is described by a pair of position and direction vectors, \mathbf{p}_i^M and \mathbf{n}_i^M ($i = 1, \dots, N^M$); those vectors represent the coordinates of a point on the line and the direction vector of the line defined in the fiducial frame coordinate system respectively.

2.2 Detection of Cylindrical Markers on MRI

After a 3D or multi-slice MR image of the fiducial frame has been obtained, each individual marker of the frame is automatically segmented on the MR image using the 3D multi-scale line filter [8]. To distinguish the fiducial frame from other anatomical structures, we propose the following filtering steps. First, the line filter is applied to the image of the fiducial frame to highlight the 3D lines that have the same width as the cylindrical markers. The filter can target 3D lines of a specific width by σ_f , the standard deviation of the isotropic Gaussian function used to estimate the partial second derivatives. The filtered image is binarized with a threshold. At this stage, only the voxels within the line structures are

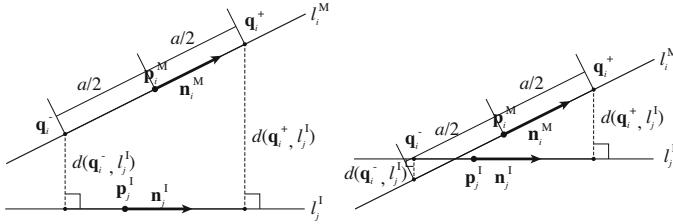


Fig. 1. We defined a distance function between one of the lines in the model of the fiducial frame, l_i^M , and one of the lines extracted from the image, l_j^I , using the distances from two points \mathbf{q}_i^- and \mathbf{q}_i^+ on line l_i^M to line l_j^I . \mathbf{q}_i^- and \mathbf{q}_i^+ are defined by point \mathbf{p}_i^M , direction vector \mathbf{v}_i^M and the distance to \mathbf{p}_i^M , $a/2$. The distance function gives zero only if the two lines match. Although the distance function depends on how \mathbf{p}_i^M is chosen, it does not depend on the location of \mathbf{p}_j^I along line l_j^I . Therefore, the distance function is insensitive to translation along line l_j^I during the registration process.

labeled '1', while the remaining voxels are labeled '0'. The voxels within the lines are then relabeled so that each segment has a unique voxel value. Each segment is examined based on its volume and dimensions. If the volume in a given segment is within a pre-defined range $[V_{\min}, V_{\max}]$, the length and width of the segment are assessed by computing the principal eigenvector of the distribution of the voxels in the segment. The segment is identified as a cylindrical marker only if its length along the principal eigenvector is close to the physical length of the markers. Once the segment is identified as a cylindrical marker l_j^I , the centroid of the segment is calculated as \mathbf{p}_j^I , and the principal eigenvector as \mathbf{n}_j^I ($j = 1, \dots, N^I$).

2.3 Registration of the Two Line Sets

Once the markers are identified as a line set $\{l_1^I, \dots, l_{N^I}^I\}$ on the MR image, the line set in the model $\{l_1^M, \dots, l_{N^M}^M\}$ is registered to $\{l_1^I, \dots, l_{N^I}^I\}$. The challenge here is that the transformation that registers the model to the MR image cannot be determined analytically, because an one-to-one correspondence between $\{l_1^M, \dots, l_{N^M}^M\}$ and $\{l_1^I, \dots, l_{N^I}^I\}$ has not been established. To address this challenge, we developed an approach similar to the Iterative Closest Line (ICL) [9]. The ICL is a point cloud registration algorithm alternative to the Iterative Closest Points (ICP) [10]; whereas the ICP registers two point clouds by iteratively associating points in the two clouds by nearest-neighbor criteria, the ICL registers them by associating linear features extracted from the point clouds. Unlike ICL, we compute the translation and rotation at once rather than computing them separately. To achieve this, we define a distance function, which becomes zero when two given lines match (Fig. 1). The two points on line l_i^M are defined by $\mathbf{q}_i^+ = \mathbf{p}_i^M + a\mathbf{n}_i^M/2$ and $\mathbf{q}_i^- = \mathbf{p}_i^M - a\mathbf{n}_i^M/2$, where the distance between the two points is a . The distances from those points to line l_j^I are:

$$d(\mathbf{q}_i^+, l_j^I) = \|(\mathbf{q}_i^+ - \mathbf{p}_j^I) - \{(\mathbf{q}_i^+ - \mathbf{p}_j^I) \cdot \mathbf{n}_j^I\} \mathbf{n}_j^I\| \quad (1)$$

$$d(\mathbf{q}_i^-, l_j^I) = \|(\mathbf{q}_i^- - \mathbf{p}_j^I) - \{(\mathbf{q}_i^- - \mathbf{p}_j^I) \cdot \mathbf{n}_j^I\} \mathbf{n}_j^I\| \quad (2)$$

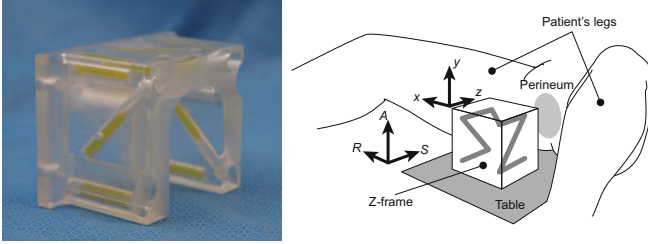


Fig. 2. A configuration of the Z-frame, which has been used in our clinical trial [4]. The Z-frame has seven rigid tubes with 7.5 mm inner diameters and 30 mm length filled with a contrast agent (MR Spots, Beekley, Bristol, CT) placed on three adjacent faces of a 60 mm cube, thus forming a Z-shaped enhancement in the images.

If we define the error function for line l_i^M and line l_j^I as:

$$E(l_i^M, l_j^I) = d(\mathbf{q}_i^+, l_j^I) + d(\mathbf{q}_i^-, l_j^I) \quad (3)$$

the error function between line l_i^M and the line set identified on the MRI, $L^I = \{l_1^I, \dots, l_{N^I}^I\}$, can be defined as:

$$E(l_i^M, L^I) = \min_{j \in \{1, \dots, N^I\}} E(l_i^M, l_j^I) \quad (4)$$

Finally, the linear transformation is computed by optimizing E using the same iterative approach as in ICP.

2.4 Experimental Setup

We evaluated the performance of the proposed fiducial detection algorithm using an existing fiducial frame called Z-frame that has been used for registration of a needle-guide template in our clinical study [4]. The existing registration algorithm estimates the position and orientation of the Z-frame with respect to a given 2D image plane based on distances between hyper-intensity dots, where the cylindrical markers intersect the slice plane. Although the automatic registration algorithm for the Z-frame works well for CT images [11], it encounters two problems when used for MRI. First, it often requires manual masking on the input image to exclude anatomical structures that lead to misidentification of tubes. Second, the existing algorithm is specialized for the Z-frame's shape and does not allow for different shapes. The rationales for using the Z-frame in our evaluation are as follows: 1) it allows direct comparison of registration accuracy with the established method; 2) it allows retrospective tests using existing clinical data, which provide realistic image features e.g. noise and structures other than the fiducial frame, including the patient anatomy.

Phantom Study. We fixed an acrylic base with a scale on the patient table of a 3 Tesla MRI scanner (MAGNETOM Verio 3T, Siemens AG, Erlangen, Germany)

to give known translations and rotations to the Z-frame. The scale allows the Z-frame to be placed at 0, 50, 100, 150, and 200 mm horizontally off the isocenter of the imaging bore, and tilted 0, 5, 10, 15, and 20 degrees horizontally from the \mathbf{B}_0 field. We evaluated the accuracy of the Z-frame registration, while translating the Z-frame along the Z-frame's X- and Y-axes and rotating around the X-, Y- and Z-axes i.e., *roll*, *pitch* and *yaw*, respectively (Fig. 2). The translation along the Z-axis was not considered, since the scanner can position the subject to its isocenter by moving the table. For the acquisition of the 3D images, we used the 3D Fast Low Angle Shot (FLASH) imaging sequence (TR/TE: 12 ms/1.97 ms; acquisition matrix: 256×256 ; flip angle 45° ; field of view: 160×160 mm; slice thickness: 2 mm; receiver bandwidth: 400 Hz/pixel; number of averages: 3). For each translation and rotation, eight sets of 3D images were acquired. The existing and proposed detection and registration methods were applied.

Clinical Study Using Existing Data. MRI data of the Z-frame were obtained during clinical MRI-guided prostate biopsies performed under a study protocol approved by the Institutional Review Board. Three-dimensional images of the Z-frame acquired at the beginning of each case were collected in 50 clinical cases, where the Z-frame was used to register the needle guide template. We performed automatic registration of the Z-frame using the proposed method. The results were visually inspected by overlaying the Z-frame model on the MRI.

For both studies, we used the medical image computing software, 3D Slicer 4.1 [12] running on a workstation (Apple Mac Pro, Mac OS X 10.7.1, CPU: Dual 6-Core Intel Xeon 2.66 GHz, Apple Inc., Cupertino, CA).

3 Results

Phantom Study. The parameters for the multi-scale line filter [8] were as follows: $\sigma_f = 3.0$, $\alpha_1 = 0.5$, $\alpha_2 = 2.0$. We used threshold for the Hessian matrix = 13.0, $[V_{\min}, V_{\max}] = [300\text{mm}^3, 2500\text{mm}^3]$, and minimum length of principal axis = 10 mm. Registration of the Z-frame on all MR images was successfully completed without tuning the parameters. Fig. 3 shows the errors between translations and rotations of the Z-frame estimated from the proposed registration method and measured on the scale. The average time for computation was 4.3 seconds per image. Table 1 shows a comparison between the registration accuracy of the proposed algorithm and that of the existing algorithm.

Table 1. Comparison between the registration accuracy of the existing (original) algorithm and the proposed algorithm using the Mann-Whitney U test

	X (mm)	Y (mm)	Roll (deg)	Pitch (deg)	Yaw (deg)
Original	-1.08 ± 0.80	-1.44 ± 1.83	-0.70 ± 0.97	-1.55 ± 1.55	0.04 ± 0.05
Proposed	-1.00 ± 0.73	-0.38 ± 0.44	-1.41 ± 1.06	-0.87 ± 0.66	0.01 ± 0.13
<i>p</i> -value	0.5	0.005	0.01	0.1	9.0×10^{-6}

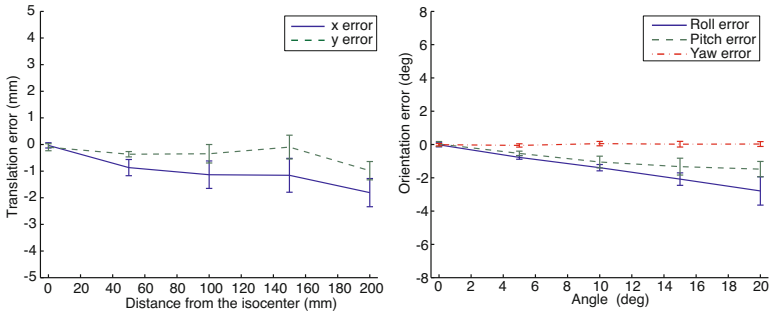


Fig. 3. The plots shows the mean and standard deviations of the translational registration errors when the fiducial frame is placed at 0, 50, 100, 150, and 200 mm horizontally off the isocenter and the rotational registration errors when it was tilted 0, 5, 10, 15, and 20 degrees around the X, Y and Z axis of the frame from its original position

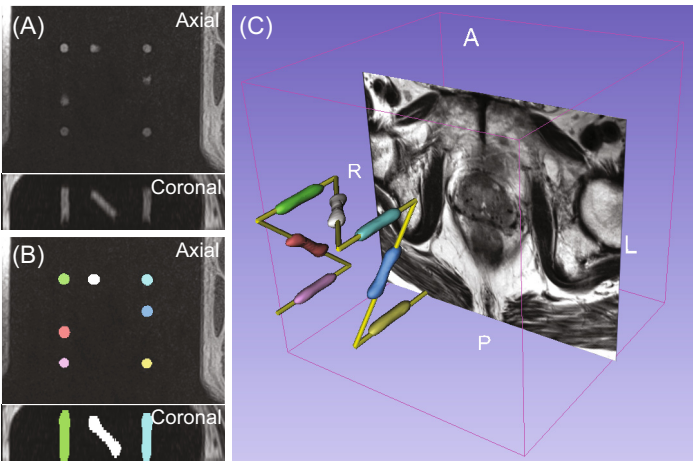


Fig. 4. (A) The original 3D MR image of the Z-frame acquired in the clinical study presents the sections of the cylindrical markers and the thighs of the patient. (B) The segmented markers are overlaid onto the original MR image. The segmented area was relabeled so that each segment has a unique voxel value presented as unique color. (C) The surface models of the segmented markers are shown with the model of the fiducial frame and an axial slice of T2-weighted prostate MRI in the 3D space.

Clinical Study Using Existing Data. The same parameters were used in the clinical study. Visual inspection of the results (Fig. 4) showed that, the Z-frame was successfully registered in 49 cases (98%). In one case, threshold values for minimum and maximum volume of markers had to be adjusted to achieve successful registration. The average computation time was 5.6 seconds.

4 Discussion and Conclusion

In this paper, we proposed a novel method for robust automatic fiducial frame detection and registration that can be applied to a variety of fiducial frame designs. The phantom study demonstrated that the proposed method was capable of registering the model of the fiducial frame to the MRI with an accuracy of 1.00 ± 0.73 mm and 1.41 ± 1.06 degrees. The clinical study demonstrated that the method was sufficiently robust to detect the fiducial frame with a success rate of 98% without any manual operation.

The use of cylindrical markers is essential to the proposed method. Our assumption is that automatic extraction of 3D linear features from cylindrical markers on the input image is more robust than that of spherical markers or sections of cylindrical markers because the Hessian matrix can selectively highlight the linear structures with a specific width, and once the linear structures are extracted, several criteria e.g. volume and size in primary and secondary axes, can be applied to filter out unwanted structures. Moreover, thanks to the approach's use of lines instead of points, the method is less prone to detection error due to MR signal defects than the other approaches that rely on simple threshold. In practice, signal defects are often caused by bubbles in capsules of liquid-based MR-visible markers. However, the signal defects can still impact the registration accuracy in our approach, because a line is identified as the eigenvectors of the voxel distribution in the segmented markers. This might explain why the registration error of the proposed method was significantly higher than the existing method in Roll but not in the other directions; for the existing technique, only the slices without any signal defect in the markers were manually selected, whereas the proposed method relies on the entire 3D image. Krieger *et al* proposed the use of template matching to minimize the effect of bubbles [5].

The proposed method provides several advantages over other methods for fully automated device-to-image registration. First, it only relies on passive markers and does not require any embedded coil or MR pulse sequence to enhance the signal from the markers. Second, the algorithm does not assume any particular frame design for automatic detection and registration. The only requirement for the fiducial frame design is the use of more than three cylindrical markers asymmetrically arranged. Such flexibility allows automatic detection and registration of a wide variety of needle guide devices. Third, the algorithm does not require any modification of its implementation in order to be adapted to a particular fiducial frame design. It only requires modifying a model of the frame and parameters, which can be provided as a configuration file. Therefore, even developers who are not specialized in image processing can design and implement device-to-image registration. Those advantages help developers to design needle guide devices with less effort and fewer constraints.

In conclusion, we propose a novel method for robust automatic fiducial frame detection and registration that can be used for a variety of fiducial frame configurations for device-to-image registration in MRI-guided interventions. The phantom and clinical studies demonstrate that the method provides accurate and robust automatic detection and registration of fiducial frames.

Acknowledgements. This work is supported by the National Institute of Health (NIH) (R01CA111288, P01CA067165, P41RR019703, P41EB015898, R01CA124377, R01CA138586), Center for Integration of Medicine and Innovative Technology (CIMIT 11-325), Siemens Seed Grant Award.

References

1. Roehl, K.A., Antenor, J.A., Catalona, W.J.: Serial biopsy results in prostate cancer screening study. *J. Urol.* 167(6), 2435–2439 (2002)
2. Hambrock, T., Futterer, J.J., Huisman, H.J., Hulsbergen-vandeKaa, C., van Basten, J.P., van Oort, I., Witjes, J.A., Barentsz, J.O.: Thirty-two-channel coil 3T magnetic resonance-guided biopsies of prostate tumor suspicious regions identified on multimodality 3T magnetic resonance imaging: technique and feasibility. *Invest. Radiol.* 43(10), 686–694 (2008)
3. de Oliveira, A., Rauschenberg, J., Beyersdorff, D., Semmler, W., Bock, M.: Automatic passive tracking of an endorectal prostate biopsy device using phase-only cross-correlation. *Magn. Reson. Med.* 59(5), 1043–1050 (2008)
4. Tokuda, J., Tuncali, K., Iordachita, I., Song, S.E., Fedorov, A., Oguro, S., Lasso, A., Fennessy, F.M., Tempany, C.M., Hata, N.: In-bore setup and software for 3T MRI-guided transperineal prostate biopsy. *Phys. Med. Biol.* 57(18), 5823–5840 (2012)
5. Krieger, A., Iordachita, I., Guion, P., Singh, A.K., Kaushal, A., Menard, C., Pinto, P.A., Camphausen, K., Fichtinger, G., Whitcomb, L.L.: An MRI-compatible robotic system with hybrid tracking for MRI-guided prostate intervention. *IEEE Trans. Biomed. Eng.* 58(11), 3049–3060 (2011)
6. Yakar, D., Schouten, M.G., Bosboom, D.G., Barentsz, J.O., Scheenen, T.W., Futterer, J.J.: Feasibility of a pneumatically actuated MR-compatible robot for transrectal prostate biopsy guidance. *Radiology* 260(1), 241–247 (2011)
7. Zangos, S., Melzer, A., Eichler, K., Sadighi, C., Thalhammer, A., Bodelle, B., Wolf, R., Gruber-Rouh, T., Proschek, D., Hammerstingl, R., Muller, C., Mack, M.G., Vogl, T.J.: MR-compatible assistance system for biopsy in a high-field-strength system: initial results in patients with suspicious prostate lesions. *Radiology* 259(3), 903–910 (2011)
8. Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G., Kikinis, R.: Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Med. Image Anal.* 2(2), 143–168 (1998)
9. Alshawa, M.: ICL: Iterative closest line a novel point cloud registration algorithm based on linear features. *Ekscentar* 10, 53–59 (2007)
10. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14(2), 239–256 (1992)
11. Susil, R.C., Anderson, J.H., Taylor, R.H.: A single image registration method for CT guided interventions. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 798–808. Springer, Heidelberg (1999)
12. Gering, D.T., Nabavi, A., Kikinis, R., Hata, N., O'Donnell, L.J., Grimson, W.E., Jolesz, F.A., Black, P.M., Wells, W.M.: An integrated visualization system for surgical planning and guidance using image fusion and an open MR. *J. Magn. Reson. Imaging* 13(6), 967–975 (2001)

Robust Intraoperative US Probe Tracking Using a Monocular Endoscopic Camera

Uditha L. Jayarathne², A. Jonathan McLeod²,
Terry M. Peters^{1,2}, and Elvis C.S. Chen¹

¹ Robarts Research Institute, London, Ontario, Canada

² Western University, London, Ontario, Canada

{ujayarat, jmcleod, tpeters, chene}@robarts.ca

Abstract. In the context of minimally-invasive procedures involving both endoscopic video and ultrasound, we present a vision-based method to track the ultrasound probe using a standard monocular video laparoscopic instrument. This approach requires only cosmetic modification to the ultrasound probe and obviates the need for magnetic tracking of either instrument. We describe an Extended Kalman Filter framework that solves for both the feature correspondence and pose estimation, and is able to track a 3D pattern on the surface of the ultrasound probe in near real-time. The tracking capability is demonstrated by performing an ultrasound calibration of a visually-tracked ultrasound probe, using a standard endoscopic video camera. Ultrasound calibration resulted in a mean TRE of 2.3mm, and comparison with an external optical tracker demonstrated a mean FRE of 4.4mm between the two tracking systems.

1 Introduction

Many procedures can be performed using endoscopy as a viable alternative to open surgery. Such minimally-invasive approaches can reduce recovery time, length of hospital stay, and morbidity. During such procedures, surgeons employ an endoscopic camera to view the organ surface and an endoscopic ultrasound (US) probe to visualize structures within the organ. In a typical surgical configuration, the video and US images are presented separately and in 2D. The surgeon must, therefore, perform spatial reasoning to mentally map the US image onto the video. Furthermore, the 2D nature of these images results in decreased depth perception.

Navigated endoscopy incorporates a spatial tracking device to infer the pose of the US probe relative to the camera, allowing US images to be registered to, and fused with, the video. The fused image is more intuitive, and lowers the cognitive load by eliminating the mental transform between the two images, which may improve hand-eye coordination of the surgeon. Examples of such systems include tumour resections using magnetic [4], robotic [9], photoacoustic [3], and vision-based [12] tracking systems, with Langø et al. [8] providing a comprehensive overview. Systems that employ extrinsic tracking devices increase cost, impact surgical workflow, require additional sterilization, and introduce other

limitations such as line-of-sight and metal interference issues for optical and magnetic tracking solutions respectively. Intrinsic tracking using the endoscopic camera to perform spatial measurements [12] does not incur these limitations, but accuracy, robustness, and the need to modify standard surgical instrument may be of concern. In particular, vision-based tracking in an endoscopic environment may be subject to lighting conditions, erroneous feature detection, and occlusion.

In this paper, a vision-based tracking system using a standard surgical laparoscopic camera is presented. A 3D marker was designed and rigidly attached to the semi-cylindrical back surface of a standard linear laparoscopic US probe, providing a set of features that can be reliably detected in the video sequence. Based on an Extended Kalman Filter framework, the proposed system is capable of tracking the 3D marker in 6 degrees of freedom, in the presence of spurious and/or missing features, even under strong specular lighting conditions. The tracking capability of the proposed system is demonstrated by performing an US probe calibration, along with both visual and quantitative validation procedures.

2 Methods and Materials

Given a set of known 3D model points M , and a set of 2D observed features U as imaged by the camera, the process of 2D-to-3D registration can be treated as a Perspective-n-Point (PnP) problem. In general, a PnP solution requires the correspondence between the 2D and 3D features to be established, as well as the optical properties of the imaging system to be known [13].

In practice, spurious features are often detected in the image, and features may be missing due to occlusion or strong specular lighting conditions. Therefore, to ensure robust operation of PnP algorithms, it is advantageous to address both the correspondence and pose estimation problems simultaneously. The proposed method employs an Extended Kalman Filter (EKF) framework [11] to constrain the search for feature correspondence. Given a calibrated camera and a 3D marker, possible poses of the 3D marker with respect to the camera are represented as a Gaussian Mixture Model (GMM). Uncertainty relating to the pose of the marker is propagated from the 3D model space to the 2D image space using EKF equations, constraining the search space for feature correspondence in a sequential fashion. A globally convergent PnP algorithm is applied to further refine the estimation established with at least 4 correct correspondences.

2.1 Hardware Setup

An “X-Corner” fiducial pattern (Fig. 1) was rigidly attached to the curved back-surface of a linear endoscopic US probe (UST-5536-7.5, Aloka, Japan) to serve as the 3D marker. A local 3D coordinate system of the pattern was defined, with the locations of the black-and-white intersections (i.e. X-corners) accurately determined using a measuring microscope (STM6-LM, Olympus, Japan).

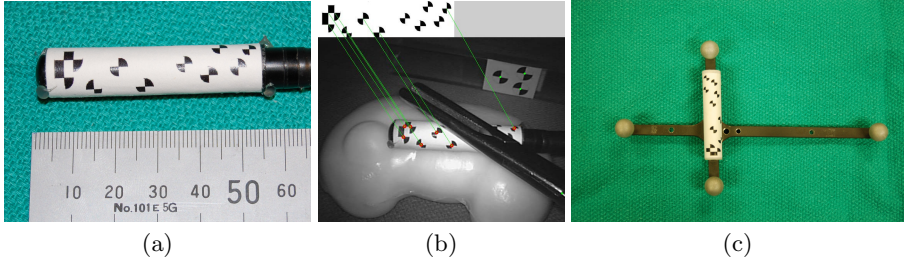


Fig. 1. (a) a 3D X -corner pattern affixed to a linear US probe, (b) the proposed method is able to establish the 2D to 3D correspondence even in the presence of spurious (top right) and missing features (those occluded by surgical grasper), and (c) a surrogate marker designed for validation using an optical tracking system

To locate the X -corners in the video image, a corner detection algorithm [1] specifically designed to detect X -corners was employed. Compared to generic corner and edge detectors [7], this algorithm tends to detect less clutter and is more computationally efficient. Once detected, the locations of X -corners are further refined to sub-pixel accuracy [2]. The camera (Surgical laparoscope, Olympus) was calibrated in a standard fashion using a planar pattern [2]. Both video and US images were captured at an image size of 640×480 pixels (Morphis, Matrox, Canada), and the video was corrected for both tangential and radial distortion.

2.2 Simultaneously Solving for the Pose and Correspondence

The pose of the US probe is represented as a 6D vector: 3 components representing the rotation in Rodrigues' form [2] and 3 representing the translation. It is assumed that the US probe can be freely displaced within the viewing frustum of the endoscopic camera, with a depth ranging from 5cm to 20cm. The orientation of the US probe mimicked realistic surgical conditions. Under these range constraints, pose samples were simulated and a Gaussian Mixture Model (GMM) was learned [5] offline to represent the 6D pose space. The learned GMM was then used to provide the initial pose and correspondence of the US probe by minimizing the following error function [11]:

$$Error(p) = \sum_{(m,u) \in Matches} \|u - Proj(p; m)\| + \tau |NotDetected| \quad (1)$$

where u is a detected X -corner, $Proj(p; m)$ is the projection of a model point m using the pose p , τ is a penalty term for unmatched points, and $|NotDetected|$ denotes the cardinality of the undetected feature set. Uncertainty in pose space is propagated to image space using EKF equations and used to constrain the matching between the detected X -corners and model points. Once a correspondence is found, the pose and the associated uncertainty are updated using EKF equations to further reduce the search space for subsequent correspondence

matching. This process is repeated until at least 4 correct correspondences are established, by which time the initial pose estimation given by the GMM is updated in response to the found correspondence. This estimated pose is refined by solving the generic PnP problem with a globally convergent algorithm [10].

The GMM, which provides an accurate pose prior but is computationally expensive, is nevertheless used to provide the initial pose estimation for the first video frame. For all subsequent frames, the pose provided by the PnP algorithm at frame i is employed as a motion prior for frame $i + 1$. Standard EKF equations are used to propagate the pose and the associated covariance throughout the video sequence.

2.3 Laparoscopic Ultrasound Calibration

The tracked US probe was calibrated [6] using a Z -phantom with 2 line fiducials arranged in a Z pattern. The Z -phantom was filled with polyvinyl chloride-plastisol (PVC) compound, which serves as a clear tissue-mimicking medium (Fig. 2(b)). A rectangular checkerboard pattern was rigidly attached to the planar surface of the phantom, and its pose was determined using OpenCV [2]. Once the homologous features were determined in both the image and camera space, the calibration was solved using the standard Orthogonal Procrustes algorithm.

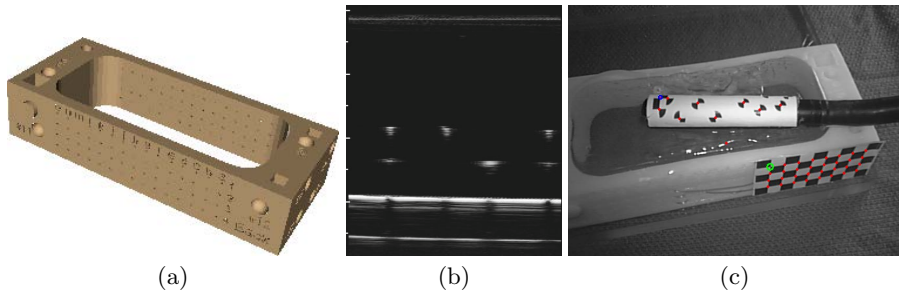


Fig. 2. (a) a CAD model of the Z -phantom (designed by PERK <http://perk.cs.queensu.ca>), (b) an US image showing 2 Z -patterns, and (c) pose of the US probe/ Z -phantom was determined using the proposed method and OpenCV [2], respectively

2.4 Validation Using Optical Tracker

A surrogate 3D marker, similar to that attached to the US probe, was constructed and rigidly attached to an optical Dynamic Reference Body (DRB) (Fig. 1c). The 3D marker and the optical DRB were carefully coregistered using a calibrated stylus. This setup was simultaneously tracked using both the proposed method and an optical tracking system (OTS) (Vicra, NDI, Canada). A total of 75 measurements were made within the depth range between 92mm to 99mm, typical of an endoscopic intervention.

3 Results

A total of 17 US/video image pairs were acquired for calibration. In each US image, the two fiducials, corresponding to the cross line of the Z-phantom (Fig. 2) were manually identified to obtain a total of 34 homologous points in both the US image and the 3D space. The calibration was accomplished using the Orthogonal Procrustes algorithm and resulted in a mean Fiducial Registration Error (FRE) of 2.2mm with a standard deviation of 0.9mm. A leave-one-out cross validation was performed where one of the 17 image pairs was removed prior to the calibration and the two fiducials on this image were then used as target points. This cross validation was repeated 17 times with different validation images, resulting in a mean Target Registration Error (TRE) of 2.3mm and a standard deviation of 1.0mm. Table 1 summarizes the result.

Table 1. US probe calibration result using a total 17 tracked US images

	mean error (mm)	standard deviation (mm)
Calibration FRE	2.2	0.9
leave-one-out TRE	2.3	1.0

Once US calibration is established, it is used to fuse the US image onto the endoscopic video. For the purpose of visual validation, two phantoms were constructed. The first comprises a spine vertebra semi-submerged into a PVC block; the second, a surgical needle inserted through a block of PVC. Figure 3(b) clearly shows that the outline of the spine forms a continuous contour, verifying that both visual tracking and US calibration are performing as expected. The proposed method is able to track the US probe in the presence of feature occlusion (Fig. 3(c,d)), as well as under ambient and endoscopic lighting conditions (Fig. 3(e,f)). Note that in the fused image, the needle seen in US is in alignment with the ends of the needle protruding from the phantom.

A total of 75 measurements of the surrogate marker were taken simultaneously using the endoscopic camera and the OTS. The tracking accuracy was evaluated at a set of 3D points lying on a virtual US fan created using the calibration obtained from the previous experiment. The 3D location of each pixel in the US fan was determined in both the OTS and the camera coordinate systems, allowing the two tracking systems to be coregistered. The FRE of this registration incorporates tracking errors from the two tracking systems as well as the registration error between the 3D marker and the optical DRB. Assuming the OTS is the gold standard, and the obtained registration is valid over the region of interest, the FRE represents the tracking error of the proposed method propagated to the US fan.

Figure 4 depicts the mean FRE of the camera coregistration using US image pixels as fiducials. The mean FRE exhibits a radial pattern: pixels close to the

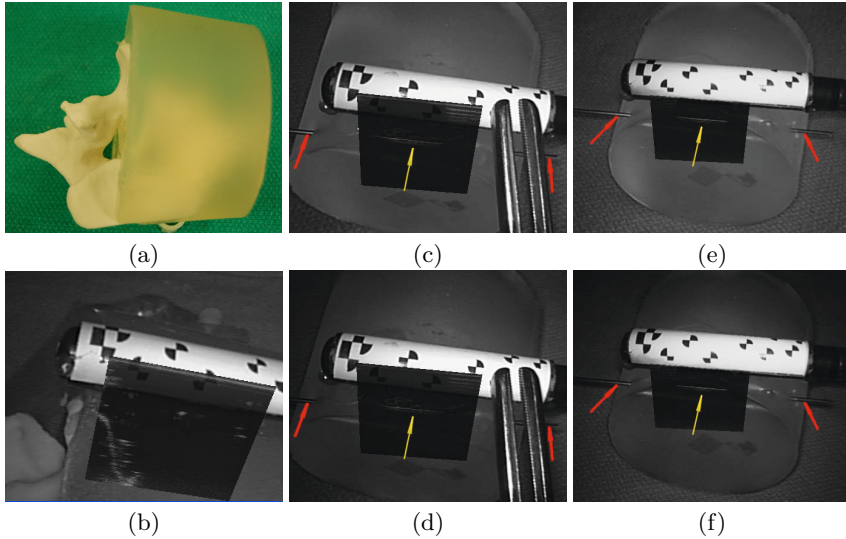


Fig. 3. Visual validation of tracking and US calibration: (a) a spine phantom, (b) US overlay showing a continuous contour of the spine outline under endoscopic lighting conditions. Needle phantom with (c,e) ambient lighting, (c,d) occlusion by surgical tool, (e,f) no occlusion, and (d,f) under endoscopic lighting conditions. Red arrows point to the ends of the needle while the yellow arrow points to its image in US. Note the dramatic improvement in the visual perception of the US beam with respect to the US probe, when the probe occludes the upper part of the US image, compared to the more naïve rendering of simply overlaying the US image on the video scene.

origin of the 3D marker exhibit a smaller mean FRE, which gradually increases as the pixel location is moved further away from the origin. This radial pattern suggests that the main contribution of the tracking error is from the rotational error of pose estimation. Over the acquired 75 measurements, the mean FRE of all pixels was 4.4mm, with a standard deviation of 3.3mm. The per-pixel mean FRE ranges from a minimum of 4.2mm to a maximum of 4.7mm. In our experience, this error is over-estimated due to the coregistration between the checker pattern and the optical DRB. Nevertheless, these results provide an upper-bound on the error introduced by the proposed method. The visual validation results suggest that the actual tracking error is substantially lower.

4 Discussion and Conclusion

A vision-based tracking procedure using a standard surgical monocular camera is proposed. By tracking a 3D pattern that is rigidly attached to a surgical tool, this approach operates under endoscopic lighting conditions and in the presence of occlusion. As with other vision-based systems, the limitation of this method is that it requires a fixed-focus camera that adheres to the pin-hole

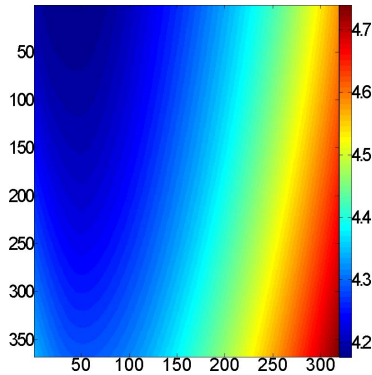


Fig. 4. Per-pixel mean FRE (in mm) of the camera-OTS coregistration shown in US image space. The x and y axis are the US image axis, with the top-left corner being closest to the origin of the 3D marker.

camera model. In contrast to other systems, this approach only requires trivial, cosmetic modification of the tracked object (endo-ultrasound probe).

One advantage of this approach is the ability to track a marker defined on a curved, rather than a planar surface, and can thus be adapted to a wide range of surgical tools. The current implementation tracks X -corners, but any feature that can be reliably detected in a video sequence can be used. Modifications to surgical tools are therefore minimized, and since no extrinsic tracking system is used, additional cost is minimal.

This method was used to calibrate a tracked US probe, and US image overlay gives a visually consistent augmentation. The reported US calibration FRE and TRE are in the order of 2mm which, in our experience, is comparable to results obtained using a magnetic tracking system. An optical tracking system was used to validate the accuracy of the proposed system. Using a surrogate marker to represent that attached to the US probe, simultaneous tracking by both systems was performed, allowing the local coordinate systems of the two trackers to be coregistered. Using the measurements from the OTS as the gold standard, camera tracking proved to be consistent with the OTS. The error of camera tracking exhibits a radial pattern, suggesting that the main contributor of the error is the rotational component of the pose estimation. The reported mean FRE is 4.4mm with a standard deviation of 3.3mm. This FRE is a combined system error including the co-calibration error of the 3D marker and DRB, which provides an upper-bound for the tracking error of the proposed system. Further analysis is warranted to provide a more accurate estimate of this error.

The proposed system currently performs at about 5 frames per second, but can be dramatically improved through more efficient implementation. Future work includes 1) pattern design of the 3D marker to optimize tracking accuracy and range of tracking angle, 2) integration into standard endoscopic procedure,

3) improvements to the algorithm to track multiple surgical instruments, and 4) extending this algorithm to track intrinsic visual features of the surgical tools, eliminating the need for extrinsic 3D markers entirely.

Acknowledgments. This work was supported by Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, and Canadian Foundation for Innovation. The authors thank Mohammad Ali Tavallaei and John Moore for their laboratory assistance.

References

1. Bennett, S., Lasenby, J.: ChESS - Quick and Robust Detection of Chess-board Features. Computing Research Repository (CoRR) abs/1301.5491 (2013)
2. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
3. Cheng, A., Kang, J.U., Taylor, R.H., Boctor, E.M.: Direct 3D Ultrasound to Video Registration Using Photoacoustic Effect. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 552–559. Springer, Heidelberg (2012)
4. Cheung, C.L., Wedlake, C., Moore, J., Pautler, S.E., Peters, T.M.: Fused Video and Ultrasound Images for Minimally Invasive Partial Nephrectomy: A Phantom Study. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 408–415. Springer, Heidelberg (2010)
5. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. *IEEE Trans. on PAMI* 24(3), 381–396 (2002)
6. Gobbi, D.G., Comeau, R.M., Lee, B.K., Peters, T.M.: Correlation of pre-operative MRI and intra-operative 3D ultrasound to measure brain tissue shift. In: Medical Imaging. Proc. SPIE 3982, pp. 77–84 (2000)
7. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of Fourth Alvey Vision Conference, pp. 147–151 (1988)
8. Langø, T., Vijayan, S., Rethy, A., Våpenstad, C., Solberg, O., Mårvik, R., Johnsen, G., Hernes, T.N.: Navigated Laparoscopic ultrasound in abdominal soft tissue surgery: technological overview and perspectives. *International Journal of Computer Assisted Radiology and Surgery* 7(4), 585–599 (2012)
9. Leven, J., et al.: DaVinci Canvas: A Telerobotic Surgical System with Integrated, Robot-Assisted, Laparoscopic Ultrasound Capability. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 811–818. Springer, Heidelberg (2005)
10. Lu, C.P., Hager, G.D., Mjolsness, E.: Fast and globally convergent pose estimation from video images. *IEEE Trans. on PAMI* 22(6), 610–622 (2000)
11. Moreno-Noguer, F., Lepetit, V., Fua, P.: Pose Priors for Simultaneously Solving Alignment and Correspondence. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 405–418. Springer, Heidelberg (2008)
12. Pratt, P., Di Marco, A., Payne, C., Darzi, A., Yang, G.-Z.: Intraoperative Ultrasound Guidance for Transanal Endoscopic Microsurgery. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 463–470. Springer, Heidelberg (2012)
13. Wu, Y., Hu, Z.: PnP Problem Revisited. *Journal of Mathematical Imaging and Vision* 24(1), 131–141 (2006)

Automatic Detection of Multiple and Overlapping EP Catheters in Fluoroscopic Sequences

Fausto Milletari, Nassir Navab, and Pascal Fallavollita

Chair for Computer Aided Medical Procedures, Technische Universität München, Germany
fausto.milletari@gmail.com, nassir.navab@tum.de,
fallavol@in.tum.de

Abstract. We propose a method to perform automatic detection of electrophysiology (EP) catheters in fluoroscopic sequences. Our approach does not need any initialization, is completely automatic, and can detect an arbitrary number of catheters at the same time. The method is based on the usage of blob detectors and clustering in order to detect all catheter electrodes, overlapping or not, within the X-ray images. The proposed technique is validated on 1422 fluoroscopic images yielding a tip detection rate of 99.3% and mean distance of 0.5mm from manually labeled ground truth centroids for all electrodes.

1 Introduction

Sudden cardiac death (SCD) is an unexpected death due to cardiac-related complications occurring in a short amount of time. SCD accounts for approximately 325,000 deaths per year in the United States alone. The trend of increasing SCD in developing or developed nations is thought to reflect a change in dietary and lifestyle habits in these nations. It is estimated that SCD claims more than 7,000,000 lives per year worldwide [1]. Most cases of SCD are related to cardiac arrhythmias. The most common electro physiological mechanisms leading to SCD are tachyarrhythmias such as ventricular fibrillation (VF) or ventricular tachycardia (VT). These disorders are frequently treated by radiofrequency (RF) catheter ablation. The precise localization of the arrhythmogenic site and positioning of the RF catheter at that site are problematic. These shortcomings can both reduce the efficacy of the procedure and increase the surgery time to several hours, especially with complex arrhythmias. To shorten the duration of RF catheter ablation and increase efficiency for treatment, commercial mapping systems providing 3D volume and color display of the cardiac chamber and electrical activation sequence have been developed. A review of these is presented in [2].

State-of-the-Art: In recent research practice, the medical imaging community has refocused its efforts to detect catheters directly in X-ray images. Many different types of catheters are used during EP procedures, each having specific configuration electrodes (e.g. size, shape). These electrodes are used for the measurement of electrical signals within the heart and also for the delivery of radiofrequency energy during treatment. Accurate and robust localization of catheters in the X-ray images

can provide enhanced functionality during procedures for guidance and also for post-procedural analysis. A crucial application of catheter localization is to record the position of the ablation catheter-tip in X-ray and map it onto the 3D roadmap during ablation therapies. Electrode detection methods must be robust enough to be used routinely during clinical procedures. Fallavollita *et al.* developed a catheter tip detection algorithm based on thresholds of the fluoroscopic images; this failed in low contrast images [3]. A technique for tracking and detecting the ablation catheter in X-ray images was first proposed by Franken *et al.* but the computational cost was relatively high making the method not applicable in clinic [4]. Coronary Sinus and ablation catheter detections were first proposed in [5-6]. Multiple user interaction and parameter fine-tunings were necessary to meet the quality of the X-ray image. Employing respiration and motion compensation methods may succeed in overcoming some of the above challenges [7]. Recently, Brost *et al.* developed a model-based lasso catheter tracking algorithm in biplane X-ray fluoroscopy [8-9]. However, the tracking required re-initialization and user interaction. Wen *et al.* successfully tracked one catheter in a cardiac cycle and required user-initialization in selecting tip electrodes [10-11]. The only work addressing multiple catheter tip-detection is presented in [12]. Here the authors require user interaction for their detections using a geodesic framework. In conclusion, all of the above works are excellent and demonstrate to the community how challenging the task is in achieving automatic and robust methods for electrode detection, specifically for the catheter-tip positions.

Contributions: In an effort to build upon published literature, we propose a method that considers all of the key challenges associated with catheter detections. Our automatic method: (i) detects the tip of the catheter; (ii) detects the other catheter electrodes; (iii) detects multiple catheters; and (iv) considers the overlapping scenario. We believe that the combination of these distinguish this paper making it unique in our community. Our proposed method does not rely on prior knowledge from previous X-ray frames. The aim is to obtain clinically acceptable results— an assigned error value of 2 mm for cardiac applications [13].

2 Methodology

First, our goal is to find all potential electrode candidates in X-ray. We perform these detections without the requirement of user interactivity or algorithm re-initialization. Then we aim at obtaining the highest number of candidate electrodes that correspond to real electrodes and to subsequently filter the outliers. Of the 1422 fluoroscopic images used in the evaluation, we randomly selected one X-ray frame from each dataset (a total of 20) to empirically define the algorithm parameters (Table 1).

2.1 Blob Detections

A blob detector formulation is used to detect electrodes. The electrode appearances are not always the same due to foreshortening and projective effects. They can appear

larger or smaller and their shape can change from rectangular \rightarrow elliptical \rightarrow circular over consecutive frames. It should be noted that for an individual X-ray image, the appearance of the electrodes belonging to the same catheter are very similar: if one of them appears as a circle it is very likely that the others share the same appearance. The candidate electrodes are obtained from a blobness measure influenced by non-maximum suppression. This blobness measure is implicitly greater than one since we are using a scale-space approach to detect electrodes at different scales. In other words, a catheter tip electrode will appear larger than the other electrodes.

Table 1. Algorithm parameter settings

Parameter	Purpose	Value	Unit	Used During
G_σ	Gaussian variance - noise reduction	1	-	preprocessing
S_σ	Start sigma for scale space creation	4	-	preprocessing
$Scales$	Number of levels (octaves) - scale space	1	-	preprocessing
$N_{samples}$	Number samples per octave - scale space	2	-	preprocessing
B_{th}	Blob detector threshold	10^{-6}	-	detection
T_{th}	Top Hat filter threshold	0.04	-	outliers rejection
S_{min}	Electrodes candidate recovery similarity value	0.85	%	electrodes recovery
$RecoveryNeigh$	Recovery neighborhood	25	px	electrodes recovery
NMS	Non maxima suppression neighborhood	3	px	electrodes recovery
P_{min}	Minimum P value for a cluster to allow blob recovery	10^{-4}	-	electrodes recovery
$ClusteringNeigh$	Clustering neighborhood	60	px	electrodes clustering
$Brightness_{min}$	Minimum brightness candidate electrodes	0.6	-	outliers rejection

We investigated two different methods that yield different performances both w.r.t. the detection rate and execution time. These approaches are based respectively on the usage of a ‘‘laplacian of gaussian’’ (LoG) and a ‘‘difference of gaussian’’ (DoG). The (LoG) blob detector [14] is a non-separable linear filter capable of finding blob-like structures while having low responses to edge-like structures. For each X-ray image it is necessary to run three linear filters and to evaluate the blobness measure:

$$Blobness_{LoG}(x, y, t_0) = t_0(L_{xx}L_{yy} - L_{xy}^2) \quad (1)$$

where L_{xx} , L_{yy} , L_{xy} are respectively the convolution of the fluoroscopic image $I(x, y)$ with G_{xx} , G_{yy} , G_{xy} being the second derivatives of the gaussian filter and $t_0 = \sigma^2$ is used for normalization purposes equal to the variance of the gaussian filter.

The (DoG) blob detector [15] is an approximation of the ‘‘laplacian of gaussian’’ filter and is based only on the usage of gaussian filters that are linearly separable. A scale-space [16] representation of the image is obtained by filtering the image with a gaussian kernel using increasing variances. The difference between two neighboring scale-space images is taken and this latter result is used as a blobness measure. The mathematical formulation for the 2D gaussian filter is:

$$G(u, v, t_0) = \frac{1}{\sqrt{2\pi t_0}} e^{-(u^2+v^2)/(2t_0)} \quad (2)$$

and the blobness measure becomes:

$$Blobness_{DoG}(x, y, t_0) = I(x, y) * G(x, y, kt_0) - I(x, y) * G(x, y, t_0) \quad (3)$$

This detector has a significant response in correspondence to edge-like features in the image and thus yields more outliers (i.e. higher false positives) when compared to the LoG detector.

2.2 Rejecting Outliers via Spatial and Geometric Constraints

False positive candidate electrodes exist. To eliminate these, a Top-Hat filter is implemented which discards candidates that do not fulfill spatial and geometric constraint characteristics of an electrode. Since we are looking for “quasi circular” candidates mimicking electrodes a structuring element with a circular diameter of 15 pixels is used. This immediately removes the majority of outliers. Second, since an electrode is metallic (and thus radiopaque), it appears as a very dark cluster in the X-ray image. Thus, candidate electrodes appearing as a bright cluster are rejected. From this, the blobness measure can be used to distinguish between catheter tips and other electrodes (tip-electrodes have a stronger measure as they are larger in size).

2.3 Clustering and Catheter Detection Recovery

Clustering is then implemented to select and group those candidate electrodes that are more likely to be part of a specific catheter. We perform clustering to:

1. Reject isolated electrode detections that were not filtered out in Section 2.2.
2. Evaluate an "importance" measure $P_i = \sum_j S(b_j c_i)$ for each cluster i where b_j are the blobs of the cluster c_i . The strength measure $S(.,.)$ is just the blobness response as it was produced by the blob detector. We will use the cluster "importance" measure in the blob recovery stage.

For a randomly selected *unvisited* blob we create a new cluster and we search the blob's neighborhood. If there are other neighbors we will add them to the current cluster and we will label the current blob as *visited*. Next, we visit the neighborhood of the candidates that we just added. Clusters containing just one blob will be deleted (they cannot be catheters). The size of the neighborhood search is specified in Table 1. **Note:** it does not matter if electrodes belonging to two or more catheters are grouped together in one cluster. As long as each one of the electrodes belonging to each catheter is inside the same cluster our method will account for this as described in the subsection 2.4.

Detection recovery is performed next. For blob clusters that have a measure P_i higher than a threshold, and for each of the candidate electrodes belonging to that cluster, we search their neighborhood to find similar structures. If the cluster i is very likely to contain a good number of true electrodes or tips (we decide this from P_i), we perform Normalized Cross Correlation in a small window around each one of the candidate electrodes to identify similar electrodes (i.e. recall that electrodes in a specific catheter share similar shape). An example of the improvement brought by the detection recovery is shown in Figure 1.

2.4 Catheter Detection and Scoring Criteria

We define a catheter as a grouping of electrodes. Prior to a cardiac ablation procedure, the electrophysiologists know the exact number of catheters that will be used. From this we formulate catheter hypotheses using the candidate electrodes detected. Then, to select the optimal global hypothesis for a specific catheter, we explore the space of possible solutions using the following greedy algorithm:

1. Knowing that N catheters are present in X-ray we select $N + 2$ electrodes that have the strongest blobness measure. It's very likely that N of them will be catheter tips.
2. Starting from each of the detected tips, generate the catheter hypotheses as follows: (a) recursively create paths that include electrode candidates; (b) the created paths must be more or less straight; (c) paths should not violate any catheter geometrical constraints (length, bending angle) and be more or less straight.
3. Formulate the global hypotheses for each N -uple of catheters from (2) that do not share candidate electrodes.
4. Select the best global hypothesis by privileging those that contain straight and long catheters as their components.

It is not recommended to select a catheter as the ensemble of electrodes that lie on the straightest and longest path not violating the catheter geometrical constraints. This will fail since some electrodes from catheter $N+1$ might be erroneously inserted in the path of catheter N – if the catheters are overlapping or really close to each other this effect will be predominant.

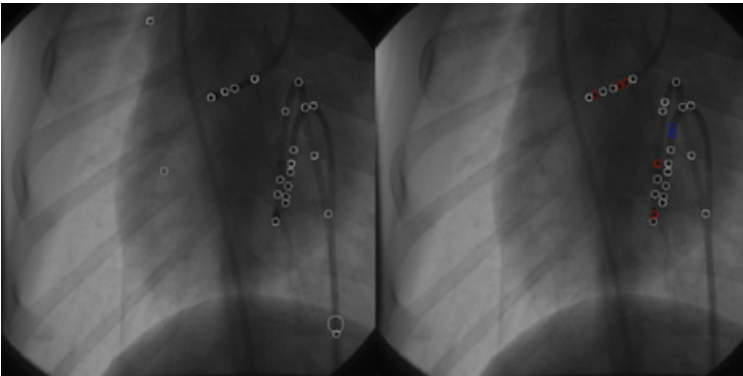


Fig. 1. (Left) blobs before clustering and recovery. (Right) blobs after clustering and recovery; blue - false positive detections induced by the recovery method. red – true positive detection induced by the recovery method.

Scoring Criteria via the Greedy Algorithm: for each candidate catheter the score is the number of electrodes it comprises. These electrodes cannot be in positions which are very far from the tip ($< 150\text{px}$ selected) or that force an impossible bending angle of the catheter ($< 20^\circ$ selected) as stated above in solution 2. Since the maximum number of catheters in our dataset is three, we evaluate the global hypotheses in order to find the best triple of candidate catheters. The best triple is the one that comprises the longest possible candidate catheters, each one starting from one tip, that don't share any of the electrodes (e.g. the intersection of the electrodes belonging to two different candidates is always the null set). The evaluation of the best global hypotheses is done taking for each generated triple, the product of the scoring functions of the candidate catheters contained in that triple. The triple achieving the highest score is then selected. Two examples of catheter detection are presented in Figure 2.

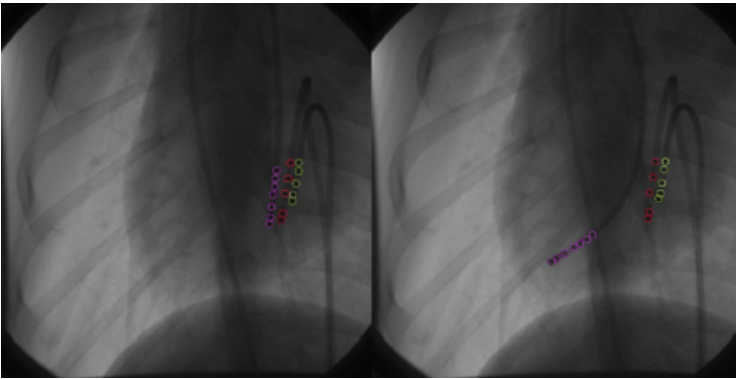


Fig. 2. Sample images depicting accurately the three catheters and their electrodes

3 Results

The method was evaluated using canine specimen laid on its right side on a fluoroscopy table (*Integris Allura, Philips Inc.*). A reference and a pacing catheter (7-French) were inserted into the right ventricle, close to the septal wall. Finally, a standard 8-French ablation catheter was inserted from the femoral vein into the left ventricle (LV) and mapped at different sites within the ventricle. A total of 20 Anterior/Posterior datasets were acquired. The image sizes are 512×512 with a pixel spacing of 0.44 mm. The C-arm energy was varied between 60-90kV to ensure variability within the data. In total, for one X-ray image there were 17 electrodes (3 tips + 14 other electrodes). A total of 1422 fluoroscopic frames were analyzed (i.e. 4266 tip electrodes; 19908 other electrodes; 24174 total electrodes). The centroids of the electrodes for each X-ray image were annotated manually by an expert observer. The mapping catheter is labeled as ‘cat 3’, the pacing catheter as ‘cat 2’ and the reference catheter as ‘cat 1’.

Detection Accuracy: Figure 3-left depicts the performances for the electrode detection using a LoG blob detector. On the horizontal axis we placed the catheters (cat1, cat2, cat3), on the vertical axis the percentage of detection of the respective electrodes (values between 0 and 1). Without the catheter recovery algorithm the detection results are: (0.68, 0.61, 0.85) and improved to (0.78, 0.78, 0.93) when performing catheter recovery. This is an 11% improvement. Similarly, Figure 3-right depicts the performances for the electrode detection using a DoG blob detector. Without the catheter recovery algorithm the detection results are: (0.81, 0.82, 0.83) and improved to (0.86, 0.88, 0.93) when performing catheter recovery. This is a 7% improvement.

Detection Precision: The mean distance of all electrodes from ground truth using a LoG blob detector with catheter recovery is 1.05 ± 0.33 pixels (0.46 ± 0.14 mm). Similarly, the mean distance of all electrodes from ground truth using a DoG blob detector with catheter recovery is 1.10 ± 0.52 pixels (0.48 ± 0.22 mm). The 95th and 99th percentile errors were 0.66 mm and 0.78 mm respectively. These are well within cardiac clinical tolerances [13].

Catheter-Tips: A total of 4236 tip electrodes were automatically annotated yielding a detection rate of 99.3%. The missed detections were due to blobs being isolated away from the catheters— we observed that both the catheter tip and background of the X-ray image had similar grayscale intensity.

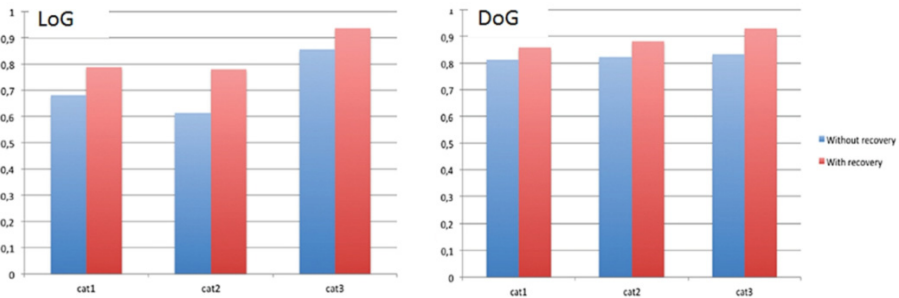


Fig. 3. Electrode detections per catheter (in %)

Outliers Per Image: Using catheter recovery, the mean number of outliers using the LoG blob detector is 10 compared to 53 for DoG. We conclude that the LoG blob detector is a more practical solution since the number of outliers is smaller; however it yields slightly less precise results compared to DoG.

Comparison to Literature: No direct numerical comparison can be made since the evaluation data, algorithm parameters, and types of catheters used differ between all published works in state-of-the art.

Processing: The algorithm was prototyped in C++ and CUDA. It has a runtime of 0.1 (LoG) and 0.3 (DoG) seconds per X-ray image using an Intel Core 2 Duo 1,86 GHz computer.

Future Work: There is room to investigate this method under various clinical conditions and different C-arm fluoroscopy devices. The variety in image quality in clinical cases is due to the variability in patient size, the variability in the image content with the presence of additional or implanted devices that were not used in our animal experiment. These will be accounted for as well as incorporating tracking in improving results [17]. Ultimately, achieving automatic detections can simplify 3D reconstruction of electrodes using single or multi-view approaches [18-19].

4 Conclusion

We introduced an automatic and robust method for multiple and overlapping catheter detection. Our method was evaluated on 1422 fluoroscopic sequences achieving 99.3% tip-detection accuracy. As expected, resolving the overlap issue and obtaining accurate electrode detections is difficult. We presented first results for these.

References

1. Mehra, R.: Global public health problem of sudden cardiac death. *Journal of Electrocardiology* 40(6), 118–122 (2007)
2. Fallavollita, P.: The Future of Cardiac Mapping. In: Breijo-Marquez, F.R. (ed.) *Cardiac Arrhythmias - New Considerations*. In *Tech* (2012), doi:10.5772/31068, ISBN: 978-953-51-0126-0
3. Fallavollita, P., Savard, P., Sierra, G.: Fluoroscopic navigation to guide RF catheter ablation of cardiac arrhythmias. In: *Proc. 26th EMBS*, vol. 1, pp. 1929–1932 (2004)
4. Franken, E., Rongen, P., van Almsick, M., ter Haar Romeny, B.: Detection of electrophysiology catheters in noisy fluoroscopy images. In: Larsen, R., Nielsen, M., Sporning, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 25–32. Springer, Heidelberg (2006)
5. Ma, Y., King, A.P., Gogin, N., Rinaldi, C.A., Gill, J., Razavi, R., Rhode, K.S.: Real-time respiratory motion correction for cardiac electrophysiology procedures using image-based coronary sinus catheter tracking. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part I*. LNCS, vol. 6361, pp. 391–399. Springer, Heidelberg (2010)
6. Ma, Y., Gao, G., Gijbbers, G., Rinaldi, C.A., Gill, J., Razavi, R., Rhode, K.S.: Image-based automatic ablation point tagging system with motion correction for cardiac ablation procedures. In: Taylor, R.H., Yang, G.-Z. (eds.) *IPCAI 2011*. LNCS, vol. 6689, pp. 145–155. Springer, Heidelberg (2011)
7. Babenko, B., Yang, M.H., Belongie, S.: Visual tracking with online multiple instance learning. In: *CVPR*, pp. 983–990 (2009)
8. Brost, A., Liao, R., Hornegger, J., Strobel, N.: 3-D respiratory motion compensation during EP procedures by image-based 3-D lasso catheter model generation and tracking. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part I*. LNCS, vol. 5761, pp. 394–401. Springer, Heidelberg (2009)

9. Brost, A., Liao, R., Hornegger, J., Strobel, N.: Respiratory motion compensation by model-based catheter tracking during EP procedures. *Med. Img. Analy.* 14(5), 695–706 (2010)
10. Wu, W., Chen, T., Wang, P., Zhou, S.K., Comaniciu, D., Barbu, A., Strobel, N.: Learning-based hypothesis fusion for robust catheter tracking in 2D X-ray fluoroscopy. In: *CVPR*, pp. 1097–1104 (2011)
11. Wu, W., Chen, T., Strobel, N., Comaniciu, D.: Fast tracking of catheters in 2D fluoroscopic images using an integrated CPU-GPU frame work. In: *ISBI*, pp. 1184–1187 (2012)
12. Yatziv, L., Chartouni, M., Datta, S., Sapiro, G.: Toward Multiple Catheters detection in Fluoroscopic Image Guided Interventions. *T-ITB* 16(4), 770–781 (2012)
13. Esteghamatian, et al.: Real-time 2D-3D MR cardiac image registration during respiration using extended Kalman filter predictors. In: *9th ICSP*, pp. 1325–1328 (2008)
14. Lindeberg, T.: Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *IJCV* 11(3), 283–318 (1993)
15. Lowe, D.G., Murphy, C.: Distinctive Image Features from Scale-Invariant Key-points. *IJCV* 60(2), 91–110 (2004)
16. Lindeberg, T.: Scale-space. In: *Encyclopedia of Computer Science and Engineering*, vol. 4, pp. 2495–2540. John Wiley and Sons (2009)
17. Heibel, H., Glocker, B., Groher, M., Pfister, M., Navab, N.: Interventional tool tracking using discrete optimization. *IEEE Trans. Med. Imaging* 32(3), 544–555 (2013)
18. Fallavollita, P.: Is Single-View Fluoroscopy Sufficient in Guiding Cardiac Ablation Procedures? *International Journal of Biomedical Imaging* 2010, Article ID 631264, 13 pages (2010), doi:10.1155/2010/631264
19. Fallavollita, P.: Acquiring Multiview C-Arm Images to Assist Cardiac Ablation Procedures. *European Association for Signal Processing: Journal on Image and Video Processing* 2010, Article ID 871409, 10 pages (2010), doi:10.1155/2010/871409

Validation of Catheter Segmentation for MR-Guided Gynecologic Cancer Brachytherapy

Guillaume Pernelle^{1,2}, Alireza Mehrtash², Lauren Barber², Antonio Damato³, Wei Wang³, Ravi Teja Seethamraju³, Ehud Schmidt², Robert A. Cormack², Williams Wells², Akila Viswanathan², and Tina Kapur²

¹ Technische Universität München

² Brigham & Women's Hospital and Harvard Medical School

³ Siemens Healthcare

Abstract. Segmentation of interstitial catheters from MRI needs to be addressed in order for MRI-based brachytherapy treatment planning to become part of the clinical practice of gynecologic cancer radiotherapy. This paper presents a validation study of a novel image-processing method for catheter segmentation. The method extends the distal catheter tip, interactively provided by the physician, to its proximal end, using knowledge of catheter geometry and appearance in MRI sequences. The validation study consisted of comparison of the algorithm results to expert manual segmentations, first on images of a phantom, and then on patient MRI images obtained during MRI-guided insertion of brachytherapy catheters for the treatment of gynecologic cancer. In the phantom experiment, the maximum disagreement between automatic and manual segmentation of the same MRI image, as computed using the Hausdorff distance, was 1.5 mm, which is of the same order as the MR image spatial resolution, while the disagreement between automatic segmentation of MR images and “ground truth”, manual segmentation of CT images, was 3.5mm. The segmentation method was applied to an IRB-approved retrospective database of 10 interstitial brachytherapy patients which included a total of 101 catheters. Compared with manual expert segmentations, the automatic method correctly segmented 93 out of 101 catheters, at an average rate of 0.3 seconds per catheter using a 3GHz Intel Core i7 computer with 16 GB RAM and running Mac OS X 10.7. These results suggest that the proposed catheter segmentation is both technically and clinically feasible.

Keywords: validation, segmentation, catheter, MRI.

1 Introduction

Gynecologic malignancies, which include cervical, endometrial, ovarian, vaginal and vulvar cancers, cause significant mortality in women worldwide. In the United States, the number of gynecologic cancers has been increasing in recent years, while the death rate has remained relatively steady at about 35%

of incidence [1]. The standard-of-care treatment for many primary and recurrent gynecologic cancers consists of chemoradiation (concurrent chemotherapy and external-beam radiation) followed by brachytherapy. In contrast to external-beam radiation treatment, in which a linear accelerator aims radiation beams at the pelvis from outside the body, in high dose rate (HDR) brachytherapy, sources that deliver high doses of radiation are placed directly inside the cancerous tissue using intracavitary applicators or interstitial applicators with catheters.

MRI is used routinely in the diagnosis of cervical cancer due to its increased tumor-to-normal-tissue contrast, scaled by noise (CNR) relative to CT. There is an increasing interest in expanding the role of MRI beyond diagnosis and into HDR treatment planning because of early results indicating it may lead to more precise treatment of the tumor and a reduction in the radiation dose to healthy tissue [2]. However, the deployment of MRI based gynecologic cancer brachytherapy treatment planning is not without challenges. In addition to the expense involved with running an MR-based clinical practice compared to the current standard-of-care, CT, there are companion technical, and specifically image analysis challenges, that have been acknowledged by international radiotherapy societies. Specifically, the GYN GEC ESTRO guidelines note that due to the steep brachytherapy dose gradients, catheter identification errors can lead to major dose deviations in both the target tumor, as well as neighboring tissues. While the source channels are well visualized in CT images, the task is more challenging and error prone when using MR images [3]. In a typical treatment, the radiation oncologist places several catheters using a transperineal approach, spaced about a centimeter apart, and these catheters can bend as they perforate stiff tissues along the insertion path. The “gold-standard” CT catheter visualization method involves placing copper wires, which have large CT absorption cross-sections, into the catheters at the end of the insertion process, with the resulting CT images segmented in brachytherapy treatment planning clinics. In contrast, distinguishing catheters from other signal voids in MRI is challenging, and requires dedicated MRI sequences that provide magnetic susceptibility artifacts that are controlled in their dimensions and directions integrated with sequence-specific image analysis. This paper presents a catheter segmentation method and validates its results in phantoms and clinical cases.

1.1 Related Work

Catheter artifact segmentation from MRI has primarily been pursued in the context of MR-guided interventions such as biopsy or radio-frequency ablation. In contrast to x-ray based imaging, where the material of the catheter is the main factor that influences its visibility, the success of this segmentation task when using MRI is heavily dependent on the MR imaging parameters, the sensitivity of the imaging sequence to magnetic susceptibility effects, and the direction of the catheter relative to the static magnetic field [5,6]. Quantifying these differences in the catheter tip locations from MR sequences optimized for susceptibility imaging, based on the shaft orientation, demonstrated a 4-5mm difference when the direction of the catheter was parallel vs. perpendicular to the direction of

the static magnetic field [7]. Once satisfactory imaging sequences and parameters are selected, the segmentation of these catheters is typically performed using a Hough transform, and it is frequently suggested in the literature that this step could use improvement because catheters tend to bend as they are inserted into the body. Methods that enhance vessel-like structures using eigenvalues of the Hessian matrix have been successful for pre-processing contrast-enhanced images (MRA, CTA) [8,9] but the results were not very useful for our MR images, which were acquired without contrast. While not developed for MRI (or 3D imagery), a method for catheter detection that is closest in approach to ours was reported in [10] for segmentation of catheters from 2D B-mode ultrasound images. It relies on the use of a Hough Transform to provide a set of points along the catheter path, and then polynomial regression to fit a curve to the catheter data, thus adequately capturing the bending of the catheter.

2 Methods and Materials

MR Imaging Sequence and Catheter Appearance. Three MRI sequences – Two-dimensional (2D) T2-weighted Fast Spin Echo (FSE), Three-dimensional (3D) FSE (Siemens SPACE), and 3D balanced steady state free precession (3D bSSFP) - were used to acquire the patient data in a 3 Tesla MRI (Siemens Healthcare, Erlangen, Germany). The tungsten-alloy filled, MRI-safe plastic catheters appear as signal voids of differing size, with the 2D FSE and 3D SPACE providing smaller artifact dimensions. The 2D T2-weighted FSE parameters were TR/TE = 3000/120 msec, $0.2 \times 0.3 \times 2.0 \text{mm}^3$. The SPACE parameters were TR/TE = 3000/160 msec, $0.4 \times 0.4 \times 1.0 \text{mm}^3$. 3D b-SSFP MRI sequence with TR/TE = 5.8/2.9 msec, $0.6 \times 0.6 \times 1.6 \text{mm}^3$ resolution, was used for rapid imaging (1.5 min/volume).

Phantom Construction. To validate the geometry of the extracted catheters, a phantom as shown in Figure 1, was constructed using commercially available transparent gel wax. An obturator, the long cylindrical component of an interstitial applicator assembly that holds the catheters in place, was placed at the center of a transparent plastic container, and gel wax was melted and poured around it. A Syed-Neblett template was then affixed orthogonal to the obturator at the edge of the plastic container. Catheters were inserted to mimic clinical scenarios, with some inserted straight, others bent, and some touching each other.

Image-Based Catheter Segmentation. The goal of the algorithm is to segment catheters of diameter 1.6mm , length upto 240mm , that may be straight or bent in configuration. In essence, starting with a manually provided catheter tip, the segmentation algorithm iteratively searches the image for a direction that maximizes the likelihood of 1.6mm -diameter signal voids in a conic region, and fits Bézier curves to the end points of these segments. Implementation of the algorithm is described next.

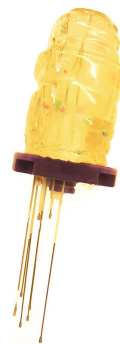


Fig. 1. Gel Wax Phantom

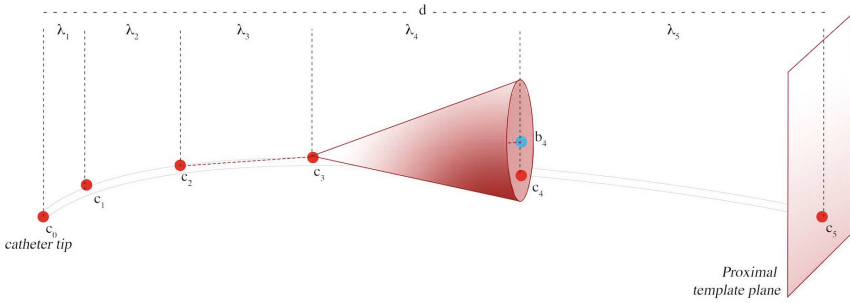


Fig. 2. Segmenting the catheter, one line-segment at a time. The user-defined catheter tip c_0 , and the plane, P , of the template where all catheters enter the body. The algorithm first finds points along the catheter by searching in conic regions of increasing height and radius, and then fits a Bézier curve to these.

Segmentation Algorithm Implementation

1. As illustrated in Figure 2, each catheter is represented as a curve with 6 control points. c_i denotes the i^{th} control point on the catheter; c_0 is the distal end of the catheter or the catheter tip, as provided by the user, and c_5 lies on the proximal template plane, P , defined by the user
2. d is the length of the segment from tip c_0 to the proximal plane P .
3. λ_i is the height of the search cone for each successive control point. λ_i values were chosen to be increasing in magnitude, and the increases related to the Fibonacci sequence¹ as follows:

$$\lambda_i = \begin{cases} \lambda_{i-1} + \lambda_{i-2} & \text{if } i \geq 2 \\ \frac{1}{19} \cdot d & \text{if } i = 0 \\ \frac{2}{19} \cdot d & \text{if } i = 1 \end{cases} \quad (1)$$

4. b_i is the center of the base of the search cone for the i^{th} point. These centers are computed as follows:

$$b_{i+1} = c_i + \lambda_i \cdot \frac{\overline{c_i c_{i-1}}}{\|c_i c_{i-1}\|} \quad (2)$$

5. r_i is the radius of the i^{th} search cone and is defined as an increasing series similar to λ_i .
6. c_i is computed by a search; it is the end point of the line segment that starts at c_{i-1} and has the minimal value for the intensity line integral along its length, among all choices in the search cone defined by b_i , r_i , and c_{i-1} .

After the control points are obtained, a 5^{th} degree Bézier curve is computed as a linear combination of the Bernstein basis polynomial, in which the first

¹ This particular method of choosing λ values delivered the computational efficiency required by the problem but other choices for the series would be valid as well.

and last control points define the extremity of the curve while the interior points pull the curve toward them:

$$B(t) = \sum_{i=0}^5 \binom{5}{i} (1-t)^{5-i} t^i c_i \quad t \in [0, 1]. \quad (3)$$

3 Validation Experiments, Metric, and Results

The segmentation method was validated in three steps. The first two steps utilized the phantom, and the third used patient images. In the phantom experiments, first the disagreement between automatic and manual segmentation of the 12 catheters from the same MR image was computed. Second, this disagreement computation was repeated between automatic segmentation of catheters in MR images and “gold-standard” obtained from manual segmentation in CT. In the third experiment, the disagreement between manual and automatic segmentations on MR images of patients was computed.

The metric used to quantify the accuracy of the catheter segmentation in each case was the symmetric Hausdorff distance (HD) [11]. If X, Y are two non-empty subsets of a metric space (E, δ) the HD is defined by

$$d_H(X, Y) = \max\left\{ \sup_{x \in X} \inf_{y \in Y} \delta(x, y), \sup_{y \in Y} \inf_{x \in X} \delta(x, y) \right\} \quad (4)$$

It should be noted that the Hausdorff distance between two catheters is high even when they disagree only along a short fraction of their lengths. This is an important criterion for this particular clinical application.

MR-MR Phantom Result. The maximum HD between automatic and manual segmentation across all 12 catheters in the same MR images was 1.5mm. This was close to the resolution of the MR scan.

MR-CT Phantom Result. The maximum HD between automatic segmentation of MR images and “ground truth”, as determined from manual segmentation of CT images for the 12 catheters was 3.5mm. The mean HD was 2.3mm and σ was 0.5mm. A visual illustration is provided in Figure 3. The RMS error of catheter tip locations from the registration of MR and CT images was 1mm. Rigid registration was performed using markers embedded in the phantom. We assume that the remainder 1mm error was due to the geometry differences between MR and CT.

Patient MRI Results. The HD between manual and automatic segmentations was computed on 101 catheters from 10 patients. The catheter tips provided in manual segmentations were used to initialize the automatic segmentation, and an axial plane was defined to mark the proximal end of the catheters. A catheter segmentation was classified as correct if this HD was less than 2mm. 2mm was chosen as the classification target by rounding up the 1.6mm catheter diameter; in the absence of apriori knowledge of the orientation of the device relative to

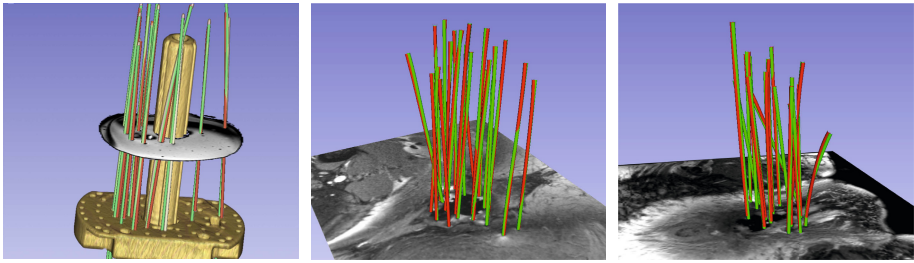


Fig. 3. Phantom MRI/CT. **Fig. 4.** Patient MRI. **Fig. 5.** Another patient MRI differences between HD differences between MRI. HD differences between manual MRI (green) & manual (green) and automatic CT segmentation (red): $3.5mm$ (max), $2.3mm(\mu)$, $0.25mm^2(\sigma^2)$. automatic (red) segmentation (red): $1.23mm$ (max), $1.01mm(\mu)$, $0.01mm^2(\sigma^2)$. automatic (red) segmentation (red): $1.36mm$ (max), $1.01mm(\mu)$, $0.01mm^2(\sigma^2)$.

the magnetic field or the imaging planes, the susceptibility artifact is always somewhat greater than the physical device dimensions. The method correctly localized 93 out of 101 catheters in an average time of 0.3 seconds per catheter on a 3GHz Intel Core i7 computer with 16 GB RAM and running Mac OS X 10.7. Figures 4 and 5 illustrate two cases in which all catheters were correctly identified. Table 1 summarizes validation statistics for each of the 10 cases.

Table 1. Summary of Patient MRI validation. Each row corresponds to a patient; it shows the catheter count, the max, the mean, and the variance of HD across all catheters, the count of catheters that were incorrectly segmented, and statistics on HD (HD*) after the outlier catheters are removed.

Case ID	Catheter count	MaxHD mm	MeanHD mm	VarHD mm ²	Outliers count	MaxHD* mm	MeanHD* mm	VarHD* mm ²
1	7	1.08	0.99	0.00	0	1.08	0.99	0.00
2	14	6.50	1.81	2.91	2	1.56	1.15	0.04
3	8	8.93	2.74	9.53	2	1.00	0.93	0.01
4	15	1.00	0.96	0.00	0	1.00	0.96	0.00
5	16	10.43	2.19	10.33	2	1.23	1.01	0.00
6	5	1.00	0.97	0.00	0	1.00	0.97	0.00
7	7	1.20	1.06	0.01	0	1.20	1.06	0.01
8	6	1.92	1.33	0.19	0	1.92	1.33	0.19
9	9	1.22	1.05	0.01	0	1.22	1.05	0.01
10	14	17.56	3.28	29.03	2	1.67	1.19	0.09

4 Conclusions

This study validated the segmentation of interstitial catheters from MRI images based on a novel method, and demonstrated that this method was able to

properly identify most of the catheter positions in clinically reasonable processing times.

5 Discussion and Future Work

Dependence on Tip Initialization. A common failure mode was observed in patient segmentations, especially when a large number of catheters were used. When the user provided a tip position, which was not located sufficiently close to the shaft of the desired catheter, the results varied based on the configuration of neighbouring catheters. Figure 6 illustrates a case where the user-provided tip resulted in a segmentation (red) that terminated in a neighbouring valley of the objective function (i.e. it mistakenly followed the path of a neighboring catheter), while the correct manual segmentation of the catheter is shown in green. In Figure 7, the catheter tip was initialized correctly, and the segmentation (red) closely matched the (green) manual segmentation. To address this weakness in the method, randomly perturbed restarts will be investigated. In addition, we believe that we can explicitly model the relation between the geometry of the catheter tip artifact and its orientation in the static magnetic field, and plan to investigate this for automatic tip detection.

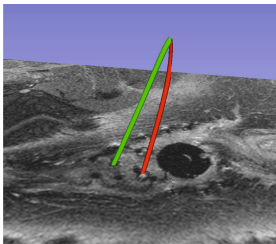


Fig. 6. Failure Mode

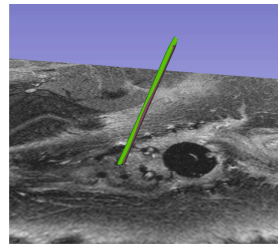


Fig. 7. Reinitialized Tip

From Segmentation to Patient Impact. It should be noted that the correlation of segmentation error with radiation dose for our patient cohort is at this time unknown, and will be analyzed in the future. Given the uncertainties associated with dose metrics in brachytherapy, it is unlikely that dose difference due to segmentation errors lower than 2mm will have a clinical impact. However, neither the dosimetry literature nor published guidelines currently provide tolerances for catheter segmentation error. In order to gauge the benefits of this study to clinical practice, a comparison is needed between radiation plan dose to tumor and organs at risk using the automated catheter segmentation from MR and the standard-of-care catheter segmentation created from CT images. If these two doses turn out to be within the dose uncertainty margins routinely employed in treatment planning, then a contribution will have been made to clinical care; the need of a treatment planning CT will be obviated for cases where appropriate

MR imagery is available at the conclusion of the catheter placement procedure. If the MR-based dose turns out to be a significant improvement over CT, then a case will have been made for performing MR based treatment planning instead of CT.

Acknowledgement. This work was supported by NIH Awards R03EB013792, P41EB015898, and U54EB005149. The authors would like to thank Dr. Yi Gao for discussions on segmentation algorithms, Dr. Steve Pieper for guidance in the implementation on the 3D Slicer software platform, and Dr. Ron Kikinis for advice on workflow that has resulted in an application that is easy to use in a clinical setting.

References

1. Cancer facts & figures 2012, <http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/cancer-facts-figures-2012>
2. Viswanathan, A.N., Szymonifka, J., Tempany-Afdhal, C., Cormack, R.A.: A prospective trial of real-time magnetic resonance-guided catheter placement in interstitial gynecologic brachytherapy. *Brachytherapy* (Epub. 2013)
3. Poetter, R., Haie-Meder, C., Van Limbergen, E., et al.: Recommendations from gynaecological (GYN) GEC ESTRO working group. *Radiother Oncol.* 78(1), 67–77 (2006); Epub 2006 January 5. PubMed PMID: 16403584
4. Cleary, K., Peters, T.M.: Image-guided interventions: technology review and clinical applications. *Annu. Rev. Biomed. Eng.* 12, 119–142 (2010)
5. Lewin, J.S.: Interventional MR imaging: concepts, systems, and applications in neuroradiology. *AJNR Am. J. Neuroradiol.* 20(5), 735–748 (1999); Review. PubMed PMID: 10369339
6. Song, S.E., Cho, N.B., Iordachita, I.I., Guion, P., Fichtinger, G., Kaushal, A., Camphausen, K., Whitcomb, L.L.: Biopsy needle artifact localization in MRI-guided robotic transrectal prostate intervention. *IEEE Trans. Biomed. Eng.* 59(7), 1902–1911 (2012); PubMed PMID: 22481805
7. DiMaio, S.P., Kacher, D.F., Ellis, R.E., Fichtinger, G., Hata, N., Zientara, G.P., Panych, L.P., Kikinis, R., Jolesz, F.A.: Catheter Artifact Localization in 3T MR Images. *Stud. Health Technol. Inform.* 119, 120–125 (2006); PMID: 16404029
8. 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)
9. Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Kikinis, R.: Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Medical Image Analysis* 2(2), 143–168 (1998)
10. Okazawa, S.H., Ebrahimi, R., Chuang, J., Rohling, R.N., Salcudean, S.E.: Methods for segmenting curved catheters in ultrasound images. *Med. Image Anal.* 10(3), 330–342 (2006); Epub 2006 March 7. PubMed PMID: 16520082
11. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(9), 850–863 (1993), doi:10.1109/34.232073

A Novel High Intensity Focused Ultrasound Robotic System for Breast Cancer Treatment

Taizan Yonetsuji¹, Takehiro Ando¹, Junchen Wang¹, Keisuke Fujiwara², Kazunori Itani², Takashi Azuma¹, Kiyoshi Yoshinaka³, Akira Sasaki¹, Shu Takagi¹, Etsuko Kobayashi¹, Hongen Liao¹, Yoichiro Matsumoto¹, and Ichiro Sakuma¹

¹ Graduate School of Engineering, The University of Tokyo, Japan

² Hitachi Aloka Medical, Ltd., Japan

³ Advanced Industrial Science and Technology, Japan

Abstract. High intensity focused ultrasound (HIFU) is a promising technique for cancer treatment owing to its minimal invasiveness and safety. However, skin burn, long treatment time and incomplete ablation are main shortcomings of this method. This paper presents a novel HIFU robotic system for breast cancer treatment. The robot has 4 rotational degrees of freedom with the workspace located in a water tank for HIFU beam imaging and ablation treatment. The HIFU transducer combined with a diagnostic 2D linear ultrasound probe is mounted on the robot end-effector, which is rotated around the HIFU focus when ablating the tumor. HIFU beams are visualized by the 2D probe using beam imaging. Skin burn can be prevented or alleviated by avoiding long time insonification towards the same skin area. The time cost could be significantly reduced, as there is no need to interrupt the ablation procedure for cooling the skin. In addition, our proposed robot control strategies can avoid incomplete ablation. Experiments were carried out and the results showed the effectiveness of our proposed system.

1 Introduction

Breast cancer is a type of cancer originating from breast tissue and most likely occurs in women. It accounts for 22.9% of all cancers in women worldwide in 2008 which has been thought as the biggest health threat to women among all cancers. The treatments for breast cancer include surgery, medications, radiation and/or immunotherapy, among which, surgery (e.g. mastectomy, quadrantectomy) is the most reliable treatment to increase the cure rate and prevent recurrence. However, the removal or partial removal of breast significantly affects the QOL (Quality of Life) in patients. Therefore, breast-conserving surgery has been drawing great attentions in the biomedical engineering community. As a type of breast-conserving surgery, high intensity focused ultrasound (HIFU) has become available for breast cancer treatment. HIFU can focus ultrasound at its focal point and direct acoustic energy into the focus. In HIFU treatment, the focused acoustic energy in human body is absorbed by tissue and is converted into heat so as to ablate tumors. Since only the focus area at a depth inside the body is heated while leaving the superficial skin intact, HIFU is considered as

a non-invasive or minimally invasive medical procedure for precise tumor ablation such as breast cancer, liver cancer, prostate cancer, etc.

Apart from the accuracy in position control of the HIFU focus, another two concerns about HIFU treatment are skin burn and ablation monitoring [1]. Although most of the acoustic energy is directed to the focus during the insonification, there is still a risk of skin burn if long-time high acoustic energy is delivered for effective ablation. In order to avoid skin burn, the ablation process has to be interrupted periodically waiting for heat dissipation of the skin, which increases the treatment time. In addition, HIFU itself cannot provide the image of the ablation area and it is usually used with other imaging modalities for identifying the target and monitoring the ablation process. Magnetic Resonance-guided HIFU (MRgHIFU) [2] and ultrasound-guided HIFU (USgHIFU) [3] are two most common types of HIFU treatment systems. The former uses intra-operative MR images and the latter uses ultrasound images to monitor the ablation process. Although intra-operative temperature monitoring (updated every several seconds) is possible by using MR thermometry, the limitations of MRgHIFU still include the oversize of the MR scanner, the complicated electronic control for HIFU focusing, and the high manufacture and treatment cost. In contrast, USgHIFU has promising potentials in terms of its “real” real-time imaging, compactness and low cost. Some robotic systems for HIFU treatment were also proposed to enable automatic and precise HIFU transducer positioning [4, 5]. However, these studies only focus on the transducer positioning by robotic systems. The robots are only used as automatic transducer holders. The mentioned concerns on skin burn and long treatment time remain unsolved.

In this paper, we propose a novel HIFU robotic system for non-invasive breast cancer treatment with effective ablation and without skin burn. A HIFU transducer for ablation and a diagnostic linear 2D ultrasound probe for real-time imaging (imaging of both the HIFU beams and the target tumor) are combined together and attached to the robot end-effector. The large movement of the HIFU focus is executed by the robot arm and the small adjustment by electronic control, which only needs 56 piezoelectric elements to constitute the HIFU transducer. Skin burn is prevented by varying insonification region of the breast skin while keeping the focus (ablation area) beneath the skin unchanged or moving under control according to our proposed ablation strategies. The robotic system is designed to work under water for ultrasound compatibility.

2 System Design

2.1 System Overview

In HIFU treatment, the gap between the HIFU transducer and the human body must be filled with water or gels for ultrasound propagation. Although a water bag may be used for this purpose, the bag will touch against the breast and cause breast deformation. In addition, the HIFU transducer also has to come in contact with the bag causing our concept of rotary insonification to be unfeasible. Thus the workspace of our robot is designed under water. Because the skin burn is caused by heat accumulation

on the skin surface covered by the insonification, it is possible to reduce the heat accumulation by moving the insonification region while keeping the focus unchanged. Based on this idea, we designed a novel articulate robot that can perform rotary insonification under water. As illustrated in Fig. 1, our system consists of an operation table on which a patient lies in the prone position, a robotic arm performing rotary insonification, a HIFU transducer that is integrated with a diagnostic ultrasound probe, a diagnostic ultrasound machine to which the probe is connected, and some control units. The spatial relationship between the diagnostic probe and the HIFU probe is calibrated so that the beams emitted by the HIFU probe can be imaged by the diagnostic probe (beam imaging), which enables visualizing both the focus (intersection of HIFU beams) and the tumor on the ultrasound machine. The robotic arm enables precise focus positioning. Furthermore, our novel ablation strategy can reduce the heat accumulation and treatment time while performing efficient ablation.

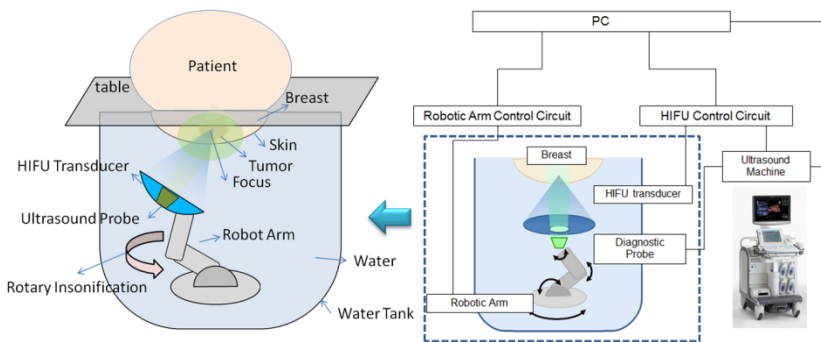


Fig. 1. System overview

2.2 Integrated HIFU Transducer

The function of a HIFU transducer is to focus HIFU beams at its focal point so that most acoustic energy is delivered to the focus to ablate tumors. By moving the focus, a volume thus can be ablated. However, the focus is below the skin and is invisible to surgeons, which makes it difficult to locate the HIFU focus to the target tumor. Unlike the diagnostic ultrasound, the HIFU transducer has no imaging functionality. First, the visualization of the HIFU beams should be solved.

We developed an integrated HIFU transducer that enables accurate and registration-free focus localization. The integrated HIFU transducer consists of a multi-channel piezoelectric element array for HIFU beam generation, a diagnostic ultrasound probe for real-time imaging, and peripheral circuit. The transducer has a diameter of 100 mm with a focal length of 100 mm. The irradiation frequency is 2 MHz. The diagnostic ultrasound probe is fixed at the center of the transducer as shown in Fig. 2(a). HIFU beams are imaged using the diagnostic probe by receiving reflected acoustic waves that are emitted from the HIFU transducer. The principle and the imaged HIFU beams are shown in Fig. 2(b). More detailed information about beam imaging can be found in our previous paper [6]. By this way, the target tumor and the

HIFU beams can be simultaneously visualized in real-time on the diagnostic ultrasound machine. Since all spatial information regarding the focus and the tumor is visible to surgeons, registration becomes unnecessary. The integrated transducer is attached to the robot end-effector and the HIFU focus is adjusted and located to the target tumor by visual feedback. Afterwards, the robot is driven to perform rotary insonification.

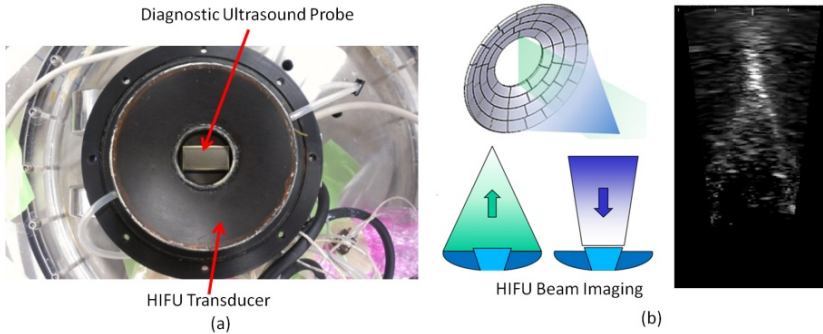


Fig. 2. Integrated transducer (a) Picture of integrated transducer (b) HIFU beam imaging

2.3 Robot Arm

We developed a chain-driven articulated robot with 4 degrees of freedom that works in a water tank. Fig. 3(a) shows the design drawing and the Denavit–Hartenberg (DH) model of our robot, and Fig. 3(b) shows its physical setup. It has two serial links with the same length and four joints, which are inside the water tank. The driving motors with encoders are placed under the water tank. Because the driving force of each joint is transmitted by gears and chains, which can work under water, only the motor shafts need water sealing. The diameter of the water tank is 600 mm and the height including motor units is 775 mm, which can be easily moved by a cart.

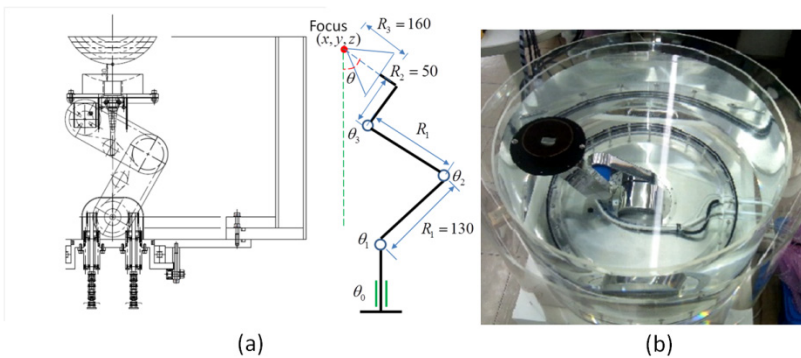


Fig. 3. Robot arm (a) Design drawing and DH model (b) Physical setup

In order to locate the focus in the 3D space within the water tank, the robot is controlled using inverse kinematics. The rotation angle of each joint can be calculated:

$$\begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right) \\ \arccos\left(\frac{z-(R_2\sin\theta_3+R_3\cos\theta_3)}{r}\right) + \arccos\left(\frac{r}{2R_1}\right) \\ \arccos\left(\frac{z-(R_2\sin\theta_3+R_3\cos\theta_3)}{r}\right) - \arccos\left(\frac{r}{2R_1}\right) \\ \theta \end{pmatrix} \quad (1)$$

Where (x, y, z, θ) represents the target pose of the transducer (see Fig. 3(a)), $(\theta_0, \theta_1, \theta_2, \theta_3)$ are the rotation angles of each joint, $r = x^2 + y^2 + z^2 + R_2^2 + R_3^2 + 2(R_2\sqrt{x^2 + y^2} - R_3z)\sin\theta_3 - 2(R_3\sqrt{x^2 + y^2} + R_2z)\cos\theta_3$, $R_1 \sim R_3$ are link lengths with $R_1 = 130 \text{ mm}$, $R_2 = 50 \text{ mm}$, and $R_3 = 160 \text{ mm}$.

2.4 Rotary Insonification Ablation

In this study, we propose a rotary insonification procedure that could prevent the skin burn without losing ablation efficiency. Since the skin burn is caused by the lasting insonification towards the same skin area, it is possible to reduce the heat accumulation by varying the insonification region of acoustic energy during the ablation. There are two modes of rotary insonification as illustrated in Fig. 4: 1. Rotate the transducer around the vertical axis passing through the focus (Fig. 4(a)). 2. Rotate the transducer around the vertical axis deviated from the focus (Fig. 4(b)). In Mode 1, the insonification region of the superficial skin varies over time while keeping the focus unchanged; in Mode 2, the focus also moves (rotation) in order to cover a large ablation volume. Both of the two modes can alleviate the heat accumulation by avoiding the stationary lasting insonification towards the same skin area.

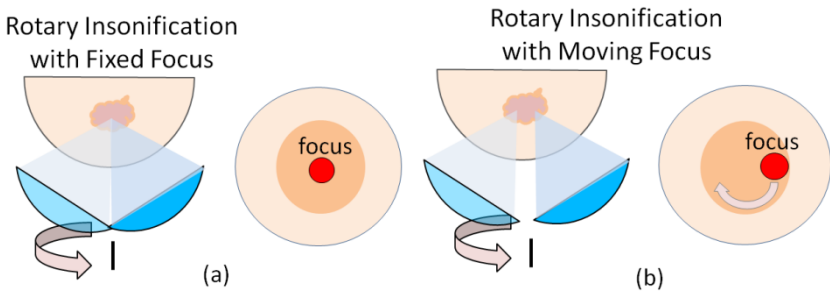


Fig. 4. Rotary insonification (a) With fixed focus (b) With moving focus

3 Experiments

3.1 Evaluation of Robot Accuracy

The rotation accuracy of each robot joint was evaluated. An optical tracking system (Polaris, NDI) was used to produce ground truth. A 6DOF optical marker was

attached to the end-effector of the robot. Rotation accuracy was evaluated by comparing the actual rotation angles reported by Polaris with the input values. The evaluation was performed for ten times every 10 degrees for each joint and the angle errors were calculated. Table 1 summarizes the statistic results of rotation angle errors. $\Delta\theta_i$ represents the rotation angle error of joint i . (+) means forward rotation; (-) means backward rotation. The errors for forward and backward rotation are different owing to the backlash of the chains. STD stands for standard deviation.

Table 1. Rotation angle error (degree)

	$\Delta\theta_0$	$\Delta\theta_1(+)$	$\Delta\theta_1(-)$	$\Delta\theta_2(+)$	$\Delta\theta_2(-)$	$\Delta\theta_3(+)$	$\Delta\theta_3(-)$
Average	0.53	0.22	0.97	0.56	1.32	0.12	1.04
STD	0.098	0.29	1.03	0.71	0.95	0.16	1.01
Maximum	0.69	0.67	2.38	1.60	2.79	0.31	2.75

3.2 Evaluation of Skin Burn

In order to evaluate the extent of the skin burn, temperature of the skin surface was estimated. As demonstrated in Fig. 5(a), a thermosensitive liquid crystal (TLC) sheet was placed on an acoustic absorbent that simulates the breast skin. Since the TLC sheet can sense temperature between 50~60°C with accuracy 1°C by color change, the temperature can thus be measured by a camera. The focus was adjusted at the distance of 20 mm under the TLC sheet. Ablation was performed during the same interval using both our rotary insonification method and a stationary insonification method. Pictures were taken using the camera as experimental results to compare the temperature distribution indicated by colors.

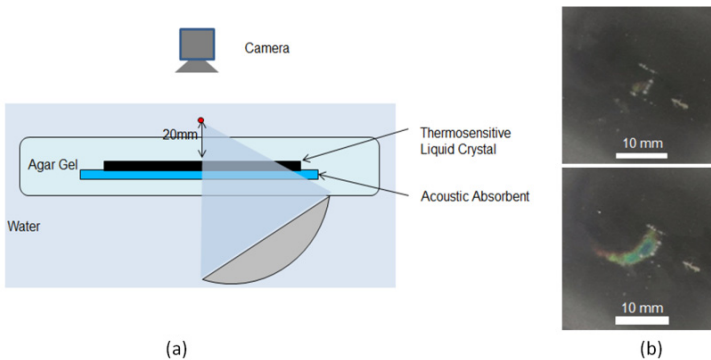


Fig. 5. Skin burn evaluation (a) Experiment design (b) Experimental results

Temperate results are shown in Fig. 5(b). The top is the result of our proposed ablation method and the bottom is the result of traditional one. It is obvious that our method successfully suppressed the heat accumulation and led to low skin temperature. The stationary insonification method heated the TLC up to over 60°C indicated

by the blue region. On the other hand, although very small region became over 50°C, our method succeeded in reducing the high temperature region.

3.3 Evaluation of Ablation

Ablation performance was evaluated using a breast phantom consisting of chicken breast meat and agar gel (Fig. 6). The phantom was made by firstly dissolving graphite and agar powder into deaerated water with the concentration of 1.5% and 4% respectively. Subsequently, the solution was put into a hemispherical mold with a diameter of 150 mm and cooled until it was solidified. Lastly, the chicken breast meat was added into the gel and was covered by pouring the solution once again onto the meat. The final breast phantom has similar shape and physical property with a real one. Here the chicken meat serves as the ablation target. The addition of graphite is to simulate the tissue scattering properties of the ultrasound.



Fig. 6. Breast phantom

A single spot was ablated using rotary insonification Mode 1 described previously. Joint 0 was rotated 360 degree with the speed of 9 degree/sec while emitting HIFU beams of 4 W/ch. For comparison, the ablation was also performed under the same condition but without rotating the transducer. Next, a larger region was ablated continuously using rotary insonification Mode 2, where the focus was moved in a circle while emitting HIFU beams of 4 W/ch. Joint 0 was rotated 360 degree with the speed of 3 degree/sec, which means that the overall treatment time was 120s. The ablation using the traditional method was also performed, where seven spots were ablated separately in order to cover the entire ablation volume with 40s lasting insonification and waiting 80s for cooling on each spot. The overall treatment time was 840s, which was significantly longer than the proposed method.

Fig. 7(a) shows the results of the single spot ablation. The ablation area was about 12×4 mm (left) with the stationary insonification, 8×4 mm (right) with our method. Fig. 7(b) shows the results of the region ablation. The left is the ablation results of the stationary insonification on seven spots. The right is the results of our proposed method. We can see that our method enabled almost homogeneous ablation to avoid incomplete ablation. In contrast, the traditional method only ablated spots one by one to cover the entire target volume, resulting in inhomogeneous and incomplete ablation.

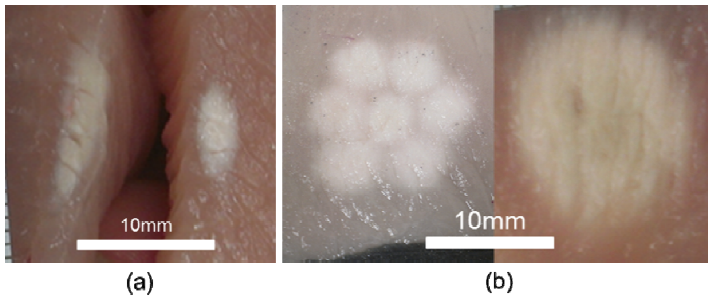


Fig. 7. Ablation evaluation (a) One spot ablation (traditional vs. proposed method) (b) Large region ablation (traditional vs. proposed method)

4 Conclusion

Our robot has a novel structure that works under water for ultrasound compatibility. Compared with the traditional ablation method that performs stationary insonification, our low-cost robotic system can achieve flexible rotary insonification without losing focus. The advantages of this approach include reducing the heat accumulation of the breast skin, decreasing the treatment time and enabling homogeneous region ablation. Future work includes ultrasound image-based visual servo control of the robot and HIFU beam autofocusing.

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

References

1. Zhang, L., Zhu, H., Jin, C., et al.: High-intensity Focused Ultrasound (HIFU): Effective and Safe Therapy for Hepatocellular Carcinoma Adjacent to Major Hepatic Veins. *Eur. Radiol.* 19, 437–445 (2009)
2. Gianfelice, D., Khiat, A., Boulanger, Y., et al.: Feasibility of Magnetic Resonance Imaging Guided Focused Ultrasound Surgery as an Adjunct to Tamoxifen Therapy in High-risk Surgical Patients with Breast Carcinoma. *J. Vasc. Interv. Radiol.* 14, 1275–1282 (2003)
3. Vaezy, S., Shi, X., Martin, R., et al.: Real-time Visualization of High-intensity Focused Ultrasound Treatment Using Ultrasound Imaging. *Ultrasound Med. Biol.* 27(1), 33–42 (2001)
4. Kheng, N., Sing, N., Phee, L., Cheng, C.: A HIFU Robot for Transperineal Treatment of Prostate Cancer. In: Proceedings of 7th International Conference on Control, Automation, Robotics and Vision 2002, vol. 2, pp. 560–565 (2002)
5. Pather, S., Davies, B.L., Hibberd, R.D.: The Development of a Robotic System for HIFU Surgery Applied to Liver Tumours. In: Proceedings of 7th International Conference on Control, Automation, Robotics and Vision 2002, vol. 2, pp. 572–577 (2002)
6. Yonetsuji, T., Fujiwara, K., Itani, K., et al.: HIFU Beam Imaging Based on Scattering Signals from Focal Area. In: Proceedings of CIRP Conference on BioManufacturing 2013, pp. 282–285 (2013)

Cell Orientation Entropy (CO_rE): Predicting Biochemical Recurrence from Prostate Cancer Tissue Microarrays

George Lee¹, Sahirzeeshan Ali², Robert Veltri³, Jonathan I. Epstein³,
Christhunesa Christudass³, and Anant Madabhushi^{2,*}

¹ Rutgers, The State University of New Jersey. Piscataway, NJ, USA

² Case Western Reserve University, Cleveland, OH, USA

³ The Johns Hopkins Hospital, Baltimore, MD, USA

Abstract. We introduce a novel feature descriptor to describe cancer cells called Cell Orientation Entropy (CO_rE). The main objective of this work is to employ CO_rE to quantitatively model disorder of cell/nuclear orientation within local neighborhoods and evaluate whether these measurements of directional disorder are correlated with biochemical recurrence (BCR) in prostate cancer (CaP) patients. CO_rE has a number of novel attributes that are unique to digital pathology image analysis. Firstly, it is the first rigorous attempt to quantitatively model cell/nuclear orientation. Secondly, it provides for modeling of local cell networks via construction of subgraphs. Thirdly, it allows for quantifying the disorder in local cell orientation via second order statistical features. We evaluated the ability of 39 CO_rE features to capture the characteristics of cell orientation in CaP tissue microarray (TMA) images in order to predict 10 year BCR in men with CaP following radical prostatectomy. Randomized 3-fold cross-validation via a random forest classifier evaluated on a combination of CO_rE and other nuclear features achieved an accuracy of $82.7 \pm 3.1\%$ on a dataset of 19 BCR and 20 non-recurrence patients. Our results suggest that CO_rE features could be extended to characterize disease states in other histological cancer images in addition to prostate cancer.

1 Introduction

In this paper, we developed a new approach to quantitatively characterize prostate cancer (CaP) morphology via cell orientation entropy (CO_rE) and thereby attempt to predict biochemical recurrence (BCR), a strong marker for presence of recurring cancer following radical prostatectomy (RP) treatment. BCR is defined by a detectable persistence of prostate specific antigen (PSA) of 0.2 ng/mL following RP. Nearly 60,000 patients undergo RP treatment for CaP each year, and

* Research reported in this publication was supported by the Department of Defense W81XWH-12-1-0171 and the National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, R43EB015199-01, and R03CA143991-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

for 15-40% of RP patients, BCR occurs within 5 years [1]. Gleason scoring (GS) is a qualitative system (2-10) which uses gland morphology to grade CaP aggressiveness and is representative of the clinical standard for predicting BCR. High GS 8-10 cases have been found to be correlated with BCR and presence of aggressive disease and often secondary treatment is provided to accompany RP based on the identification of high GS. Meanwhile, patients with GS 6 typically have a very low incidence of BCR and would not indicate a need for secondary treatment. Unfortunately, outcomes of intermediate GS 7 cancers can vary considerably, and statistical tables suggest a 5-year BCR-free survival rate as low as 43% in these men [2]. As such, predicting BCR in GS 7 cases is an important and largely unsolved problem with significant clinical and therapeutic implications.

While pathologists have traditionally used microscopic evaluation of histological tissue to determine the extent and severity of cancer, the recent advent of digital whole slide scanners has allowed for the development of quantitative histomorphometry (QH) for automated evaluation of histological tissue. The main idea behind these QH methods is to model the appearance of tumor morphology on histopathology via shape, textural, and spatio-architectural descriptors. While qualitative cancer grading remains by far the single most important prognostic measure of aggressive disease, it is subjective and prone to inter-reviewer variability among pathologists [3].

Many researchers have attempted to develop automated, computerized grading algorithms to address the problems of inter-reviewer variability in cancer grading and thereby improve classification accuracy [4,5,6,7,8]. Jafari-Khouzani et al. [6] examined the role of image texture features based on co-occurrence matrices for the purpose of automated CaP grading. However, these matrices are based on pixel intensity and lack direct biological significance. Tabesh et al. [5] also looked at color, texture, and structural morphology to evaluate prostate histopathology in terms of grading. However, complex spatial relationships between structures are not investigated.

Graph tessellations of cell nuclei using Voronoi or Delaunay graphs aim to describe the spatial interactions between nuclei in the tissue and have previously been found to be predictive of CaP grade [4]. However, these features are derived from fully connected graphs, whose edges traverse across epithelial and stromal regions. By connecting globally, fully connected graphs tend to dilute the contribution of the tumor morphologic features specific to the cancer epithelium. Therefore, global graphs are not sensitive to local cell organization, which may be critical in characterizing tumor aggressiveness.

Analysis of local subgraphs, which unlike global graphs (e.g. Voronoi and Delaunay) that aim to capture a global architectural signature for the tumor, can allow for quantification of local interactions within flexible localized neighborhoods. Bilgen et al. [7] constructed different types of cell graphs for evaluating breast cancer. In [8], Veltri et al. investigated nuclear morphology using a descriptor called nuclear roundness variance. Cell morphology was found to exceed Gleason scoring for predicting CaP aggressiveness.

In this paper, we present a new set of QH features, cell orientation entropy (COreE), which aim to capture the local directional information of epithelial cancer cells. CaP is fundamentally a disease of glandular disorganization and the resulting breakdown in nuclei orientation is related to its grade [9]. Epithelial cells align themselves with respect to the glands, and thus display a coherent directionality. However, cancerous prostate glands are less well formed, resulting in a more chaotic organization and orientation of the surrounding nuclei.

COreE attempts to model this difference between cancerous and benign regions via a novel scheme, unique to digital pathology image analysis. Firstly, it is the first rigorous attempt to quantitatively model cell orientation and explore the linkage between cell orientation and CaP aggressiveness. Secondly, while previous work has focused on global graph networks for characterizing tumor architecture, COreE employs subgraphs to construct local cell networks and thereby quantify second order statistics based on co-occurrence matrices of cell orientations. While co-occurrence matrices are commonly used to describe image textures [10], by quantifying second order statistics of image intensities, this is the first instance of the use of the co-occurrence matrix to evaluate local, higher order interactions of nuclear orientations. These second order local statistical features of nuclear orientation yield a rich set of descriptors for distinguishing the different CaP tumor classes.

2 Cell Orientation Entropy (COreE)

2.1 Automated Cell Segmentation

We employed an energy based segmentation scheme presented in [11] to detect and segment a set of cell/nuclei $\gamma_i, p \in \{1, 2, \dots, n\}$, where n is the total number of nuclei found. This segmentation scheme is a synergy of boundary and region-based active contour models that incorporates shape priors in a level set formulation with automated initialization based on watershed. The energy functional of the active contour is comprised of three terms. The combined shape, boundary and region-based functional formulation [11] is given below:

$$F = \underbrace{\beta_s \int_{\Omega} (\phi(\mathbf{x}) - \psi(\mathbf{x}))^2 |\nabla \phi| \delta(\phi) d\mathbf{x}}_{\text{Shape+boundary force}} + \underbrace{\beta_r \int_{\Omega} \Theta_{in} H_{\psi} d\mathbf{x} + \int_{\Omega} \Theta_{out} H_{-\psi} d\mathbf{x}}_{\text{Region force}} \quad (1)$$

where $\beta_s, \beta_r > 0$ are constants that balance contributions of the boundary based shape prior and the region term. $\{\phi\}$ is a level set function, ψ is the shape prior, $\delta(\phi)$ is the contour measure on $\{\phi = 0\}$, $H(\cdot)$ is the Heaviside function, $\Theta_r = |I - u_r|^2 + \mu |\nabla u_r|^2$ and $r \in \{in, out\}$.

The first term is the prior shape term modeled on the prostate nuclei, thereby constraining the deformation achievable by the active contour. The second term, a boundary-based term detects the nuclear boundaries from image gradients. The third term drives the shape prior and the contour towards the nuclear boundary based on region statistics.

2.2 Calculating Cell Orientation

To determine the directionality for each cell γ_i , we perform principal component analysis on a set of boundary points $[x_i, y_i]$ to obtain the principal components $Z = [z_1, z_2]$. The first principal component z_1 describes the directionality of the cell in the form of the major axis $z_1 = \langle z_1^x, z_1^y \rangle$, along which the greatest variance occurs in the nuclear boundary. The principal axis z_1 is converted to an angle $\bar{\theta}(\gamma_i) \in [0^\circ 180^\circ]$ counterclockwise from the vector $\langle 1, 0 \rangle$ by $\bar{\theta}(\gamma_i) = \frac{180^\circ}{\pi} \arctan\left(\frac{z_1^y}{z_1^x}\right)$.

2.3 Local Cell Subgraphs

Pairwise spatial relationships between cells are defined via sparsified graphs. A graph $G = \{V, E\}$, where V represents the set of n nuclear centroids $\gamma_i, \gamma_j \in V, i, j \in \{1, 2, \dots, n\}$ as nodes, and E represents the set of edges which connect them. The edges between all pairs of nodes γ_i, γ_j are determined via the probabilistic decaying function

$$E = \{(i, j) : r < d(i, j)^{-\alpha}, \forall \gamma_i, \gamma_j \in V\}, \tag{2}$$

where $d(i, j)$ represents the Euclidean distance between γ_i and γ_j . $\alpha \geq 0$ controls the density of the graph, where α approaching 0 represents a high probability of connecting nodes while α approaching ∞ represents a low probability. $r \in [0, 1]$ is an empirically determined edge threshold.

2.4 Calculating Second Order Statistics for Cell Orientation

The objects of interest for calculating CORe features are the cell directions given by a discretization of the angles $\bar{\theta}(\gamma_i)$, such that $\theta(\gamma_i) = \omega \times \text{ceil}(\frac{\bar{\theta}}{\omega})$, where ω is a discretization factor. Neighbors defined by the local cell subgraphs G , allow us to define neighborhoods for each cell. For each $\gamma_i \in V$, we define a neighborhood \mathcal{N}_i , to include all $\gamma_j \in V$ where a path between γ_i and γ_j exists in graph G .

An $N \times N$ co-occurrence matrix \mathcal{C} subsequently captures angle pairs which co-occur in each neighborhood \mathcal{N}_i , such that for each \mathcal{N}_i ,

$$\mathcal{C}_{\mathcal{N}_i}(a, b) = \sum_{\gamma_i, \gamma_j} \sum_{a, b=1}^N \begin{cases} 1, & \text{if } \theta(\gamma_i)=a \text{ and } \theta(\gamma_j)=b \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where $N = \frac{180}{\omega}$, the number of discrete angular bins. We then extract second order statistical features (Contrast energy, Contrast inverse moment, Contrast average, Contrast variance, Contrast entropy, Intensity average, Intensity variance, Intensity entropy, Entropy, Energy, Correlation, Information measure 1, Information measure 2) from each co-occurrence matrix $\mathcal{C}_{\mathcal{N}_i}(a, b)$. Selected formulations are described in Table 1. Mean, standard deviation, and range of Θ across all \mathcal{N}_i constitute the set of 39 CORe features.

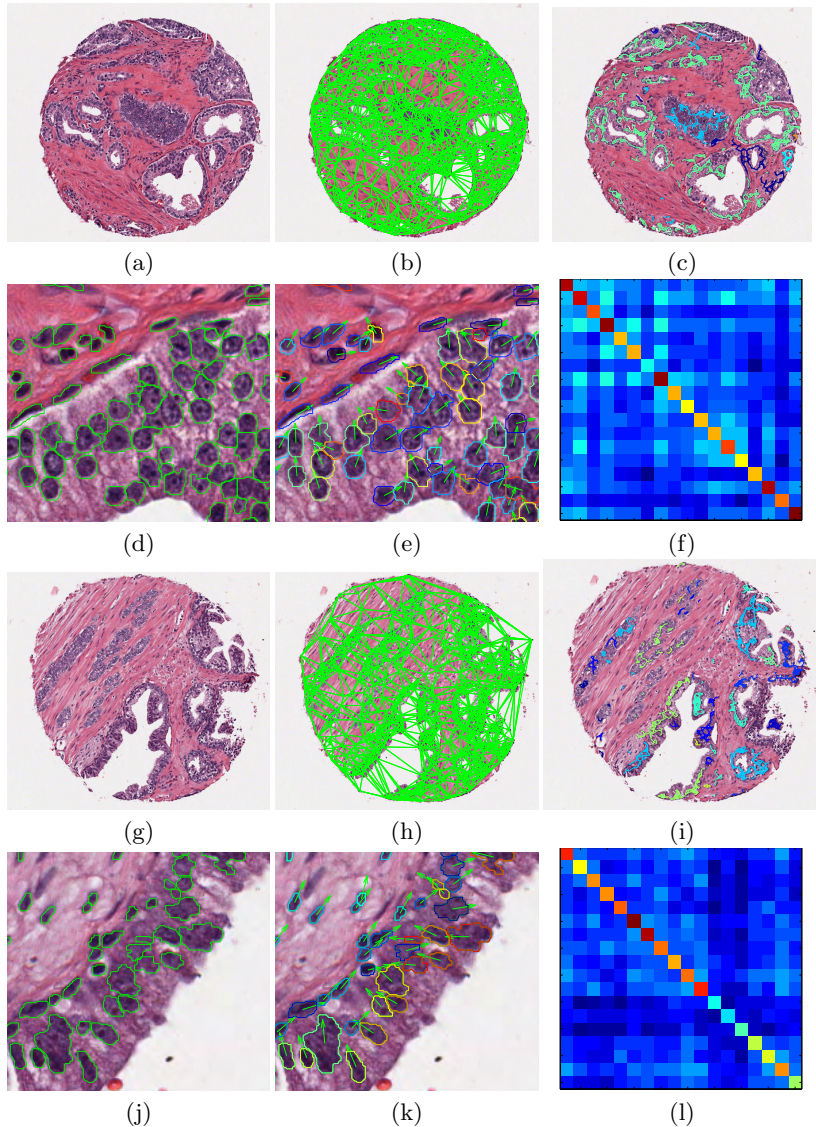


Fig. 1. Prostate TMAs pertaining to (a)-(f) BCR and (g)-(l) NR case studies. Nuclei are used as nodes for calculation of (b),(h) Delaunay graphs. Automated segmentation (d),(j) defines the nuclear boundaries and locations from the TMA image. (e),(k) Cell orientation vectors are calculated from the segmented boundaries (illustrated via different boundary colors). (c),(i) Subgraphs are formed by connecting neighboring cells. CORE features calculate contrast in the cell orientation (with dark regions showing more angular coherence and bright regions showing more disorder). Summation of the co-occurrence matrices provide a visual interpretation of disorder, where (f) shows brighter co-occurrence values in the off-diagonal cells, suggesting higher co-occurrence of nuclei of differing orientations compared to (l).

Table 1. Representative COre features

COre Feature (θ)	Description
Entropy	$\sum_{a,b} -\mathcal{C}(a,b) \log(\mathcal{C}(a,b))$
Energy	$\sum_{a,b} \mathcal{C}(a,b)^2$
Correlation	$\sum_{a,b} \frac{(a-\mu_a)(b-\mu_b)\mathcal{C}(a,b)}{\sigma_a\sigma_b}$
Contrast (variance)	$\sum_{a,b} a-b ^2 \mathcal{C}(a,b)$

3 Experimental Design

3.1 Prostate Cancer Tissue Microarray Data

While COre is extensible towards the histological analysis of other pathological diseases, we have chosen prostate cancer (CaP) as a test case for this initial work. Our dataset comprised of histologic image samples in the form of tissue microarray (TMA) cores from 19 CaP patients who experienced BCR within 10 years of RP, and from 20 patients who did not (NR). Patients were matched for GS 7 and tumor stage 3A. CaP tissue included in the TMAs were selected and reviewed by an expert pathologist. For this study, each of 39 patients was represented by a single randomly selected 0.6mm TMA core image, chosen from a set of 4 TMA cores taken for that patient.

3.2 Comparative Methods for Evaluating COre

We compared the efficacy of COre features with previously studied nuclear features. The shape of individual nuclei has previously been shown to be prognostic of GS [8,12]. The set of 100 cell morphology features representing mean, standard deviation of nuclear size and shape are summarized in Table 2.

Nuclear/cell architecture refers to the spatial arrangement of cells in cancerous and benign tissue. 51 architectural image features describing the nuclear arrangement were extracted as described in [12]. Voronoi diagrams, Delaunay Triangulation and Minimum Spanning Trees were constructed on the digital histologic image using the nuclear centroids as vertices (See Table 2).

For all feature sets, the nuclear segmentations from Section 2.1 were used to calculate the cell boundaries and centroids. In total, we investigated the performance of 4 feature cohorts: (1) 100 features describing cell morphology, (2) 51 features describing cell architectures, (3) 39 features describing cell orientation entropy (COre), and (4) the combined feature set spanning cohorts (1-3).

3.3 Random Forest Classifier

In this study, we demonstrate the efficacy of including COre features for improving classification accuracy and area under the receiver operating characteristic curve (AUC) in predicting BCR in CaP patients from prostate TMAs. Randomized 3-fold cross validation was performed on the top 10 most informative features selected via Student *t*-test for each of 4 feature cohorts defined in Section 3.2. Classification was performed using a random forest classifier.

Table 2. Summary of 151 nuclear morphologic features

Cell Morphology	#	Description
	100	Area Ratio, Distance Ratio, Standard Deviation of Distance, Variance of Distance, Distance Ratio, Perimeter Ratio, Smoothness, Invariant Moment 1-7, Fractal Dimension, Fourier Descriptor 1-10 (Mean, Std. Dev, Median, Min / Max of each)
Cell Architecture		Description
Voronoi Diagram	12	Polygon area, perimeter, chord length: mean, std. dev., min/max ratio, disorder
Delaunay Triangulation	8	Triangle side length, area: mean, std. dev., min/max ratio, disorder
Minimum Spanning Tree	4	Edge length: mean, std. dev., min/max ratio, disorder
Nearest Neighbors	27	Density of nuclei, distance to nearest nuclei

4 Results and Discussion

Figure 1 reveals the ability of the COrE features to capture the differences in angular disorder across localized cell networks and illustrates the differences between the BCR and NR cases in terms of the COrE features.

In Table 3, we can summarize the performance of feature descriptors describing cell architecture and cell morphology which appear to have a maximum BCR prediction accuracy of 79.9%. However, by inclusion of novel cell orientation entropy (COrE) features, the overall classifier accuracy improves to 82.7%. Similar improvements are also observed in terms classification AUC. This reflects the utility of COrE features as a valuable prognostic measurement for predicting BCR in conjunction with previously described nuclear morphologic features.

Classifier improvement following inclusion of COrE features suggests that many of the new COrE features are non-correlated with previously defined cell architectural and morphological feature sets. This distinction is illustrated in Figure 1, where we observe the differences between COrE features compared with those obtained from Voronoi and Delaunay graphs. These graphs span across stromal and epithelial regions, while COrE features are limited to sub-graphs in localized regions. It is also important to note that the combination of COrE and nuclear morphologic features clearly and significantly outperform the clinical standard of pathologist grade, which classified all cases as GS 7.

Table 3. 100 runs of 3-fold Random Forest Classification

	Architecture	Morphology	COrE	Arch + Morph + COrE
Accuracy	71.2 \pm 4.2%	79.9 \pm 3.7%	74.6 \pm 4.1%	82.7 \pm 3.1%
AUC	0.641 \pm 0.054	0.773 \pm 0.042	0.688 \pm 0.063	0.809 \pm 0.037

5 Concluding Remarks

In this work, we presented a new feature descriptor, cell orientation entropy (CORe), for quantitative measurement of local disorder in nuclear orientations in digital pathology images. We demonstrated high accuracy and improvement in predicting BCR in 39 CaP TMAs via the use of CORe features. While CORe features did not outperform other quantitative histomorphometric measurements such as nuclear shape and architecture significantly, the combination of nuclear shape, architectural and CORe features boosted classifier accuracy in identifying patients at risk for BCR following radical prostatectomy. More significantly, the combination of CORe and other image based features significantly outperformed pathologist derived GS, which is 50% for GS 7, and is further known to have at best moderate inter-observer agreement ($\kappa = 0.47-0.7$) [3]. In future work, we aim to evaluate the applicability of CORe features in other disease sites such as breast cancer.

References

1. Trock, B., et al.: Prostate cancer-specific survival following salvage radiotherapy vs observation in men with biochemical recurrence after radical prostatectomy. *JAMA* 299(23), 2760–2769 (2008)
2. Han, M., et al.: Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J. Urol.* 169(2), 517–523 (2003)
3. Allsbrook Jr., W., et al.: Interobserver reproducibility of gleason grading of prostatic carcinoma: general pathologist. *Hum. Pathol.* 32(1), 81–88 (2001)
4. Christens-Barry, W., Partin, A.: Quantitative grading of tissue and nuclei in prostate cancer for prognosis prediction. *Johns Hopkins Apl Technical Digest* 18, 226–233 (1997)
5. Tabesh, A., et al.: Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Transactions on Medical Imaging* 26(10), 1366–1378 (2007)
6. Jafari-Khouzani, K., Soltanian-Zadeh, H.: Multiwavelet grading of pathological images of prostate. *IEEE Trans. on Biomedical Engineering* 50(6), 697–704 (2003)
7. Bilgin, C., et al.: Cell-graph mining for breast tissue modeling and classification. *IEEE Eng. in Med. and Biol. Soc., EMBS*, 5311–5314 (2007)
8. Veltri, R.W., et al.: Nuclear roundness variance predicts prostate cancer progression, metastasis, and death: A prospective evaluation with up to 25 years of follow-up after radical prostatectomy. *The Prostate* 70(12), 1333–1339 (2010)
9. Epstein, J.I.: An update of the gleason grading system. *Journal Of Urology (the)* 183(2), 433 (2010)
10. Haralick, R., et al.: Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics* (6), 610–621 (1973)
11. Ali, S., Madabhushi, A.: An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Transactions on Medical Imaging* 31(7), 1448–1460 (2012)
12. Doyle, S., et al.: Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer. *BMC Bioinformatics* 13(1), 282 (2012)

Robust Selection-Based Sparse Shape Model for Lung Cancer Image Segmentation

Fuyong Xing^{1,2} and Lin Yang^{1,2}

¹ Division of Biomedical Informatics, Department of Biostatistics

² Department of Computer Science, University of Kentucky, KY 40506, USA

Abstract. Accurate cellular level segmentation of lung cancer is the prerequisite to extract objective morphological features in digitized pathology specimens. It is of great importance for image-guided diagnosis and prognosis. However, it is challenging to correctly and robustly segment cells in lung cancer images due to cell occlusion or touching, intracellular inhomogeneity, background clutter, etc. In this paper, we present a novel segmentation algorithm combining a robust selection-based sparse shape model (top-down) and an efficient local repulsive balloon snake deformable model (bottom-up) to tackle these challenges. The algorithm has been extensively tested on 62 cases with over 6000 tumor cells. We experimentally demonstrate that the proposed algorithm can produce better performance than other state-of-the-art methods.

1 Introduction

Efficient and accurate cell segmentation on pathology images for lung cancer can provide diagnosis and prognosis support for improved characterization and personalized treatment, but it remains as a challenging problem due to cell occlusion/touching, densely clustering, size variation, intracellular inhomogeneity, etc. Segmentation methods that only rely on bottom-up information may not be sufficient to handle images exhibiting large inhomogeneities like lung cancer. Supervised learning [11], multi-reference level set [4], hierarchical partial matching [17], and various shape prior models [7,3,8,16,14] were proposed to address these challenges and achieved excellent performance. Recently, Zhang *et al.* [19] modeled shape prior using a sparse shape composition, which produces accurate organ-level segmentation in lung and liver images. Compared with the traditional PCA-based active shape model [7], sparse representation-based shape model can preserve local details [19].

Sparse shape model using all training shapes as the dictionary is inefficient for a large dataset. KSVD [1] is a popular dictionary learning algorithm which is also used in [19], however it is not designed as a discriminative and selection-based dictionary learning method with respect to classification and segmentation. In this work, we propose a novel and robust selection-based dictionary learning algorithm for cellular level shape modeling for lung cancer tumor cells. Different from KSVD, this method directly selects the most representative cellular shapes from the training dataset as dictionary bases. The robustness of the dictionary learning method is achieved by minimizing an integrated square error with a

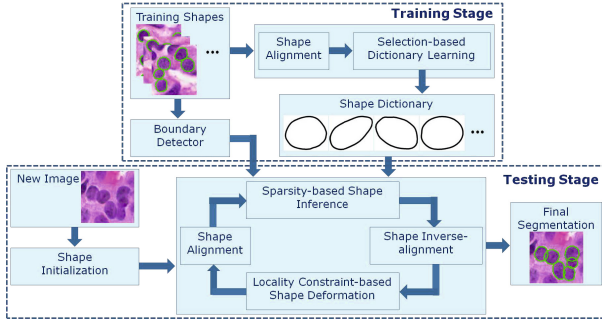


Fig. 1. The flow chart of the proposed algorithm

sparse constraint. In addition, an efficient online dictionary update scheme is designed to adaptively update dictionary bases. Finally, a novel local repulsive balloon snake is presented to fast deform shapes considering the requirement for touching cell segmentation.

2 Methodology

Figure 1 shows the framework of the whole segmentation algorithm. In the training stage, a cell boundary detector is learned using the method in [18]. The training cell shapes, after alignment with Procrustes analysis [7], are utilized to train a compact dictionary using the proposed selection-based dictionary learning method. In the testing stage, the seeds of the cells are automatically detected using a voting-based seed detection algorithm presented in [12], and these detected local minima (seeds) are used to initiate an adaptive H-minima transformation to obtain the initial shapes [5], one per cell. The algorithm alternately performs shape deformation using an efficient local repulsive balloon snake, and shape inference using the shape prior derived from the sparse shape model.

Robust Selection-Based Sparse Shape Model: In this paper, cell shape $v \in R^{2m}$ is represented by the concatenated 2D coordinates of m landmarks which are automatically detected by the rules: 1) The two endpoints of the major axis of the shape are selected as major landmarks. 2) All the other landmarks are interpolated along the shape. Given N cell shapes $\{v_i\}$ aligned by Procrustes analysis [7], sparsity-based shape modeling aims to find a compact shape dictionary $B = [b_1 \ b_2 \ \dots \ b_K]$ ($\{b_k \in R^{2m}\}$ are bases) and a sparse coefficient α such that any aligned shape v can be represented with a few bases: $v = B\alpha + \epsilon$, where ϵ is the residual. Calculating α is called sparse coding, while dictionary learning can be formulated as

$$\min_{B, \{\alpha_i\}} \sum_{i=1}^N \|v_i - B\alpha_i\|_2 + \lambda \|\alpha_i\|_1, \quad (1)$$

For a large cell shape dataset, it is intuitive to select a subset of the data as a shape repository that can sufficiently represent the whole dataset. This summarization can help remove outliers that are not the true representatives of the

dataset and reduce the computational time for runtime optimization due to the decreased object-space dimension. Based on these considerations, we propose a novel selection-based dictionary learning method for sparse representation by minimizing a locality-constrained integrated squared error (ISE):

$$\min_{\theta} J(\theta) = \min_{\theta} \left[\int g(\epsilon|\theta)^2 d\epsilon - \frac{2}{N} \sum_{i=1}^N g(\epsilon_i|\theta) \right] + \lambda \sum_{i=1}^N \sum_{k=1}^K |\alpha_{ik}| \|v_i - b_k\|^2, \tag{2}$$

s. t. $1^T \alpha_i = 1, \forall i,$

where $\epsilon_i = v_i - B\alpha_i$ and $\alpha_i = [\alpha_{i1} \ \alpha_{i2} \ \dots \ \alpha_{iK}]^T$. $g(x|\theta)$ is a parametric model with parameter θ . The first two terms forms the integrated squared error, which is robust to outliers [13]. The last term constrains local representation of bases with weighted sparse codes, and is used to encourage each cell to be sufficiently represented by its neighboring dictionary bases for similarity preserving, which is essential in the sparse reconstruction. The constraint $1^T \alpha_i = 1, \forall i$ ensures the shift-invariance. The residual is modeled with multivariate normal distribution: $\epsilon_i \sim N(0, \sigma^2 I_{2m})$. In this way $g(\epsilon_i|\theta) = \xi \phi(\epsilon_i|0, \sigma^2 I_{2m})$, where ξ denotes the percentage of the inlier shapes that need to be estimated and ϕ is the probability density function of multivariate normal distribution. Based on (2), the dictionary B and sparse coefficients $\{\alpha_i\}$ can be calculated by estimating $\theta = \{\xi, B, \alpha_1, \alpha_2, \dots, \alpha_N, \sigma^2\}$.

Equation (2) can be solved by performing dictionary basis selection and sparse coding alternatively. As $J(\theta)$ in (2) is differentiable, projection based-gradient descent is utilized for minimization. We update the bases $\{b_k\}$ by directly selecting shapes within each iteration. For sparse coding, we keep the dictionary fixed. Based on the sparse reconstruction criterion, the sparse coding objective function can be rewritten as:

$$\min_{\{\alpha_i\}_{i \in A}} \left[\sum_{i \in A} \|v_i - B\alpha_i\|^2 + \lambda \sum_{k=1}^K |\alpha_{ik}| \|v_i - b_k\|^2 \right], \quad \textit{s. t.} \quad 1^T \alpha_i = 1, i \in A, \tag{3}$$

where A is the set of indices corresponding to estimated inlier shapes. Locality-constrained linear coding (LLC) [15] is applied to (3) for sparse coding, where the neighboring bases are defined in terms of the Euclidean distances between the shape and dictionary bases.

Let B^0 be the initial dictionary where the bases are randomly selected from the dataset, LLC is used in (3) to compute the current coefficients $\{\alpha_i\}$. Meanwhile, the active set A is updated with the indices corresponding to the $N \cdot \xi$ shapes with the smallest reconstruction errors $\|\epsilon_i\|_2^2$ in each iteration. At the t -th step, $\theta^t = \{\xi^t, B^t, \alpha_1^t, \alpha_2^t, \dots, \alpha_N^t, \sigma^{t2}\}$, and $B^t = [b_1^t, b_2^t, \dots, b_K^t]$, denote the gradient of $J(\theta)$ in (2): $\frac{\partial J(\theta)}{\partial b_k}$, $\frac{\partial J(\theta)}{\partial \sigma^2}$, and $\frac{\partial J(\theta)}{\partial \xi}$, as J_{b_k} , J_{σ^2} and J_{ξ} , respectively. The basis b_k is updated by selecting the shape v_l which has the largest correlation between the displacement and the current b_k^t :

$$COR(x_l, b_k^t, J_{b_k}) = \frac{(v_l - b_k^t)^T (-J_{b_k})}{\|(v_l - b_k^t)\|_2 \| -J_{b_k} \|_2}, \tag{4}$$

Let A^t represent a set of indices corresponding to the current estimated inliers, the current reconstruction error E^t and the ISE error F^t are defined as

$$E^t = \frac{\sum_{i=1, i \in A^t}^N \|\epsilon_i^t\|_2^2}{\sum_{i=1}^N I(i \in A^t)}, \quad F^t = \xi t^2 \phi(0|0, 2\sigma^2 I_{2m}) - \frac{2\xi t}{N} \sum_{i=1}^N \phi(\epsilon_i^t|0, \sigma^2 I_{2m}), \quad (5)$$

where $I(x)$ is the indicator function. Assume that E_{min}^t is the current reconstruction error, E_{rep}^t is the reconstruction error after replacing the k -th basis with v_l , F_{min}^t is the current ISE error, and F_{rep}^t is the ISE error after replacing the k -th basis with v_l , then the replacement will be performed only if $E_{min}^t > E_{rep}^t$ and $F_{min}^t > F_{rep}^t$. The σ^2 and ξ are updated in the negative gradient directions:

$$(\sigma^{t+1})^2 = \sigma^{t2} - \Delta h_{\sigma^2} \cdot J_{\sigma^2}, \quad \xi^{t+1} = \xi^t - \Delta h_{\xi} \cdot J_{\xi}, \quad (6)$$

where Δh_{σ^2} and Δh_{ξ} represent the learning rates.

Sparse Shape Model Update: Our algorithm supports efficient dictionary update. Recall that each basis b_k corresponds to one row $\alpha^{(k)}$ of the sparse coefficient matrix, and one basis is significant if it is used to represent many shapes in the sparse reconstruction such that $\alpha^{(k)}$ will have many nonzero elements with relatively large values. Therefore, we define the significance of basis b_k as

$$sig(b_k) = C \sum_{i=1}^N \alpha_i^{(k)} e^{-\frac{\|\epsilon_i\|_2^2}{\sigma^2}}, \quad (7)$$

where C denotes the normalized constant, and $\alpha_i^{(k)}$ is the i -th element of $\alpha^{(k)}$ which is the k -th row of the sparse coefficient matrix $G = [\alpha^{(1)T} \alpha^{(2)T} \dots \alpha^{(K)T}]^T$. The learned shape dictionary can be only updated when the current reconstruction error in (5) is larger than a threshold by replacing the current least significant basis with the most important one generated from the new images. This represents the case that cell shapes change dramatically.

Shape Deformation: For shape deformation, we propose a novel repulsive balloon snake (RBS) with locality constraint. We introduce a local repulsive term into the model in [6] to handle touching cells so that each point on the i -th shape v_i will be driven by both its own internal/external forces, and the extrinsic repulsive forces calculated from its neighboring contours (shapes). Practically, each cell is often surrounded by a limited number (M) of adjacent cells, and only its neighboring cells make dominant repulsive force contributions to its shape deformation. This suggests that we can deform shape v_i in its *local coordinate system* for speedup, and it can be implemented by simply using v_i 's M ($M \ll N$) nearest neighbors V_i . Therefore, we formulate a local RBS to find a balance between internal $F_{int}(v_i)$ [6] and external $F_{ext}^{LRB}(v_i)$ forces: $F_{int}(v_i) + F_{ext}^{LRB}(v_i) = 0$, where $F_{ext}^{LRB}(v_i)$ is defined as:

$$F_{ext}^{LRB}(v_i) = \gamma \mathbf{n}_i(s) + \omega \sum_{j \in S_i} \int_0^1 f(d_{ij}(s, u)) \mathbf{n}_j(u) du + \eta p(v_i(s)), \quad (8)$$

where $S_i = \{j : v_j \in V_i, j \neq i\}$. $d_{ij}(s, u) = \|v_i(s) - v_j(u)\|_2$ is the Euclidean distance between $v_i(s)$ and $v_j(u)$. $f(d_{ij}(s, u))$ is preferably designed so that its

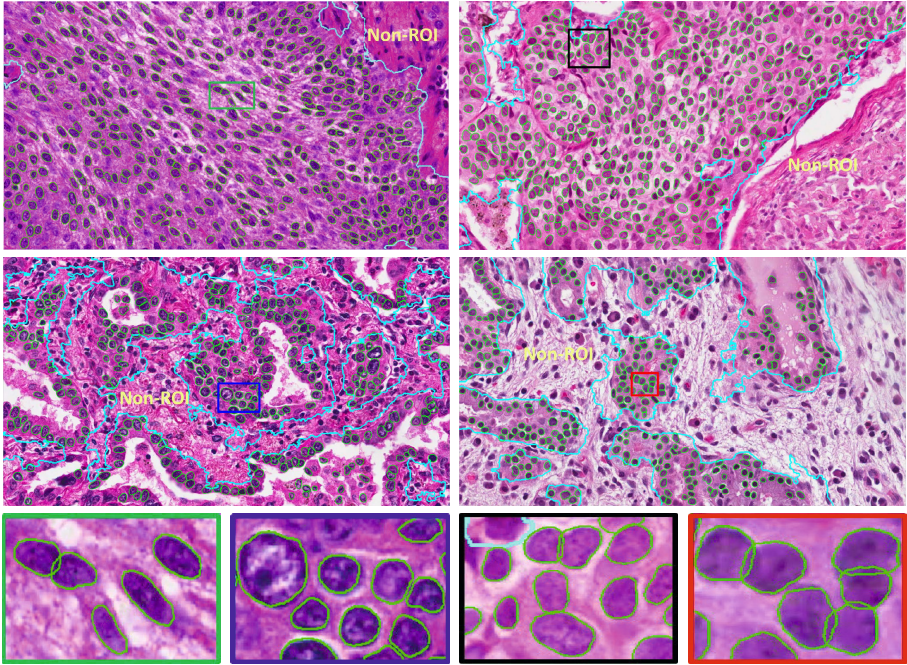


Fig. 2. Segmentation results using the proposed algorithm on regions of interest (ROIs) of the lung cancer pathology specimens. The ROIs and non-ROIs are separated by cyan curves. Several patches are zoomed in for better illustration.

value decreases as the distance from point $v_i(s)$ to $v_j(u)$, and it controls the magnitudes of repulsive forces. In the implementation $f(x)$ is simply chosen as: $f(x) = x^{-2}$. $p(v_i(s))$ represents the edge detector which is chosen as the learned boundary detector [18] (the non-learning based edge detector [6] can also be utilized). Equation (8) reduces the computation complexity significantly from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \cdot M)$ due to $M \ll N$.

3 Experiments

The algorithm is extensively tested on the whole slide H&E-stained images of non-small cell lung cancer that contains 62 cases. There are hundreds of cells in each case, and in total the dataset has over 6000 tumor cells. The ground truth segmentation is manually labeled and confirmed by three pathologists. 20 cases out of 62 are used for training with the rest for testing. The algorithm is implemented with Matlab, and the parameters are chosen based on cross-validation and fixed in the experimental stage: $\lambda = 0.0001$, $K = 300$ in (2,3) and $\gamma = 0.8$, $w = 0.5$, $\eta = 5$ in (8).

Segmentation Performance Analysis I: We design both qualitative and quantitative experiments to evaluate the proposed segmentation algorithm.

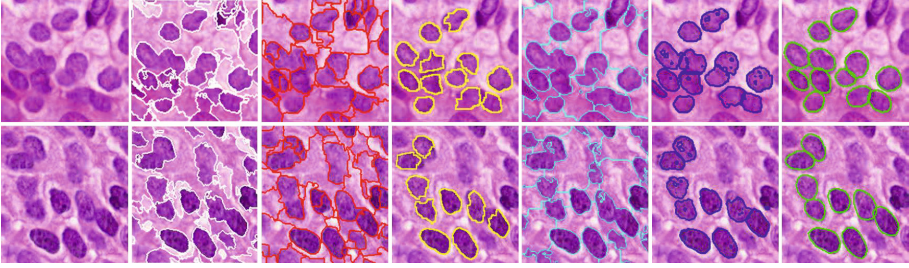


Fig. 3. Comparative segmentation using different methods on two sample patches (rows 1 and 2). From left to right: original image, MS, ISO [10], MWS, GCC [2], RLS [12], and ours. MWS, RLS, and ours use the same initialization.

In Figure 2, hundreds of cells on four image patches (1712×952) are accurately segmented within the regions of interest (ROIs). The ROIs are obtained using an Adaboost learning method [9] with textron feature, and the ROIs represent those regions in which pathologists are more interested, such as epithelium regions. The comparative segmentation results on two randomly selected image patches using the proposed algorithm and the other five methods, including mean shift (MS), isoperimetric (ISO) [10], marker-based watershed (MWS), graph-cut and coloring (GCC) [2], and repulsive level set (RLS) [12] are presented in Figure 3. It is clear that this novel algorithm does take advantage of object topology constraints and shape prior, and therefore it can handle partial occlusion, inhomogeneous intensity, and size variations. Please note that the proposed algorithm is designed to preserve the touching cell shapes in order to facilitate the subsequent morphological feature extraction. For quantitative analysis with the pixel-wise segmentation accuracy, we calculate the precision P and recall R : $P = \frac{|sr \cap gt|}{|sr|}$, $R = \frac{|sr \cap gt|}{|gt|}$ where sr represents the segmentation result and gt denotes the ground truth. The final segmentation results are presented in Table 1. The 80% column denotes the sorted 80% highest accuracy. As one can tell, the proposed method produces the best quantitative segmentation accuracy.

Table 1. PIXEL-WISE SEGMENTATION ACCURACY

	$P.M.*$	$P.V.$	$P.80\%$	$R.M.$	$R.V.$	$R.80\%$
MS	0.74	0.08	0.96	0.79	0.03	0.89
ISO [10]	0.71	0.09	0.99	0.81	0.02	0.92
MWS	0.89	0.05	1.0	0.77	0.01	0.86
GCC [2]	0.79	0.06	0.99	0.77	0.02	0.88
RLS [12]	0.84	0.02	0.95	0.85	0.01	0.92
Ours	0.90	0.01	0.98	0.89	0.01	0.96

* $P.M./R.M.$, $P.V./R.V.$, and $P.80\%/R.80\%$ are mean, variance, and sorted 80% highest accuracy of precision/recall, respectively.

Segmentation Performance Analysis II: To analyze the performance of the selection-based sparse shape model, we compare our shape modeling method with: mean shape model (MSM) generated with Procrustes analysis, traditional shape model using PCA [7], general sparse shape model using KSVD [1], and shape inference by solving the LASSO optimization problem. All these methods are tested with the same initialization. The results are displayed in Figure 4 in

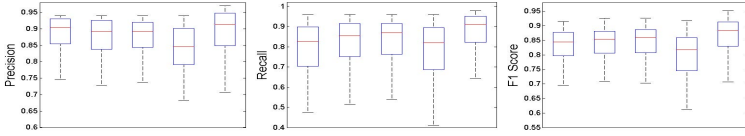


Fig. 4. Box plot for quantitative comparison (precision, recall, and F1 score). In each panel, x -axis denotes 5 shape models (from left to right): MSM, PCA, KSVD, LASSO, and Our method.

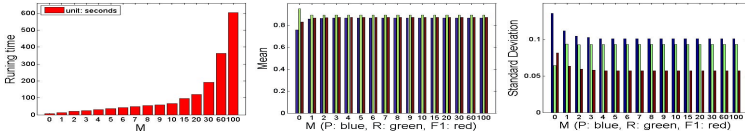


Fig. 5. The performance with respect to the number of nearest neighbors M

terms of precision P , recall R , and F1 score ($F1 = \frac{2PR}{P+R}$). MSM and LASSO give the worst performance in terms of recall, while shape models based on MSM, PCA, and KSVD have relatively higher precision. Our model produces better performance with respect to the criteria. Please note that one advantage of the proposed dictionary learning method is its robustness to outliers.

Running Time Analysis: The Matlab execution time of this novel algorithm depends on M in (8). We evaluate the execution time of the segmentation algorithm with $\{M = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 60, 100\}$ on one 704×704 image patch that contains $N = 101$ detected tumor cells, as shown in Figure 5. $M = 0$ corresponds to no repulsive term considered in (8). The recall is relatively high since all shapes deform independently without any repulsion until they reach the maximum number of iterations, therefore they can almost cover the whole cells. However, the precision of boundary segmentation is extremely low. The execution time is 602 seconds when $M = N - 1$, which represents the case that the repulsive forces from all other cells in the image are calculated. When $M = 5$, also used in our experiments, the execution time is 35 seconds. As shown in Figure 5, the segmentation accuracy does not have a significant variation when $M > 5$, since the square of distance in $f(x)$ of (8) greatly penalizes those contours far away. Practically, one can select a proper M by providing a rough estimation to speed up the algorithm instead of using $M = N - 1$.

4 Conclusion

In this work, we present a novel cellular image segmentation algorithm for lung cancer, which combines the bottom-up image appearances with top-down shape priors. The shape repository is learned with a novel locality-constrained, selection-based dictionary learning algorithm, and a local repulsive balloon snake is proposed for adaptive shape deformation.

Acknowledgement. This research is funded by Kentucky Lung Cancer Research Program (KLCR) pilot award. The project described is also supported by the National Center for Research Resources, UL1RR033173, and the National Center for Advancing Translational Sciences, UL1TR000117.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing over-complete dictionaries for sparse representation. *TSP* 54(11), 4311–4322 (2006)
2. Al-Kofahi, Y., Lassoued, W., Lee, W., Roysam, B.: Improved automatic detection and segmentation of cell nuclei in histopathology images. *TBME* 57(4), 841–852 (2010)
3. Ali, S., Madabhushi, A.: An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *TMI* 31(7), 1448–1460 (2012)
4. Chang, H., Han, J., Spellman, P.T., Parvin, B.: Multireference level set for the characterization of nuclear morphology in glioblastoma multiforme. *TBME* 59(12), 3460–3467 (2012)
5. Cheng, J., Rajapakse, J.C.: Segmentation of clustered nuclei with shape markers and marking functions. *TBME* 56(3), 741–748 (2009)
6. Cohen, L.D.: On active contour models and balloons. *CVGIP: Image Understanding* 53(2), 211–218 (1991)
7. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. *CVIU* 61(1), 38–59 (1995)
8. ElBaz, M.S., Fahmy, A.S.: Active shape model with inter-profile modeling paradigm for cardiac right ventricle segmentation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 691–698. Springer, Heidelberg (2012)
9. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55(1), 119–139 (1997)
10. Grady, L., Schwartz, E.L.: Isoperimetric graph partitioning for image segmentation. *TPAMI* 28(1), 469–475 (2006)
11. Kong, H., Gurcan, M., Belkacem-Boussaid, K.: Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. *TMI* 30(9), 1661–1677 (2011)
12. Qi, X., Xing, F., Foran, D.J., Yang, L.: Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *TBME* 59(3), 754–765 (2012)
13. Scott, D.W.: Parametric statistical modeling by minimum integrated squared error. *Technometrics* 43, 274–285 (2001)
14. Shi, Y., Qi, F., Xue, Z., Chen, L., Ito, K., Matsuo, H., Shen, D.: Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. *TMI* 27(4), 481–494 (2008)
15. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: *CVPR*, pp. 3360–3367 (2010)
16. Wilms, M., Ehrhardt, J., Handels, H.: A 4D statistical shape model for automated segmentation of lungs with large tumors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 347–354. Springer, Heidelberg (2012)

17. Wu, Z., Gurari, D., Wong, J.Y., Betke, M.: Hierarchical partial matching and segmentation of interacting cells. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 389–396. Springer, Heidelberg (2012)
18. Zhan, Y., Dewan, M., Zhou, X.S.: Cross modality deformable segmentation using hierarchical clustering and learning. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 1033–1041. Springer, Heidelberg (2009)
19. Zhang, S., Zhan, Y., Metaxas, D.N.: Deformable segmentation via sparse shape representation and dictionary learning. *MIA* 16(7), 1385–1396 (2012)

Flash Scanning Electron Microscopy

Raphael Sznitman, Aurelien Lucchi, Marco Cantoni,
Graham Knott, and Pascal Fua

École Polytechnique Fédérale de Lausanne, Switzerland
`firstname.lastname@epfl.ch`

Abstract. Scanning Electron Microscopy (SEM) is an invaluable tool for biologists and neuroscientists to study brain structure at the intra-cellular level. While able to image tissue samples with up to $5nm$ isotropic resolution, image acquisition is prohibitively slow and limits the size of processed samples. In this work, we propose a novel approach to speeding up imaging when looking for specific structures. Unlike earlier methods, we explicitly balance the conflicting requirements of spending enough time scanning potential regions of interest to ensure that all targets are found while not wasting time on unpromising regions. This is achieved by using a Markov Random Field to model target locations and optimizing scanning locations by using a Branch-and-Bound strategy. We show that our approach significantly outperforms state-of-the-art methods to locate mitochondria in brain tissue.

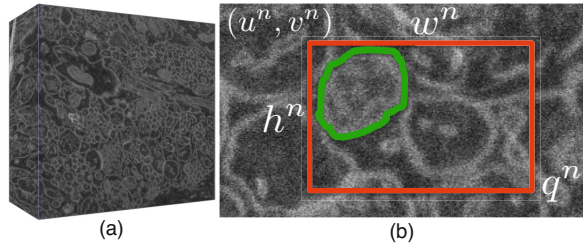
1 Introduction

Modern neuroscience has greatly benefited from Scanning Electron Microscope (SEM) technology. With its ability to image with nm and isotropic resolution, SEMs allow detailed visualization of intra-cellular structures as depicted in Fig. 1. This has already produced important advances in the study of neuro-degenerative diseases [1,2] and is expected to play a critical role in further understanding of brain function [3,4].

However, the very slow scanning speed of SEMs stands in the way of that promise because it severely limits the size of tissue samples that can be imaged. To image a volume, the electron beam has to scan the tissue surface multiple times, accumulate pixel information at each pass, and then average pixel samples to provide a clear and coherent image of the tissue surface. The top layer, or *slice*, is then removed using a diamond cutter or Focused Ion Beam and the process repeated for the newly visible layer. This process may take days for volumes as small as $10\mu m^3$. This is particularly burdensome for hypothesis testing procedures where the goal is to perform many different experimental manipulations and verify their effects on specific properties of tissue samples, such as counting the number of mitochondria or synapses per μm^3 .

In this paper, we therefore address the problem of speeding up SEM imaging for the purpose of detecting and measuring specific intra-cellular structures. While the majority of related research has focused on providing automatic segmentation and labeling tools [5,6,7], very few have addressed this throughput

Fig. 1. (a) Volume generated from slice by slice SEM scanning. (b) Sample slice image with a mitochondrion outline drawn in green. In red, the scanned region $q^n = (u^n, v^n, w^n, h^n)$ at time step n .



issue. Among these, hypothesize-and-scan strategies [7,8] save time by first scanning the image fewer times than typically required, using the resulting noisy image to hypothesize regions very likely to contain target structures, and re-scanning these regions. The procedure is repeated until only parts of the image that contain target structures are imaged with maximum quality. In [7], this is achieved by using a cascade of detectors to sequentially identify promising image regions. This requires correctly setting numerous parameters in the cascade and, as we will show in our experiments, is challenging to tune. By contrast, the scanning strategy of [8] is optimal in theory, but only when there is a single target in the field of view, an assumption that is systematically violated in practice and leads to inefficiencies. Furthermore, only results on simulated data are reported in these two papers [7,8].

Part of the problem with these earlier approaches is that they do not explicitly optimize for scanning time. The speed-ups are only byproducts of performing fewer scans at uninteresting locations. By contrast, in this work, we propose a hypothesize-and-scan approach that avoids the aforementioned problems by explicitly formulating the selection of regions to be scanned and re-scanned as finding the optimal compromise between maximizing the probability of finding target structures and minimizing the cost of scanning. In practice, this is done by using a Markov Random Field (MRF) to model the likelihood of targets being at different locations and performing mean-field approximate inference to estimate Maximum Posterior Marginals (MPM). Finding the optimal region to scan can then be solved exactly and yields regions to be re-scanned that are both compact and likely to contain targets. This strategy is simple to implement and we demonstrate significant accuracy and speed improvements on real brain tissue.

2 Flash Scanning for Target Discovery

The goal of our *Flash Scanning* approach is to scan as many times as necessary *all* the target structures and the rest as sparsely as possible. In essence, this means simultaneously satisfying two conflicting goals—minimizing the size of the scanned regions, while maximizing the number of targets within them—that is, finding a Pareto-Optimum.

To this end, for each slice, we first scan each pixel once and use a mean-field approximation to estimate the probability of a pixel belonging to a target

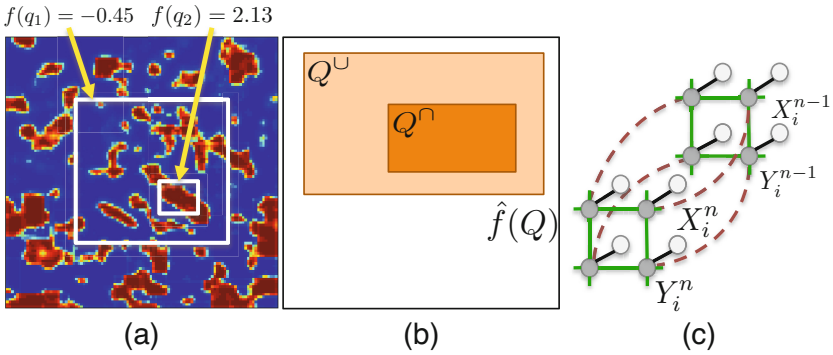


Fig. 2. (a) Visual representation of $P(Y_i^n | X^n, Y^{n-1})$, with high values in red and low values in blue. Two rectangular regions, q_1 and q_2 , are shown and their corresponding f scores. (b) Branch-and-Bound: depiction of regions Q^U and Q^N when optimizing the bounding function $\hat{f}(Q)$. (c) MRF depiction with 4-connected neighbourhood and dependencies on previous slices.

structure, based both on learned texture measures and the strong correlation between slices. We then find a rectangle that maximizes the integral of the probability density within it, minus its area multiplied by a fixed coefficient. In essence, the multiplicative coefficient selects one among many possible Pareto-Optima because the scanning time is proportional to the rectangle area. We then solve our optimization problem exactly using a branch-and-bound approach, re-scan the optimal rectangle, update our probabilities and iterate. This process is visually depicted in the supplementary video associated with this paper.

2.1 Notation

For a single slice, S , let $n = 0, \dots, N$ be discrete time steps of our iterative process. Let I^n be the reconstructed image of the slice at step n , with $I^0 = 0$ at each pixel location. Our goal is to scan the slice such that we reconstruct clearly all targets by time step N .

At each step n , we must hence direct the electron beam to scan a rectangular region of the slice **once** and denote this region $q^n = (u^n, v^n, h^n, w^n)$, as depicted by Fig. 1(b). Let C_p be the number of times pixel p has been scanned and let C_{\max} be the maximum number of times a pixel can be scanned. For each scanned region q^n , we acquire the corresponding new pixel intensities $\{s_p^n\}_{p=1}^{|q^n|}$, and incorporate them into I^n by weighted averaging for all $p \in q^n$, $I_p^n = I_p^{n-1}(1 - \alpha_p) + s_p^n \alpha_p$, where $\alpha_p = 1/C_p$. Furthermore, let I^n be decomposed into $r \times r$ image patches and let $Y_i^n \in \{-1, 1\}$ be a random variable denoting the presence of a target in image patch i at iteration n . Also, let $X_i^n \in \{-1, 1\}$ be a random image observation at location i and computed from the image I^n .

2.2 Objective and Optimization

Our goal at each time step n is to compute the next region q^{n+1} to scan in order to image as fast as possible all the targets with the highest accuracy. In particular, we want regions with high target probability to be scanned many times, while avoiding scanning large regions that are irrelevant. For this reason, we take q^{n+1} to be

$$q^{n+1} = \arg \max_{q \in \mathcal{Q}} f(q) = \arg \max_{q \in \mathcal{Q}} \sum_{i \in q} P(Y_i^n = 1 | X^n, Y^{n-1}) \mathbb{1}_{\{C_i \leq C_{\max}\}} - \lambda |q|, \quad (1)$$

where $P(Y_i^n = 1 | X^n, Y^{n-1})$ is the conditional probability that a target is at patch Y_i^n , $|q|$ is the area of region q , \mathcal{Q} is the space of all rectangular patch regions and $\lambda > 0$ is a user parameter that balances accuracy and speed. In addition, $\mathbb{1}_{\{C_i \leq C_{\max}\}}$ is 1 as long as $C_i \leq C_{\max}$ and 0 there after, hence penalizing regions that have been over-scanned.

To give some intuition for the meaning of Eq. (1), Fig. 2(a) depicts the conditional probability at each location i , two candidate regions (q_1, q_2) and their respective f scores. Observe that the objective favors regions very likely to contain targets, while penalizing large ones that are slow to scan, which results in densely packed target regions.

Given Eq. (1) is additive in $P(Y_i^n = 1 | X^n, Y^{n-1})$, we may use a Branch-and-Bound optimization strategy [9] to solve the global optimum exactly. The idea is to evaluate bounds on subsets of regions $Q \subset \mathcal{Q}$ in a divide-and-conquer approach. This is achieved by defining a bounding function $\hat{f} : \mathcal{Q} \rightarrow \mathbb{R}$ for subsets of regions. We write

$$\hat{f}(Q) = \sum_{i \in Q^{\cup}} P(Y_i^n = 1 | X^n, Y^{n-1}) \mathbb{1}_{\{C_i \leq C_{\max}\}} - \lambda |Q^{\cap}|, \quad (2)$$

where Q^{\cup} is the union of all rectangles in Q and Q^{\cap} is the intersection of all rectangles in Q , as depicted in Fig. 2(b). Clearly, \hat{f} is a proper bounding function as $f(q) = \hat{f}(Q)$ when $Q = q$, and $\hat{f}(Q) \geq \max_{q \in Q} f(q)$.

3 Implementation

Given the optimization problem of Eq. (1), we now discuss how we compute the image observations X_i^n and the posterior distributions $P(Y_i^n = 1 | X^n, Y^{n-1})$, as well as our Flash Scanning algorithm.

Target Detectors and Observation Model. The observation X_i^n , describing the evidence of a target at location i , is computed directly from the reconstructed image I^n . As in [5], we use simple feature extraction with SVM classification to describe this evidence. Specifically, for node i , we compute 1) a feature dedicated to node i and 2) a feature from the 8-connected neighbours of node i . For 1) we compute for image patch i , the histogram of intensity (10 bins), the co-occurrence matrix (64 bins) and concatenate the two. For 2), we extract the

same features as in 1) from each neighbour of i , adding intensity histograms together and co-occurrences together. These are then normalized by the number of neighbours and concatenated, yielding a second 10+64 bin feature. Both 74-bin feature vectors are then concatenated and the estimator, X_i^n is the output of a trained RBF-kernel binary SVM (target=1 or background=-1) evaluated with the extracted 148 bin feature vector. Given that image noise varies as a function of how many scans have been performed at a location, a dedicated c-scan SVM is used when estimating regions observed with c scans. Note that other features may improve performances and in effect shift the pareto-optimum higher.

Target Model and Posterior Update. Given we are interested in inferring each of the variables Y^n , we use a Markov Random Field (MRF) to describe the joint probability distribution of all the target locations Y^n and observations X^n . We write

$$P(X^n, Y^n | Y^{n-1}) = \frac{1}{Z} \prod_i P(Y_i^n | Y_i^{n-1}) \prod_{j \in \mathcal{N}_i} P(Y_i^n Y_j^n) \prod_i P(X_i^n | Y_i^n), \quad (3)$$

where \mathcal{N}_i is a 4-connected neighbourhood of node i (depicted in Fig. 2(c)) and Z is a normalizing constant. This model allows us to describe the likelihood of a given patch of being a target, while taking into account it's local neighbourhood. Here, we have assumed that X_i^n is conditionally independent of Y_i^{n-1} given Y_i^n , and $P(Y^n | Y^{n-1}) = \prod_i P(Y_i^n | Y_i^{n-1}) \prod_{j \in \mathcal{N}_i} P(Y_i^n Y_j^n)$. This MRF model is convenient as it allows us to represent where all the targets lie on a current slice, as a function of occupancy in previous slices.

Within this MRF model, $P(Y_i^n | X^n, Y^{n-1})$ is known as the Maximum Posterior Marginal (MPM). To compute these for all Y_i^n , we use the Mean-Field [10] approximation method, which involves setting $P(Y_i^n | X^n, Y^{n-1}) = \mu_i$, where μ_i is initially 0 and then iteratively updating

$$\mu_i = \mu_i \beta + (1 - \beta) \tanh\left(\sum_{j \in \mathcal{N}_i} \mu_j + P(Y_i^n | X_i^n) + P(Y_i^n | X_i^n, Y_i^{n-1})\right), \quad (4)$$

where β is a damping rate and $P(X_i^n | Y_i^n) \propto (1 + \exp(g(X_i^n)))^{-1}$ with $g(X_i^n)$ being the output of the SVM. In our experiments, we set $\beta = 0.5$, iterate 10 times for each node and set $P(Y_i^n | Y_i^{n-1}) = \exp(Y_i^n Y_i^{n-1} / T)$ and $P(Y_i^n, Y_j^n) = \exp(Y_i^n Y_j^n / T)$, where $T = 5$ is a temperature.

Algorithm: The proposed algorithm is summarized in Alg. 1. For each slice, the user provides an initial probability distribution for each node $P(Y_i^0)$, λ and the total number of iterations permitted N . Both the final MPM and the scanned image I^N are returned by our procedure. For a given slice, we first scan each pixel location once (line 1), then iterate lines 3-6 N times. For subsequent slices, we first apply temporal smoothing of the form $P(Y_i) = P(Y_i)\alpha + 1/2(1 - \alpha)$, where $\alpha = 0.75$ in our experiments, and then use the smoothed probability as the prior for the next slice. As in [7], this is to account for the 3D nature of the targets and promote regions that have been heavily scanned in previous locations, as they are likely to be in nearby locations. For the first slice, we set $P(Y_i^0) = 1/2$.

Algorithm 1. Flash Scanning ($P(Y^0), \lambda, N$)

-
- 1: Scan the entire slice one.
 - 2: **for** $n = 1, \dots, N$ **do**
 - 3: Select region to scan: $q^n = \arg \max_{q \in \mathcal{Q}} f(q)$
 - 4: Scan region q^n and update I^n
 - 5: Compute target estimators: $\{X_i^n\}_{i \in q^n}$ from I^n
 - 6: Compute MPMs using Eq. (4)
 - 7: **end for**
 - 8: **return** $P(Y^N)$ and I^N
-

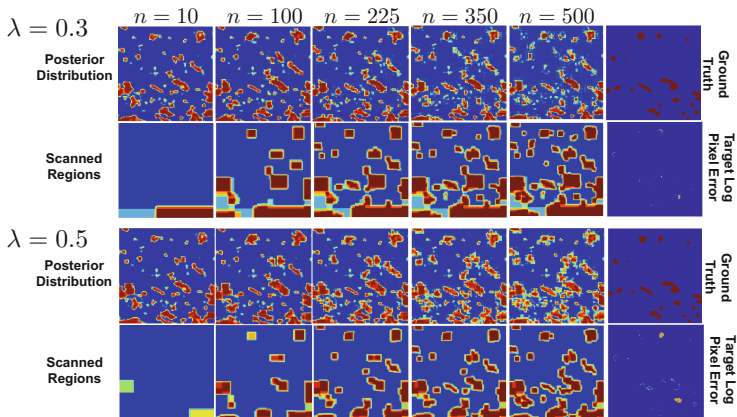


Fig. 3. Flash Scanning iterations for a slice when $\lambda = 0.5$ and 0.3 . In each case, we show the evolution of the posterior distribution and the scanned locations for different values of n . At $n = 500$, we show the log-pixel errors for mitochondria locations (yellow are high errors and blue are low errors).

4 Experiments and Results

To test our approach we imaged 120, $5nm$ thick slices using a Zeiss NVision40 FIBSEM microscope, such that each slice consisted of a 1280×1280 image with $5nm$ pixel resolution. Each slice was scanned a total of 10 times to produce 10 noisy 1280×1280 images, *i.e.* $C_{\max} = 10$, and we collected the individual pixel values from each scan using a Fibics scanning head [11].

As in [7,8], we use mitochondria as our target structures. Mitochondria pixels were hand labeled in the entire 120-image stack by an expert. The first 20 images were used to train the different $c = \{1, \dots, C_{\max}\}$ scan SVM classifiers as described in Sec. 3 and the rest for testing purposes. Each Y_i node of the MRF consisted of an image patch of size 12×12 pixels. Computations were performed offline using a 2.3GHz PC. In practice, steps 3 and 6 of Alg. 1 take approximately 0.2 seconds, which is negligible considering the days of scanning performed during typical experiments.

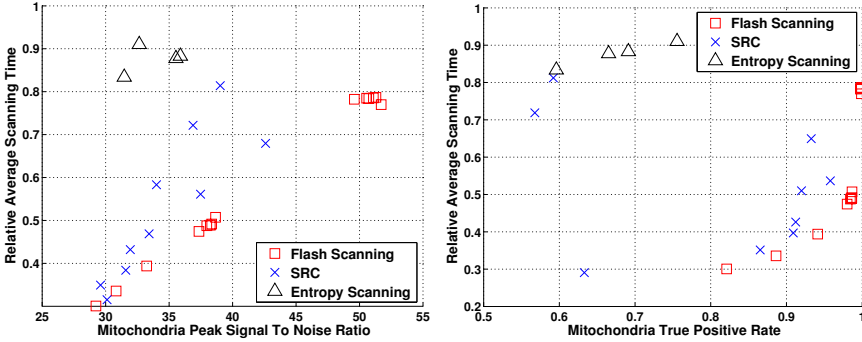


Fig. 4. Scanning time, PSNR and TPR performances of Flash Scanning and earlier methods. For each method, a data point corresponds to a set of parameters used. In general, Flash scanning consistently outperforms earlier approaches by being faster and more accurate at scanning mitochondria.

Evaluation: In Fig. 3, we show a sequence of iterations of the Flash Scanning algorithm for a given slice with parameters $N = 500$ and $\lambda = \{0.3, 0.5\}$. We show the evolution of the MPMs and the number of scans dedicated to each region of the space as n approaches 500 (large values in red, small values in blue). We also show the ground truth locations of mitochondria for the evaluated slice, as well as the log pixel error between the reconstructed images and the maximally scanned images at mitochondria locations.

From the figure, we see that our objective Eq. (1) concentrates scanning on target locations and generally avoids scanning irrelevant regions. We also see that when $\lambda = 0.5$, scanned regions are smaller and more compact than when $\lambda = 0.3$. This is consistent with Eq. (1) that imposes heavier costs for scanning large windows for large values of λ . In other words, larger values of λ means faster scanning at the risk of missing some target structures. This effect is clearly visible at $n = 500$ where some mitochondria have not been scanned when $\lambda = 0.5$ but have been when $\lambda = 0.3$. The full sequence of iterations can be viewed in our supplementary video.

Comparison: We evaluated our method against the SRC [7] and the Entropy Scanning (ES) [8] algorithms. For each tested method, including ours, we evaluated the scanning time relative to scanning the entire block with 10 scans, the Peak Signal-To-Noise Ratio (higher values being better) at mitochondria locations and the True Positive Rate (TPR) for regions scanned maximally. To show the effect of using different parameters, we ran our algorithm with different parameter values, $\{N \in (300, 1000), \lambda \in (0.3, 0.6)\}$. We did the same for both SRC and ES, while using the same SVM classifiers for all three methods. The results are shown in Fig. 4.

Flash scanning outperforms SRC, which itself does much better than ES. More specifically, for the tested set of parameters, Flash scanning significantly improves the speed and accuracy compared to SRC. For example, at roughly half traditional scanning times, our method has an average PSNR of over 37,

while SRC is of under 34. Similarly, for a PSNR of 37, SRC is 8% slower than our method. Note, that these differences are statistically significant, and similar observations can be made with respect to the reported TPR results. Note that similar results were also attained by performing experiments on synthetic data generated as described in [7].

5 Conclusion

We have presented a new method for efficiently scanning tissue samples using SEMs when imaging predefined targets. Our method uses an intuitive objective function that selects scanning regions that are likely to contain targets, but which are small in order to avoid taking too long to scan. In addition, our optimization is solved exactly and use a MRF to allow localization of multiple targets simultaneously. Our method requires far fewer parameters to tune compared to earlier approaches and out-performs them in terms of speed and accuracy. In the future, we plan to investigate how to integrate the cost of beam reposition when selecting regions to scan.

Acknowledgements. This work was supported in part by the ERC grant MicroNano.

References

1. Knott, G., Marchman, H., Wall, D., Lich, B.: Serial Section Scanning Electron Microscopy of Adult Brain Tissue Using Focused Ion Beam Milling. *Journal of Neuroscience* 28(12), 2959–2964 (2008)
2. Van-Rijnsoever, C., Oorschot, V., Klumperman, J.: Correlative light-electron microscopy combining live-cell imaging and immunolabeling of ultrathin cryosections. *Nature Methods* 5, 973–980 (2008)
3. Alivisatos, P., Chun, M., Church, G., Greenspan, R., Roukes, M., Yuste, R.: The brain activity map project and the challenge of functional connectomics. *Neuron* 74, 970–974 (2012)
4. HBP: EU human brain project (2013), <http://www.humanbrainproject.eu/>
5. Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P.: Supervoxel-Based Segmentation of Mitochondria in EM Image Stacks with Learned Shape Features 31(2), 474–486 (2011)
6. Laptev, D., Vezhnevets, A., Dwivedi, S., Buhmann, J.M.: Anisotropic ssTEM image segmentation using dense correspondence across sections. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 323–330. Springer, Heidelberg (2012)
7. Sznitman, R., Lucchi, A., Pjescic-Emedji, N., Knott, G., Fua, P.: Efficient Scanning for EM Based Target Localization. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 337–344. Springer, Heidelberg (2012)
8. Sznitman, R., Lucchi, A., Frazier, P., Jedynek, B., Fua, P.: An optimal policy for target localization with application to electron microscopy. In: ICML (2013)
9. Lampert, C., Blaschko, M., Hofmann, T.: Efficient Subwindow Search: A Branch and Bound Framework for Object Localization. *PAMI* 31, 2129–2142 (2009)
10. Murphy, K.: *Machine Learning: A Probabilistic Perspective*. MIT (2012)
11. Fibics (2012), <http://www.fibics.com>

Superpixel Classification Based Optic Cup Segmentation

Jun Cheng^{1,*}, Jiang Liu¹, Dacheng Tao², Fengshou Yin¹,
Damon Wing Kee Wong¹, Yanwu Xu¹, and Tien Yin Wong^{3,4}

¹ Institute for Infocomm Research, Agency for Science,
Technology and Research, Singapore

{jcheng, jliu, fyin, wkwong, yaxu}@i2r.a-star.edu.sg

² University of Technology, Sydney, Australia
dacheng.tao@uts.edu.au

³ Singapore Eye Research Institute, Singapore

⁴ National University of Singapore, Singapore
ophwty@nus.edu.sg

Abstract. In this paper, we propose a superpixel classification based optic cup segmentation for glaucoma detection. In the proposed method, each optic disc image is first over-segmented into superpixels. Then mean intensities, center surround statistics and the location features are extracted from each superpixel to classify it as cup or non-cup. The proposed method has been evaluated in one database of 650 images with manual optic cup boundaries marked by trained professionals and one database of 1676 images with diagnostic outcome. Experimental results show average overlapping error around 26.0% compared with manual cup region and area under curve of the receiver operating characteristic curve in glaucoma detection at 0.811 and 0.813 in the two databases, much better than other methods. The method could be used for glaucoma screening.

1 Introduction

Glaucoma is a chronic eye disease in which the optic nerve is progressively damaged. It is the second leading cause of blindness, and is predicted to affect around 80 million people by 2020 [1]. Although glaucoma cannot be cured, its progression can be slowed down by treatment. Therefore, it is important to detect it in time. The cup-to-disc ratio (CDR) [2] is a widely used indicator for glaucoma. A larger CDR indicates a higher risk of the disease. In current clinical practice, the CDR is measured manually. However, manual measurement is subjective, time consuming and expensive for large scale screening. Recently, automated CDR measurement from 2D fundus images starts to receive some attention. In 2D images, the optic disc can be divided into two distinct zones, namely, a central bright zone called the optic cup (in short, cup) and a peripheral region called

* This work is funded by Singapore A*STAR SERC Grant (092-148-0073).

the neuroretinal rim (in short, rim), as shown in Fig. 1. The CDR is computed as the ratio of the vertical cup diameter to the vertical disc diameter clinically. Accurate segmentation of the cup is essential for CDR measurement. In this paper, we focus on automatic cup segmentation only.

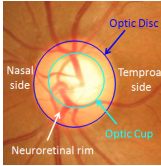


Fig. 1. Major structures of the optic disc



Fig. 2. An example with no pallor

Cup segmentation in 2D fundus images is a challenging task as 2D images do not carry depth information, which is the primary indicator of the cup boundary. In these images, one landmark for cup detection is the area of maximum color contrast inside the optic disc, i.e., pallor. Another landmark is the vessel bends at the cup boundary. Several methods have been proposed for cup segmentation from 2D images. In [3], thresholding is used. It is essentially based on pallor information. However, in many subjects from screening, there is no obvious pallor to mark the cup boundary. Fig. 2 shows such an example. In [4], [5], vessel bends are used. The challenge is to correctly identify vessel bends. Xu *et al.* proposed a sliding window and regression based method [6]. However, the sliding window strategy requires heavy computational cost. Recently, Yin *et al.* [7] developed a deformable model based approach, where the initialization of the cup boundary is based on pallor combined with prior knowledge of cup.

One main challenge in cup segmentation is to determine the cup boundary in images with weak or no pallor. In such scenarios, it is difficult to determine the cup based on intensity only. Although vessel bends are potential landmarks, they suffer from high false alarm. In addition, these methods often require a set of heuristic parameters which raises the concern of the robustness of the methods. In this paper, we propose superpixel classification for cup segmentation. The proposed method embeds the prior knowledge into superpixel classification by using location feature to overcome the absence of pallor. In addition, other challenges such as uneven illuminations within and between images have also been newly discussed and explored. Contrast enhanced intensities, mean removed intensities and center surround statistics are proposed to overcome these challenges accordingly. Based on the detected cup, CDR is computed for glaucoma detection.

The rest of paper is organized as follows. In Section 2, we introduce superpixel classification based cup segmentation including the generation of superpixels, the extraction of feature from superpixels for the classification and the classification of superpixels for cup estimation. Section 3 shows the experimental results including the accuracy for cup segmentation, the CDR and the glaucoma detection. Discussions and conclusions are presented in the last section.

2 Methodology

The proposed superspixel classification based method comprises: an over-segmentation step to divide the disc image into superspixels; a feature extraction step to compute features from each superspixel; a classification step to determine each superspixel as cup or non-cup (rim) and to estimate the boundary. Fig. 3 shows the procedure. Superspixel is an image patch which is better aligned with intensity edges than a rectangular patch. It captures redundancy in the image and reduces the complexity of subsequent processing. This paper uses the simple linear iterative clustering [8] algorithm (SLIC) to aggregate nearby pixels into superspixels in retinal fundus images. Compared with other superspixel methods, SLIC is fast, memory efficient and has excellent boundary adherence. It is also simple to use with only one parameter, i.e., the number of desired superspixels k . Next, features from superspixels are extracted to classify them as cup or non-cup.

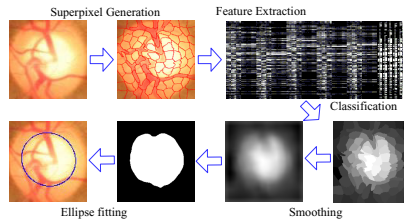


Fig. 3. Superspixel based optic cup segmentation

2.1 Feature Extraction

Many features such as color, appearance, gist, location and texture can be extracted for superspixel classification. To compute good features for cup segmentation, it is important to understand the main challenges in the task. The first challenge is that the contrast between rim and cup varies largely from one image to another. Therefore, it is beneficial to enhance the image contrast. In this paper, we apply histogram equalization on the red r , green g , and blue b channels to get three contrast enhanced intensity maps $HE(r), HE(g), HE(b)$, where $HE(\cdot)$ denotes the function of histogram equalization. The second challenge is the illumination changes from one image to another, i.e., some images look darker and some brighter. To overcome this issue, the mean removed intensity maps are also computed as $r - \bar{r}, g - \bar{g}, b - \bar{b}$, where \bar{x} denotes the mean value in channel x . Combining them with the contrast enhanced intensity maps, we obtain six intensity based maps $I = [I_1, \dots, I_6]$. The first type of features from a superspixel are the mean intensities MI of pixels within the superspixel. Mathematically, MI for j^{th} superspixel SP_j with n_j pixels is computed as: $MI(i) = \frac{1}{n_j} \sum_{(x,y) \in SP_j} I_i(x,y)$, for $i = 1, \dots, 6$.

Unbalanced illumination across the optic disc within the image is another challenge. For example, the temporal side of the disc is often brighter than the nasal side. In this paper, we compute center surround statistics (CSS) to overcome this issue. CSS are useful as they are computed from the difference between image maps, which are less sensitive to unbalanced illumination across the disc. To compute CSS, nine spatial scale dyadic Gaussian pyramids are generated with a ratio from 1:1 (level 0) to 1:256 (level 8). Then center-surround operation between center finer levels $c = 2, 3, 4$ and surround coarser levels $s = c + d$, with $d = 3, 4$ is applied to obtain six maps computed at levels of 2-5, 2-6, 3-6, 3-7, 4-7, and 4-8 from an image channel [9]. Because of the scale difference, surround levels are interpolated to be the same size as the corresponding center levels, and then they are subtracted to generate the center surround difference maps. In this paper, we compute the maps from r , g and b channels to get $6 \times 3 = 18$ maps. The CSS features are then computed as the mean and variance of these maps within superpixels. Denoting M_i , $i = 1, 2, \dots, 18$, as the i^{th} map, the feature CSS consists of the mean M_μ and variance M_σ of maps within the superpixels, i.e., $CSS = [M_\mu \ M_\sigma]$, where M_μ and M_σ are computed as follows:

$$M_\mu(i) = \frac{1}{n_j} \sum_{(x,y) \in SP_j} M_i(x,y), \quad M_\sigma(i) = \frac{1}{n_j} \sum_{(x,y) \in SP_j} (M_i(x,y) - M_\mu(i))^2 \quad (1)$$

Combining CSS with MI yield a feature with $18 \times 2 + 6 = 42$ dimensions. Besides the above, blood vessels from the cup and rim often look quite similar. In order to differentiate the vessels from the two regions, MI and CSS from neighboring superpixels are also included as context features in the classification of current superpixel. By choosing proper desired number of superpixels, a superpixel from vessels often have two or more non-vessel neighbors. Therefore, the vessels from the cup and rim can be better differentiated from their neighbors. In this paper, four neighboring superpixels for current superpixel SP_j are obtained and denoted as SP_{j_1} , SP_{j_2} , SP_{j_3} and SP_{j_4} . SP_{j_1} is determined as the first superpixel by moving out of the current superpixel horizontally to the left from its center. Similarly, SP_{j_2} , SP_{j_3} and SP_{j_4} are determined by moving right, up and down.

Since intensities and CSS features rely on color changes from cup to rim, they do not work well for images with very weak or no color changes from the cup to rim. To obtain a reliable result in such a scenario, we make use of the prior knowledge that the cup is usually the center part of the disc. For this purpose, we further include the distance D between the center of superpixel SP_j and the center of optic disc as location information. Mathematically, D is computed as

$D = \sqrt{\left(\frac{x_c - x_j}{h}\right)^2 + \left(\frac{y_c - y_j}{w}\right)^2}$, where (x_c, y_c) denotes the coordinate of the disc centre, (x_j, y_j) denotes the coordinate of the centre of SP_j , h and w denotes the height and width of the disc, respectively. By including the location information, the prior knowledge would be automatically embedded into classification. The final feature has $42 \times 5 + 1 = 211$ dimensions.

2.2 Superpixel Classification

The support vector machine (SVM) is used for the two-class classification in our application. The LIBSVM [10] with linear kernel is used in our experiments as we find that the nonlinear mapping using the radial basis function kernel does not improve much the performance. In the training step, we randomly obtain the same number of superpixels from the cup and the rim. Two-fold cross-validation is adopted to determine the SVM parameters. In the testing, instead of directly using the binary classification results from LIBSVM, we obtain the output values from the SVM decision function and smooth them with a mean filter, as illustrated in Fig. 3. The smoothed decision values are then used to get the binary decisions for all pixels by the default threshold used in SVM. The largest connected object is obtained and its boundary is used as the raw estimation. The best fitted ellipse [11] is computed as the cup boundary. The ellipse fitting here is beneficial to overcome the noise introduced by vessels at the inferior and superior sector of the cup.

3 Experimental Results

Two data sets are used. The first set includes 650 different eye images from Singapore Malay Eye Study (SiMES). Manual disc and cup boundaries are available for the 650 images. The 650 images are randomly divided into 325 images for training and 325 images for testing. The second set includes 1676 images of a population based data from Singapore Chinese Eye Study (SCES). Only diagnostic outcome is available for SCES. All SCES data are used in evaluating the performance of glaucoma screening using the proposed method. The disc boundaries are obtained following the method in [12]. Among the 2326 eyes, 168 of SiMES and 46 of SCES eyes are diagnosed as glaucomatous by ophthalmologists. The overlapping error E is computed as one evaluation metric to examine the difference between automated and manual cup: $E = 1 - \frac{Area(S \cap M)}{Area(S \cup M)}$, where S and M

Table 1. Mean overlapping error μ_E at different parameters

	no filter	5 × 5	10 × 10	15 × 15	20 × 20	25 × 25
50	29.3%	28.4%	27.7%	27.2%	26.8%	26.6%
100	27.1%	26.5%	26.3%	26.1%	26.0%	26.1%
200	26.6%	26.2%	25.9%	25.9%	25.8%	25.8%
400	26.6%	26.1%	26.0%	25.9%	25.8%	25.9%

Table 2. Performance by various methods

Measurement	μ_E	Mean CDR error μ_δ		AUC	
		SiMES		SiMES	SCES
Data set	SiMES	glaucoma	healthy		
Threshold [3]	53.5%	0.141	0.129	0.638	0.574
ASM [7]	31.3%	0.121	0.088	0.716	0.756
Regression [6]	28.4%	0.133	0.091	0.729	0.767
$MI + CSS$	28.3%	0.108	0.084	0.734	0.694
$MI + D$	27.0%	0.074	0.102	0.757	0.768
$CSS + D$	26.8%	0.102	0.078	0.782	0.789
$MI + CSS + D$ (Proposed)	26.0%	0.100	0.075	0.811	0.813
Manual	-	-	-	0.839	-

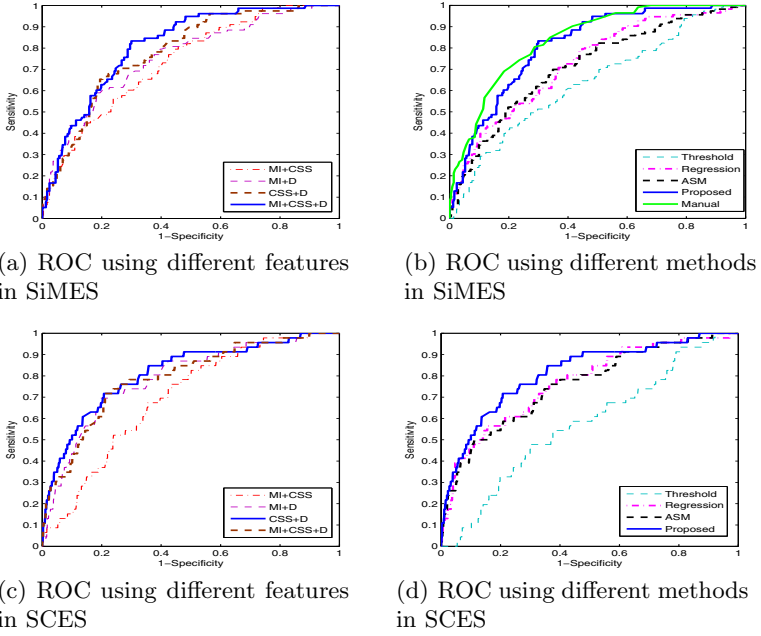
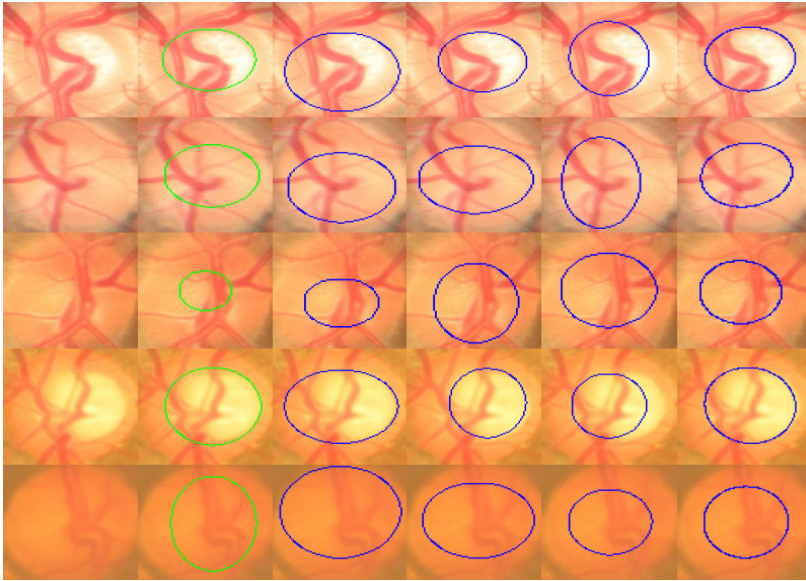


Fig. 4. ROC of glaucoma detection



(a) Original (b) Manual (c) Threshold (d) ASM (e) Regression (f) Proposed

Fig. 5. Sample results. From top to bottom, the manual CDR for the five examples from top to bottom are: 0.54, 0.55, 0.35, 0.68 and 0.83. The CDR by the proposed method are 0.55, 0.55, 0.54, 0.65, and 0.63 respectively.

denote the segmented and the manual cup respectively. In addition, the CDR accuracy is also evaluated using CDR error, computed as $\delta = |CDR_m - CDR|$, where CDR_m denotes the manual CDR. The glaucoma detection accuracy is evaluated using the area under curve (AUC) of the receiver operating characteristic (ROC) curve.

The first set of experiments was carried out to evaluate the performance under different parameters. Table 1 shows the mean overlapping error μ_E for the number of superpixels k at 50, 100, 200 and 400 in combined with different filter sizes from 5×5 to 25×25 as well as the case without a filter. The results show $k = 50$ is too small for this application while there is mild improvement when k increases from 100 to 400. It is shown that the results with a filter are better than that without. The size of filter only affects the performance very slightly from 15×15 to 25×25 . Since a smaller k requires less computation, we adopt $k = 100$ with a 20×20 filter in subsequent tests.

In order to evaluate how the features MI , CSS and D affect the cup segmentation, we conducted the tests using different combinations of the three types of features. Besides the mean overlapping error μ_E , we also compute mean CDR error μ_δ and AUC of ROC curve in glaucoma detection to evaluate the performance. In addition, we have also compared the proposed method with the threshold method [3], the ASM method [7] and the regression method [6]. Table 2 shows mean overlapping error μ_E , the mean CDR error μ_δ and the AUC of the ROC curves. Fig. 4 shows the ROC curves. Results show that MI works well for glaucomatous subjects while relatively poor for healthy subjects. CSS works better for healthy subjects. D is an important feature that improves the performance significantly. Fig. 5 shows the cup boundaries for five samples by the proposed method and three other methods. From the results, we can see that previous methods can work well when pallor is obvious, but their performance in discs without obvious pallor is less accurate. The proposed method over-estimates very small cups ($CDR < 0.4$) and under-estimates very large cups ($CDR > 0.8$) when the pallor is very weak or absent, e.g., the third and last samples. This is because there are very few such very small or very large cups in the data set. When D is included in the feature space where most discs have medium cups, the trained classifier is dominated by the majority medium cups. This limitation is reasonable as these are the most difficult cases even for human eyes. Although the bias exists, the obtained CDRs for very large cups are still high and the CDRs for very small cups are still small, as shown in Fig. 5. Therefore, we obtain good glaucoma detection results.

4 Discussions and Conclusions

In this paper, we present a superpixel classification based cup segmentation for glaucoma detection. It is shown that MI is a good feature as majority of glaucomatous subjects often have relatively obvious pallor. CSS is useful to reducing difficulty due to the uneven illumination. It is less sensitive to the illumination change. The location information further improves the results as it incorporates the prior knowledge of the cup.

The accuracy of our current method is much better than previous methods [3], [6], [7]. The glaucoma detection accuracy by the proposed method is only slightly below the manual CDR. From the discussions with clinicians, it is good enough for a large-scale glaucoma screening program. However, there are still many aspects for improvement in the proposed methods. For example, the proposed method under-estimates the very large cups while over-estimating the very small cups when pallor is very weak or absent. In addition, CDR based screening also has its limitations. Therefore, combining CDR with other factors are expected to further improve the performance. In the future, we would explore the integration of other factors to improve diagnosis outcomes toward a more reliable and efficient glaucoma screening system.

References

1. Quigley, H.A., Broman, A.: The number of people with glaucoma worldwide in 2010 and 2020. *Br. J. Ophthalmol.* 90(3), 262–267 (2006)
2. Damms, T., Dannheim, F.: Sensitivity and specificity of optic disc parameters in chronic glaucoma. *Invest. Ophthalm. Vis. Sci.* 34, 2246–2250 (1993)
3. Joshi, G.D., Sivaswamy, J., Karan, K., Krishnadas, R.: Optic disk and cup boundary detection using regional information. In: *Proc. IEEE Int. Symp. Biomed. Imag.*, pp. 948–951 (2010)
4. Joshi, G.D., Sivaswamy, J., Krishnadas, S.R.: Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE Trans. on Medical Imaging* 30, 1192–1205 (2011)
5. Wong, D.W.K., Liu, J., Lim, J.H., Li, H., Wong, T.Y.: Automated detection of kinks from blood vessels for optic cup segmentation in retinal images. In: *Proc. SPIE*, vol. 7260, p. 72601J (2009)
6. Xu, Y., Xu, D., Lin, S., Liu, J., Cheng, J., Cheung, C.Y., Aung, T., Wong, T.Y.: Sliding window and regression based cup detection in digital fundus images for glaucoma diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part III. LNCS*, vol. 6893, pp. 1–8. Springer, Heidelberg (2011)
7. Yin, F., Liu, J., Wong, D.W.K., Tan, N.M., Cheung, C., Baskaran, M., Aung, T., Wong, T.Y.: Automated segmentation of optic disc and optic cup in fundus images for glaucoma diagnosis. In: *IEEE Int. Symp. on Computer-Based Medical Systems*, pp. 1–6 (2012)
8. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pat. Ana. Mach. Intel.* 34, 2274–2282 (2012)
9. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259 (1998)
10. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
11. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least squares fitting of ellipses. *IEEE Trans. on Pat. Anal. and Mach. Intel.* 21(5), 476–480 (1999)
12. Cheng, J., Liu, J., Xu, Y., Yin, F., Wong, D.W.K., Tan, N.-M., Cheng, C.-Y., Tham, Y.C., Wong, T.Y.: Superpixel classification based optic disc segmentation. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) *ACCV 2012, Part II. LNCS*, vol. 7725, pp. 293–304. Springer, Heidelberg (2013)

Learning from Partially Annotated OPT Images by Contextual Relevance Ranking

Wenqi Li¹, Jianguo Zhang¹, Wei-Shi Zheng²,
Maria Coats³, Frank A. Carey³, and Stephen J. McKenna¹

¹ CVIP, School of Computing, University of Dundee, Dundee, UK

² School of Information Science and Technology, Sun Yat-sen University, China

³ Ninewells Hospital & Medical School, University of Dundee, UK

Abstract. Annotations delineating regions of interest can provide valuable information for training medical image classification and segmentation methods. However the process of obtaining annotations is tedious and time-consuming, especially for high-resolution volumetric images. In this paper we present a novel learning framework to reduce the requirement of manual annotations while achieving competitive classification performance. The approach is evaluated on a dataset with 59 3D optical projection tomography images of colorectal polyps. The results show that the proposed method can robustly infer patterns from partially annotated images with low computational cost.

1 Introduction

Optical Projection Tomography (OPT) microscopy is a relatively new 3D imaging modality [7]. It has an effective resolution of $5\mu\text{m}$ to $10\mu\text{m}$ and is ideally suited for specimens between 0.5 mm and 10 mm. Recently OPT has been used to image colorectal polyps. The analysis of these images is currently performed visually and the classification of polyps exhibits variability depending on the experience and awareness of the experts [4]. We investigate automated analysis of polyp regions to assist pathologists in colorectal cancer diagnosis.

To model the underlying patterns of image regions, accurate annotations are desirable. However the volumetric images of polyps are large (1024^3 voxels); while high resolution brings us considerable detail, difficulty arises in obtaining annotations. In our dataset, a polyp typically extends across $700 \sim 800$ slices and about 0.5 billion voxels in total. Fully delineating 3D regions slice by slice is tedious and time-consuming.

In this paper we investigate an alternative approach based on partial, sparse, incomplete annotations. We propose a learning framework for partially annotated OPT images, for the task of classifying dysplastic changes in colorectal polyps. More specifically, the objective in this paper is to discriminate between image patches that contain low-grade dysplasia (LGD) and image patches that contain invasive cancer. This is a first step towards the goal of automated polyp analysis.

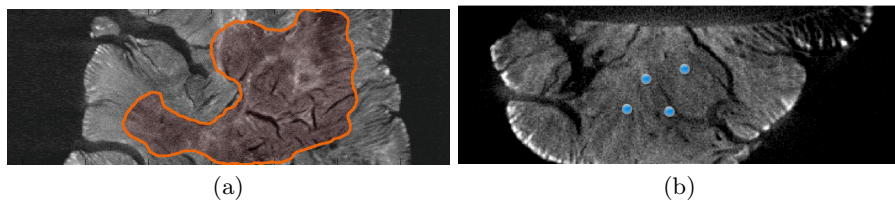


Fig. 1. OPT colorectal polyp images with (a) region annotation and (b) partial annotations

Different forms of partial annotation can be appropriate for different image modalities and applications. In this paper, we consider partial annotations consisting of just one click or a few clicks in the 3D polyp region of interest (as shown in Fig. 1(b)) as an alternative to the stronger annotation shown in Fig. 1(a). The annotation effort required is quite different. Our goal is to reduce the annotation efforts while achieving good classification performance. In addition, learning should scale well making it suitable for high-resolution volumetric images.

In [4] local features for patch and region classification of OPT images were compared. Here we focus on the model learning aspect of the task. Our method falls into the broad category of weakly supervised classification. At one extreme of this category, annotation is performed only at the image level in which case Multiple Instance Learning (MIL) has been adopted. In MIL, a sample is classified as positive if at least one of the instances is classified as positive. Dundar et al. [2] proposed a large margin based approach for pathology slides. It shared some similarity to this work however the prediction was at image level. In [8] MIL was adapted to classify and segment histopathology images. Doyle et al. [1] applied active learning to detect cancer regions with histopathology annotations. Our approach is to leverage spatial annotation but to keep annotation simple, sparse and thus fast to perform.

2 Methods

In supervised classification settings, locations outside annotated regions are usually ignored during training because the corresponding class labels are considered unknown. However, for images annotated with a partial annotation protocol, the annotations carry information about the class membership at unannotated locations. We refer to 3D windows as *patches*. Patches in the training set at locations with annotated (known) class labels are referred to as *reference patches*. Patches near to them (in terms of displacement or distance in feature space) are referred to as *candidate patches*. In this paper, we consider an extreme form of partial annotation consisting of point locations obtained via mouse clicks. We introduce our definition of contextual relevance, based on which we then propose a ranking model for classification.

2.1 Labeling Patches' Confidence

First we assign confidence labels to candidate patches. Consider a reference patch \mathbf{S}_r sampled at an annotated location \mathbf{z}_r , labeled as $y_r \in \{1, -1\}$. The patch \mathbf{S}_k sampled at location \mathbf{z}_k will have a lower confidence label y_k . y_k can be set to:

$$y_k = a(\mathbf{S}_k, \mathbf{S}_r) y_r, \quad (1)$$

where $a(\cdot, \cdot) \in (0, 1]$ is a measurement of affinity between two image patches. The absolute value of y_k can be viewed as a confidence measurement.

As patches sampled at locations near to each other usually belong to the same class, the reference patch of \mathbf{S}_k can be set as the nearest annotated patch \mathbf{S}_r . Affinity $a(\cdot, \cdot)$ is set as a Gaussian function with regard to spatial displacement of \mathbf{S}_k and \mathbf{S}_r in the image and a scaling parameter σ , i.e.:

$$a(\mathbf{S}_k, \mathbf{S}_r) = \exp\left(-\frac{\|\mathbf{z}_k - \mathbf{z}_r\|^2}{\sigma^2}\right). \quad (2)$$

Another way to define $a(\cdot, \cdot)$ is to consider similarity in image feature space. Assuming feature \mathbf{x}_r is extracted from reference patch \mathbf{S}_r and \mathbf{x}_k from \mathbf{S}_k , we can alternatively define the function as:

$$a(\mathbf{S}_k, \mathbf{S}_r) = \exp\left(-\frac{\|\mathbf{x}_k - \mathbf{x}_r\|^2}{\delta^2}\right). \quad (3)$$

Note that $a(\cdot, \cdot)$ can be extended to use multiple reference patches. Here we assign only one (the nearest) reference patch for each candidate patch.

2.2 Contextual Relevance Ranking Model

Let $\mathbf{x}_i \in \mathbb{R}^d, i = 1, \dots, N$ denote a feature vector extracted from an image patch indexed by i . We assign the label $y_i \in \{1, -1\}$ if the i th feature vector is from a reference patch; otherwise we set y_i according to formula (1). We form the ranking model by optimizing a regularized margin-based problem:

$$\min_{\mathbf{w}, b} \quad \frac{\lambda}{2} \|\mathbf{w}\|^2, \quad (4)$$

$$\text{s.t.} \quad \frac{1}{y_i} (\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall \mathbf{x}_i \in \mathbf{X}, \quad (5)$$

$$\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j \geq R_{ij}, \quad \forall \mathbf{x}_i \in \mathbf{X}^+, \mathbf{x}_j \in \mathbf{X}^-, \quad (6)$$

where $\mathbf{X}^+ = \{\mathbf{x}_k : 0 < y_k < 1\}$ and $\mathbf{X}^- = \{\mathbf{x}_k : -1 < y_k < 0\}$. R_{ij} is the pairwise contextual relevance of two patches \mathbf{S}_i and \mathbf{S}_j :

$$R_{ij} = \frac{a(\mathbf{S}_i, \mathbf{S}_{ir}) a(\mathbf{S}_j, \mathbf{S}_{jr})}{a(\mathbf{S}_i, \mathbf{S}_{ir}) + a(\mathbf{S}_j, \mathbf{S}_{jr})}, \quad (7)$$

where the patches \mathbf{S}_{ir} and \mathbf{S}_{jr} are the reference patches of \mathbf{S}_i and \mathbf{S}_j respectively.

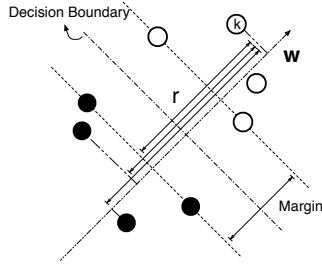


Fig. 2. Geometric interpretation of contextual relevance ranking model. \mathbf{w} is the weight vector; point k is a feature vector extracted from a reference patch. r refers to the differences of ranking score between data point k and points in the other class (projected along the direction of \mathbf{w}). Constraints in formula (5) were designed for minimizing classification error; constraints in formula (6) were designed for optimizing ranking difference r .

Constraints (5) are for all feature vectors in the training set. Note that in (5) features from candidate patches y_k are loosely constrained compared to their reference patches $y_r \in \{-1, 1\}$ because $|y_k| \in (0, 1)$. Constraints (6) are rankings of a pair of patches from two images with regard to their contextual relevance. We argue that patches sampled nearer to annotated locations (in terms of image location or location in feature space) should be classified with a larger score, i.e., further away from decision boundaries. The constraints keep the projected distance between any data point with high magnitude in y and the data in the opposite class large. In the case that pair $(\mathbf{x}_i, \mathbf{x}_j)$ is labeled with certainty, i.e., $y_i = \pm 1$ and $y_j = \pm 1$, the pairwise constraint (6) vanishes due to constraint (5). Fig. 2 illustrates the geometric interpretation of this model.

Given a fixed training set, the optimization problem can be transformed into dual form of \mathbf{w} by constructing a new feature set with $\frac{(\mathbf{x}_i - \mathbf{x}_j)}{R_{ij}}$. Then this can be solved by any SVM dual form solver, e.g. LIBSVM, SVM^{light}. However, this method is very slow and constructing feature set $\frac{(\mathbf{x}_i - \mathbf{x}_j)}{R_{ij}}$ across all the pairwise constraints is infeasible for our problem because the set of candidate patches is large. Here we tackle the primal form directly with a recently proposed efficient stochastic gradient method, SAG [6]. This method enables us to learn features online and with minimal storage cost.

To solve the optimization problem we minimize function (4) while controlling constraint violations in (5) and (6). Combining them together the risk function $J(\mathbf{w}, b)$ on training features $\{\mathbf{x}_i\}_{i=1}^N$ and labels $\{y_i\}_{i=1}^N$ can be written as:

$$J(\mathbf{w}, b) = \frac{\lambda}{2} \|\mathbf{w}\|^2 + \frac{1}{N^+ N^-} \sum_{\mathbf{x}_i \in \mathbf{X}^+} \sum_{\mathbf{x}_j \in \mathbf{X}^-} (f_i + f_j + C g_{ij}); \quad (8)$$

$$\text{where: } f_i = f\left(\frac{1}{y_i}(\mathbf{w}^T \mathbf{x}_i + b)\right), \quad f_j = f\left(\frac{1}{y_j}(\mathbf{w}^T \mathbf{x}_j + b)\right), \quad (9)$$

$$g_{ij} = g(\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j - R_{ij}). \quad (10)$$

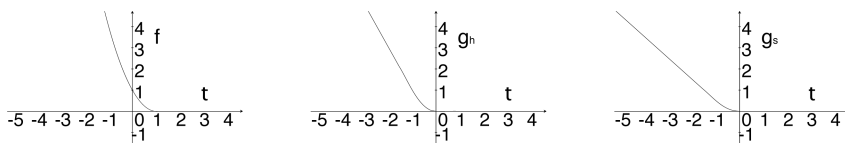


Fig. 3. Demonstration of loss functions (left to right): (1) squared hinge loss $f(t)$, (2) Huber loss $g_h(t)$, (3) smoothed hinge loss $g_s(t)$

The loss term $f(\cdot)$ corresponding to constraints (5) is squared hinge loss:

$$f(t) = \max(0, 1 - t)^2; \tag{11}$$

the loss term $g(\cdot)$ corresponding to constraints (6) can be a Huber loss function or a smoothed hinge loss function:

$$g(t) = \begin{cases} 0 & \text{if } t \geq 0 \\ -2t - 1 & \text{if } t < -1 \\ t^2 & \text{otherwise} \end{cases}, \tag{12} \quad g(t) = \begin{cases} 0 & \text{if } t \geq 0 \\ -t - \frac{1}{2} & \text{if } t < -1 \\ \frac{1}{2}t^2 & \text{otherwise} \end{cases}. \tag{13}$$

In risk function $J(\mathbf{w}, b)$, parameter λ is the regularization strength; $C \geq 0$ controls the trade-off between classification errors and ranking errors; N^+ and N^- are the number of positive and negative samples respectively. Fig. 3 illustrates the loss function used in $J(\mathbf{w}, b)$. Squared hinge loss is much more sensitive to outliers and large errors than the smoothed hinge loss and Huber loss. It is applied to patch classification to ensure the risk function is sensitive to every training label. The latter two functions are choices for the pairwise ranking errors. OPT images of colorectal polyps usually involve large intra-class variations so we expect the pairwise outliers would not dominate the risk function. There are other loss functions that meet our requirements [5]. We chose these convex and smooth functions as they can be efficiently integrated into SAG.

To apply SAG methods for minimizing $J(\mathbf{w}, b)$ iteratively, at each iteration \mathbf{w} is updated with an average of the gradient of a randomly selected training pair $(\mathbf{x}_i, \mathbf{x}_j)$ and most recently computed gradients of the other training pairs. At the $(k+1)$ th iteration the updating rule with a small step size α_k has the form:

$$\mathbf{w}^{k+1} = (1 - \alpha_k \lambda) \mathbf{w}^k - \frac{\alpha_k}{N^+ N^-} \sum_{\mathbf{x}_i \in \mathbf{X}^+} \sum_{\mathbf{x}_j \in \mathbf{X}^-} grad_{ij}^k, \tag{14}$$

where for the training pair (i^k, j^k) , we set:

$$grad_{ij}^k = \begin{cases} f'_i + f'_j + Cg'_{ij} & \text{if } (i, j) = (i^k, j^k) \\ grad_{ij}^{k-1} & \text{otherwise} \end{cases}. \tag{15}$$

For the bias term b , we simply extend each feature vector with one bias component (from \mathbf{x} to $[\mathbf{x}; b]$) in each iteration. This method has an exponential convergence rate and with a few implementation tricks (described in [6]) we reduce the storage cost to $\mathcal{O}(N^+ N^-)$. This allows the method to scale to large datasets.

3 Experiments

Data. OPT images from 59 patients acquired using ultraviolet light and Cy3 dye were used in this study. Each image was of one colorectal polyp specimen and had $1024 \times 1024 \times 1024$ voxels with aspect ratio 1 : 1 : 1. For each volumetric image, 3D regions were annotated by a trained pathologist with labels of dysplastic change. In 30 images, regions judged to consist entirely of low-grade dysplasia (LGD) were annotated. In the other 29 images, regions judged to consist entirely of invasive cancer were annotated. During the manual annotation process, the pathologist was asked to roughly indicate significant regions instead of exhaustively delineating all the regions.

Experimental Setup. We evaluated two aspects of the proposed model in terms of patch classification performance: (1) the ability to utilise unlabelled patches, compared with not using unlabelled patches, and using unlabelled patches naively (standard SVM); (2) the choice of loss function and affinity measurement.

In the experiments we applied 10-fold cross-validation. The dataset was randomly split into 10 folds (about 3 cancer and 3 LGD images per fold); 10 iterations of training and testing were performed such that within each iteration one fold was used as test set. The performance was averaged over the 10 iterations.

Test patches were randomly sampled from annotated regions in the test sets (about 1400 patches per fold). The partial annotation process was simulated by randomly sampling point locations within the pathologist-annotated regions. Candidate patches were randomly sampled outside the annotated regions in the training set. With reference and candidate patches, three types of models were trained:

- **T1:** training with only reference patches, using standard SVM. (*SVM-ref*)
- **T2:** training with both reference and candidate patches, using standard SVM. Labels of candidate patches can be assigned with either feature-based or location-based affinity (formula (2) or (3)). (*SVM-fea* and *SVM-loc*)
- **T3:** using our proposed model with both reference and candidate patches. We evaluated four combinations of different loss functions (formula (12) and (13)) and affinities (formula (2) and (3)). (*Prop-hub-fea*, *Prop-hub-loc*, *Prop-squ-fea*, and *Prop-squ-loc*)

We started with a training set with only 1 reference patch and 20 candidate patches for each training image (training set size: 1, 140). The models were learned using T1, T2 and T3 methods respectively and the classification performance was evaluated on the test patches. Then we added more reference patches and their associated candidate patches. At each iteration 1 reference and about 6 candidate patches per image were added. Such iterations were repeated 30 times till there were 1,590 reference patches in the training set. At the final iteration the number of training patches was about 12,000.

The size of each image patch was set to $21 \times 21 \times 21$ voxels. For feature extraction we used Bag of Words with Random Projection since this achieved

highest classification accuracies in [4]. The dimensionality of each feature vector was 200. Each feature was normalized to zero mean and unit variance.

In all standard SVM evaluations we used the LIBLINEAR [3] solver that solves the L_2 regularized squared loss primal problem (with regularization parameter searched from 10^{-7} to 10^7 and $eps = 0.01$). In our proposed method, C searched from 10^{-10} to 10^{-5} , $\lambda = \frac{1}{N+N^-}$, $b = 0$, and the stochastic gradient step size was set to 0.004. The scaling factors were estimated from standard deviation of all distances ($\|\mathbf{z}_k - \mathbf{z}_r\|$ or $\|\mathbf{x}_k - \mathbf{x}_r\|$) between reference patches and candidate patches in the training set ($\sigma = 158.1$ in formula (2), $\delta = 7071.1$ in formula (3)).

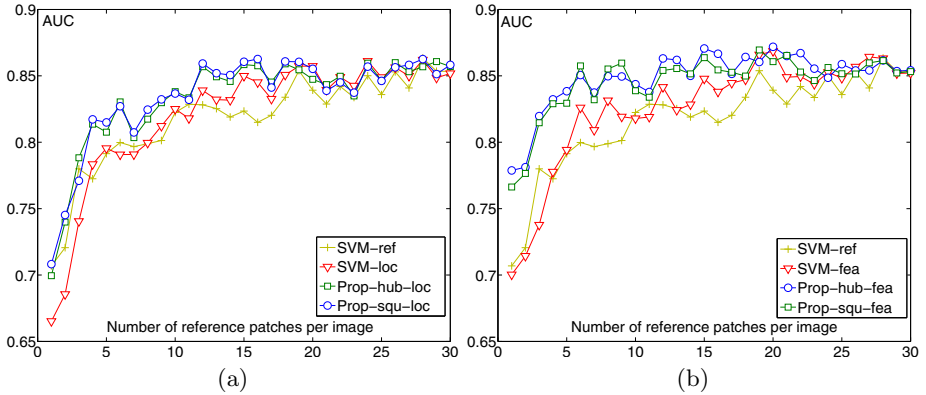


Fig. 4. AUC values depending on number of reference patches with (a) location-based affinity measurement, and (b) feature-based affinity measurement

Results. Area Under ROC Curve (AUC) obtained when classifying patches as LGD or invasive cancer was used as a performance measure. Fig. 4 shows AUC values depending on the number of reference patches per training image. We list the AUC values depending on number of reference patches per image in Table 1.

In Fig. 4, with more than 10 reference patches per image (530 reference patches, about 3,000 training patches in total) the classification performances of all the methods saturated. With both location-based and feature-based affinity the SVM-loc method showed the same or slightly higher AUCs than the SVM-ref method. This indicates that simply feeding uncertain patches to standard SVM does little to help patch classification performance. The information presented in uncertain patches was not utilized effectively by standard SVM. The proposed methods performed relatively well with small training sets indicating that they were making effective use of the unannotated patches. AUCs of all methods converged to similar values when number of reference patches reaches 30 per image.

For both affinity-based experiments, the proposed models with Huber loss and smoothed hinge loss showed almost the same AUCs. However our grid search of parameters showed that the best parameters C are quite different ($C = 10^{-4}$ for Huber loss and $C = 10^{-2}$ for smoothed hinge loss).

Table 1. Performance comparison between standard SVM and proposed model. AUC values(%) \pm standard errors depending on the number of reference patches per image.

affinity patches	Location-based				Feature-based			
	1	2	5	10	1	2	5	10
SVM-ref	71 \pm 2.9	72 \pm 1.4	79 \pm 2.3	82 \pm 2.1	71 \pm 2.9	72 \pm 1.4	79 \pm 2.3	82 \pm 2.1
SVM	67 \pm 2.2	69 \pm 3.0	80 \pm 2.2	83 \pm 2.3	70 \pm 2.8	71 \pm 2.5	79 \pm 2.8	82 \pm 1.7
Prop-hub	70 \pm 2.2	74 \pm 2.0	81 \pm 1.9	84 \pm 1.6	78 \pm 2.7	78 \pm 2.4	84 \pm 2.6	85 \pm 1.9
Prop-squ	71 \pm 2.3	75 \pm 1.8	81 \pm 1.8	84 \pm 1.9	77 \pm 3.0	78 \pm 2.7	83 \pm 2.7	84 \pm 1.9

4 Conclusions

We have proposed a learning model for partially annotated images. The experiment on a dataset of 59 OPT images showed that it is able to robustly learn from patches with uncertain labels, achieving high classification accuracies while reducing the annotation effort. At the same time our model can be efficiently evaluated with only $\mathcal{O}(N^+N^-)$ in storage cost. Therefore it is suitable for high-resolution, volumetric datasets.

Acknowledgement. This work is partially supported by the Dundee Cancer Centre (DCC) Development Fund and an RSE-NSFC Joint Project (RSE Reference: 443570/NNS/INT).

References

1. Doyle, S., Monaco, J., Feldman, M., Tomaszewski, J., Madabhushi, A.: An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC Bioinformatics* 12(1), 1–14 (2011)
2. Dundar, M., Badve, S., Raykar, V., Jain, R., Sertel, O., Gurcan, M.: A multiple instance learning approach toward optimal classification of pathology slides. In: *ICPR*, pp. 2732–2735 (2010)
3. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research* 9, 1871–1874 (2008)
4. Li, W., Zhang, J., McKenna, S.J., Coats, M., Carey, F.A.: Classification of colorectal polyp regions in optical projection tomography. In: *ISBI* (2013)
5. Rennie, J.D., Srebro, N.: Loss functions for preference levels: Regression with discrete ordered labels. In: *Proceedings of the IJCAI Multidisciplinary Workshop on Advances in Preference Handling*, pp. 180–186 (2005)
6. Roux, N.L., Schmidt, M., Bach, F.: A stochastic gradient method with an exponential convergence rate for finite training sets. In: *NIPS 25*, pp. 2672–2680 (2012)
7. Sharpe, J., Ahlgren, U., Perry, P., Hill, B., Ross, A., Hecksher-Sørensen, J., Baldock, R., Davidson, D.: Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science* 296(5567), 541–545 (2002)
8. Xu, Y., Zhu, J.Y., Chang, E., Tu, Z.: Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: *CVPR*, pp. 964–971 (2012)

Phenotype Detection in Morphological Mutant Mice Using Deformation Features

Sharmili Roy¹, Xi Liang², Asanobu Kitamoto³, Masaru Tamura⁴,
Toshihiko Shiroishi⁴, and Michael S. Brown¹

¹ School of Computing, National University of Singapore
sharmili@comp.nus.edu.sg

² National ICT Australia (NICTA), Australia

³ National Institute of Informatics, Japan

⁴ National Institute of Genetics, Japan

Abstract. Large-scale global efforts are underway to knockout each of the approximately 25,000 mouse genes and interpret their roles in shaping the mammalian embryo. Given the tremendous amount of data generated by imaging mutated prenatal mice, high-throughput image analysis systems are inevitable to characterize mammalian development and diseases. Current state-of-the-art computational systems offer only differential volumetric analysis of pre-defined anatomical structures between various gene-knockout mice strains. For subtle anatomical phenotypes, embryo phenotyping still relies on the laborious histological techniques that are clearly unsuitable in such big data environment. This paper presents a system that automatically detects known phenotypes and assists in discovering novel phenotypes in μ CT images of mutant mice. Deformation features obtained from non-linear registration of mutant embryo to a normal consensus average image are extracted and analyzed to compute phenotypic and candidate phenotypic areas. The presented system is evaluated using C57BL/10 embryo images. All cases of ventricular septum defect and polydactyly, well-known to be present in this strain, are successfully detected. The system predicts potential phenotypic areas in the liver that are under active histological evaluation for possible phenotype of this mouse line.

1 Introduction

Completion of the human genome project brought comprehension of location and sequence of each human gene. Owing to the 99% genetic homology between mouse and human, mouse has been chosen as the principal study model to annotate genetic sequence with its functional information [1], [2]. Gene targeting technology is being actively employed by many international organizations to generate mutant mouse lines by knocking out each of the approximately 25,000 mouse genes (i.e., systematically removing each gene one by one and growing the mouse). High-throughput phenotypic assessment systems are necessary to systematically analyze and interpret the genetic information generated by

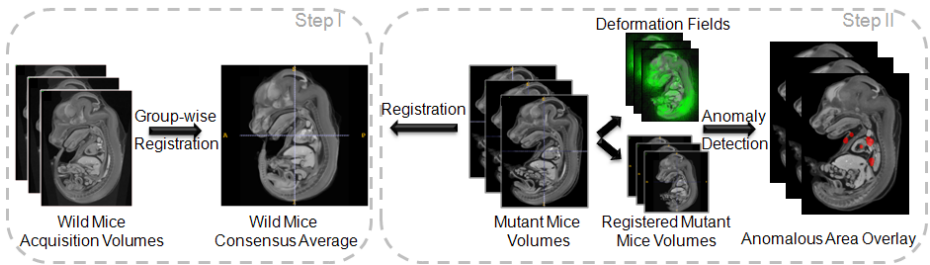


Fig. 1. Defect detection consists of two steps. A mean of the normal mouse group is computed in the first step. In the second step, mutant group is registered to the normal mean and the resulting deformations are analyzed to detect defects.

these large-scale mutagenesis programs. A significant proportion of the generated strains are embryonic lethal resulting in the shift towards prenatal phenotyping.

The research community focusses on semi-automatic analysis of anatomical volumetric variation in various mouse strains using representative average images [3], [4], [5]. Although evaluation based on average images may be beneficial for an initial examination, phenotypes that are randomized in position and texture such as the intestines and developing trabeculae of the heart [6] or subtle structural organ failures without large volume changes cannot be characterized using this technique. Another body of work focusses on detection of defects via model-based segmentation [7] or by better data visualization using tissue staining [8], [9]. Segmentation techniques fail if the defect characteristics are unknown or if the anatomy is hard to label such as bone joints. Enhanced visualization, while useful, still requires long expert hours to interpret the data.

In this paper, we present a generalized defect detection framework that automatically computes candidate phenotypic areas without using atlas, segmentation or any defect specific features. Instead, our approach uses deformation fields that are widely used to study anatomical variations [3], [4], [10], [11]. We extract various features from deformation fields obtained by registering mutant mice to a normal mean and combine them to detect coarse, subtle as well as randomized defects (Fig. 1). Statistical characteristics of deformation fields have been previously studied to detect gross defects in mice brain using multi-modality images [11]. Our approach, however, targets a single imaging modality and successfully handles both subtle as well as significantly differing anatomy.

2 Methods

2.1 Sample Preparation and Imaging Protocol

This study is performed using C57BL/10 mutant mice generated at the National Institute of Genetics, Japan. Post organogenesis, growth and development in the embryo starts at ~ 14.0 days post-coitum (dpc). Image registration cannot be applied at stages earlier than this due to unformed or absent organs. Further, at

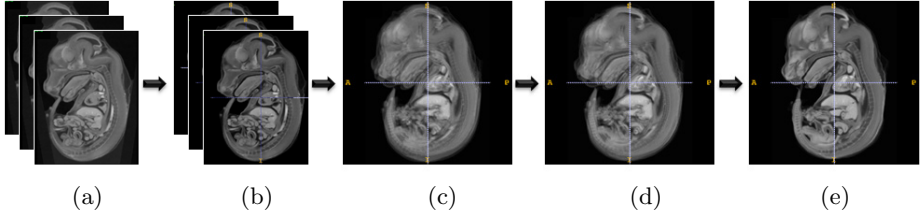


Fig. 2. This figure illustrates the steps in the computation of normal mean image. (a) Acquisition volume (b) extracted normalized embryo images (c)-(e) consensus average images at rigid, affine and B-Spline registration stages respectively.

a relatively mature stage such as 15.5 dpc accurate registration of abdomen is difficult to achieve due to variation in intestinal position and crowding within the abdominal cavity [4]. Therefore, embryo samples at 14.5 dpc were collected. A total of 14 embryos were used out of which 3 were normal and 11 had chromosomal aberrations. 11 mutant embryos consisted of 3 homozygotes generated by inbreeding C57BL/10 mice and 8 heterozygotes obtained by cross breeding C57BL/10 and normal littermates. The samples were washed in phosphate buffered saline and fixed in 4% paraformaldehyde until imaging. Before scanning, embryos were soaked in 1 : 3 mixture of lugol solution and double distilled water. Scan was carried out in Scanxmate-E090S 3D μ CT system (Comscantecno, Japan) with the embryos fitted in 1.5 milliliters eppendorf tube fixed using wet paper. Keeping the X-ray source at 60kVp and 130mA, each specimen was rotated 360° in steps of 0.36° generating 1000 projections of 640×480 pixels. The 3D μ CT data was reconstructed at an isotropic resolution of $9.5 \times 9.5 \times 9.5 \mu\text{m}^3$.

2.2 Normal Mouse Consensus Average Image

Embryo pixels are extracted from the acquisition volumes using Gaussian mixture modeling, thresholding and mathematical morphology (Fig. 2(a) and (b)). Pixel intensity ranges are normalized and a standard group-wise registration routine consisting of rigid, affine and B-Spline registration stages is initiated. Rigid registration corrects differing orientations of individual embryos by choosing a reference and spatially aligning the rest to it. Averaging the rigid registered images results in a blurry reference (Fig. 2(c)) that is not biased towards the geometry of the initial reference because this step does not affect the geometry of the subjects [5]. Embryos are then registered to the blurry reference via affine transformation and the reference is updated (Fig. 2(d)). As a final step B-Spline based non-linear registration is applied to locally align the affine registered embryos to the reference. The non-linear registration is formulated with a similarity energy function comprising of mutual information [12] and a rigidity penalty [13]. 10 iterations of this registration are applied in a multi-resolution fashion where the control point spacing gradually reduces to 8 voxels.

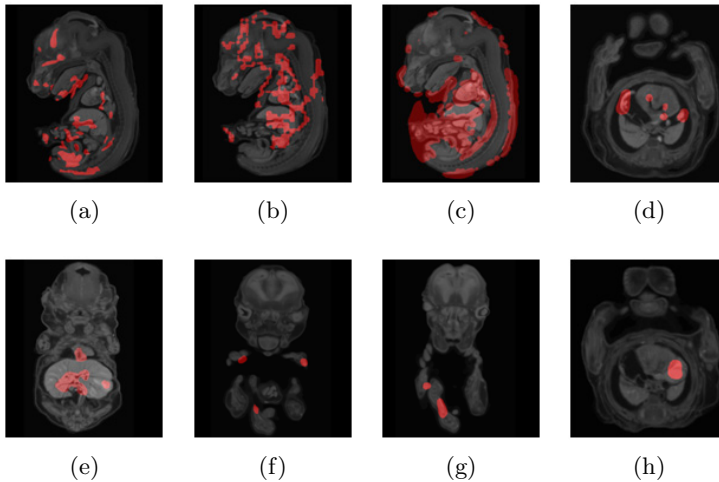


Fig. 3. (a),(b),(c) Jacobian, stress and intensity variance masks overlaid on mutant image respectively (d),(e) defective areas identified by $(I_{IV} \cap I_J)$ (f),(g) areas generated by $(I_{IV} \cap I_S)$ (h) defective areas captured by $(I_J \cap I_S)$

The reference image is updated after each iteration leading to the final consensus average (Fig. 2(e)). Elastix toolbox is used to implement this registration scheme [14].

2.3 Deformation Features and Masks for Defect Detection

To detect defects in mutant mice, they are registered to the normal average image using the same three-stage registration pipeline as above except that in each stage the reference is always kept fixed to the normal average. Registration of each mouse results in the corresponding deformation field. These deformation fields are used to compute 3D Jacobian maps using determinant of local Jacobian matrix at each voxel [10]. Jacobian determinant greater than one represents voxel expansion and less than one represents voxel compression. Jacobian of deformation is a popular tool to study inter-group structural differences [10], [11]. We apply Jacobian determinant in phenotyping by computing a Jacobian mask, I_J , one for each mutant mouse, that selects voxels at which Jacobian determinant is δ units away from one (Fig. 3(a)). δ is kept 0.5 for experiments. We realized, however, that Jacobian determinant fails for defects where volume changes are minimal. Further, we find that Jacobian introduces numerous false positives by highlighting areas that are found to be non-defective by the phenotyping experts, thus resulting in low precision. Performance of I_J in detecting known defects namely, Ventricular Septum Defect (VSD) and polydactyly is summarized in Table 1. Detection specificity for both the defects is very low with I_J .

To capture subtle defects with low volumetric changes, we compute another deformation feature that we call deformation stress. Deformation stress (D_s) is

Table 1. This table compares VSD and polydactyly detection accuracy (in %) of various features. VSD is assumed detected if the ventricular area is highlighted.

	I_J	I_S	$(I_J \cap I_S)$ F_1	$(I_{IV} \cap I_J)$ F_2	$(I_{IV} \cap I_S)$ F_3	$(F_2 \cup F_3)$ F_4	$(F_1 \cup F_4)$ F_5
VSD Sensitivity	88.8	100.0	88.8	77.7	77.7	77.7	100.0
VSD Specificity	50.0	100.0	100.0	100.0	100.0	100.0	100.0
Polydactyly Sensitivity	76.9	84.6	61.5	46.1	76.9	76.9	92.3
Polydactyly Specificity	48.4	80.6	87.1	90.3	87.1	87.1	87.1

computed by dividing the volume into small blocks and measuring the entropy of deformation direction inside each block.

$$D_S(v) = - \sum_{u \in B(v)} p(\theta(u)) \log(p(\theta(u))); \quad (1)$$

$B(v)$ in Equation 1 represents the block in which voxel v lies and $\theta(u)$ is the displacement direction at voxel u . For experiments the volume was divided into cuboids of size 8 voxels. Using D_S we compute a mask, I_S , that chooses voxel blocks that have high entropy of deformation direction and hence are undergoing incoherent deformation. For experiments I_S selected the top 50% blocks that exhibited high deformation entropy. Fig. 3(b) shows an example of I_S and Table 1 enlists its performance in detecting known phenotypes. Some false positives are introduced due to inclusion of sources and sinks in the deformation field.

Since I_J and I_S individually fail to detect all defects and both introduce false positives, a simple combination of the two does not give satisfactory results (Table 1). In practice multiple mice from a mouse line are imaged before phenomic analysis is performed. We introduce this group information in defect detection by calculating voxel-wise intensity variance (V_{IV}) across the group of mutant mice that are registered to the normal mean.

$$V_{IV}(v) = \frac{1}{N_M - 1} \sum_{i=1}^{N_M} (M_i(v) - N_{Avg}(v))^2. \quad (2)$$

N_M , M_i and N_{Avg} in Equation 2 are mutant mouse population size, i^{th} registered mutant mouse image and the normal consensus average respectively. From V_{IV} we compute a mask I_{IV} by selecting the top 50% voxels that have high intensity variance and hence low registration accuracy (Fig. 3(c)).

We find that many false positives introduced by I_J and I_S are pruned when these features are combined with I_{IV} . Table 1 lists the accuracies when the detection criterion is $(I_{IV} \cap I_J)$, $(I_{IV} \cap I_S)$ or both combined. From experiments on C57BL/10 mutant mice we find that $(I_{IV} \cap I_J)$ mainly captures VSD and unusually wide liver lobe junctions. Figs. 3(d) and 3(e) show some detection results obtained by this factor. $(I_{IV} \cap I_S)$ captures polydactyly and unnatural position and deformations of tail and limbs (Figs. 3(f) and 3(g)).

It is possible to have inaccurately registered morphological structures that are not selected by the intensity variance mask. Uniform body cavities (dark

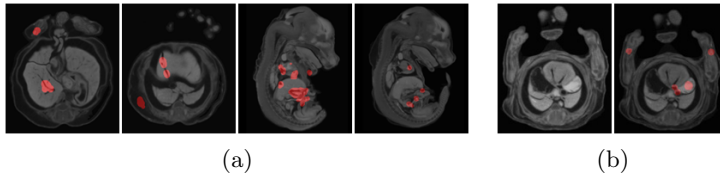


Fig. 4. (a) Defect detection results in the liver lobe junctions, heart and intestine of C57BL/10 mice (b) the left and right images depict a healthy heart and the misjudged defect respectively

regions) or muscular organs (like liver lobes and heart atria) are some examples where we find that intensity variance is low due to spatially uniform intensity values even though there is a abnormality. $(I_J \cap I_S)$ addresses regions where both Jacobian and stress are high irrespective of the intensity variance. By adding this term in the detection rule we are able to detect spatially uniform defective regions. Some secondary phenotypes like enlarged heart atrium due to high blood pressure induced by VSD are captured by this term as shown in Fig. 3(h).

Combining the three terms we propose the defect detection rule as

$$I_{Defect} = (I_{IV} \cap I_J) \cup (I_{IV} \cap I_S) \cup (I_J \cap I_S). \quad (3)$$

Table 1 enlists the performance of this detection rule. Number of detected regions can be readily increased or decreased by relaxing or tightening the thresholds while generating the Jacobian, deformation stress and intensity variance masks. Simple morphological operations like dilation and erosion are applied as noise reduction measures to clean up the detection results.

3 Results

Complete phenomic analysis of a mouse strain is a very tedious and slow process. C57BL/10 strain is still under investigation and hence full phenotypic characteristic consisting of all phenotypic defects is yet unknown. Therefore, even though we can evaluate the detection rule (Equation 3) in terms of precision, a formal evaluation of recall is not possible. VSD and polydactyly are two established genetic defects in C57BL/10 mice. We compute sensitivity and specificity of the detection algorithm with respect to these two defects.

When evaluated over the mutant database of 3 homozygote and 8 heterozygote embryos, the algorithm detected all cases of VSD without generating any false positives. Out of the 13 cases of polydactyly 12 were successfully detected and 1 was missed. The missed case belonged to the only mouse in the database that had its umbilical chord removed resulting in registration errors at the nearby areas. 4 false positive polydactyl cases were reported in situations where due to high proximity toes of both the feet seemed fused in the 3D renderings.

To evaluate the rest of the detected areas, a user study was conducted with phenotyping experts having long experience in mouse imaging and phenotyping.

The expert comments were very encouraging and they noted that all the regions detected by the algorithm had biological significance. Some of the areas were due to genetic defects, some due to non-genetic defects and some due to organogenesis or procedural interventions. After careful histological examination, it was found that majority (57%) of the total detected regions belonged to genetic defects. Apart from known phenotypes, the algorithm detected areas in the liver lobe junctions that are candidate areas for potential phenotype of this mouse strain and are under active phenotypic evaluation.

14% output regions were detected due to malformed body cavities. Though these regions do not represent defects due to genetic makeup, they still signify biological malformations. Another 8% regions were noted to be due to genesis and extensive developmental remodeling of gonads at this gestational stage. The rest of the output was attributed to blood clots, randomized umbilical chord regions and pancreatic genesis.

When the detection rule was applied to wild-type mice, some areas were reported. These areas represent blood clots, umbilical chord, malformed body cavities and pancreatic and gonadic organogenesis. One false positive was generated for heart septal defect in a case where low spatial intensity variance makes the judgement hard even for an unexperienced human eye (Fig. 4(b)). With further image processing it is possible to improve the detection accuracy by neglecting the high intensity blood clots and masking out umbilical chord regions.

4 Discussion and Conclusion

We have presented a generic deformation based defect detection framework for 3D μ CT images of mutant mice. Our system has the potential to greatly enhance phenotyping throughput by automatically detecting all known phenotypes. Unlike other algorithms designed to detect specific known defects, our system also highlights candidate novel defects that may not be readily recognized by human experts due to absence of significant visual features. Owing to voxel-by-voxel analysis, defects are localized to sub-structures and those affecting multiple structures are visualized collectively. Though our evaluation database is small, the results clearly establish the potential of the proposed system in patterning defects. Our framework can be easily adapted to examine other 3D scan images amenable to registration. We acknowledge that the registration method may affect the detection results, however, the registration scheme used in the paper is widely employed in mice phenotyping [3], [4], [5]. Deformation field resulting from only the non-linear registration step is used for defect detection. The detection performance is found to be fairly robust to parameter variation.

Since the proposed framework is independent of the defect features, classification of defects into those that are genetically induced and those that are not is out of scope for the current system. Currently we provide frequency of occurrence as an indicator of whether or not a defect is genetic. As an example, since VSD and polydactyly are detected in all homozygote embryos, the probability of these defects being genetic is reported to be 100%. Similar probabilities

are assigned to all detected regions. In future we plan to use advanced image processing and statistical techniques to device classifiers that can differentiate between genetic and non-genetic defects.

References

1. Collins, F.S., Rossant, J., Wurst, W.: A mouse for all reasons. *Cell* 128, 9–13 (2007)
2. Mouse Genome Sequencing Consortium: Initial Sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562 (2002)
3. Zamyadi, M., Baghdadi, L., Lerch, J.P., Bhattacharya, S., Schneider, J.E., Henkelman, R.M.: Mouse embryonic phenotyping by morphometric analysis of MR images. *Physiol. Genomics* 42A, 89–95 (2010)
4. Wong, M.D., Dorr, A.E., Walls, J.R., Lerch, J.P., Henkelman, R.M.: A novel 3D mouse embryo atlas based on micro-CT. *Development* 139(17), 3248–3256 (2012)
5. Cleary, J.O., Modat, M., Norris, F.C., Price, A.N., Jayakody, S.A., Martinez-Barbera, J.P., Greene, N.D.E., Hawkes, D.J., Ordidge, R.J., Scambler, P.J., Ourselin, S., Lythgoe, M.F.: Magnetic resonance virtual histology for embryos: 3D atlases for automated high-throughput phenotyping. *Neuroimage* 54(2), 769–778 (2011)
6. Nieman, B.J., Wong, M.D., Henkelman, R.M.: Genes into geometry: imaging for mouse development in 3D. *Curr. Opin. Genet. Dev.* 21(5), 638–646 (2011)
7. Norris, F.C., Modat, M., Cleary, J.O., Price, A.N., McCue, K., Scambler, P.J., Ourselin, S., Lythgoe, M.F.: Segmentation propagation using a 3D embryo atlas for high-throughput MRI phenotyping: comparison and validation with manual segmentation. *Magn. Reson. Med.* 69(3), 877–883 (2013)
8. Degenhardt, K., Wright, A.C., Horng, D., Padmanabhan, A., Epstein, J.A.: Rapid 3D phenotyping of cardiovascular development in mouse embryos by micro-CT with iodine staining. *Circ. Cardiovascular Imaging* 3(3), 314–322 (2010)
9. Cleary, J.O., Price, A.N., Thomas, D.L., Scambler, P.J., Kyriakopoulou, V., McCue, K., Schneider, J.E., Ordidge, R.J., Lythgoe, M.F.: Cardiac phenotyping in ex vivo murine embryos using μ MRI. *NMR Biomed.* 22(8), 857–866 (2009)
10. Xie, Z., Yang, D., Stephenson, D., Morton, D., Hicks, C., Brown, T., Bocan, T.: Characterizing the regional structural difference of the brain between tau transgenic (rTg4510) and Wild-Type Mice using MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 308–315. Springer, Heidelberg (2010)
11. Nieman, B.J., Flenniken, A.M., Adamson, S.L., Henkelman, R.M., Sled, J.G.: Anatomical phenotyping in the brain and skull of a mutant mouse by magnetic resonance imaging and computed tomography. *Physiol. Genomics* 24(2), 154–162 (2006)
12. Mattes, D., Haynor, D.R., Vesselle, H., Lewellyn, T.K., Eubank, W.: Nonrigid multimodality image registration. In: *Proc. SPIE*, vol. 4322, pp. 1609–1619 (2001)
13. Staring, M., Klein, S., Pluim, J.P.W.: A rigidity penalty term for nonrigid registration. *Med. Phys.* 34(11), 4098–4108 (2007)
14. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* 29(1), 196–205 (2010)

Efficient Reconstruction-Based Optic Cup Localization for Glaucoma Screening

Yanwu Xu¹, Stephen Lin², Damon Wing Kee Wong¹, Jiang Liu¹, and Dong Xu³

¹ Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore

² Microsoft Research Asia, P.R. China

³ School of Computer Engineering, Nanyang Technological University, Singapore

Abstract. We present a reconstruction-based learning technique to localize the optic cup in fundus images for glaucoma screening. In contrast to previous approaches which rely on low-level visual cues, our method instead considers the input image as a whole and infers its optic cup parameters from a codebook of manually labeled reference images based on their similarity to the input and their contribution towards reconstructing the input image. We show that this approach can be formulated as a closed-form solution without any search, which leads to highly efficient and 100% repeatable computation. Our tests on the *ORIGA* and *SCES* datasets show that the performance of this method compares favorably to those of previous techniques while operating at faster speeds. This suggests much promise for this approach to be used in practice for screening.

1 Introduction

Glaucoma is a disease of the optic nerve and a leading cause of blindness worldwide. Due to the lack of visual symptoms or discomfort in the early stages, studies have shown that more than 90% [1] of those with glaucoma are unaware that they have the disease. Since visual loss in glaucoma is irreversible and permanent, there is a strong need to detect this “silent thief of sight” as early as possible for immediate intervention.

During glaucoma progression, the death of ganglion nerve cells often leads to changes in the appearance of the optic nerve head, also known as the optic disc. In particular, there is a deepening and expansion of the excavation in the optic disc, known as the optic cup, as glaucoma advances, leading to an increase in the cup-to-disc ratio (CDR) as illustrated in Fig. 1. The CDR is a major consideration of clinicians in assessing glaucoma from direct observation [2]. However, manual assessment is labor-intensive, highly reliant on the training and expertise of the examiner, and not cost-efficient for screening.

These practical issues have motivated the development of computer-aided techniques for glaucoma screening using images from digital fundus cameras, which have become widely available in healthcare settings. Various automated methods based on CDR estimation have been presented, with the optic cup localized either by classifying pixels as part of the cup or rim (the disc area outside the cup) [3][4], or through an analysis of sub-regions (*e.g.*, defined by sliding windows [5] or superpixels [6]). These methods typically perform their classification or analysis based on explicit low-level visual cues such as intensity [4][5] or vessel kinks [7], but the reliance on low-level cues can make a method less robust to imaging noise and low contrast edges.

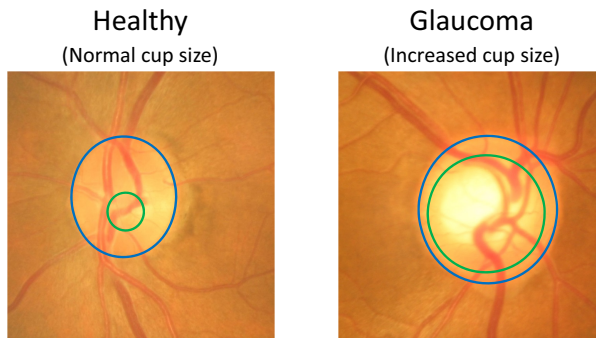


Fig. 1. Effect of glaucoma on cup-to-disc ratio (CDR): optic disc (blue) and cup (green)

In contrast to these methods, our technique identifies the optic cup within an input image in a holistic, reconstruction-based manner. From a codebook of reference images with expert-labeled cup regions, our method computes an optimal linear reconstruction of the input image from the reference images in a manner that gives greater emphasis to references more similar to the input. The linear reconstruction weights are then applied to the cup parameters of the corresponding reference images to estimate the cup region in the input image. Through this use of manually-labeled cup localization examples from reference images, our reconstruction-based technique can directly take advantage of clinician knowledge without having to rely on low-level feature processing.

This holistic approach to optic cup localization leads to higher accuracy than the current state-of-the-art optic cup detection method [6]. Furthermore, our formulation of this method has a one-step closed-form solution which results in significant computational savings. Our approach is inspired by medical image segmentation techniques that employ linear reconstruction [8] but require additional processing steps (*e.g.*, nearest neighbor search) and are less effective for sparse sets of reference images. The accuracy and efficiency of our reconstruction-based technique holds much promise for the development of practical automated/assisted glaucoma diagnosis systems.

2 Methodology

In our proposed localization framework, we adopt a left-eye frame of reference, with fundus images from right eyes horizontally flipped to be aligned with this reference frame. The optic disc is first segmented by using existing methods such as Active Shape Model (ASM) [9]. After normalization of the segmented disc, we localize the optic cup with our proposed reconstruction-based method. The risk of glaucoma is finally assessed using the calculated CDR value.

2.1 Disc Segmentation and Normalization

To segment the optic disc from a retinal fundus image, we employ either ASM [9] trained on codebook images or Template Matching [10]. The purpose of using two

different disc segmentation methods is so that we can compare the effect of different disc segmentation results on our optic cup localization. Following [9] and [10], we use only information from the green channel to localize the optic cup. Each segmented disc is normalized to a standardized circle with a radius of 128 pixels, and then each pixel value is linearly scaled to $[0, 1]$. Finally, the mean value over all the pixels in the optic disc is subtracted from each pixel to remove the influence of illumination variation among images.

2.2 Codebook Generation

In this work, our codebook is generated by random sampling from the manually-labeled images. To prevent bias, these samples are selected with an even distribution over the range of CDR values, which is divided for our dataset into three intervals: large-cup ($\text{CDR} > 0.6$, 72 images), mid-cup ($0.4 < \text{CDR} \leq 0.6$, 418 images), and small-cup ($\text{CDR} \leq 0.4$, 160 images). One-third of the codebook is thus generated by taking random samples from each of the three CDR intervals.

2.3 Closed-Form Reconstruction

From the codebook, we have n reference (ground truth) disc images $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{k \times n}$ and their corresponding ground truth cups $C = \{c_1, c_2, \dots, c_n\} \in \mathbb{R}^{4 \times n}$, where $c_i = \{u_i, v_i, a_i, b_i\}^T$ denotes the cup descriptors of the i -th reference disc, u_i, v_i are the coordinates of the cup center, and a_i is the cup radius and b_i is the associated CDR after disc normalization to a radius of 1. For a given test disc $y \in \mathbb{R}^{k \times 1}$, we want to recover optimal linear reconstruction coefficients $w \in \mathbb{R}^{n \times 1}$, $|w| = 1$, that minimize the reconstruction error $\|y - Xw\|^2$. After w is obtained, the cup localization of y is estimated as Cw .

Since reconstruction of cup parameters is more accurate from images more similar to the test disc, we also include a cost term in the objective function that penalizes the use of references that are less similar to the test image. Let us denote the costs for the reference discs in X as the vector $d = \{d_1, d_2, \dots, d_n\}^T \in \mathbb{R}^{n \times 1}$, where d_i is the cost of using x_i for reconstruction. The overall cost term can then be expressed as $\|d \odot w\|^2$ where \odot denotes the Hadamard product. Combining this cost term with the reconstruction error gives the following objective function:

$$\arg \min_w \|y - Xw\|^2 + \lambda \|d \odot w\|^2, \quad \text{s.t. } |w| = 1. \quad (1)$$

This objective can be minimized in closed form using the Lagrange multiplier method, without the need for iterations:

$$w = \frac{1}{\mathbf{1}^T (\hat{X}^T \hat{X} + \lambda D^T D) \mathbf{1}} (\hat{X}^T \hat{X} + \lambda D^T D)^{-1} \mathbf{1}, \quad \hat{X} = (\mathbf{1} \otimes y - X), \quad (2)$$

where $D = \text{diag}(d)$ and \otimes denotes the Kronecker product. For simplicity, we define in our implementation the cost d_i as the Gaussian distance between the test disc y and the i -th reference disc x_i , *i.e.*,

$$d_i = e^{-\frac{\|y - x_i\|^2}{2\sigma_f^2}}, \quad (3)$$

where σ_f is a parameter that accounts for imaging noise.

3 Experiments

We evaluated the cup localization accuracy and CDR-based diagnosis performance of our proposed method. Comparisons are presented of our reconstruction-based approach to state-of-the-art pixel [4], sliding window [5] and superpixel [6] based methods, as well as to other reconstruction-based approaches. In addition, we compare our system against the current clinical standard for glaucoma detection using intra-ocular pressure (IOP) and to CDR values from expert graders. The effects of different optic disc segmentation algorithms and codebook sizes on our system are also examined.

3.1 Evaluation Criteria

For cup localization, we use the two evaluation criteria in [5] and [6], *i.e.*, non-overlap ratio (m_1) and absolute CDR error (δ), defined as

$$m_1 = 1 - \frac{\text{area}(E_{dt} \cap E_{gt})}{\text{area}(E_{dt} \cup E_{gt})}, \quad \delta = \frac{|D_{dt} - D_{gt}|}{2} \quad (4)$$

where E_{dt} denotes a detected cup region, E_{gt} denotes the ground-truth cup ellipse, D_{dt} is the vertical diameter of the detected cup, and D_{gt} is the vertical diameter of the ground-truth cup.

For glaucoma diagnosis, the area under the ROC (receiver operation characteristic) curve (AUC) is used for performance evaluation. The ROC is plotted as a curve which shows the tradeoff between sensitivity (P_+) and specificity (P_-), defined as

$$P_+ = \frac{TP}{TP + FN}, \quad P_- = \frac{TN}{TN + FP}, \quad (5)$$

where TP and TN denote the number of true positives and true negatives, respectively, and FP and FN denote the number of false positives and false negatives, respectively.

3.2 Experimental Setup

Two datasets, namely ORIGA [11] and SCES [12], were used in the experiments to validate both cup localization and glaucoma diagnosis accuracy. The *ORIGA* dataset, with ground truth disc and cup labelings as well as clinical glaucoma diagnoses, is comprised of 168 glaucoma and 482 normal images from studies of a Malay population. It is used in our work to validate both cup localization and glaucoma diagnosis. The Chinese population-based *SCES* dataset consists of 1676 images, of which 46 ($\sim 3\%$) are glaucoma cases. Since the *SCES* dataset provides only clinical diagnoses, it will be used only to assess the diagnostic performance of our system.

3.3 Effect of Different Disc Segmentation Methods on Our Method

We investigated the effect of different disc segmentation methods in our framework. The same parameters (codebook size $n = 99$, regularization weight $\lambda = 1000$, and Gaussian distance parameter $\sigma_f = 4$) were used for all the methods, and the results are

Table 1. Performance comparisons using different disc segmentation algorithms

Segmentation algorithms		Evaluation criteria	
Reference discs	Test discs	m_1	δ
Template matching	Template matching	0.225	0.071
ASM	ASM	0.227	0.073
Manual	Manual	0.229	0.074
Manual	Template matching	0.230	0.075
Manual	ASM	0.231	0.075
ASM	Template matching	0.229	0.074
Template matching	ASM	0.232	0.076

listed in Table 1. From the results, we observe that the proposed framework is robust and not very sensitive to the disc segmentation method. Slightly better performance was achieved when the reference discs and test discs were obtained using the same method. We note that using the same automatic segmentation method gives better results than manual labeling, likely because of greater consistency in the automatic methods.

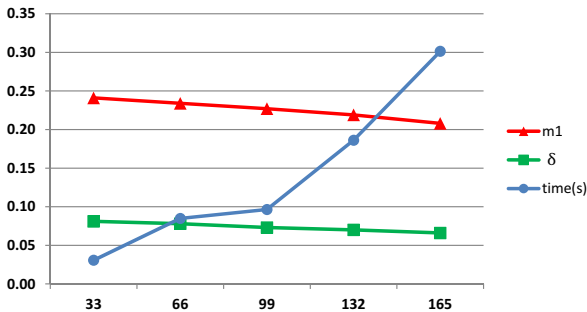
3.4 Cup Localization Comparisons

We compared our reconstruction-based approach to state-of-the-art pixel, window and superpixel based methods. Listed in Table 2, the results show that the proposed method achieves significant improvements in cup localization accuracy, which indicates the advantage of a holistic, reconstruction-based approach over techniques that analyze low-level visual cues. Substantial gains are made in terms of speed as well. With a codebook of 99 reference images, the proposed method locates cups in 0.1s, more than ten times faster than the comparison methods: pixel-based [4] (1.5s), window-based [5] (360s), and superpixel-based [6] (1.7–20.2s).

To support our particular reconstruction-based formulation, we also compare it to other possible reconstruction approaches, based on sparse coding [13], Locally Linear Embedding (LLE) [14], and Locality Preserving Projections (LPP) [15]. The codebook used for our method (with 99 reference images) is also used for these other reconstruction techniques. The sparse coding regularization weight was set to 100; LLE was implemented using the 11 nearest neighbors, and LPP was applied with an 11-dimensional subspace. These settings were found to approximately maximize the performance of these techniques, which is also reported in Table 2. The performance of the sparse coding method indicates limitations in using only a small number of reference images and not taking their similarity to the test image into account. Its optimization runs at 6.4s per image. Like our method, LLE considers both reference image similarity and reconstruction error, but does so in separate steps (first finding nearest neighbors, then reconstructing from them). This is shown to give lower accuracy than optimizing them jointly in our method. As with sparse coding, a reason for this is the sparse reference data, which makes it difficult to determine a priori which or how many reference images are needed for accurate reconstruction. A benefit of LLE, however, is its computational efficiency, running at 0.02s per image. The LPP method is also very fast

Table 2. Cup localization performance comparisons on *ORIGA* dataset

Method	m_1	δ
reconstruction based	0.225	0.071
<i>pixel based [4]</i>	0.474	0.149
<i>window based [5]</i>	0.284	0.096
<i>supapixel based [6]</i>	0.267	0.081
<i>error reduction relative to [4]</i>	52.5%	52.3%
<i>error reduction relative to [5]</i>	20.8%	26.0%
<i>error reduction relative to [6]</i>	15.7%	12.3%
<i>sparse coding</i>	0.408	0.158
<i>LLE</i>	0.247	0.081
<i>LPP</i>	0.377	0.150
<i>reconstruction-based without similarity cost</i>	0.271	0.091

**Fig. 2.** Cup localization accuracy and speed comparison with different codebook sizes

(0.012s per image), but loses much information in the dimensionality reduction and is very sensitive to alignment errors in the training and testing samples.

Table 2 also shows that leaving out the similarity-based cost term from our objective function (by setting $\lambda = 0$) yields a result that is worse (with significance tests of m_1 : $p < 0.001$, cdr : $p < 0.001$). This indicates that giving higher weights to more similar references is indeed better.

3.5 Influence of Codebook Size on Accuracy and Processing Time

For reconstruction-based approaches, codebook size is a key performance factor in terms of both accuracy and speed. We examined the cup localization accuracy and processing time of the proposed method with different codebook sizes $N \in \{33, 66, 99, 132, 165\}$ and the remaining images used for testing. As shown in Fig. 2, accuracy in terms of m_1 and δ improves with larger codebook sizes. The consequent increases in computation time still leave this method much faster than previous cup localization techniques, and in the following subsection we use all 650 images from *ORIGA* as reference images in evaluating glaucoma diagnostic performance on the *SCES* dataset.

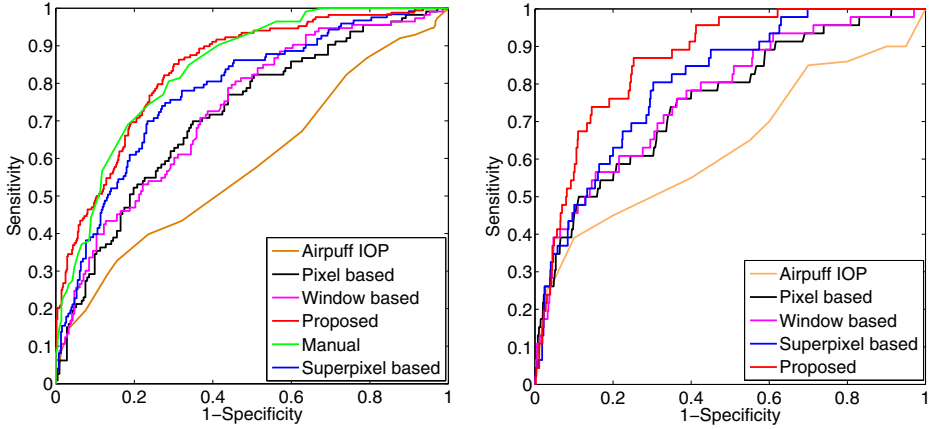


Fig. 3. Glaucoma diagnosis performance. Left: *ORIGA* dataset. Right: *SCES* dataset.

3.6 Glaucoma Diagnosis

To assess the glaucoma diagnosis accuracy of our method, we compare its predictions to the ground truth diagnoses of ophthalmologists and to state-of-the-art techniques. In addition, we compare to the current standard of care for glaucoma detection using IOP, as well as CDR grading results from an expert grader. For testing on *ORIGA*, a random selection of $N = 99$ images is used to generate the codebook and the remaining 551 images are used for testing. Testing on *SCES* was carried out using a codebook trained using the 650 images from *ORIGA*. We note that the reference images from *ORIGA* are of a different ethnicity from that of *SCES*.

The results are shown in Fig. 3. One can observe that the proposed reconstruction-based method outperforms previous automatic methods and IOP. The results are also close to those of the expert grader on the *ORIGA* dataset. The AUC values of our method on *ORIGA* and *SCES* are 0.823 and 0.860, respectively. However, the corresponding sensitivity values are only 58.0% and 73.9% on *ORIGA* and *SCES*, respectively, at an observation specificity of 85%. This indicates that other clinical diagnosis cues besides just CDR should be explored to further improve accuracy.

4 Conclusion

For CDR-based cup localization in glaucoma diagnosis, we have proposed a one-step reconstruction-based method that is computationally efficient and relatively insensitive to the disc segmentation method used. On tests with the *ORIGA* dataset to evaluate cup localization accuracy, our system achieves a 22.5% non-overlap ratio (m_1) against the ground truth and a 0.071 absolute CDR error (δ), a significant improvement over current state-of-the-art methods. Tests on the *SCES* dataset show that our method achieves an AUC value of 0.86, a sensitivity of 73.9% at a specificity of 85%. In future work, we plan to improve this reconstruction technique by introducing a learning algorithm to generate more effective codebooks for larger datasets, and also will investigate using data from multiple modalities to improve system performance.

References

1. Shen, S., Wong, T.Y., Foster, P., Loo, J., Rosman, M., Loon, S., Wong, W., Saw, S.M., Aung, T.: The Prevalence and Types of Glaucoma in Malay People: the Singapore Malay Eye Study. *Invest. Ophthalmol. Vis. Sci.* 49(9), 3846–3851 (2008)
2. Jonas, J., Budde, W., Panda-Jonas, S.: Ophthalmoscopic Evaluation of the Optic Nerve Head. *Survey of Ophthalmology* 43, 293–320 (1999)
3. Abramoff, M., Alward, W., Greenlee, E., Shuba, L., Kim, C., Fingert, J., Kwon, Y.: Automated Segmentation of the Optic Disc from Stereo Color Photographs Using Physiologically Plausible Features. *Invest. Ophthalmol. Vis. Sci.* 48(4), 1665–1673 (2007)
4. Wong, D.W.K., Lim, J.H., Tan, N.M., Zhang, Z., Lu, S., Li, H., Teo, M., Chan, K., Wong, T.Y.: Intelligent Fusion of Cup-to-Disc Ratio Determination Methods for Glaucoma Detection in ARGALI. In: *Int. Conf. Engin. in Med. and Biol. Soc.*, pp. 5777–5780 (2009)
5. Xu, Y., Xu, D., Lin, S., Liu, J., Cheng, J., Cheung, C.Y., Aung, T., Wong, T.Y.: Sliding Window and Regression based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part III. LNCS*, vol. 6893, pp. 1–8. Springer, Heidelberg (2011)
6. Xu, Y., Liu, J., Lin, S., Xu, D., Cheung, C.Y., Aung, T., Wong, T.Y.: Efficient Optic Cup Detection from Intra-image Learning with Retinal Structure Priors. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 58–65. Springer, Heidelberg (2012)
7. Wong, D.W.K., Liu, J., Tan, N.M., Yin, F., Wong, T.Y.: Automatic Detection of the Optic Cup Using Vessel Kinking in Digital Retinal Fundus Images. In: *IEEE Int. Symp. Biomed. Imaging*, pp. 1647–1650 (2012)
8. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de-Solorzano, C.: Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data. *IEEE Trans. Med. Im.* 28(8), 1266–1277 (2009)
9. Yin, F., Liu, J., Ong, S.H., Sun, D., Wong, D.W.K., Tan, N.M., Baskaran, M., Cheung, C.Y., Aung, T., Wong, T.Y.: Model-based Optic Nerve Head Segmentation on Retinal Fundus Images. In: *IEEE Int. Conf. Engin. in Med. and Biol. Soc.*, pp. 2626–2629 (2011)
10. Cheng, J., Liu, J., Wong, D.W.K., Yin, F., Cheung, C.Y., Baskaran, M., Aung, T., Wong, T.Y.: Automatic Optic Disc Segmentation with Peripapillary Atrophy Elimination. In: *IEEE Int. Conf. Engin. in Med. and Biol. Soc.*, pp. 6224–6227 (2011)
11. Zhang, Z., Yin, F., Liu, J., Wong, D.W.K., Tan, N.M., Lee, B.H., Cheng, J., Wong, T.Y.: Origa-Light: An Online Retinal Fundus Image Database for Glaucoma Analysis and Research. In: *IEEE Int. Conf. Engin. in Med. and Biol. Soc.*, pp. 3065–3068 (2010)
12. Sng, C.C., Foo, L.L., Cheng, C.Y., Allen Jr., J.C., He, M., Krishnaswamy, G., Nongpiur, M.E., Friedman, D.S., Wong, T.Y., Aung, T.: Determinants of Anterior Chamber Depth: the Singapore Chinese Eye Study. *Ophthalmology* 119(6), 1143–1150 (2012)
13. Olshausen, B.A., Field, D.J.: Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature* 381(6583), 607–609 (1996)
14. Roweis, S., Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290(5500), 2323–2326 (2000)
15. Cao, Y., Yuan, Y., Li, X., Turkbey, B., Choyke, P.L., Yan, P.: Segmenting Images by Combining Selected Atlases on Manifold. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part III. LNCS*, vol. 6893, pp. 272–279. Springer, Heidelberg (2011)

Automatic Detection of Blue-White Veil by Discrete Colour Matching in Dermoscopy Images

Ali Madooei, Mark S. Drew, Maryam Sadeghi, and M. Stella Atkins

School of Computing Science
Simon Fraser University
amadooei@cs.sfu.ca,
<http://www.cs.sfu.ca/~amadooei>

Abstract. Skin lesions are often comprised of various colours. The presence of multiple colours with an irregular distribution can signal malignancy. Among common colours under dermoscopy, blue-grey (blue-white veil) is a strong indicator of malignant melanoma. Since it is not always easy to visually identify and recognize this feature, a computerised automatic colour analysis method can provide the clinician with an objective second opinion. In this paper, we put forward an innovative method, through colour analysis and computer vision techniques, to automatically detect and segment blue-white veil areas in dermoscopy images. The proposed method is an attempt to mimic the human perception of lesion colours, and improves and outperforms the state-of-the-art as shown in our experiments.

1 Introduction

Early detection of melanoma is paramount to patients' prognosis towards greater survival. The challenges involved in clinical diagnosis of early melanoma have provoked increased interest in computer-aided diagnosis systems through automatic analysis of skin lesion images.

Dermoscopy is a popular imaging technique, widely used by clinicians for examining and monitoring pigmented skin lesions [1]. "Inevitably, most efforts in computerizing diagnosis of melanoma lean towards automatic analysis of dermoscopy images" [2]. Dermoscopy unveils visual features in pigmented lesions that are not discernible by the naked eye.

Among dermoscopic features, colour has a substantial role. For instance, one of the most significant indicators of malignant melanoma is the blue-white veil (a.k.a. blue-grey veil) feature (with sensitivity of 51% and a specificity of 97% [3]).

Fig. 1-(a,c) show dermoscopy images of melanoma with blue-white veil feature; the blue-white veil regions are structureless areas of confluent blue pigment with a ground-glass haze (as if the image were out of focus there) [4].

The colour blue (under dermoscopy) indicates melanin localized within deeper parts of the skin [5]. Blue-white veil is found associated with hyperkeratinisation over dense amounts of melanin in the dermis [6]. While uniform blue-white structures may also be observed over blue nevus (benign) lesions, asymmetrical and irregular blue-white veil usually indicates malignant melanoma.

There exists extensive literature on skin lesion image analysis (see [8] for a survey), but here we focus on a study by Celebi et al. [7] which specifically targets automatic

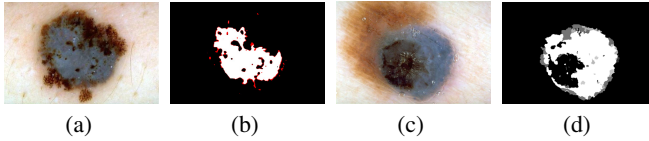


Fig. 1. (a,c): Melanoma images with Blue-white Veil; (b): veil mask by Alg. 1; The red areas are indicating disagreement between Alg. 1 and [7]; (d): veil probability map by Alg. 2

blue-white veil detection¹. Their approach involves pixel classification using explicit thresholding, where the threshold values are induced by a trained decision tree.

This paper puts forward an alternative method, by incorporating colour analysis and computer vision techniques, to automatically detect and segment blue-white veil regions in dermoscopy images. The proposed method is described next.

2 Method

We describe two approaches for automatic detection and segmentation of blue-white veil regions in dermoscopy images. The first approach is a makeover of the method [7] which performs equally well with substantively less computation. The second approach, our main contribution, is an innovative method which attempts to mimic human interpretation of lesion colours under dermoscopy. The latter incorporates colour analysis and computer vision techniques and, as will be shown, outperforms the former method in our experimental trials.

2.1 Blue-White Veil Detection by Thresholding

Celebi et al. [7] used a set of 105 images selected from [4], consisting of 43 images containing sizeable blue-white veil areas with the remaining 62 free of this feature. For each image a number of small circular regions that contain either veil or non-veil pixels were manually determined by a dermatologist and used for training.

A decision tree classifier with C4.5 [10] induction algorithm was employed to classify each pixel in the training stage into two classes: veil and non-veil. Among the 18 different colour and texture features included,² only two features appeared in the induced decision rules: The classification was conducted by thresholding on a normalized-blue channel ($B / (R + G + B)$) and relative-red feature (defined as $R - \bar{R}_s$ where \bar{R}_s is the mean of red channel values for *healthy* skin areas only).

Decision trees are simple to use and easy to understand, yet they impose disadvantages and shortcomings. For instance, their good detection rate usually arrives at the expense of high false positives. In our experiments to reproduce the method [7], using

¹ The use of blue-white veil feature has been reported in some commercially available computer-aided diagnosis (CAD) systems that do use colour information (see e.g. [9]). To our knowledge however, the only study that reports a method, experimental procedure, and results specifically pertaining to the feature being examined here is the one by Celebi et al.

² The description of features – as well as the feature extraction process – is omitted for space considerations. The interested reader is referred to [7] for details.

the training images,³ we arrived at sensitivity of 78.51% and specificity of 97.03%, which demonstrates our earlier argument about this method.

Also, caution must be taken in defining the decision boundaries (threshold values); the colour values are highly dependent on the acquisition technique and the imaging setup. For example, if illumination changes, or in cases with shadows, shading, and other colour-degrading factors, thresholding methods might fail ungracefully.

The findings of Celebi et al. indicate that blue-white veil colour data has a restricted range of values and does not fall randomly in a given colour feature space. This indicates that their method can benefit from the choice of colour representation. To investigate this, we have reproduced their training experiment where each pixel has been represented by its corresponding coordinates in RGB, nRGB, L*a*b*, YIQ, HSV, and MHG [11] colour spaces. These colour models were chosen since each represents a different family of colour spaces.

Our investigation revealed that by using normalized blue, and the L* channel (of L*a*b*) one can reproduce the results obtained by Celebi. Our finding further showed that we can replace L* with $Lum = R + G + B$ and obtain equally good results. The latter colour feature is considerably faster to compute, particularly to replace relative-red feature in the original method. Note that the computation of the relative-red feature requires a search for normal (healthy) skin colour in the background. This search also constrains the method's performance subject to the accuracy of the lesion border detection algorithm and skin colour filter used.

We further investigated the efficacy of employing alternative classification techniques. In our trials, we used Single Gaussian, Mixture of Gaussian, Bayesian, Logistic Regression, and support vector machine (SVM). Unsurprisingly, the SVM produced the best results, however, the improvements were negligible with respect to added computational cost and complexity (data not shown for space considerations).

In conclusion, we first propose a revised thresholding-based method (Algorithm 1) which is based on but improves on Celebi's approach (in term of computational cost and complexity). Testing Algorithm 1 on training images, we achieved sensitivity of 78.22% and specificity of 96.76%, which is very similar to the performance of [7]. Fig. 1-(b) illustrates the veil mask generated by this method.

Algorithm 1. – Blue-white veil detection by thresholding

```

1: Load a skin lesion image
2: for each pixel do
3:    $Lum = R + G + B$ 
4:    $nB = B/Lum$ 
5:   if  $nB \geq 0.3$  and  $0.6 \leq Lum \leq 2$  then
6:     Classify pixel as veil
7:   else
8:     Classify pixel as non-veil
9:   end if
10: end for

```

³ To conduct the experiments the entire training image-set were used, whereas training has been done on the manually selected regions as described.

An important insight to Celebi's study is that texture features were found irrelevant to detection of blue-white veil data. Celebi et al. used the classical statistical texture measures of [12] (entropy, contrast, correlation) derived from the grey level co-occurrence matrix (GLCM) of an image. We further investigated this by using two more popular descriptors in the texture classification literature: LBP [13] and MR8 [14]. Our investigation agrees with the findings in [7], attesting that texture descriptors are not discriminating features for detection of blue-white veil structure.

2.2 Blue-White Veil Detection by Discrete Colour Matching

In this section we intend to analyse the problem under study, from a further level of abstraction. That is, going beyond describing a given colour by merely indicating the location of the corresponding colours in the considered colour space.

The scheme followed in Alg. 1 and also in [7] and many others is entirely based on data analysis: it is meant for computational convenience and not to model human perception and interpretation of colour. We propose an alternative approach, as described shortly, to mimic the colour assessment performed by human observers. The alternative is an attempt to find a more perceptually and semantically meaningful way of describing colours under dermoscopy. To this aim, we ask the question of how a dermatologist identifies the presence of certain colours under dermoscopy.

Colour *naming* is to some degree subjective. However we can safely say that observers are influenced by the colours they saw previously. Indeed, a dermatologist needs training to be able to identify the blue-white veil feature. Thus, we propose to identify this feature by colour matching to a *discrete* set of colours best describing "blue-white veil". Here, we are interested in asking the question how we would name the blue-white veil colour and which colours are in this colour family.

To answer the latter questions, we analysed the veil data by mapping their colour values to the Munsell colour system [15]. In colorimetry, the Munsell colour system is one of the most fundamental colour-modellings. The Munsell system offers both perceptual and quantitative colour definitions. The perceptual definition (given in the form of a book with printed colour patches) is appropriate for the use of artists such as painters and designers, whereas the quantitative definition provides measurement standards that are appropriate for technical and scientific use.

Quantitatively, a colour in Munsell space is defined by its hue, value (lightness), and chroma (colour purity). Hue is given by a number between 0 and 10 which denotes the five principal hues: Red, Yellow, Green, Blue, and Purple, along with 5 intermediate hues (e.g., YR) halfway between adjacent principal hues. Each of these 10 hues is then broken into 10 sub-steps, so there are a total of 100 hues with integer prefixes. Value, indicating how light a colour is, is a number between 0 (signifying black) and 10 (signifying white). For any given Munsell hue and value, chroma starts at 0 (grey) and extends to a variable positive number, which represents the most saturated colour of that hue and value.

Since there is no direct conversion from standard RGB to Munsell colour quantities, we consider finding the best (closest) match between each pixel colour triple and the equivalent colour value representation of each Munsell colour patch. For colour matching, we consider working in the $L^*a^*b^*$ colour space (a.k.a. CIELAB) which is an *approximately* perceptually uniform colour model. That is, colours are organized in



Fig. 2. Colour Palette of most frequent colours identified as blue-white veil

such a way that colour differences agree more consistently with human visual perception. Thus the difference between two colours can be measured using e.g. the Euclidean distance metric. Note that the common sRGB colour space does not yield this property. In addition, CIE colour spaces are device independent, which make them suitable for colour matching and colour comparison.

Moreover, to achieve the best generalization, each colour pixel value is replaced with the colour value of its ‘superpixel’ representation. “Superpixels group pixels into perceptually meaningful atomic regions, which can be used to replace the rigid structure of the pixel grid” [16]. Superpixels also preserve information over scales and sampling resolutions, which captures image redundancy.⁴

Next, we match each colour of veil data to its closest colour patch in the Munsell system using the nearest neighbour (NN) searching technique. Interestingly, the 146,353 pixels under analysis mapped to only 116 of the totality of 2352 Munsell colour patches available in our look-up table.⁵ Among these, 98% of the veil data is described by only 80 colour patches. Fig. 2 shows these 80 colours organized on a palette according to their frequency (in a descending fashion, from left to right).

We also analysed non-veil data by the same principle. The 254,739 pixels from non-veil areas mapped to 129 Munsell colour patches, among which only 3 patches were overlapping with the 116 veil patches. These 3 contribute (all together) to less than 2% of veil data and are not considered among the 80 patches in the blue-white veil colour palette.

The colour palette (Fig. 2) can be used (much like dermatologist prior knowledge) to extract blue-white veil regions from an input image according to a nearest neighbour approach.⁶ From the training data we can also define the posterior probability of a pixel being veil given its colour in terms of the likelihood of observing its associated colour patch given the class label and the prior probability of classes.

Matching each pixel of each image in the fashion described above could be computationally expensive. To reduce this cost, we can segment colours in any given image and instead match the colour vector of the centroid of each cluster (segment) to colours in the colour palette. By colour segmentation, we decompose an image into visually homogeneous regions and effectively reduce the number of basic colour regions, while preserving salient features of the overall image. This is particularly of interest to us because in dermoscopy images there are a handful of distinct colour regions (such as

⁴ Here, we compute superpixels according to the SLIC [16] algorithm. SLIC is a simple and efficient method based on a spatially localized version of k-means clustering.

⁵ The look-up table is available at <http://www.wallkillcolor.com/mutable.csv>

⁶ The use of a colour palette is in part inspired by the work of Seidenari et al. [17].

blue-white veil) where the presence of a feature significantly affects the diagnosis while the information within such a region is often less important and can be neglected.

Colour image segmentation can be done in an automated fashion using e.g. the mean-shift method. Mean shift (MS) is a fast and robust mode-seeking algorithm; it has been used for unsupervised image segmentation where the number and the shape of the image clusters are unknown. In this study we use the EDISON [18] software.

Algorithm 2 summarizes the method described here. Testing Alg. 2 on training images, we achieved sensitivity of 83.36% and specificity of 98.04%, which outperforms Alg. 1 as well as the method [7]. Fig. 1-(d) illustrates the veil mask generated by this method (note the probabilistic [non-binary] output).

Algorithm 2. – Blue-white veil detection by discrete colour matching

```

1: Load a skin lesion image
2: Convert from sRGB to CIELAB (Assumption: light is D65)
3: Segment using EDISON (parameters are SpatialBandWidth=7, RangeBandWidth=6.5, and
   MinimumRegionArea=0.01*nRows*nColumns.)
4: for each segmented region do
5:   Find the best match from colour palette
6:   if The best match is within the tolerance (threshold) distance then
7:     Classify region as veil
8:     Assign  $P(Veil|segmentcolour) = P(colourpatch|Veil) \times P(Veil)$ 
9:   else
10:    Classify region as non-veil
11:    Assign  $P(Veil|segmentcolour) = 0$ 
12:   end if
13: end for

```

3 Experiments

We tested our proposed methods on a set of 223 images selected from [4] by Celebi et al. [7]. This set consists of 173 images containing blue-white veil areas and a remaining 50 free of this feature. For this set of images, since the lesion border is provided with the ground truth, we can run the method [7] and compare the classification results, as presented in Table 1. Note that, as claimed, the proposed method Alg. 1 has produced reasonably similar results to [7] whereas it is substantively easier to compute. Moreover, the proposed method Alg. 2 performs best and in particular has higher recall (sensitivity), which is of significant importance for clinical examinations.

Accurate detection and segmentation of blue-white veil feature can be useful for computer analysis of skin lesion images. In clinical assessments however, presence or absence of this feature is associated with diagnosis. Accordingly, in a different experiment we aimed to determine only the presence (or absence) of veil feature in a set

Table 1. Classification results – Proposed methods vs. [7]

Method	n	Precision	Recall	F-Score	Specificity
Celebi et al. [7]	223	0.70	0.65	0.68	0.97
Proposed Method 1	223	0.67	0.65	0.66	0.97
Proposed Method 2	223	0.70	0.71	0.70	0.97

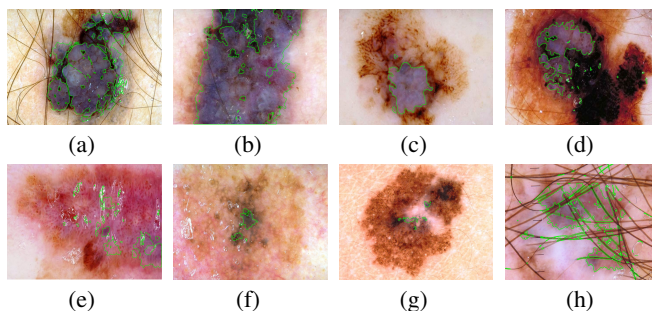


Fig. 3. (a-d): Easy images; (e-h): challenging images – The green border indicates veil areas detected by Algorithm 2. Note, the binary mask here is created by setting a threshold on the distance metric. The threshold was found empirically from training data (and set to 100 units).

of 300 images taken from various sources. The image set is divided to two subsets of 200 ‘easy’ and 100 ‘challenging’ images. An image is considered challenging if the blue-white veil area was too small, too pale, occluded, or had variegated colour.

Our experiment produced accuracy of 75% and 58% for Alg. 1 on easy and challenging sets respectively. Using Alg. 2 these results were boosted to 87% and 67% accordingly. Note that, in order to exclude very small areas which are without clinical relevance, a minimum value⁷ for areas of veil region was considered. Fig. 3 illustrates the output of Alg. 2 on some of the images in each set.

It is to be noted that the proposed method can easily be extended to account for detecting other colour features in dermoscopy images. In fact, it can be seen as a framework for colour based detection and assessment problems of similar nature. Common colours under dermoscopy are light brown, dark brown, black, blue, blue-grey, red, yellow, and white [5]. Since it is not always easy to visually distinguish these, our method (once extended to detect these *common* colours) can be effectively used as a colour clustering and colour naming engine.

The blue-white veil colour palette can be used for training purposes, as well as a reference tool (such as designers’ colour chart) for dermoscopy trainees. For that matter, a colour palette of other common colours can be generated as well. Note that the proposed method can automatically generate colour palettes given labelled training data. The advantage of this is twofold since statistical data can be extracted from training data and associated with colour patches on colour palettes. In fact, one can use this method to extract a colour map of lesions with common colours (under dermoscopy) and link it to statistical data to e.g. associate a feature with a diagnosis.

Finally, this method can aid clinical and laboratory investigations to e.g. confirm the high diagnostic relevance of presence or absence of colour and colour related features.

4 Conclusion

We have presented two schemas for automatic detection of blue-white veil feature in dermoscopy images. We first propose a revised thresholding-based method with results

⁷ Threshold is set to 0.5% of image area.

comparable to state of the art, with much reduced computation. The second approach, our main contribution, sets out an innovative method that attempts to mimic human interpretation of lesion colours. The latter outperforms prior art and moreover introduces a perceptually and semantically meaningful new approach which can serve as a scaffolding for new colour investigations in dermoscopy, for example for detection and recognition of *common* colours under dermoscopy.

Future work includes extending the proposed method to detect other *common* colours under dermoscopy, and demonstrating the usefulness of the feature under study in detection of early melanoma in a computerized automatic fashion.

References

1. Kaminska-Winciorek, G., Spiewak, R.: Tips and tricks in the dermoscopy of pigmented lesions. *BMC Dermatology* 12(1) (August 2012)
2. Madooei, A., Drew, M.S., Sadeghi, M., Atkins, M.S.: Automated pre-processing method for dermoscopic images and its application to pigmented skin lesion segmentation. In: 20th Color and Imaging Conference, pp. 158–163 (2012)
3. Menzies, S.: *Atlas of Surface Microscopy of Pigmented Skin Lesions*. McGraw Hill (2003)
4. Argenziano, G., et al.: *Interactive atlas of dermoscopy*. Edra Medical Publishing & New Media, Milan (2000)
5. Soyer, H.P., et al.: Dermoscopy of pigmented skin lesions. *EJD* 11(3), 270–276 (2001)
6. Adams, J.: Dermoscopic blue hue and blue-white veil in skin lesions indicate different diagnoses. *Am. J. Dermatopathol.* (23), 463–469 (2001)
7. Celebi, M.E., et al.: Automatic detection of blue-white veil and related structures in dermoscopy images. *Computerized Medical Imaging and Graphics* 32(8), 670–677 (2008)
8. Korotkov, K., Garcia, R.: Computerized analysis of pigmented skin lesions: a review. *Artificial Intelligence in Medicine* 56(2), 69–90 (2012); PMID: 23063256
9. Menzies, S., et al.: The performance of SolarScan. *Archives of Dermatology* (11), 1388–1396 (2005); PMID: 16301386
10. Quinlan, J.R.: *C4. 5: programs for machine learning*, vol. 1. Morgan Kaufmann (1993)
11. Madooei, A., Drew, M.S., Sadeghi, M., Atkins, M.S.: Intrinsic melanin and hemoglobin colour components for skin lesion malignancy detection. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 315–322. Springer, Heidelberg (2012)
12. Haralick, R., Shapiro, L.: *Computer and Robot Vision*, p. 459. Addison-Wesley (1992)
13. Ojala, T., et al.: Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal.* 24(7), 971–987 (2002)
14. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. *Int. J. of Computer Vision* 62(1), 61–81 (2005)
15. Landa, E.R., Fairchild, M.D.: Charting color from the eye of the beholder. *American Scientist* 93(5), 436–443 (2005)
16. Achanta, R., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(11), 2274–2282 (2012)
17. Seidenari, S., Pellacani, G., Grana, C.: Computer description of colours in dermoscopic melanocytic lesion images reproducing clinical assessment. *British J. of Dermatology* 149(3), 523–529 (2003)
18. Georgescu, B., Christoudias, C.M.: The edge detection and image segmentation (EDISON) system (2003), Code available at <http://coewww.rutgers.edu/riul/research/code/EDISON/>

Separation of Benign and Malignant Glands in Prostatic Adenocarcinoma

Sabrina Rashid¹, Ladan Fazli², Alexander Boag³, Robert Siemens³,
Purang Abolmaesumi¹, and Septimiu E. Salcudean¹

¹ Department of Electrical and Computer Engineering,
University of British Columbia, Vancouver, BC, Canada

² The Vancouver Prostate Center,
University of British Columbia, Vancouver, BC, Canada

³ Kingston General Hospital, Kingston, ON, Canada

Abstract. This paper presents an analysis of the high resolution histopathology images of the prostate with a focus on the evolution of morphological gland features in prostatic adenocarcinoma. Here we propose a novel technique of labeling individual glands as malignant or benign. In the first step, the gland and nuclei objects of the images are automatically segmented. Individual gland units are segmented out by consolidating their lumina with the surrounding layers of epithelium and nuclei. The nuclei objects are segmented by using a marker controlled watershed algorithm. Two new features, Number of Nuclei Layer (N_{NL}) and Ratio of Epithelial layer area to Lumen area (R_{EL}) have been extracted from the segmented units. The main advantage of this approach is that it can detect individual malignant gland units, irrespective of neighboring histology and/or the spatial extent of the cancer. The proposed algorithm has been tested on 40 histopathology scenes taken from 10 high resolution whole mount images and achieved a sensitivity of 0.83 and specificity of 0.81 in a leave-75%-out cross-validation.

Keywords: Nuclei layer, epithelial layer, prostatic adenocarcinoma.

1 Introduction

Prostate cancer is one of the most frequently diagnosed cancers and ranks second among the cancer related deaths of men worldwide [1]. Analysis of the histopathology specimens of prostate is an important step for prostate cancer diagnosis and treatment planning.

The tissue features of these histopathology images are the key indicators of prostate cancer. Among the different types of prostate cancer, the most common one is the prostatic adenocarcinoma, cancer pertaining to the gland units of the prostate. Pathologists determine the extent of this cancer by carefully evaluating the changes in the gland morphology. The gland is the main histopathological structural unit in prostate. Fig. 1 shows the structure of a normal gland unit. It mainly comprises a lumina of irregular shape, a layer of epithelial cells, and

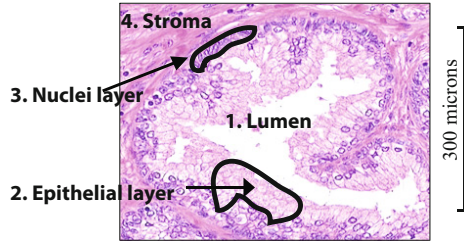


Fig. 1. Illustration of the histopathology components associated with a gland unit: 1. Lumen, 2. Epithelial layer, 3. Nuclei, and 4. Stroma.

nuclei surrounding the lumina. The unit is supported by a surrounding fibromuscular stroma. When the slides are stained using a Hematoxylin and Eosin (H&E) solution, the nuclei turn dark blue and the epithelial layer and stroma turn into different shades of purple to pink.

The recent literature on computerized diagnosis of prostate cancer quantizes the morphological and architectural features associated with the gland units for cancer detection and grading. The most commonly used features on analyzing histopathology specimens are related to the size and shape of gland lumina, nuclei shape and density [2], [3], [4], [5]. Though gland size and shape do contain information about the abnormality of prostate tissue, this feature is not exclusive to cancerous tissue only. In case of other prostate anomalies such as, atrophy and benign prostatic hyperplasia the gland size and shape resembles that of cancerous ones [6]. Apart from gland-based features, some approaches exploit overall image features such as, energy and entropy of multiwavelet coefficient of the image [7], fractal dimension [8] and so on. But these features are not specific to each gland and do not capture the gland specific features that are clinically used for cancer classification.

Therefore, more decisive features are needed for effective separation of individual glands. Here, we propose two novel features based on which glands can be classified: i) Number of Nuclei Layer (N_{NL}), and ii) Ratio of Epithelial layer area to Lumen area (R_{EL}). Fig. 2 illustrates the change of appearance between benign and malignant glands. In benign glands there are multiple layers of nuclei surrounding the gland unit with relatively thinner epithelial layer compared to lumen area. On the other hand, in malignant units there are usually a single layer of nuclei surrounding the gland unit with a thick epithelial layer. Our proposed features N_{NL} and R_{EL} quantize these two properties of malignant glands. In a recent literature, Nguyen et al. [3] achieved an accuracy of 0.79 in classification of benign and malignant glands by exploiting region-specific features such as percentage of nuclei pixels, gland shape, crowdedness etc. In comparison to that, our proposed features are strictly gland specific and involve i) pixel labeling, ii) segmentation of each nuclei in gland, and iii) finding the number of layers of nuclei for each gland from angle-dependent histograms. The advantage

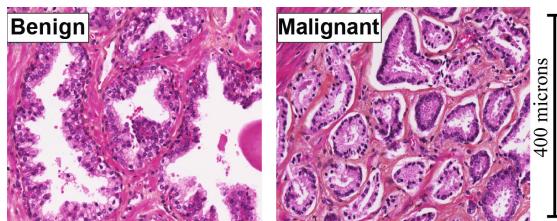


Fig. 2. Visual comparison between benign and malignant prostate glands

of this technique is that it can detect a malignant gland irrespective of the region properties. In cases where malignant glands are present in close proximity of benign glands, this approach might provide a more sensitive cancer annotation compared to the approaches that use region-dependent image features [3]. The proposed technique has been evaluated on 40 histopathology scenes extracted from 10 whole mount images with a resolution of $0.5 \mu\text{m}$ per pixel and achieved 0.83 and 0.81 of sensitivity and specificity, respectively, in a leave-75%-out cross-validation experiment.

This paper is organized as follows: the methodology of the complete gland classification algorithm is presented in Section 2. In Section 3 the experimental results of the proposed algorithm are presented. Finally, Section 4 presents concluding remarks and suggestions for future work.

2 Methodology

2.1 Segmentation of the Gland

The gland segmentation algorithm has been partially adopted from another work of Nguyen et al. [9]. In the first step, each pixel in the image is categorized into one of these four categories: i) Gland lumen, ii) epithelial layer, iii) nuclei, and iv) stroma. Small training patches of each class have been selected to train the classifier for pixel labeling. The classification is based on the color information of these histological objects in the two chromaticity layer ‘a*’ and ‘b*’ of the $L^*a^*b^*$ color space.

For pixel labeling, we have used linear discrimination analysis instead of the Voronoi tessellation based approach from [9]. The main drawback of Voronoi tessellation approach is that when the number of training samples is large, the classification time for each testing data point is very high compared to that of linear discriminant analysis [10]. Therefore, when the number of testing samples are very large the reported nearest neighborhood approach will be very expensive to compute. After having the labeled image, the complete gland units are constructed by iteratively consolidating lumen objects with the surrounding epithelial layer and nuclei pixels [9] (see Fig. 3c).

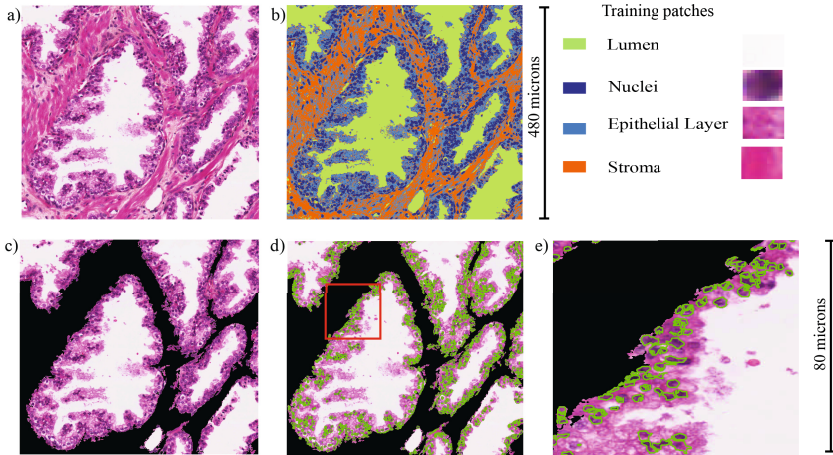


Fig. 3. Gland segmentation. a) A sample histopathology scene, b) Labeled image, c) Segmented gland units after consolidation of lumina with surrounding epithelial layer and nuclei objects, d) Segmented nuclei objects overlaying on the segmented gland units, and e) enlarged view of the segmented nuclei objects.

2.2 Segmentation of the Nuclei Units

Segmentation of nuclei is performed by employing a marker controlled watershed algorithm [11] followed by a Support Vector Machine (SVM) based object classification. The segmentation function used in the watershed algorithm is the gradient image of the inverted grayscale image of the input scene. This gradient image is modified by placing regional minima in the marked pixels of foreground and background objects of the image. The foreground markers are determined by finding the regional minima in the image after the morphological opening and closing by reconstruction operations. To determine the background marker, the same operation is performed with the inverted image. Along with the nuclei objects, the segmented objects often include some other histological objects, i.e., crystalloids inside glands, darkly stained stroma/epithelial object due to nonuniform staining and so on.

The non-nuclei objects are then filtered out by employing a SVM classifier. The features used in this classification are the mean intensity, entropy, and the standard deviation of the segmented objects. The classifier is trained using a linear kernel and manually selected training samples of nuclei and non-nuclei objects. Fig. 3d) and e) depicts the segmented nuclei objects that are part of the gland unit. Since our main focus is to extract the features related to gland morphology, the segmented nuclei units that float in the stromal region are not considered for further analysis.

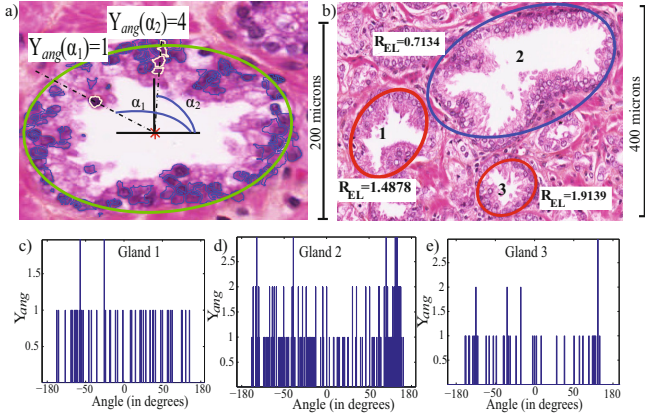


Fig. 4. a) Graphical illustration of Y_{ang} calculation. b) Sample histopathology scene with a single benign (marked with blue ellipse) and two malignant gland units (marked with red ellipses). c), d), and e) illustrate the different appearances of the histograms (Y_{ang}) of benign and malignant glands.

2.3 Extraction and Classification of the Features

Number of Nuclei Layer, N_{NL} : To determine the number of nuclei layers pertaining to each gland, at first the segmented nuclei objects are paired with the corresponding gland unit that minimizes the distance between the centroid of the nuclei and the gland lumen boundary. For each of the combined gland-nuclei object, an ellipse is fit around it. The angular location of each of the nuclei is evaluated by calculating the angle of the connecting line of the gland centroid and corresponding nuclei centroid (see Fig. 4a). Then the feature N_{NL} is evaluated from the histogram Y_{ang} of angular locations of nuclei. Customized bin spacing has been utilized to account for glands of different sizes. The bin spacing for the histogram is evaluated as $360/P_g$, where P_g is the perimeter of the corresponding ellipse surrounding the gland. Then the N_{NL} is evaluated by counting the total number of instances where multiple nuclei have same angular location in the histogram and then normalizing it by dividing by P_g . Mathematically, $N_{NL} = \frac{1}{P_g} |\{n | Y_{ang}(n) \geq 1\}|$. Fig. 4c-e illustrates the different nature of histogram, Y_{ang} in case of benign and malignant glands. As can be observed from the figure, the benign histogram provides more instances of multiple nuclei having same angular location.

Ratio of Epithelial Layer area to Lumen Area, R_{EL} : This feature is evaluated by simply taking the ratio of the epithelial layer area to lumen area of the gland. In case of malignant glands, fast multiplication of cells lead the epithelial layer to invade more in to gland lumina. As a result, the ratio gets larger in case of malignant gland units.

After the feature extraction we choose optimum thresholds on the features, $\tau_{N_{NL}}$ and $\tau_{R_{EL}}$ for the classification of benign and malignant glands. We classify a gland (G_i) as malignant when the parameters fulfill the following criteria, $Label_{G_i} = \{Malignant | N_{NL}(G_i) < \tau_{N_{NL}}, R_{EL}(G_i) > \tau_{R_{EL}}\}$. These threshold parameters are tuned by performing Receiver Operator Characteristics (ROC) analysis in a leave-75%-out experiment (discussed in the following section).

3 Experiments and Results

The proposed algorithm has been evaluated on 40 different histopathology scenes containing a total of 2145 glands. These scenes have been extracted from 10 whole mount histopathology images obtained from eight radical prostatectomy patients. These whole mount histopathology images are digitized at $20\times$ magnification ($0.5 \mu m$ per pixel) with an Aperio scanner. Cancerous regions in these images are annotated by an expert pathologist. These annotations are used as the gold standard to evaluate the performance of the proposed algorithm.

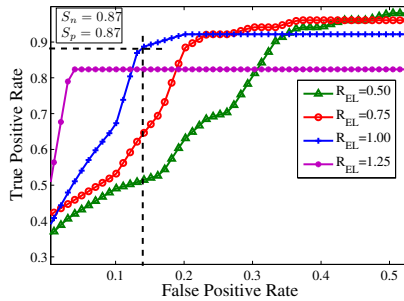


Fig. 5. a) ROC of our algorithm for four discrete values of R_{EL} . At the optimum operating point the algorithm achieves, sensitivity $S_n = 0.87$ and specificity $S_p = 0.87$.

The performance of the algorithm is influenced by the choice of the parameter value of R_{EL} and N_{NL} . We tune the parameters by performing ROC analysis on randomly selected 25% of the glands in the dataset. The ROC curve of the classifier is generated by varying the parameter N_{NL} as $\{0, 0.08, 0.16, \dots, 4\}$. To determine the effect of varying R_{EL} on the classifier performance the following operation has been performed: for each choice of R_{EL} in the set $R_{EL} = \{0, 0.25, \dots, 2.5\}$

Table 1. AUC obtained by our algorithm for different parameter values of R_{EL}

R_{EL}	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25	2.5
AUC	0.84	0.88	0.89	0.87	0.81	0.81	0.81	0.81	0.80	0.77

individual ROC curve by varying N_{NL} has been generated. We choose the thresholds $\tau_{N_{NL}}$ and $\tau_{R_{EL}}$ corresponding to the optimum operating point in the ROC curve. Then we use these thresholds on the remaining 75% of the data for gland classification. By repeating the above leave-75%-out experiment 10 times we achieved 0.83 ± 0.007 of sensitivity and 0.81 ± 0.005 of specificity. When we performed similar analysis on the entire dataset we achieved 0.87 of sensitivity and 0.87 of specificity at the optimum operating point. The corresponding values of $\tau_{R_{EL}}$ and $\tau_{N_{NL}}$ at this point are 1 and 2.24, respectively. Table 1 lists the Area Under the Curve (AUC) obtained for different values of R_{EL} . Fig. 5 illustrates the ROC curve for four different R_{EL} values .

Fig. 6 illustrates the performance of the proposed algorithm on three sample histology scenes, one entirely benign scene, one entirely malignant scene obtained from marked cancerous region by the pathologist, and one scene comprising both malignant and benign glands in close proximity of each other. Experimentally classified benign and malignant units are marked by blue and yellow ellipses, respectively. In all the examples, strong agreement between the pathologist’s marking and experimental classification of glands corroborates the effectiveness of the proposed algorithm.

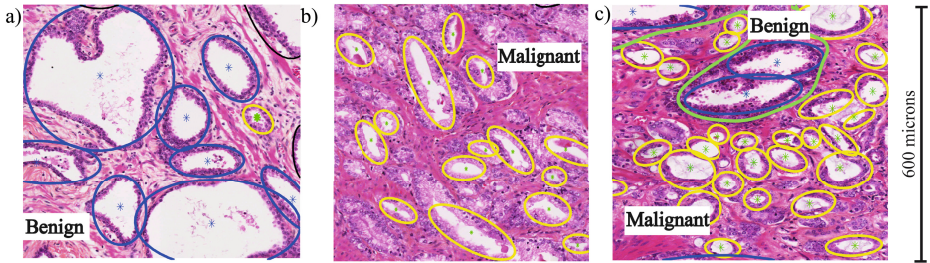


Fig. 6. Application of the proposed technique on three sample histology scenes. The yellow and blue ellipses are used to denote the malignant and benign gland units respectively. Sample scenes containing a) only benign glands, b) only malignant glands and c) both the benign and malignant glands. The green annotation mark by pathologist denotes the separation of benign and malignant glands. All the images are shown in the same scale of magnification.

4 Conclusion

In this paper, we have proposed a technique for the classification of individual benign and malignant glands based on two novel features, Number of Nuclei layers and Ratio of Epithelial layer area to Lumen area. To the best of our knowledge, this is the first work to quantify nuclei layers associated with each gland based on angular histogram. This is also the first report of individual gland labeling as malignant or benign without relying on the surrounding histology information. Since most reports on automatic cancer annotation are region based, cases with very small cancerous area might not be diagnosed by those approaches.

The application of individual gland-based technique proposed here will lead to a more sensitive cancer annotation and thus improved diagnosis of early stage prostate cancer. In the future, we plan to implement this technique on entire whole mount images. Moreover, we plan to investigate the relationship between the proposed features and long term disease progression.

Acknowledgment. Funding from the NSERC-CIHR Collaborative Health Research Projects, the BC Innovation Council NRAS Program, and a travel grant from the Institute for Computing, Information and Cognitive Systems (ICICS) at the University of British Columbia are gratefully acknowledged.

References

1. Bohring, C., Squires, T.: Cancer statistics. *CA Cancer J. Clin.* 43, 7–26 (1993)
2. 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)
3. Nguyen, K., Sarkar, A., Jain, A.K.: Structure and context in prostatic gland segmentation and classification. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part I. LNCS*, vol. 7510, pp. 115–123. Springer, Heidelberg (2012)
4. Clark, M.D., Askin, F.B., Bagnell, C.R.: Nuclear roundness factor: a quantitative approach to grading in prostate carcinoma, reliability of needle biopsy tissue, and the effect of tumor stage fore usefulness. *The Prostate* 10, 199–206 (1987)
5. Naik, S., Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A.: Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. In: *Proc. of 2nd Workshop on Micro. Image Anal. with Applications in Biology* (2007)
6. Epstein, J.I., Netto, G.J.: *Biopsy interpretation of the prostate*. Lippincott Williams & Wilkins (2007)
7. Khouzani, J.K., Zadeh, S.H.: Multiwavelet grading of prostate pathological images. In: *Proc. SPIE*, vol. 4628, pp. 1130–1138 (2002)
8. Huang, P.W., Lee, C.H.: Automatic classification for pathological prostate images based on fractal analysis. *IEEE Transactions on Medical Imaging* 28(7), 1037–1050 (2009)
9. Nguyen, K., Sabata, B., Jain, A.K.: Prostate cancer grading: Gland segmentation and structural features. *Pattern Recognition Letters* 33, 951–961 (2011)
10. Krzanowski, W.: *Principles of multivariate analysis*. Oxford Uni. Press (1996)
11. Meyer, F.: Topographic distance and watershed lines. *Signal Processing* 38(1), 113–125 (1994)

A Cross-Sectional Piecewise Constant Model for Segmenting Highly Curved Fiber Tracts in Diffusion MR Images

Brian G. Booth and Ghassan Hamarneh

Medical Image Analysis Lab, Simon Fraser University, Burnaby, Canada
bgb2@sfu.ca, hamarneh@sfu.ca

Abstract. We propose a cross-sectional piecewise constant model for the segmentation of highly curved fiber tracts in diffusion MRI scans. An “anchor curve”, obtained via tractography, provides the overall shape of the tract and allows us to examine the tract’s microstructure at the level of cross-sectional planes normal to the curve. Each cross-section is modeled as a piecewise constant image, allowing us to address changes in measured diffusion due to the curving of the tract while still capturing overall tract structure. Results on both synthetic and real data show improved segmentation quality compared to state-of-the-art methods, particularly in areas of crossing fibers.

1 Introduction

In diffusion MRI (dMRI), segmentation is often used to delineate axonal fiber tracts connecting functional brain regions [12]. Initial attempts to segment fiber tracts focused around performing streamline tractography, then defining the segment as the set of voxels that contain the streamlines (e.g., [16]). However, the goal of tractography is to capture a tract’s direction and orientation, not its width. As a result, tractography cannot capture fine details along the surface of a fiber tract, routinely leading to under-segmentation [3]. Instead of relying on a collection of 3D streamlines with an unclear encapsulating surface, segmentation algorithms that label the underlying 3D image domain are preferred for defining a tract’s volumetric region. Among these volumetric dMRI segmentation algorithms, many assume a piecewise-constant model of the image [9, 11]. However, given the fact that dMRI data contains tract orientation information, the success of these piecewise constant approaches is limited to tracts that have little curvature (e.g., corpus callosum [9]).

Segmenting highly curved tracts in dMRI scans requires extending segmentation techniques to handle a tract’s variable appearance. This can be done by either increasing the complexity of the image model (e.g., piecewise smooth [17]) or by pre-processing the dMRI scan so that a simpler segmentation model can be applied effectively. The latter approach has been more popular over the past decade with examples including segmentation based on pre-computed edge information [10] and clustering voxels using local statistics pre-computed from Parzen

windows [2]. More recently, tractography results have been used to provide global tract shape information as input to the segmentation process [3, 13, 15], allowing for the pre-processing of an image based on the orientation of the tract. This global shape information may well complement the local appearance information obtained using Parzen windowing or edge maps, yet, individually, these approaches are limited by either susceptibility to noise or lack of fidelity between the data and the image model [13].

We propose that a hybrid approach, where local appearance information is combined with global shape information, can show increased segmentation accuracy for diffusion MR images. We base this hybrid algorithm on the assumption that a tract’s cross-section (i.e., the plane perpendicular to the tract’s local direction) shows relatively constant diffusion compared to its surroundings. The piecewise constancy assumption is then justifiably applied only at a local scale while, at the global scale, cross-sectional planes are defined based on an “anchor curve” obtained from tractography. Results on both synthetic and real data show improved segmentation quality compared to state-of-the-art methods [2, 9–11, 15], particularly in areas of crossing fiber tracts.

2 Methods

Figure 1 displays the general workflow of our segmentation algorithm. Like in the work of Niethammer et al. [15], we begin by generating an anchor curve from a tractography algorithm. We also end by segmenting a simplified version of the diffusion MR image. Where we differ is in how we generate that simplified image. In [15], Niethammer et al. reorient tensors by the curvature of the anchor

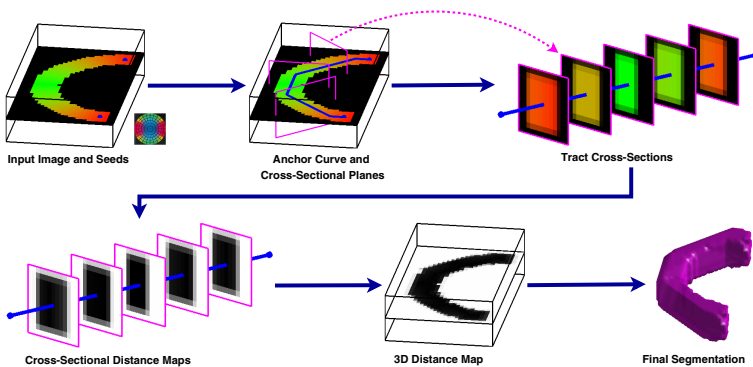


Fig. 1. Our proposed segmentation workflow. Tractography is employed to generate an anchor curve (blue) which is then used to generate cross-sections of the fiber bundle (magenta). For each cross-section, we measure diffusion dissimilarities between the points on the plane and the intersection point between the plane and the anchor curve. These dissimilarities are then interpolated back into a 3D image and a scalar segmentation algorithm provides us with the final segmentation.

curve, then assume a global piecewise-constant image model. We instead assume a piecewise-constant model only on cross-sectional planes normal to the anchor curve. The following subsections describe how we employ that model to simplify and segment diffusion MR images.

Anchor Curve Generation: Given a dMR image $\mathcal{I} : \Omega \rightarrow \mathcal{M}$ that maps a point \mathbf{x} in our image space $\Omega \subset \mathbb{R}^3$ to a diffusion representation (e.g., tensor, ODF) on a manifold \mathcal{M} , we can generate an anchor curve $\mathbf{r} : [0, 1] \rightarrow \Omega$ from various tractography algorithms. In this work, we employ the minimal path tractography algorithm of Zalesky [18] to generate \mathbf{r} . Edge weights for the graph used by Zalesky’s algorithm were computed analytically using [6].

Obtaining Tract Cross-Sections: For each point s along the given anchor curve \mathbf{r} , we compute the Frenet frame defined by the curve’s local tangent \mathbf{T} , normal \mathbf{N} , and binormal \mathbf{B} vectors

$$\mathbf{T} = \frac{\partial \mathbf{r}}{\partial s} \quad \mathbf{N} = \frac{\frac{\partial \mathbf{T}}{\partial s}}{\|\frac{\partial \mathbf{T}}{\partial s}\|} \quad \mathbf{B} = \mathbf{T} \times \mathbf{N}. \quad (1)$$

The resulting normal and binormal vectors span (and parameterize) the cross-sectional plane normal to \mathbf{r} . For points s_i on the anchor curve \mathbf{r} where $\partial \mathbf{T} / \partial s = 0$, we use the closest stable Frenet frame of \mathbf{r} rotated so that its tangent vector aligns with the tangent vector at s_i .

Given the local Frenet frame $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ and a point s on the anchor curve \mathbf{r} , we generate points $\mathbf{x} \in \Omega$ on the cross-sectional plane by sampling along the normal and binormal vectors:

$$\mathbf{x} = \mathbf{r}(s) + u * \mathbf{N} + v * \mathbf{B}. \quad (2)$$

The diffusion representation (e.g., tensor, ODF) at \mathbf{x} – and correspondingly (u, v) in the cross-sectional image space Φ_s – is linearly interpolated from the original image \mathcal{I} . This procedure produces our cross-sectional images $I_s : \Phi_s \rightarrow \mathcal{M}$ and the corresponding 3D coordinates of each cross-sectional image pixel $\Pi_s : \Phi_s \rightarrow \Omega$.

Cross-Sectional Piecewise Constancy: Our approach is based on the assumption that the diffusion data within a cross-section of a fiber bundle can be well-modeled using a piecewise-constant function. Given that a fiber tract is a collection of coherently aligned axons, we expect the diffusion within the cross-section of the tract to be similar to that at the plane’s intersection with the anchor curve. Meanwhile, we expect diffusion on that cross-sectional plane but outside the fiber tract to be different from that at the plane’s anchor curve intersection point. As a result, computing dissimilarities between the diffusion data on the anchor curve $\mathcal{I}(\mathbf{r}(s))$ and the diffusion data throughout the cross-section I_s will provide us with a scalar feature D_s that will correlate well with fiber bundle membership.

Various dissimilarity metrics can be employed, including those for second order tensors (e.g., [1]), 4th-order tensors [5], and spherical harmonic ODF representations [8]. Given a chosen metric $d(\cdot)$, we employ the following mapping

$$D_s(u, v) = \log \left(\frac{d(I_s(u, v), \mathcal{I}(\mathbf{r}(s)))}{FA(\mathcal{I}(\mathbf{r}(s)))} + \epsilon \right) \quad (3)$$

which applies a log mapping to the dMRI dissimilarities and normalizes them by the fractional anisotropy (FA) of the point intersecting the anchor tract. We found, following empirical examination, that the log-mapping leads to more Gaussian-distributed dissimilarities inside and outside the tract of interest while dividing by the anchor curve's FA helps normalize the range of dissimilarities from one cross-section to another. Note that since the anchor curve was obtained via tractography, its FA will be greater than zero.

Mapping Dissimilarities to the Image Space: Once we have the tensor distance feature D_s computed from (3) for a collection of points Π_s defined by (2), we proceed with reconstructing a 3D distance image. This task is a basic scattered data interpolation problem and we employ an approach based on radial basis functions and k-nearest neighbours. The interpolated 3D dissimilarity map \mathcal{D} is given as

$$\mathcal{D}(\mathbf{x}) = \sum_{i=1}^k \frac{\exp(-\|\Pi_{s(i)}(u(i), v(i)) - \mathbf{x}\|)}{\sum_{j=1}^k \exp(-\|\Pi_{s(j)}(u(j), v(j)) - \mathbf{x}\|)} D_{s(i)}(u(i), v(i)) \quad (4)$$

where $(u(i), v(i))$ in cross section $s(i)$ is the i^{th} nearest neighbour to \mathbf{x} in Ω .

Dissimilarity Map Segmentation: Using the local diffusion dissimilarities from (3), we have reduced our dMR image, with its variable region appearance and manifold-valued data, to a scalar image that is well modeled by a piecewise constant function. As a result, it now makes sense to employ a piecewise constant segmentation algorithm. We use a probabilistic variant of the Chan-Vese segmentation algorithm that minimizes

$$E(\mathcal{S}, \mu_{in}, \sigma_{in}, \mu_{out}, \sigma_{out}) = \alpha \int_{\partial \mathcal{S}} d\mathbf{x} + \beta \int_{\mathbf{x} \in \mathcal{S}} -\log(p(\mathbf{x}|\mu_{in}, \sigma_{in}))d\mathbf{x} + \beta \int_{\mathbf{x} \notin \mathcal{S}} -\log(p(\mathbf{x}|\mu_{out}, \sigma_{out}))d\mathbf{x} \quad (5)$$

where $\mathcal{S} \subset \Omega$ is the segmentation, $\mu_{in}, \sigma_{in}(\mu_{out}, \sigma_{out})$ represent the mean and standard deviation of distances D inside (outside) \mathcal{S} , and weights α, β regulate the trade-off between the contour regularization and image fidelity terms. We optimize the segmentation energy in (5) using the total variational approach of Bresson et al. [7]. Note that this is the same optimization scheme used by the competing approach of Niethammer et al. [15] though they use it to segment images of diffusion tensor primary eigenvectors. As with the approach of Niethammer et al., all voxels a distance greater than d_{max} from the anchor curve are set to belong to the background while the foreground segment containing the anchor curve is taken as the final segmentation.

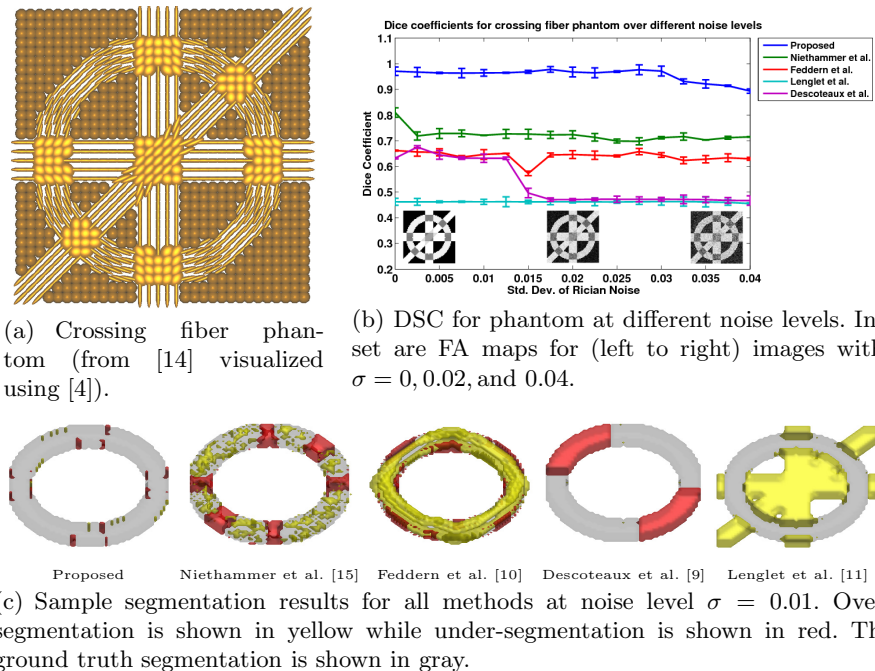


Fig. 2. Segmentation results for the ring tract in (a). Note that we obtain significantly higher Dice coefficients than competing methods as we are able to better model curved tracts and fiber crossings. Further, our approach generates consistent results across various noise levels.

3 Experimental Setup and Results

To evaluate the effectiveness of our segmentation approach, we perform two quantitative experiments, one on the synthetic phantom presented in [14] and another on 18 cingulum bundles from dMRI scans from the IXI database¹. In both cases, resulting segmentations were compared to expert-drawn manual segmentations using the Dice similarity coefficient (DSC). The segmentation algorithms from [9–11, 15] are used as comparison methods. In all cases, $k = 5$, $\epsilon = \exp(-4)$ and the log-Euclidean distance metric was used to compute the dissimilarity maps [1]. To ensure fairness of comparison between segmentation algorithms, we optimize the weights of all energy terms in all segmentation algorithms (e.g., α, β) using genetic algorithms. Results are shown for the weights that produce the maximum DSC.

Phantom Experiment: Figure 2(a) displays the middle slice of the synthetic phantom from [14]. We seek to segment the ring tract in the phantom in order to test our algorithm’s ability to handle both tract curvature and crossing regions. We further test the impact of image noise by adding Rician noise of different

¹ <http://brain-development.org>

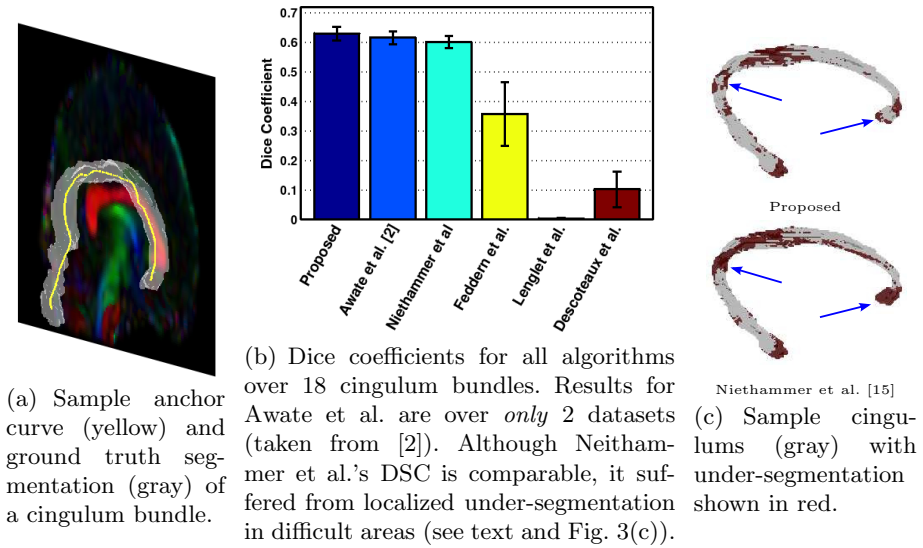


Fig. 3. Results on the segmentation of cingulum bundles from real dMRI scans. A sample is shown in (a). Note that we obtain significantly higher Dice coefficients than competing methods (largest $p = 0.0298$). For the methods that were able to segment the cingulum, we were better able to reduce under-segmentation as highlighted by the blue arrows in (c).

magnitudes to the phantom. Twenty-five noisy images are generated for each noise level and all competing segmentation methods are applied.

Figure 2(b) shows the resulting DSC for each algorithm and noise level. Note that our approach significantly outperforms those algorithms presented in [9–11, 15]. The reasons for this improvement can be seen in Figure 2(c). The approach of Niethammer et al. [15], which assumes a piecewise constant image after rotating tensors to the anchor curve’s Frenet frame, has difficulty segmenting the crossing regions where the global piecewise constant assumption does not hold. However, our assumption of cross-sectional piecewise constancy still holds in these regions, resulting in a more accurate segmentation. Further, Niethammer et al. rely only on the primary eigenvector for segmentation, leading to over-segmentation leaking into isotropic regions in which the primary eigenvector may align with those within the segment of interest. Meanwhile, the geodesic active contours approach of Feddern et al. [10] is limited by poor edge information due to image noise. Finally, the piecewise constant segmentation approaches of Descoteaux et al. [9] and Lenglet et al. [11] (the latter of which includes a geodesic active contour edge term to the segmentation energy) fail to model the tensor image appropriately, leading to poor segmentations. Our approach avoids these problems by applying the piecewise constant image model on a per cross-section basis.

Real Data Experiment: We employ 18 expertly-drawn manual segmentations of cingulum bundles from 9 dMRI scans as ground truth segmentations to test the accuracy of our algorithm on real data. Figure 3(a) shows a representative

example of the region of interest and its corresponding anchor curve. Note that the curvature of the cingulum makes a piecewise constant function a poor choice for modelling diffusion across the whole bundle.

Figure 3(b) shows the resulting DSC for all 18 cingulum segmentations. Note that our proposed approach performs better than the competing methods, even beating the average 0.615 DSC reported on comparable data for *only two* cingulum bundles in [2]. The piecewise constant approaches in [9, 11] failed to segment the cingulum. Instead, the segmentation leaked and delineated the corpus callosum seen in red in Figure 3(a). Meanwhile, the geodesic active contours approach of Feddern et al. [10] showed difficulty dealing with noisy edge information, leading to over-segmentation. The closest competing method to ours is that of Niethammer et al. [15], one sample of which is shown in Figure 3(c). Although the DSC values are somewhat comparable (Fig. 3(b)), the increased accuracy of our method was significant ($p = 0.0298$). Further, our approach showed consistent (i.e., over all 18 tracts) reduction in the amount of under-segmentation compared to [15]. This reduction, likely due to our use of a more localized image appearance model, was most pronounced around the genu and splenium (as highlighted by the blue arrows). Quantitatively, we observed a significant reduction of 10.48% in the number of under-segmented voxels (Niethammer et al. [15]: $\mu = 1306.8$ voxels, $\sigma = 362.8$ voxels; Proposed: $\mu = 1169.8$ voxels, $\sigma = 381.9$ voxels. $p = 0.018$).

4 Conclusion

We proposed herein a cross-sectional piecewise constant model for diffusion MRI segmentation, allowing us to combine local diffusion information with global shape information. Using an “anchor curve” obtained via tractography, we are able to generate cross-sections of the tract and apply the piecewise constant model at that local level. We have shown that the resulting segmentation algorithm is better capable of handling curved tracts and crossing regions than many competing methods [2, 9–11, 15]. Future work will focus on determining whether these results remain consistent if we change the diffusion model (e.g., from tensor to ODF).

Acknowledgements. BGB was supported by both IODE Canada and the Government of Alberta. GH was partially supported by NSERC. This research has been enabled by the use of computing resources provided by WestGrid and Compute/Calcul Canada.

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Mag. Res. Med.* 56, 411–421 (2006)
2. Awate, S.P., Zhang, H., Gee, J.C.: A fuzzy, nonparameteric segmentation framework for DTI and MRI analysis with applications to DTI-tract extraction. *IEEE Trans. Med. Imag.* 26(11), 1525–1536 (2007)

3. Barbieri, S., Bauer, M.H., Klein, J., Nimsky, C., Hahn, H.K.: A variational, non-parametric approach to the fuzzy segmentation of diffusion tensor images. In: Proceedings of MICCAI Workshop on Computational Diffusion MRI (CDMRI), pp. 134–145 (2010)
4. Barmpoutis, A., Vemuri, B.C., Shepherd, T.M., Forder, J.R.: Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat hippocampi. *IEEE Trans. Med. Imag.* 26(11), 1537–1546 (2007)
5. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth order tensor field estimation from DW-MRI. *NeuroImage* 45, S153–S162 (2009)
6. Booth, B.G., Hamarneh, G.: Exact integration of diffusion orientation distribution functions for graph-based diffusion MRI analysis. In: Proceedings of IEEE ISBI, pp. 935–938 (2011)
7. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J., Osher, S.: Fast global minimization of the active contour/snake model. *J. Math. Imag. Vis.* 28(2), 151–167 (2007)
8. Chiang, M.-C., Klunder, A.D., McMahon, K., de Zubicaray, G.L., Wright, M.J., Toga, A.W., Thompson, P.M.: Information-theoretic analysis of brain white matter fiber orientation distribution functions. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007. LNCS*, vol. 4584, pp. 172–182. Springer, Heidelberg (2007)
9. Descoteaux, M., Deriche, R.: High angular resolution diffusion MRI segmentation using region-based statistical surface evolution. *J. Math. Imag. Vis.* 33(2), 239–252 (2009)
10. Feddern, C., Weickert, J., Burgeth, B.: Level-set methods for tensor-valued images. In: Proc. of IEEE Workshop on Geometric and Level Set Methods in Computer Vision, pp. 65–72 (2003)
11. Lenglet, C., Rousson, M., Deriche, R.: DTI segmentation by statistical surface evolution. *IEEE Trans. Med. Imag.* 25(6), 685–700 (2006)
12. de Luis-Garcia, R., Alberola-Lopez, C., Westin, C.F.: Segmentation of Tensor Fields: Recent Advances and Perspectives. In: *Tensors in Image Processing and Computer Vision*. Springer, London (2009)
13. Melonakos, J., Niethammer, M., Mohan, V., Kubicki, M., Miller, J., Tannenbaum, A.: Locally-constrained region-based methods for DW-MRI segmentation. In: Proceedings of ICCV, pp. 1–8 (2007)
14. Nazem-Zadeh, M.R., Davoodi-Bojd, E., Soltanian-Zadeh, H.: Level set fiber bundle segmentation using spherical harmonic coefficients. *Computerized Medical Imaging and Graphics* 34, 192–202 (2010)
15. Niethammer, M., Zach, C., Melonakos, J., Tannenbaum, A.: Near-tubular fiber bundle segmentation for diffusion weighted imaging: Segmentation through frame reorientation. *NeuroImage* 45, S123–S132 (2009)
16. O'Donnell, L., Westin, C.-F.: White matter tract clustering and correspondence in populations. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005, Part I. LNCS*, vol. 3749, pp. 140–147. Springer, Heidelberg (2005)
17. Vemuri, B.C., Liu, M., Amari, S.I., Nielsen, F.: Total Bregman divergence and its applications to DTI analysis. *IEEE Trans. Med. Imag.* 30(2), 475–483 (2011)
18. Zalesky, A.: DT-MRI fiber tracking: A shortest paths approach. *IEEE Trans. Med. Imag.* 27(10), 1458–1571 (2008)

Improving DTI Resolution from a Single Clinical Acquisition: A Statistical Approach Using Spatial Prior

Vikash Gupta, Nicholas Ayache, and Xavier Pennec

INRIA Sophia Antipolis, ASCLEPIOS Project

Abstract. Diffusion Tensor Imaging (DTI) provides us with valuable information about the white matter fibers and their arrangement in the brain. However, clinical DTI acquisitions are often low resolution, causing partial volume effects. In this paper, we propose a new high resolution tensor estimation method. This method makes use of the spatial correlation between neighboring voxels. Unlike some super-resolution algorithms, the proposed method does not require multiple acquisitions, thus it is better suited for clinical situations. The method relies on a *maximum likelihood* strategy for tensor estimation to optimally account for the noise and an anisotropic regularization prior to promote smoothness in homogeneous areas while respecting the edges. To the best of our knowledge, this is the first method to produce high resolution tensor images from a single low resolution acquisition. We demonstrate the efficiency of the method on synthetic low-resolution data and real clinical data. The results show statistically significant improvements in fiber tractography.

Keywords: DTI, Super-resolution, Partial volume effect.

1 Low Spatial Resolution in Clinical DTI

DTI is a non-invasive imaging modality used for mapping the diffusion of water molecules in the brain. The diffusion process depends on the tissue microstructure and thus makes imaging white matter (WM) fibers and their organization in the brain feasible. It can be used to study the brain connectivity through a tractography analysis. In the clinics, DTI is used to study WM disorders. However, clinical DTI acquisitions have usually low spatial resolution (typically 2.0 - 5.5 mm in each direction). In practice, it is often desirable to visualize and gather fiber structure information from high resolution (HR) images. But, acquiring HR diffusion weighted images (DWI) requires either longer acquisition time or using scanners with stronger magnetic fields (7 T and 11 T), compared to the ones used in a clinical setting (1.5 T and 3.0 T). Both of these options are not suitable for clinical scenarios. So, resampling the low resolution DTI to a high resolution space is currently the only option available to clinicians. However, the present resampling techniques do not account for partial volume (PV)

effects present in the clinical images at native resolution. The PV effects in the DWI leads to an underestimation of tensors adversely affecting fiber tractography [1]. In this work, we propose a tensor estimation method, which takes into account the partial volume effect and yields a more spatially coherent tensor field. The efficacy of the method is established through significant improvements in tractography.

Super-Resolution Reconstruction

One of the first super-resolution algorithms for MRI proposed by Peled et al. uses a combination of spatially shifted single shot diffusion weighted images to create HR images [2]. This method uses eight repeated low resolution scans with a shifted field of view and different b-values each time. In another study [3], the volume acquisition is spatially shifted in the slice-selected direction and inter slice reconstruction is achieved using the Irani-Peleg’s [4] back-projection method. But this method was not extended for DTI studies. More recently, super-resolution on diffusion weighted images using multiple anisotropic orthogonal DWI scan has been proposed [5]. All these methods of super-resolution rely on multiple acquisitions of the same subject. While this may be an acceptable practice in a research environment, this would be difficult and even undesirable in a clinical setting. Unlike previous methods, we propose a HR tensor reconstruction algorithm which does not require multiple acquisitions. Our method also accounts for the PV in the low resolution DWIs, producing DTI at higher resolutions. The proposed method can be related to the regularized super-resolution method of [6], which recovers WM fiber configuration based on volume fraction of anisotropic fiber population in the neighboring voxels. However, we use anisotropic regularization to recover the full tensor information. The method is described in section 2. Experiments performed on simulated low resolution acquisitions in section 3.1 show that the fractional anisotropy (FA) image computed using HR tensor estimation is better contrasted and less blurry compared to the one computed using tensor resampling. A comparison of FA images on real data and quantitative evaluation on fiber tractography is presented in sections 3.2 and 3.3 respectively.

2 Statistical Tensor Reconstruction with Spatial Prior at Any Resolution

The diffusion tensor D is related to each DWI S_i corresponding to the encoding gradient g_i and the image with null gradient S_0 using the Stejskal-Tanner Equation [7], $S_i = S_0 e^{-bg_i^T D g_i}$ where b is the diffusion factor. D is a second order tensor (i.e., a symmetric positive definite matrix). We build our tensor estimation model based on variational methods for joint estimation and smoothing of DTI [8] in the Log-Euclidean framework.

2.1 A Discrete Signal Degradation Model: From High Resolution to Low Resolution Images

A voxel in the low resolution (LR) image x_k is composed of the voxels y_j in the HR image. The observed signal $S^{LR}(x_k)$ in the voxel x_k is the weighted sum of the intensities of the voxels y_j :

$$S^{LR}(x_k) = \sum_j \alpha_{kj} S^{HR}(y_j),$$

where $S^{HR}(y_j)$ is the signal intensity in the voxel y_j of the HR image; α_{kj} is the partial volume overlap of the HR voxel y_j in the LR voxel x_k . Based on the above model, each LR gradient image, S_i^{LR} can be written in terms of the underlying HR tensors (D^{HR}) and the S_0^{HR} (image with null gradient in high resolution):

$$S_i^{LR}(x_k) = \sum_j \alpha_{kj} S_0^{HR}(y_j) e^{-bg_i^T D^{HR}(y_j) g_i}. \quad (1)$$

For simplicity, we write S_i^{LR} as S_i and expand the term using Equation 1, when necessary.

2.2 Tensor Estimation

With the above image degradation model, we solve the inverse problem of estimating tensors in the HR space. This is an ill-posed inverse problem because for a given signal $S_i(x_k)$ there is an infinite number of combinations of $D(y_j)$ possible in Equation 1. In the Log-Euclidean framework, we define L as the matrix logarithm of D : $L = \text{Log}(D)$. Following [8], the tensor estimation can be looked upon as a variational formulation, i.e., one should minimize the following energy functional,

$$E(S_0, L) = \frac{1}{2} \text{Sim}(S_0, L) + \frac{\lambda_s}{2} \text{Reg}(S_0) + \frac{\lambda_L}{2} \text{Reg}(L), \quad (2)$$

where $\text{Sim}(\cdot)$ is the data fidelity term and $\text{Reg}(S_0)$, $\text{Reg}(L)$ are the regularization priors on S_0 and L images with the respective weights λ_s and λ_L . In [8], the tensor estimation is performed at the same resolution as that of acquisition. However, in this paper, we aim to estimate tensors at a resolution higher than the acquisition resolution. The LR acquired signal and the HR estimated tensors are linked through the degradation model described in Equation 1. Thus, the solution of the energy minimization problem is an optimal tensor field, which is spatially coherent and is consistent with the observed signal.

2.3 Similarity Criteria for Tensor Estimation

The observed signal in any DWI can be modeled as $\tilde{S}_i = S_i + \eta$, where \tilde{S}_i is the observed DWI, S_i is the true image (as in Equation 1) and η is the associated

noise model. As described in [8], the data attachment term $E(\cdot)$ corresponds to a maximum likelihood estimator (MLE) adapted to the noise model. The MLE for the probability density function of a given noise model is,

$$\text{Sim}_{ML}(\cdot) = - \sum_{i=0}^N \sum_{x_k} \log \left[p(\tilde{S}_i(x_k) | S_i(x_k)) \right], \quad (3)$$

where $S_i(x_k)$ is the same as in Equation 1 and $p(\cdot)$ is the probability density function (pdf) of the the noise model. Unlike [8], $E(\cdot)$ is a function of two variables, the HR image with null gradient, $S_0(y_j)$ and the logarithm of HR tensor field, $L(y_j)$. The maximization is achieved using the steepest descent algorithm. The gradient of $\text{Sim}_{ML}(S_0(y_j), L(y_j))$ by differentiating Equation 3 is,

$$\nabla \text{Sim}_{ML} = \left[\frac{\partial \text{Sim}_{ML}}{\partial S_0(y_j)} \quad \frac{\partial \text{Sim}_{ML}}{\partial L(y_j)} \right]^T. \quad (4)$$

Taking partial derivatives of Equation 3 with respect to $S_0(y_j)$ and $L(y_j)$,

$$\frac{\partial \text{Sim}_{ML}}{\partial S_0(y_j)} = - \sum_{i=0}^N \sum_{x_k} r(\cdot) \frac{\partial \tilde{S}_i(x_k)}{\partial S_0(y_j)}; \quad \frac{\partial \text{Sim}_{ML}}{\partial L(y_j)} = - \sum_{i=0}^N \sum_{x_k} r(\cdot) \frac{\partial \tilde{S}_i(x_k)}{\partial L(y_j)},$$

where $r(\cdot) = p'(\cdot)/p(\cdot)$ and $p'(\cdot)$ is the derivative of $p(\cdot)$ with respect to $\tilde{S}_i(x_k)$. Differentiating Equation 1, we get

$$\begin{aligned} \frac{\partial \tilde{S}_i(x_k)}{\partial S_0(y_j)} &= \alpha_{kj} \exp[-bg_i^T \exp(L(y_j))g_i]; \\ \frac{\partial \tilde{S}_i(x_k)}{\partial L(y_j)} &= -b \alpha_{kj} S_0(y_j) \exp[-bg_i^T \exp(L(y_j))g_i] \partial_{G_i}[\exp(L(y_j))], \end{aligned}$$

where $\partial_{G_i}[\exp(L(y_j))] = \partial[g_i^T \exp(L(y_j))g_i]/\partial L(y_j)$ is the directional derivative of the matrix exponential. A detailed implementation for computing $\partial_{G_i}[\exp(L(y_j))]$ is available in [8]. In the following sections, we discuss the MLE with Gaussian noise model followed by brief implementation details.

MLE with Gaussian Noise. Assuming Gaussian noise of variance σ^2 on the LR image, the pdf of the observed signal \tilde{S}_i knowing the expected signal S_i is $p(\tilde{S}_i | S_i)$. Using Equation 3, $r(\tilde{S}_i(x_k) | S_i(x_k))$ can be computed as,

$$p(\tilde{S}_i | S_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\tilde{S}_i - S_i)^2}{2\sigma^2}\right); \quad r(\tilde{S}_i | S_i) = \frac{S_i - \tilde{S}_i}{\sigma^2}, \quad (5)$$

Alternatively, the MLE with Rician noise can be computed as in [8].

2.4 Non-stationary Spatial Prior

Without any additional constraints, many HR tensor images could explain the observed low resolution DWIs. Among all these solutions, anisotropic regularization promotes smooth images in homogeneous areas while respecting the

edges. The anisotropic behavior is achieved using a ϕ -functional, i.e., $\text{Reg}(s) = \int_{\Omega} \phi(\|\nabla s\|)$, where s is the intensity (scalar or tensor) of the voxel of image Ω . In our implementation, we use $\phi(s) = 2(1 + s^2/\kappa^2) - 2$, with κ as the image normalization factor [8]. The gradient of $\text{Reg}(s)$ is

$$\nabla \text{Reg}(s) = -2\psi(\|\nabla s\|)\Delta s - 2\nabla^T(\psi(\|\nabla s\|))\nabla s, \quad (6)$$

where $\psi(s) = \phi'(s)/s$. The anisotropic regularization can also be viewed as a spatial Markov random field (MRF) where the state (intensity) of the voxel is dependent on the states of the neighboring voxels. The anisotropic smoothing of posterior probabilities is equivalent to the Maximum A Posteriori (MAP) solution of a discrete MRF, making the full criterion in Equation 2 a MAP estimator.

The proposed HR tensor estimation method is similar to the one proposed in [8], in terms of the variational formulation. However, the total energy $E(\cdot)$ is constrained through the signal degradation model in Equation 1. The novelty of the method lies in the fact that it estimates the tensors in HR while taking into account the unavoidable partial volume effects during acquisition, to produce a spatially coherent tensor field.

2.5 Implementation

The total energy $E(S_0, L)$ is minimized using the steepest descent method with line search algorithm.

$$S_0^{t+1} = S_0^t - \frac{dt_s}{2} \left[\frac{\partial \text{Sim}}{\partial S_0} + \lambda_s \nabla \text{Reg}(S_0) \right]; \quad L^{t+1} = L^t - \frac{dt_L}{2} \left[\frac{\partial \text{Sim}}{\partial L} + \lambda_L \nabla \text{Reg}(L) \right],$$

where dt_s and dt_L are the step sizes for steepest descent. The two images S_0 and L are optimized alternately, until convergence. The step-sizes are reduced by half if the total variational energy is not decreased, until the step-size are too small ($dt_s < 10^{-6}$, $dt_L < 10^{-10}$), that is the total energy E cannot be minimized any further. The algorithm is implemented in C++.

3 Comparison between Tensor Resampling and HR Tensor Estimation

3.1 Simulated LR Acquisition

Data were acquired using a GE 1.5 T scanner, with 24 encoding gradient directions (with a b-value of 700 s/mm^2). The image has $256 \times 256 \times 26$ voxels of size $0.9375 \times 0.9375 \times 5.5 \text{ mm}^3$. The images are downsampled by a factor of two in the axial plane and the tensor field is estimated. This LR tensor field is then resampled to the original image size using Log-Euclidean interpolation [9]. The HR tensor estimation is used to estimate tensors at the original resolution from the downsampled images. In Fig. 1, the fractional anisotropy (FA) maps from both tensor resampling and HR estimation methods are shown. The FA map computed from the HR tensor estimation is better contrasted and less blurry when compared with the one computed from resampling the LR tensor image.

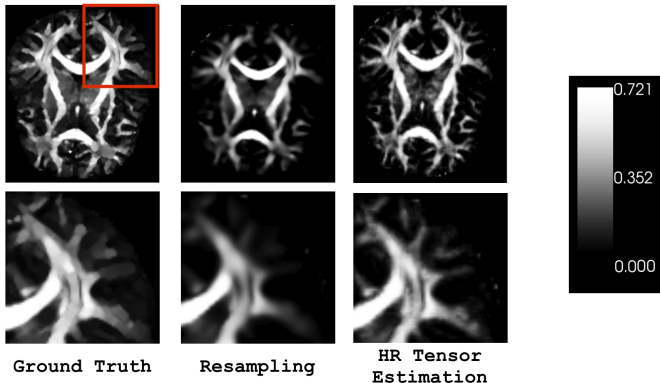


Fig. 1. Simulated LR acquisition. Top row: axial views of the ground truth image, FA computed from resampled tensor field and HR tensor estimation. The FA map computed from the HR tensor estimation method is better contrasted than the one computed from the resampled tensor field. Bottom row: zoomed region (red square).

3.2 Increase in FA on Real Data

In this section, we use a real dataset acquired with a Siemens 1.5 T scanner, with 21 encoding gradient directions and a b-value of 1000 s/mm^2 . The images have $80 \times 80 \times 40$ voxels of size $3 \times 3 \times 3.3 \text{ mm}^3$. The tensor field is estimated at the native resolution and is resampled using the Log-Euclidean framework. Several WM tracts which were not visible in the resampled images can easily be seen in the FA map, when the FA is computed from HR tensor estimation method. The HR tensor estimation algorithm presented in this paper shows an increase in FA values in the white matter regions. In Fig. 2, the FA maps are overlaid on the corresponding B0 image for better anatomical reference. In both the axial and coronal views, the external capsule and the corpus callosum can be clearly delineated with the HR tensor estimation method. We observed an increase in FA values by 43% in the WM regions (regions marked with arrows in Fig. 2) with the proposed method.

3.3 Influence on Tractography: Quantitative Evaluation

The same data set as in section 3.1 is used for a fiber tracking experiment. However, in this case the HR tensor estimation is done at 1 mm isotropic resolution and the tensor field is resampled to the same isotropic resolution. The fiber-tracking is done using the MedINRIA (www.med.inria.fr) 1.90 DTI-Track tool. All the voxels with $\text{FA} > 0.25$, are considered as seed voxels for fiber tracking. The proposed method shows a statistically significant increase in the fiber lengths, when compared with tensor resampling. A one tailed t-test on fiber lengths gives a p-value less than 2×10^{-16} . The increase in FA values in WM regions as described in section 3.2 leads to an 82% increase in the number of

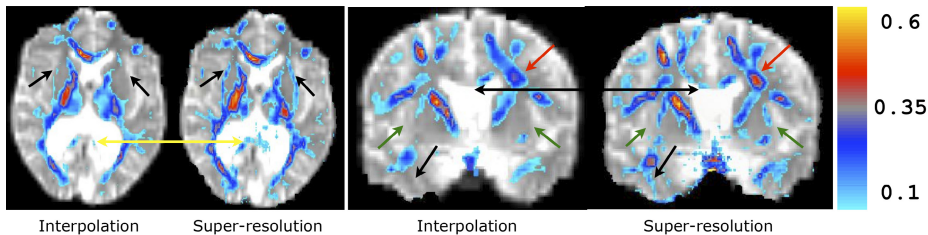


Fig. 2. Influence on FA in real clinical data: The axial and coronal views show that the external capsule and part of corpus callosum can be delineated using the HR tensor estimation. The corresponding arrows show regions with considerable increase in FA.

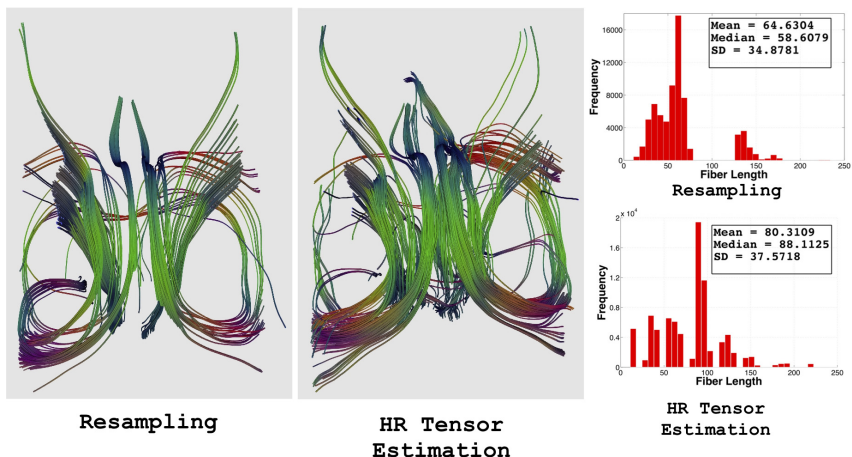


Fig. 3. Middle column shows comparatively denser fiber bundle in the fornix region for the HR tensor estimation method (superior-inferior view) compared to tensor resampling (left column). Right column: a quantitative comparison of fiber lengths.

seed voxels for HR tensor estimation and thus denser fibers. The fornix of the brain is tracked and the results are compared in Fig. 3 for tensor resampling (left column) and HR tensor estimation (middle column). There is a 25% increase in the mean length of fibers with the proposed tensor estimation method compared to tensor resampling. The right column in Fig. 3 shows the histogram for fiber length distribution.

4 Conclusions

In this paper, we tackled the problem for resampling low resolution DTI data on higher resolution for tractography or statistical analysis purposes at the population level. We propose to replace the resampling step by a DTI reconstruction

at the high resolution using a MAP estimator with a spatial prior. The method is compared with tensor resampling method on simulated low resolution data as well as on real clinical data. Results showed better contrasted and less blurry FA maps on the simulated data. We confirmed our results with experiments on the Fiber Cup [10] which shows a 6.39 % reduction in mean error in the principal diffusion direction. On the real data, a considerable increase in FA is observed, making some WM regions in parts of corpus callosum and external capsule recognizable. Fiber tracking with the HR tensor estimation shows statistically significant increase in the length and number of fibers. In the future, we would like to investigate if a better statistical power can be achieved in population based studies using the proposed method.

Acknowledgements. We would like to acknowledge Nice University Hospital (CHU) and Dr. Jean-Jacques Lemaire from Clermont-Ferrand Hospital (CHU) for providing DTI data. The work was partially supported by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 and by the French ANR Karametria project.

References

1. Vos, S.B., Jones, D.K., Viergever, M.A., Leemans, A.: Partial volume effect as a hidden covariate in DTI analyses. *Neuroimage* 55(4), 1566–1576 (2011)
2. Peled, S., Yeshurun, Y.: Superresolution in MRI: Application to human white matter fiber tract visualization by diffusion tensor imaging. *Magnetic Resonance in Medicine* 45(1), 29–35 (2001)
3. Greenspan, H., Oz, G., Kiryati, N., Peled, S.: MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging* 20(5), 437–446 (2002)
4. Irani, M., Peleg, S.: Motion analysis for image enhancement: Resolution, occlusion, and transparency. *J. of Vis. Com. and Image Rep.* 4(4), 324–335 (1993)
5. Scherrer, B., Gholipour, A., Warfield, S.K.: Super-resolution reconstruction to increase the spatial resolution of diffusion weighted images from orthogonal anisotropic acquisitions. *Medical Image Analysis* 16(7), 1465–1476 (2012)
6. Nedjati-Gilani, S., Alexander, D.C., Parker, G.J.: Regularized super-resolution for diffusion MRI. In: *Proc. of ISBI*, pp. 875–878. IEEE (2008)
7. Stejskal, E., Tanner, J.: Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J. of Chem. Phys.* 42, 288 (1965)
8. Fillard, P., Pennec, X., Arsigny, V., Ayache, N.: Clinical DT-MRI estimation, smoothing, and fiber tracking with Log-Euclidean metrics. *IEEE Transactions on Medical Imaging* 26(11), 1472–1482 (2007)
9. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Mag. Res. in Med.* 56(2), 411–421 (2006)
10. Fillard, P., Descoteaux, M., et al.: Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom. *Neuroimage* 56(1), 220–234 (2011)

Adaptively Constrained Convex Optimization for Accurate Fiber Orientation Estimation with High Order Spherical Harmonics

Giang Tran¹ and Yonggang Shi^{2,*}

¹ Dept. of Mathematics, UCLA, Los Angeles, CA, USA

² Lab of Neuro Imaging (LONI), UCLA School of Medicine, Los Angeles, CA, USA

Abstract. Diffusion imaging data from the Human Connectome Project (HCP) provides a great opportunity to map the whole brain white matter connectivity to unprecedented resolution *in vivo*. In this paper we develop a novel method for accurately reconstruct fiber orientation distribution from cutting-edge diffusion data by solving the spherical deconvolution problem as a constrained convex optimization problem. With a set of adaptively selected constraints, our method allows the use of high order spherical harmonics to reliably resolve crossing fibers with small separation angles. In our experiments, we demonstrate on simulated data that our algorithm outperforms a popular spherical deconvolution method in resolving fiber crossings. We also successfully applied our method to the multi-shell and diffusion spectrum imaging (DSI) data from HCP to demonstrate its ability in using state-of-the-art diffusion data to study complicated fiber structures.

1 Introduction

With the advance of diffusion weighted MR imaging techniques from the Human Connectome Project (HCP) [1, 2], large scale datasets acquired using sophisticated sampling schemes are becoming publicly available. This provides unprecedented opportunities for mapping the white matter fiber structure with higher spatial and angular resolutions. The vast amount of data, however, also poses significant challenges for data analysis algorithms that have focused mostly on conventional, single-shell acquisition schemes. In this work, we propose a novel method for analyzing diffusion images with arbitrary acquisition schemes by accurately reconstructing the fiber orientation distribution (FOD). We demonstrate our method can achieve superior angular resolution and resolve fiber structures on both simulated and *in vivo* data from the HCP.

The diffusion tensor model is practically the most popular method for studying major fiber bundles with diffusion imaging data[3]. For the mapping of whole brain connectivity, however, the tensor model is vastly insufficient as complicated fiber crossings occur frequently throughout the brain. To overcome this difficulty,

* This work was in part supported by NIH grants K01EB013633, R01MH094343, and P41EB015922.

various techniques for high angular resolution diffusion imaging (HARDI) were developed [4–9]. By representing the FOD with spherical harmonics, the spherical deconvolution model has demonstrated great potential in efficiently resolving complicated fiber crossings [6, 9]. The ill-posedness of the deconvolution problem, however, has hindered the use of high order spherical harmonics to resolve fibers with small separation angles. To improve numerical stability, Laplacian or Tikhonov regularizations [7, 9] were incorporated, but these models are limited to single-shell acquisition schemes and only partially overcome the difficulty.

In this work we propose a new method for FOD estimation from data acquired with general sampling schemes, which enables us to seamlessly process HCP data collected with either the multi-shell or DSI schemes [10] illustrated in Fig. 1. At the core of our method is a constrained convex optimization problem for spherical deconvolution with adaptively chosen constraints. By adaptively selecting a minimal set of uniformly distributed constraints, our method can easily use high order spherical harmonics to reconstruct crossing fibers with very small separation angles.

In our experiments, we demonstrate on simulated data that our method outperforms a previous method [9] in reliably resolving fibers with small crossing angles. We also apply it to data from the HCP to demonstrate its generality in processing cutting-edge diffusion imaging data.

The rest of the paper is organized as follows. In section 2, we formulate the spherical deconvolution problem in the general setting of analyzing data from arbitrary acquisition schemes. The adaptively constrained convex optimization approach is then developed in section 3. Experimental results are presented in section 4. Finally conclusions are made in section 5.

2 Spherical Deconvolution Model

In this section, we develop the spherical deconvolution model for FOD reconstruction from general acquisition schemes. At each voxel, the diffusion signal at the b-value b_i and the direction u_i is denoted as $s(b_i, u_i) (i = 1, \dots, N)$. Let S denote the unit sphere, and $f : S \rightarrow \mathbb{R}^+$ the FOD. Given the single fiber response function $k(b, u, w)$ for a fiber in the direction w , the diffusion signal is expressed as the convolution of the FOD and the kernel:

$$s(b, u) = \int_S f(w)k(b, u, w)dw + n(b, u) \quad (1)$$

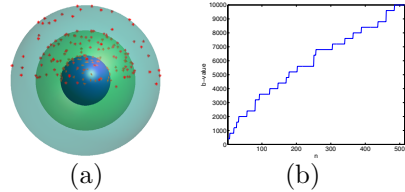


Fig. 1. An illustration of the two acquisition schemes from the HCP. (a) The multi-shell sampling scheme with 270 directions distributed over three shells with $b=1000, 2000, 3000 \text{ s/mm}^2$. (b) The DSI sampling scheme has 514 directions with b-values increasing from 400 to 10,000 s/mm^2 .

where n is noise. Instead of estimating the kernel directly from the data [6], which is difficult for general acquisition schemes such as DSI, we follow the single tensor model and represent the kernel parametrically as:

$$k(b, u, w) = e^{-b(\lambda_2 + (\lambda_1 - \lambda_2)(u \cdot w)^2)}. \quad (2)$$

where the only parameters $\lambda_1 \gg \lambda_2 = \lambda_3$ are the eigenvalues of the tensor model. These parameters can be either chosen from previous literature or computed easily from the data.

For efficient computation, the FOD is represented with the spherical harmonics up to the order L as:

$$f(w) = \sum_{l,m} x_l^m Y_l^m(w) \quad \forall w \in S \quad (3)$$

where Y_l^m is the m -th real spherical harmonics at the order $l = 0, 2, \dots, L$, and x_l^m is the coefficient for the basis Y_l^m . Note that only even order spherical harmonics are used because the FOD is symmetric on the sphere. From a signal processing perspective, high order spherical harmonics are needed if we want to accurately represent or reconstruct crossing fibers with really small separation angles. Due to numerical difficulties, typically spherical harmonics up to the order of eight were used in previous work [6, 9], which limits their capability in reliably resolving fiber crossing of small angles.

Using the Funk-Hecke theorem, we can express the diffusion signal as

$$\begin{aligned} s(b, u) &= \int_S \sum_{l,m} x_l^m Y_l^m(w) e^{-b(\lambda_2 + (\lambda_1 - \lambda_2)(u \cdot w)^2)} dw + n(b, u) \\ &= \sum_{l,m} Y_l^m(u) G_l(b, \lambda_1, \lambda_2) x_l^m + n(b, u) \end{aligned} \quad (4)$$

with $G_l(b, \lambda_1, \lambda_2)$ defined as:

$$G_l(b, \lambda_1, \lambda_2) = 2\pi \int_{-1}^1 P_l(t) e^{-b\lambda_2 - b(\lambda_1 - \lambda_2)t^2} dt, \quad (5)$$

where P_l is the Legendre polynomial of degree l .

Let \underline{s} denote the vector of diffusion signals $s(b_i, u_i)$ sampled at a discrete set of points $(b_i, u_i) (i = 1, \dots, N)$. For simplicity, we denote Y_l^m and x_l^m by Y_j and x_j , respectively, with $j = m + \frac{l^2 + l + 2}{2}$. Given the maximum order L of spherical harmonics used, the total number of basis functions is $J = (L+1)(L+2)/2$. Let $\underline{x} = [x_1, \dots, x_j, \dots, x_J]$ be the vector of coefficients for the FOD, we can write (4) in matrix form as:

$$\underline{s} = A\underline{x} + \underline{n} \quad (6)$$

where \underline{n} denotes the vector of noise, and $A = B \cdot G$ is the entry-wise product of two matrices B and G defined as follows:

$$B = \begin{pmatrix} Y_1(u_1) & Y_2(u_1) & \cdots & Y_J(u_1) \\ Y_1(u_2) & Y_2(u_2) & \cdots & Y_J(u_2) \\ \cdots & \cdots & \cdots & \cdots \\ Y_1(u_N) & Y_2(u_N) & \cdots & Y_J(u_N) \end{pmatrix} \quad G = \begin{pmatrix} G_0(b_1) & G_2(b_1) & \cdots & G_2(b_1) & \cdots & G_L(b_1) \\ G_0(b_2) & G_2(b_2) & \cdots & G_2(b_2) & \cdots & G_L(b_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ G_0(b_N) & G_2(b_N) & \cdots & G_2(b_N) & \cdots & G_L(b_N) \end{pmatrix}.$$

For each order l , the element $G_l(b_i) = G_l(b_i, \lambda_1, \lambda_2)$ is repeated $2l + 1$ times on the i -th row.

3 Adaptively Constrained Convex Optimization

In this section, we develop a novel approach for FOD estimation by solving the spherical deconvolution problem as a constrained convex optimization problem. The key idea is the adaptive selection of the set of constraints for every voxel to avoid overly constrain the solution and affect reconstruction accuracy. The constraint we impose on FOD reconstruction is motivated by its non-negativity condition. With the spherical harmonics representation up to a fixed order, however, it is impossible to completely eliminate negative values in the FOD. It is thus critical to limit the negative components to the minimal extent and ensure major fiber directions are captured. To achieve this goal, our strategy is to constraint the FOD to be non-negative on a *minimal* set of *uniformly* distributed points on the sphere.

With a remeshing algorithm [11], we build a collection of constraint sets $\mathcal{V} = \{V_1, V_2, \dots\}$, with varying number of points on the sphere, where each member $V_Q = \{v_Q^1, v_Q^2, \dots, v_Q^Q\}$ is a set of Q uniformly distributed points on the hemisphere of the unit sphere. With the spherical harmonics representation, the requirement that f should be non-negative on V_Q can be expressed as:

$$C_Q \underline{x} \geq 0 \tag{7}$$

where C_Q is a matrix of size $Q \times J$ defined as

$$C_Q = \begin{pmatrix} Y_1(v_Q^1) & Y_2(v_Q^1) & \cdots & Y_J(v_Q^1) \\ Y_1(v_Q^2) & Y_2(v_Q^2) & \cdots & Y_J(v_Q^2) \\ \cdots & \cdots & \cdots & \cdots \\ Y_1(v_Q^Q) & Y_2(v_Q^Q) & \cdots & Y_J(v_Q^Q) \end{pmatrix}.$$

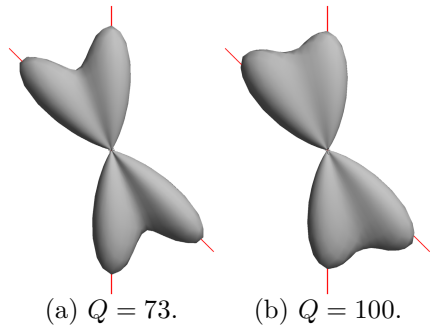


Fig. 2. FOD reconstruction from simulated diffusion data. The true fiber directions are plotted as red lines.

Given a specific constraint set $V_Q(Q = 1, 2, \dots)$, the spherical deconvolution problem for FOD reconstruction can be formulated as a constrained convex optimization problem:

$$\begin{aligned} \min \quad & \| \underline{s} - A\mathbf{x} \|^2 \\ \text{s.t.} \quad & C_Q \mathbf{x} \geq 0 \end{aligned} \quad (8)$$

Because this problem is convex, global minimum can always be found numerically with software packages such as `cvx` [12]. With the FOD being non-negative on a set of uniformly distributed points, we ensure that large negative components will not occur and most of the energy of the FOD are contributed by physically meaningful, i.e., positive, components. Because the complexities of fiber crossings are spatially varying across the brain, the number of active constraints in (8) could be different as a result. Thus it is also unreasonable to fix the number of constraints. To overcome this difficulty, we adaptively search through the constraint collection \mathcal{V} at every voxel to find the smallest Q such that the solution satisfies:

$$\frac{\int_{f(w)>0} f dw}{\int_{f(w)<0} |f| dw} > \delta. \quad (9)$$

This condition measures how successful the reconstructed FOD is able to focus its energy on positive components. For example, if we pick $\delta = 25$, we ensure more than 95% of the L1 energy of the FOD are from positive components. As a demonstration, we show in Fig. 2 the FOD reconstruction results of two fibers using simulated diffusion data from 60 directions with $b=1000$ s/mm². The maximum order of spherical harmonics used here is $L = 8$. The result in Fig. 2 (a) is obtained from adaptively determined constraints, where $Q = 73$, and the result in Fig. 2 (b) is obtained by fixing $Q = 100$. We can see the overly constrained solution is less sharp and one of its peaks is obviously misaligned with the true fiber direction.

For practical implementation, there is no need to start the search from $Q = 1$. Given a maximum order L , we can pick a constraint set from experience and start the search there. For $L = 8$, we typically start the search at $Q = 60$ and the reconstruction of the FOD in Fig. 2(a) took less than one second.

4 Experimental Results

In this section, we present experimental results to demonstrate our method on both simulated and in vivo data from the HCP. In our experience, the FOD reconstruction results from our method are very robust to the selection of the parameters in the kernel. For all experiments, we thus fix the parameters for the kernel as $\lambda_1 = 0.0017$ and $\lambda_2 = 0.0003$ following the literature [9], and the threshold in (9) as $\delta = 25$.

4.1 Simulated Data

In the first experiment, we compare with the spherical deconvolution method in [9] on simulated data. For two fibers with a crossing angle of 30° , we follow the multi-tensor model to simulate the diffusion data from a single-shell acquisition scheme of 81 directions with $b = 3000$ s/mm^2 . Rician noise was added to obtain a signal to noise ratio of 20. For every parameter selection from both methods, the experi-

ments were run 100 times to obtain the mean and standard deviation of the FOD. For the spherical deconvolution method in [9], the results with the maximum order $L = 8$ and $L = 16$ are plotted in Fig. 3 (a) and (b). At the order of $L = 8$, we chose the same regularization parameter 0.006 as in [9]. We can see that the reconstructed FOD at this order cannot resolve the fiber crossing. When the order was increased to $L = 16$, its result became highly oscillatory even though we increased the regularization parameter to 0.02. With our adaptively constrained optimization method, such oscillations were successfully suppressed and accurate results were obtained as shown in Fig. 3(c). This demonstrates the superior ability of our method in resolving fiber crossings at small angles with data from single-shell acquisition schemes.

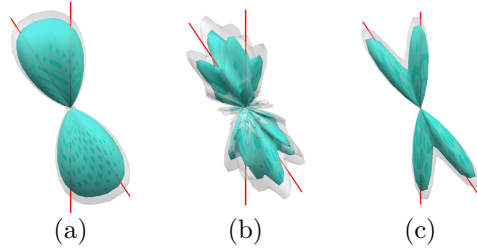


Fig. 3. A comparison of 100 runs of FOD reconstruction results from simulated diffusion data. The cyan surface is the mean FOD and the shaded surface is mean plus two standard deviation of the FOD from 100 runs. Red lines indicate true fiber directions. (a) Method in [9]: $L = 8$. (b) Method in [9]: $L = 16$. (c) Our method: $L = 16$.

4.2 Multi-shell Data from HCP

In the second experiment, we applied our method to the diffusion data of a subject from the HCP that was acquired with a multi-shell sampling scheme as illustrated in Fig. 1 (a). The reconstructed FODs of an ROI on an axial slice, which is shown in Fig. 4(b), in the right thalamus with $L = 8$ and $L = 16$ are plotted in Fig. 4 (a) and (c). All FODs are color-coded with the directions. The reconstructed FODs are consistent with known anatomical knowledge that various fibers cross the thalamus to reach different cortical regions. The top-right corner of the ROI touches the cortical spinal tract that goes from inferior to the superior part of the brain. At this location, the FOD reconstructed with our method in both Fig. 4 (a) and (c) has only one major fiber direction, which is consistent with the orientation of the cortical-spinal tract. By comparing the results in Fig. 4(a) and (c), especially regions highlighted by the dashed lines, we can clearly see that our method successfully uses high order spherical harmonics to achieve better angular resolution of crossing fibers than results reconstructed with low order spherical harmonics.

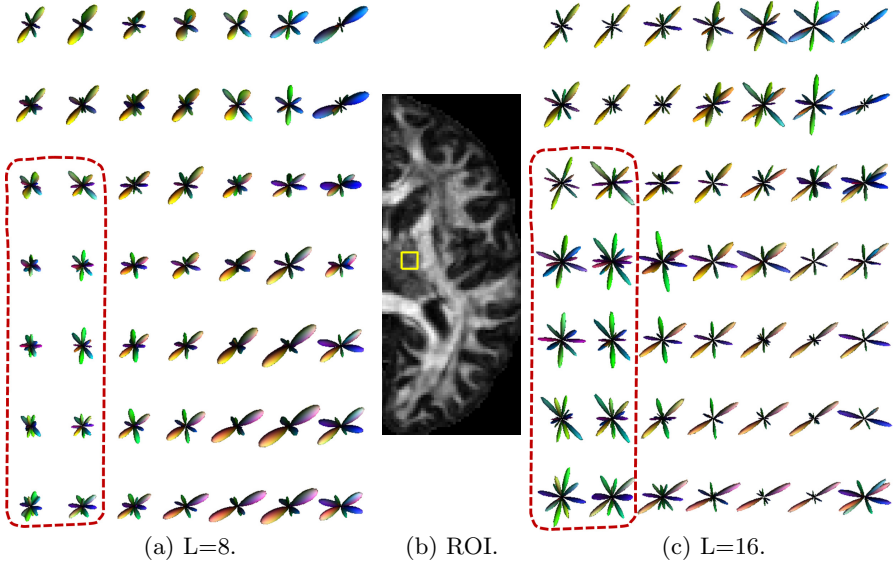


Fig. 4. FOD reconstruction results from multi-shell diffusion data of the HCP

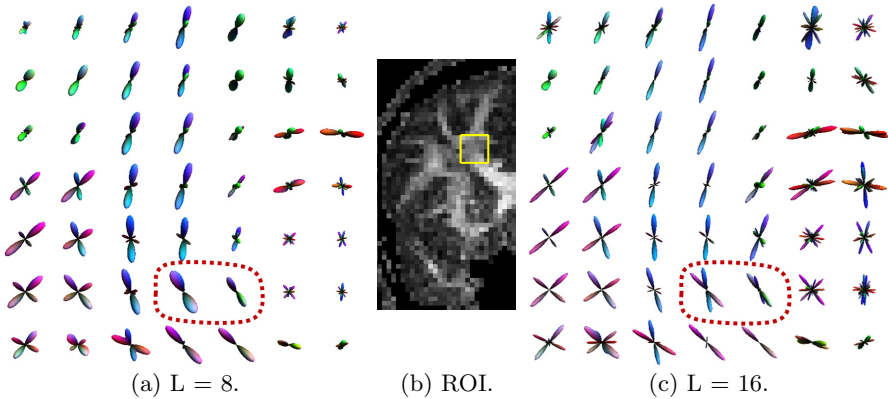


Fig. 5. FOD reconstruction results from DSI data of the HCP

4.3 DSI Data from HCP

In the third experiment, we applied our method for FOD reconstruction using DSI data of a subject from the HCP. As illustrated in Fig. 1(b), the diffusion data was acquired at 514 points in the q -space with b -values ranging from 400 to 10000 s/mm^2 . For an ROI on a coronal slice shown in Fig. 5(b) that has crossing fibers possibly from the corpus callosum, cortical spinal tract, and superior longitudinal fasciculus, we applied our method with $L=8$ and $L=16$ to compute the FODs. The reconstructed FODs are plotted in Fig. 5(a) and (c). As a demonstration, we highlighted two FODs with dashed lines in Fig. 5(a) and (c). It clearly shows the

power of our method in using high order spherical harmonics to resolve crossing fibers with small separation angles.

5 Conclusions

In this paper we developed a novel approach for the accurate reconstruction of FODs from arbitrarily sampled diffusion imaging data. By solving the spherical deconvolution as an adaptively constrained convex optimization problem, our method can robustly use high order spherical harmonics to resolve complicated fiber crossings. We demonstrated the power of our method on HCP data from both the multi-shell and DSI acquisition schemes. For future work, we will integrate our method with tractography algorithms and investigate its application in studying whole brain connectivity.

References

1. Essen, D.V., Ugurbil, K., et al.: The human connectome project: A data acquisition perspective. *NeuroImage* 62(4), 2222–2231 (2012)
2. Toga, A., Clark, K., Thompson, P., Shattuck, D., Van Horn, J.: Mapping the human connectome. *Neurosurgery* 71(1), 1–5 (2012)
3. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66(1), 259–267 (1994)
4. Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Wedeen, V.J.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine* 48(4), 577–582 (2002)
5. Behrens, T., Berg, H.J., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* 34(1), 144–155 (2007)
6. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23(3), 1176–1185 (2004)
7. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35(4), 1459–1472 (2007)
8. Jian, B., Vemuri, B.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. *IEEE Trans. Med. Imag.* 26(11), 1464–1471 (2007)
9. Descoteaux, M., Deriche, R., Knosche, T., Anwander, A.: Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Trans. Med. Imag.* 28(2), 269–286 (2009)
10. Wedeen, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine* 54(6), 1377–1386 (2005)
11. Peyré, G., Cohen, L.: Geodesic remeshing using front propagation. *Int’l Journal of Computer Vision* 69(1), 145–156 (2006)
12. CVX Research, Inc.: CVX: Matlab software for disciplined convex programming, version 2.0 beta (September 2012), <http://cvxr.com/cvx>

Fiber Continuity Based Spherical Deconvolution in Spherical Harmonic Domain

Marco Reisert¹ and Henrik Skibbe²

¹ Department of Diagnostic Radiology, Medical Physics, University Medical Center, Breisacher Street 60a, 79106 Freiburg, Germany

² Graduate School of Informatics, Kyoto University, Gokasho, 611-0011 Uji, Kyoto, Japan

Abstract. The reliable estimation of fiber orientation distributions from diffusion-sensitized magnetic resonance imaging is an important processing step for revealing structural connectivity in the human brain. Fiber orientation distributions are usually represented in spherical harmonic (SH) domain, but many modern spatial regularization techniques were not yet formulated in SH-domain such that they cannot benefit from the well known advantages. In this article we propose a novel SH-formulation of the contour enhancement propagator (fiber continuity) which enables to implement a variety of processing steps in SH-domain.

1 Introduction

Magnetic resonance imaging (MRI) has the potential to visualize non-invasively the fibrous structure of the human brain white matter [5]. Based on the directional dependency of water diffusivity in fibrous tissue it is possible to reveal underlying connectivity information. There are numerous methods for estimating orientation distributions on the basis of the diffusion-weighted MR-signal. We will focus on spherical deconvolution [8], which is one way to estimate the so called fiber orientation distributions (FOD) on the basis of the diffusion-weighted MR-signal. The idea is based on a model-driven deconvolution scheme to turn the diffusion weighted MR-signal into a FOD. Fiber orientation distributions are usually represented in spherical harmonic (SH) domain. However, sophisticated spatial regularizers based on $\mathbb{R}^3 \times S^2$ -diffusion [1,2,3,6] are difficult to implement in SH domain (apart from the simple isotropic ones [4]). They all rely on an equiareal discretization of the two-sphere S_2 . Indeed, there are implementations [3] that use spherical harmonics as an intermediate S_2 -interpolation scheme, but they cannot benefit from the well-known advantages of the spherical harmonic representation such as the compact and memory efficient storage, the analytic and efficient computations of S_2 -convolutions, and the closeness under rotations. In this article we overcome these restrictions by showing how the generators of diffusion and convection on $\mathbb{R}^3 \times S^2$ can be described and implemented in terms of spherical harmonics and how the spherical discretization as proposed in [3,6] can be avoided.

2 Theory

We are interested in solving or propagating a partial differential equation of the form $\partial_t \phi(\mathbf{r}, \mathbf{n}, t) = H\phi(\mathbf{r}, \mathbf{n}, t)$ where ϕ is a function (typically the FOD) on $\mathbb{R}^3 \times S^2$ and

H a linear differential operator in the orientation $\mathbf{n} \in S_2$ and position $\mathbf{r} \in \mathbb{R}^3$. The orientation \mathbf{n} corresponds the local fiber direction at location \mathbf{r} in the brain. In [3] it was shown that, if H generates directed diffusion/convection, then it is a linear/quadratic form in $\mathbf{n}^\top \nabla$ (convection and anisotropic diffusion) and linear in $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$ and \mathcal{J}^2 , which denotes the Laplace-Beltrami operator on the two-sphere. Our goal is to understand the action of the generator H , if the field ϕ is written in terms of a spherical harmonics expansion

$$\phi(\mathbf{r}, \mathbf{n}) = \sum_{j=0}^{\infty} \mathbf{f}^j(\mathbf{r})^\top \mathbf{Y}^j(\mathbf{n}) = \sum_{j=0}^{\infty} \sum_{m=-j}^{m=j} \frac{f_m^j(\mathbf{r})}{Y_m^j(\mathbf{n})}$$

where the spherical tensor fields \mathbf{f}^j are obtained by the projections $\mathbf{f}^j = \frac{1}{2j+1} \langle \mathbf{Y}^j, \phi \rangle$, where \mathbf{Y}^j are semi-Schmidt normalized spherical harmonics. Hence we are interested in matrix elements $\hat{H}_{j'm'}^{j'm} = \langle Y_{m'}^{j'}, H Y_m^j \rangle$ of H in spherical harmonic representation such that the propagating equation can be written as $\partial_t f_m^j(\mathbf{r}, t) = \sum_{j',m'} \hat{H}_{j'm'}^{j'm} f_{m'}^{j'}(\mathbf{r}, t)$ where $\hat{H}_{j'm'}^{j'm}$ is a differential operator in \mathbf{r} , but purely algebraic in the orientation coordinate. The spherical Laplace-Beltrami operator is well known in this representation $\langle Y_m^j, \mathcal{J}^2 \phi \rangle = -j(j+1) f_m^j$. Also for the spatial Laplacian Δ the result is trivial $\langle Y_m^j, \Delta \phi \rangle = \Delta f_m^j$ (see [7]). But for the directed convection $\mathbf{n} \cdot \nabla$ and the directed diffusion $(\mathbf{n} \cdot \nabla)^2$ generator the results are more intricate. In our companion report [7] we give the general proof for $SE(3)$ diffusion, but state here the more simple case for $\mathbb{R}^3 \times S_2$. For the convection generator one finds

$$\langle Y_m^j, (\mathbf{n} \cdot \nabla) \phi \rangle = \sum_{\substack{j'=-1,1 \\ m=m'+q}} \frac{2j'+1}{2j+1} \langle jm|j'm', 1q \rangle \langle j0|j'0, 10 \rangle \partial_q^1 f_{m'}^{j'} \tag{1}$$

where $\partial_m^j = R_m^j(\nabla)$ is the so called spherical derivative operator, which is a j -homogeneous differential as the solid harmonic $R_m^j(\mathbf{r}) = r^j Y_m^j(\mathbf{r}/r)$ is a j -homogeneous polynomial. The $\langle jm|j'm', 1q \rangle$ are the so called Clebsch Gordan coefficients, which are just real numbers. The sum runs over the indices j', m' and q , but note that due to the selection rules of the Clebsch Gordan coefficients only very few terms remain (at most 6). On the other hand the diffusion generator takes the form

$$\langle Y_m^j, (\mathbf{n} \cdot \nabla)^2 \phi \rangle = \frac{\Delta}{3} + \frac{2}{3} \sum_{\substack{j'=-2,0,2 \\ m=m'+q}} \frac{2j'+1}{2j+1} \langle jm|j'm', 2q \rangle \langle j0|j'0, 20 \rangle \partial_q^2 f_{m'}^{j'} \tag{2}$$

Again the sum runs over j', m' and q and can contain at most 10 terms.

Spatially Regularized Spherical Deconvolution. The goal of spherical deconvolution is to find a FOD ϕ such that $\|\mathbf{F}\phi - S\|^2$ is minimized. Here S is typically the diffusion weighted measurement normalized by the b0-image. The operator \mathbf{F} denotes the spherical convolution with the so called fiber response function. As \mathbf{F} is highly non-regular a regularizer is needed. In this article we consider the contour enhancement [3] regularizer (also called fiber continuity [6]) $R_{FC}(\phi) = \lambda \int (\mathbf{n} \cdot \nabla \phi)^2 d\mathbf{x}d\mathbf{n}$ that prevents

‘arbitrary’ smoothing and preserves and emphasize the fibrous nature of the data. Instead of smoothing isotropically like done in [4] the smoothing takes place only along the current fiber direction. The idea behind is that a fiber which goes through voxel \mathbf{r} with direction \mathbf{n} is very likely to go also through voxel $\mathbf{r} + \epsilon\mathbf{n}$ for some small ϵ .

The choice of regularization strength is a crucial issue. We found that too strong regularization emphasizes discretization artifacts. On the other hand too low values lead to less stable results. We found by a simple visual inspection of a simulated crossing a value of $\lambda = 0.005$ is a good trade-off between stability and accuracy.

Optimization and Implementation. In order to find the optimum of $J_{\text{reg}} = \|\mathbf{F}\phi - S\|^2 + R_{\text{FC}}(\phi)$, one has to compute its derivative and set it to zero. This leads to $(\mathbf{F}^T \mathbf{F} - \lambda(\mathbf{n} \cdot \nabla)^2)\phi = \mathbf{F}^T S$. The second term of the equation is formulated with the proposed SH representation from equation (2) and solved by an ordinary conjugate gradients (CG) scheme. The operator \mathbf{F} is simple to implement in terms of spherical harmonics, because it is a diagonal matrix. For details about the discretized sphere (512 directions) approach we refer to our companion report [7]. For both methods the CG algorithm was iterated 100 times, which was enough for convergence. For initialization a ‘zero’ FOD was used.

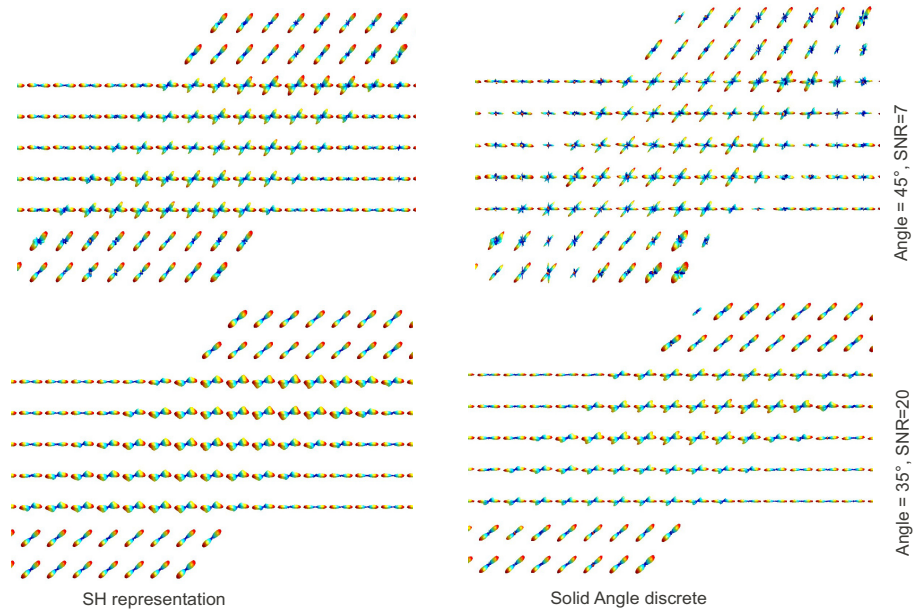


Fig. 1. Two simulated crossing situations to compare the SH representation with the angular discrete version. The simulation was performed with 64 gradient directions at a b-value of $1000s/mm^2$ and with a diffusion coefficient of $10^{-3}mm^2/s$

3 Experiments

The goal of the experiment is to compare the proposed spherical harmonic representation with the discrete angular representation of the fiber orientation distributions. Therefore, we simulated the MR-signal of a crossing region with 64 gradient directions. The signal was simulated by using the standard exponential model $S_{\mathbf{n}_{\text{fib}}}(\mathbf{n}) = e^{-bD(\mathbf{n} \cdot \mathbf{n}_{\text{fib}})^2}$, that is, no diffusion perpendicular to a fiber is assumed. We have chosen $bD = 1$, which emulates a b-value of $1000\text{s}/\text{mm}^2$ and a typical diffusion coefficient for the human brain. The generated signal was distorted by Rician noise $S_{\text{noisy}} = \sqrt{(S + n_{\text{real}})^2 + n_{\text{imag}}^2}$, where n_{real} and n_{imag} are normally distributed real numbers with standard deviation σ . The signal-to-noise ratio is defined as $\text{SNR} = 1/\sigma$, that is, the SNR is calculated with respect to the $b = 0$ measurement. The crossing was created on a 24×24 voxel grid, where the tracts of the crossing are on average 5 voxels thick. To get an impression look at Figure 1. To measure the performance of the deconvolution method, the local maxima of the estimated FODs are extracted and compared to the ground truth direction (see [7] for details). A ground truth direction is said to be detected, when it is in a range of 10° from a detected direction. To measure the performance we used precision, recall and the f-score¹. The precision rate reflects how many of the directions detected by the algorithm are true fiber directions, on the other hand, the recall rate gives the percentage of true fiber directions that were detected by the algorithm. The f-score is just a measure reflecting the trade-off between recall and precision. To generate the performance measures the simulations were repeated 100 times.

Method Comparison and Discussion. In Figure 1 we visually compare the spherical harmonic implementation for a cutoff of $L = 8$ with the discrete version with 512 directions on the sphere. We consider a crossing angle of 35° and 45° . For both situations one fiber bundle direction was chosen along the underlying Cartesian coordinate axis. For the larger crossing, which is easier to resolve, we assumed a relatively low SNR of 7. For this case one can see that the discrete version is more susceptible to noise than the spherical harmonic version, which is not astonishing due to the implicit regularization by the finite SH-cutoff of $L = 8$. For the smaller crossing angle of 35° a higher SNR of 20 was assumed. In this case a SH-representation of $L = 8$ is nearly at its limits to discriminate between both directions, while the discrete version can still well distinguish. Further, one can observe that for both methods the FODs along the horizontal Cartesian axis are sharper than along the skew axis. But this effect is more prominent for the discretized version than for the spherical harmonic representation.

In Figure 2 we show quantitative results. The crossing was simulated for varying crossing angles between 30° and 90° . Additionally we varied the absolute pose α of the crossing. For $\alpha = 0^\circ$ the horizontal bundle is aligned with underlying x-axis of

¹ Let TP be the number of successfully found ground truth directions, let FP be the number of detections that are not in a range of 10 degree to a ground truth direction, and let FN the number of ground truth direction that are not detected, then precision = $\text{TP}/(\text{TP} + \text{FP})$ and recall = $\text{TP}/(\text{TP} + \text{FN})$ and f-score = $2 \text{ precision} \cdot \text{recall}/(\text{precision} + \text{recall})$.

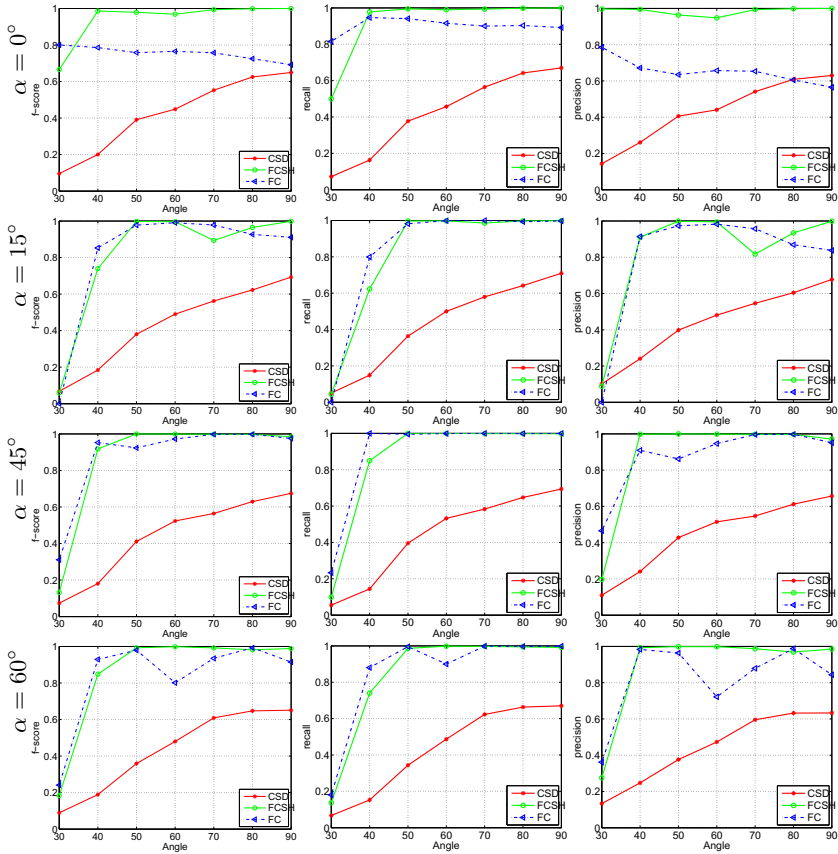


Fig. 2. Detection accuracies in terms of precision/recall and f-score for different crossing angles and absolute angles α . The measurement was simulated at a $SNR = 20$ with a value of $bD = 1$ and 64 gradient directions.

the Cartesian grid. With growing α the whole configuration is rotated clockwise. The crossing was simulated at an SNR of 20, which is a realistic scenario. As a baseline experiment we show results of the so called Constrained Spherical Deconvolution (CSD) approach [9], where an additional positivity constraint is used to obtain more stable results (original implementation obtained by the author). Apart from the CSD approach the results obviously depend on the absolute angle of the configuration due to the underlying Cartesian voxel grid. The discrete approach has advantages for small crossing angles, but suffers from a more severe dependency on the absolute orientation of the configuration and less precision. For crossing angles above 45° the SH-based approach solves the task nearly without any error.

In Figure 3 the results are further investigated by scatter plots in the ϕ, θ -plane. Each extracted local maxima is represented as a blue dot in the ϕ, θ -plane, while the true directions are depicted by solid red lines and the tolerance used for quantitative evaluation

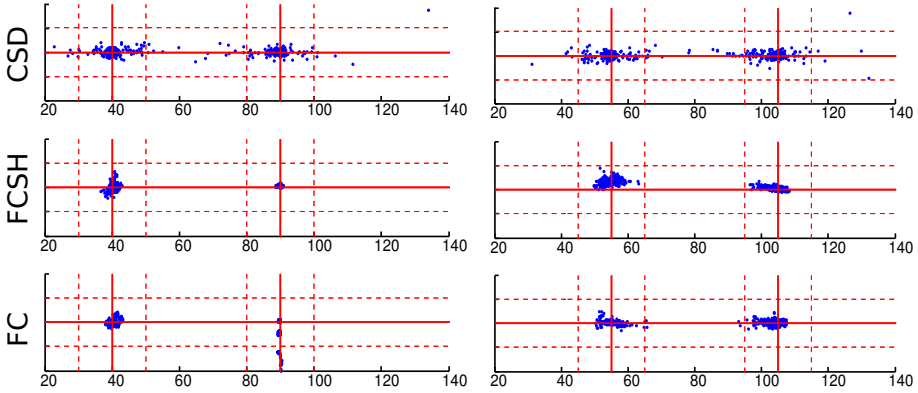


Fig. 3. Scatter plots in the ϕ - θ plane for the crossing configuration with a crossing angle of 50° . The angle ϕ is plotted along the x-axis, θ along the y-axis of the scatter plots. The intersections of the thick red lines indicate the expected ground truth directions. The dotted lines indicate the 10° detection tolerance. On the left, the results for the configuration with an absolute angle of $\alpha = 0$ are shown, on the right the results for $\alpha = 15^\circ$ are given.

by dashed red lines. With the same parameters as above a crossing of 50° was simulated and reconstructed by the three methods. The crossing was simulated twice, for an absolute angle of $\alpha = 0$, i.e. one direction is along the x-axis, and secondly for an absolute angle of $\alpha = 15^\circ$. For $\alpha = 0$ the SH-based approach is able to perfectly resolve the direction along the x-axis ($\phi = 90^\circ$), while the other direction ($\phi = 40^\circ$) is a bit more blurry. On the other hand, the discrete approach has severe problems with the $\phi = 90^\circ$ direction, which explains the lack of precision, i.e. apart from the true direction there are some additional local maxima that produce false positives. The main reason is the interplay of the 64 gradient directions and the 512 discrete directions of the FOD. The effect is reduced by an increase of measurement directions. But also the SH-based approach has problems when the number of measurement directions is too low. They are revealed for an absolute angle of $\alpha = 15^\circ$. Besides the uncertainty caused by the measurement noise one can observe a systematic bias. For example, for the direction along $\phi = 55^\circ$, the center of the distribution is shifted in θ by approximately 5° . Also for the other direction the distribution is a bit squeezed. We also found that the main reason is the low number of measurement directions. For example, for 128 gradient directions the estimated directions are unbiased. Another way to reduce the effect is to decrease the expansion cutoff of the spherical harmonic representation. For $L = 6$ and 64 gradient directions the estimates do not show a bias. To conclude the differences: both methods have problems when the number of measurements becomes too low. While the discrete approach shows scattered, multimodal distributions, the SH-based approach shows a slight systematic bias but the distributions stay unimodal.

In Figure 4 we show a real world example of the human brain. The setting of the measurement is nearly the same like in the simulations. A b-value of $1000s/mm^2$ and 61 gradient directions were used with an isotropic resolution of $2mm$. A kernel of the form $S_{\mathbf{n}_{\text{fib}}}(\mathbf{n}) = e^{-bD(\mathbf{n} \cdot \mathbf{n}_{\text{fib}})^2}$ with $bD = 1$ was used as a model for deconvolution.

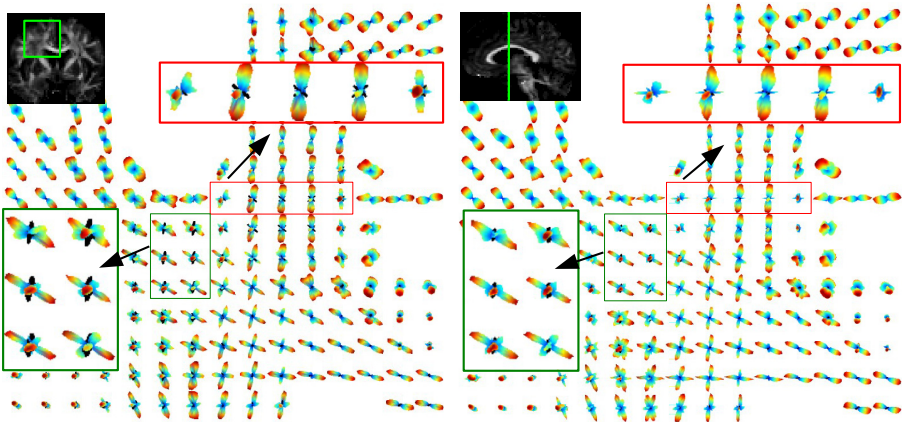


Fig. 4. A real world example of the human brain. A coronal section is shown. Left: results of the SH-based approach with $L = 8$, right: the discrete approach with 512 directions.

Figure 4 shows a coronal section in glyph representation. The green rectangle highlights a region with the largest differences. The red rectangle shows a regions where the discrete approach shows a direction which does not appear for the SH-based approach. Whether the direction is true or not is difficult to say, but the fact that it precisely points along the x-axis makes it dubious.

Memory Consumption and Running Time. The memory consumption of the SH-based and discrete approach is easy to compare. We want to consider the above real world experiment as an example. The whole volume has a size of $96 \times 96 \times 60$. Exploiting the symmetry and realness each voxels consumes $(L + 2)^2/4$ complex numbers in SH representation, resulting in $50 \cdot 8$ bytes for $L = 8$ and double precision. Overall, one volume needs $96^2 \cdot 60 \cdot 50 \cdot 8 = 220$ MB. On the other hand, in discrete representation with 512 direction needs $96^2 \cdot 60 \cdot 512 \cdot 8 = 2264$ MB, which is 10-times more compared to the SH-representation. Recall, that the conjugate gradient algorithm needs four instances of the volume in memory. The running time is more difficult to compare. The application of $(\mathbf{n} \cdot \nabla)^2$ and the spherical convolution \mathbf{H} are the important processing steps. In case of the discrete approach one has to compute for each of the 512 components six second order finite differences to get $(\mathbf{n} \cdot \nabla)^2 \phi$. For the SH-based approach one also have to compute six finite differences, but the linear combinations of them to obtain the final values are more expensive. In practice we found that one application of $(\mathbf{n} \cdot \nabla)^2$ with 256 discrete directions is comparable to an SH-based application of $(\mathbf{n} \cdot \nabla)^2$ with $L = 10$. The computation of the operator \mathbf{H} is negligible in SH-representation, while it is the bottleneck for the discrete approach. The running time can heavily differ depending on the used library. For example, the highly optimized BLAS matrix-multiplication shipped with MATLAB is 10 times faster than the standard non-optimized version. To give an example, to reconstruct the above volume with 100 CG iteration with

512 directions takes on a *Intel Xeon X7560 @ 2.27GHz* about 20 minutes with a highly optimized multiplication. For comparison, our SH-implementation for $L = 8$ takes about 10 minutes on the same machine.

4 Conclusion

We proposed a SH-formulation for a family of diffusion/convection generators that are of particular interest for the spatial regularization of FODs. The SH-formulation offers advantages in terms of running time and memory consumption and keeps approximately the performance and behavior of the original discrete implementation. While the discrete implementation tends to produce additional local maxima and spikey distributions, the distributions of the SH-implementation are more smooth, but shows a bit worse performance for smaller crossing angles.

Acknowledgements. This work was partly supported by 'Bioinformatics for Brain Sciences' under the Strategic Research Program for Brain Sciences, by the MEXT (Japan), the 'Baden-Württemberg Stiftung - Eliteprogramme for Postdocs' and by the Deutsche Forschungsgemeinschaft (DFG) Grant KI 3286/2-1. The authors would also like to thank D. Tournier for providing his original CSD implementation.

References

1. Barmpoutis, A., Vemuri, B.C., Howland, D., Forder, J.R.: Extracting tractosemas from a displacement probability field for tractography in DW-MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 9–16. Springer, Heidelberg (2008)
2. Delputte, S., Dierckx, H., Fieremans, E., D'Asseler, Y., Achten, E., Lemahieu, I.: Postprocessing of brain white matter fiber orientation distribution functions. In: ISBI 2007, pp. 784–787 (2007)
3. Duits, R., Franken, E.: Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of hardi images. *International Journal of Computer Vision* 92, 231–264 (2011)
4. Goh, A., Lenglet, C., Thompson, P.M., Vidal, R.: Estimating orientation distribution functions with probability density constraints and spatial regularity. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 877–885. Springer, Heidelberg (2009)
5. Jones, D.K. (ed.): *Diffusion MRI: Theory, Methods and Applications*. Oxford University Press (2010)
6. Reisert, M., Kiselev, V.: Fiber continuity: An anisotropic prior for odF estimation. *IEEE Trans. Med. Imaging* 30(6), 1274–1283 (2011)
7. Reisert, M., Skibbe, H.: Left-Invariant Diffusion on the Motion Group in terms of the Irreducible Representations of $SO(3)$. ArXiv e-prints (February 2012)
8. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 23(3), 1176–1185 (2004)
9. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35(4), 1459–1472 (2007)

A 4D Hyperspherical Interpretation of q -space

A. Pasha Hosseinbor¹, Moo K. Chung¹, Yu-Chien Wu²,
Andrew L. Alexander¹, and Barbara B. Bendlin¹

¹ University of Wisconsin-Madison, USA

² Dartmouth College, USA

hosseinbor@wisc.edu

Abstract. 3D q -space can be viewed as the surface of a 4D hypersphere. In this paper, we seek to develop a 4D hyperspherical interpretation of q -space by projecting it onto a hypersphere and subsequently modeling the q -space signal via 4D hyperspherical harmonics (HSH). Using this orthonormal basis, we analytically derive several quantitative indices and numerically estimate the diffusion ODF. Importantly, we derive the integral transform describing the relationship between the diffusion signal and propagator on a hypersphere. We also show that the HSH basis expends less fitting parameters than other well-established methods to achieve comparable signal and better ODF reconstructions. All in all, this work provides a new way of looking at q -space.

1 Introduction

One of the first physical applications of quantum mechanics was in solving the Schrödinger equation for the hydrogen atom. It had been solved in position-space by Schrödinger, himself, but not in momentum-space, which is related to position-space via the Fourier transform. The momentum-space solution was of interest to quantum chemists because it could potentially reveal additional quantum mechanical insights about the hydrogen atom not found in the position space solution. Nearly a decade after Schrödinger's work, V. Fock solved the Schrödinger equation for the hydrogen atom directly in momentum-space. In his classic paper [4], Fock stereographically projected 3D momentum-space onto the surface of a 4D unit hypersphere, and after this mapping was made, he was able to show that the momentum-space hydrogen orbitals could be simply expressed in terms of 4D hyperspherical harmonics (HSH), which are the multidimensional analogues of the 3D spherical harmonics.

In diffusion MRI, analogous to momentum- and position-space in quantum mechanics, the signal decay and ensemble average propagator (EAP) are Fourier transform (FT) pairs within the q -space framework:

$$P(\mathbf{k}) = \int E(\mathbf{q}) e^{-2\pi i \mathbf{q} \cdot \mathbf{k}} d^3 \mathbf{q}, \quad (1)$$

where \mathbf{k} is the displacement vector in EAP-space and \mathbf{q} is the diffusion wave-vector in signal-space. We denote $\mathbf{q} = q\mathbf{u}$ and $\mathbf{k} = k\mathbf{r}$, where \mathbf{u} and \mathbf{r} are 3D unit vectors. An interesting problem, similar to that of the hydrogen atom, is

whether a new interpretation of q -space can be obtained by stereographically projecting q -space and EAP-space onto a hypersphere.

In this paper, following the work of Fock, we seek to develop a 4D hyperspherical interpretation of q -space by focusing on four things: 1) Modeling the 3D q -space signal in terms of the 4D HSH, which is achieved by stereographically projecting 3D q -space onto the surface of a 4D hypersphere; 2) Using this single, orthonormal basis to reconstruct the diffusion orientation distribution function (ODF); 3) Computing the familiar q -space metric zero-displacement probability (P_0) and introducing a novel hyperspherical diffusivity index; and 4) Deriving the integral transform that maps from the signal-hypersphere to the EAP-hypersphere. The last point is especially significant because the integral transform describing the relationship between any two functions individually existing on some n -dimensional sphere S^n , given that the two functions are FT pairs on the $(n - 1)$ -plane, has never been derived. Lastly, we compare the HSH basis to Bessel Fourier Orientation Reconstruction (BFOR) [5], and show that HSH expansion requires less fitting parameters than BFOR to achieve comparable signal and better ODF reconstructions.

2 Methods

2.1 4D Hyperspherical Harmonics

Consider the 4D unit hypersphere S^3 existing in \mathbb{R}^4 . The Laplace-Beltrami operator on S^3 is defined as $\Delta_{S^3} = \frac{1}{\sin^2 \beta} \frac{\partial}{\partial \beta} \sin^2 \beta \frac{\partial}{\partial \beta} + \frac{1}{\sin^2 \beta} \Delta_{S^2}$, where Δ_{S^2} is the Laplace-Beltrami operator on the unit sphere S^2 . The eigenfunctions of Δ_{S^3} are the 4D HSH $Z_{nl}^m(\beta, \theta, \phi)$: $\Delta_{S^3} Z_{nl}^m = -l(l+2)Z_{nl}^m$. The 4D HSH are defined as [3]

$$Z_{nl}^m(\beta, \theta, \phi) = 2^{l+1/2} \sqrt{\frac{(n+1)\Gamma(n-l+1)}{\pi\Gamma(n+l+2)}} \Gamma(l+1) \sin^l \beta C_{n-l}^{l+1}(\cos \beta) Y_l^m(\theta, \phi), \quad (2)$$

where (β, θ, ϕ) obey $(\beta \in [0, \pi], \theta \in [0, \pi], \phi \in [0, 2\pi])$, C_{n-l}^{l+1} are the Gegenbauer (i.e. ultraspherical) polynomials, and Y_l^m are the 3D spherical harmonics. The l denotes the degree of the HSH, m is the order, and $n = 0, 1, 2, \dots$, and these three integers obey the conditions $0 \leq l \leq n$ and $-l \leq m \leq l$. The number of HSH corresponding to a given value of n is $(n+1)^2$. The HSH form an orthonormal basis on the hypersphere.

2.2 4D Stereographic Projection of q -space onto Hypersphere

In order to model the q -space signal with the HSH, we need to map 3D q -space onto a 4D hypersphere of radius ρ_0 , which is achieved via stereographic projection. The q -space coordinates are defined as $q_x = q \sin \theta \cos \phi$, $q_y = q \sin \theta \sin \phi$, and $q_z = q \cos \theta$. The coordinates of the signal-hypersphere are defined by the 4D

vector \mathbf{s} , whose components are $s_1 = p_o \sin \beta \sin \theta \cos \phi$, $s_2 = p_o \sin \beta \sin \theta \sin \phi$, $s_3 = p_o \sin \beta \cos \theta$, and $s_4 = p_o \cos \beta$. The relationship between \mathbf{q} and \mathbf{s} is then

$$s_1 = \frac{2p_o^2 q_x}{q^2 + p_o^2}, \quad s_2 = \frac{2p_o^2 q_y}{q^2 + p_o^2}, \quad s_3 = \frac{2p_o^2 q_z}{q^2 + p_o^2}, \quad s_4 = \frac{p_o(q^2 - p_o^2)}{q^2 + p_o^2} \quad (3)$$

According to Eq. (3), the center of q -space $(0, 0, 0)$ projects onto the south pole $(0, 0, 0, -p_o)$ of the hypersphere. As $q \rightarrow \infty$, the projection (s_1, s_2, s_3, s_4) moves closer to the north pole $(0, 0, 0, p_o)$. Eq. (3) establishes a one-to-one correspondence between q -space and the 4D hypersphere.

Stereographic projection exhibits two important properties. First, it is conformal, which means it preserves angles - the angles (θ, ϕ) in q -space are preserved in 4D hyperspherical space. However, stereographic projection does not preserve volume; in general, the volume of a region in the 3D plane doesn't equal the volume of its projection onto the hypersphere. In fact, the degree of volume distortion in going from a differential volume element in q -space $d^3\mathbf{q}$ to that of the hypersphere $dV = p_o^3 d\Omega$ can be shown to be

$$d^3\mathbf{q} = \left(\frac{q^2 + p_o^2}{2p_o}\right)^3 \frac{1}{p_o^3} dV = \left(\frac{q^2 + p_o^2}{2p_o}\right)^3 d\Omega = \left(\frac{p_o}{1 - \cos \beta}\right)^3 d\Omega, \quad (4)$$

where $d\Omega = \sin^2 \beta \sin \theta d\beta d\theta d\phi$.

2.3 Diffusion Signal Modeling via HSH Basis

Stereographically projecting q -space onto the hypersphere results in the q -space signal existing along the surface of the hypersphere. According to Fourier analysis, any square-integrable function defined on a sphere can be expanded in terms of the spherical harmonics. Thus, stereographic projection allows the 3D q -space signal to be expanded in terms of the HSH:

$$E_{p_o}(\beta, \theta, \phi) \approx \sum_{n=0}^N \sum_{l=0}^n \sum_{m=-l}^l C_{nlm} Z_{nl}^m(\beta, \theta, \phi), \quad (5)$$

where E_{p_o} denotes the q -space signal existing on hypersphere of radius p_o . The realness of the diffusion signal requires use of the real HSH, and so we employ a modified real basis proposed in [6] for Y_l^m .

An important axiom to state is that the q -space signal, itself, remains invariant after the mapping - that is, for a given q -space point (q_x, q_y, q_z) and its corresponding projection on the hypersphere (s_1, s_2, s_3, s_4) , $E(q_x, q_y, q_z) = E_{p_o}(s_1, s_2, s_3, s_4)$. In q -space, the diffusion signal is even i.e. $E(q_x, q_y, q_z) = E(-q_x, -q_y, -q_z)$. Evenness in q -space doesn't necessarily translate into evenness on the hypersphere. According to Eq. (3), $(-q_x, -q_y, -q_z)$ projects to $(-s_1, -s_2, -s_3, s_4)$, and so evenness in q -space is tantamount to $E_{p_o}(s_1, s_2, s_3, s_4) = E_{p_o}(-s_1, -s_2, -s_3, s_4)$ on the hypersphere, indicating that the signal is not even on the hypersphere. In other words, stereographic projection destroys

evenness. For this reason, we are free to use both the even and odd HSH. Thus, for a given truncation order N , the total number of expansion coefficients is $W = (N + 1)(N + 2)(2N + 3)/6$.

2.4 Relationship between EAP and q-space Signal on Hypersphere

Lets project 3D EAP-space onto a 4D hypersphere of radius p_o , whose coordinates are defined by the 4D vector $\mathbf{v} = \mathbf{v}(p_o, \beta', \theta', \phi')$. The Fourier relationship given in Eq. (1) between the signal and EAP does not hold true on the hypersphere. The question, then, is what integral transform maps from the signal-hypersphere to the EAP-hypersphere. We show¹, for the first time, that this integral transform is

$$P_{p_o}(\mathbf{v}) = p_o^3 \int E_{p_o}(\mathbf{s}) \frac{e^{-2\pi i[\mathbf{s} \cdot \mathbf{v} - s_4 v_4]/(1 - \cos \beta)(1 - \cos \beta')}}{(1 - \cos \beta)^3} d\Omega, \tag{6}$$

where P_{p_o} denotes the EAP existing on hypsphere of radius p_o . Eq. (6) is not one of the more familiar integral transforms encountered in mathematics literature.

2.5 HSH Metrics

A well-known q -space metric is $P_o \equiv P(\mathbf{k} = 0)$ [1, 9], which is a measure of diffusion restrictivity. $\mathbf{k} = 0$ corresponds to the south pole of the EAP-hypersphere i.e. $\beta' = \pi$. Hence using Eq. (6) and the HSH basis, we can derive a hyperspherical P_o :

$$P_o = P_{p_o}(\beta' = \pi) = \int_{\Omega \in S^3} \frac{E_{p_o}(\mathbf{s})}{(1 - \cos \beta)^3} d\Omega = \int_{\Omega \in S^3} \left(\frac{q^2 + p_o^2}{2p_o} \right)^3 E_{p_o}(\mathbf{s}) d\Omega \tag{7}$$

The integral in (7) is difficult to evaluate analytically, which is due to the non-volume-preserving nature of stereographic projection. To overcome this, we compute an uncorrected P_o by assuming q -space is uniformly projected onto the hypersphere:

$$P_{o_{unc}} = \int_{\Omega \in S^3} E_{p_o}(\Omega) dV = p_o^3 \sum_{n,l,m} C_{nlm} \int_{\Omega \in S^3} Z_{nl}^m(\Omega) d\Omega = \pi \sqrt{2} p_o^3 C_{000}, \tag{8}$$

where we use the fact that $Z_{00}^0 = \frac{1}{\pi \sqrt{2}}$. $P_{o_{unc}}$ will, naturally, suffer from volume distortion, which is corrected for by a signal weighting operation discussed in the next section.

¹ For derivation, see http://brainimaging.waisman.wisc.edu/~ameer/HSH_Suppl.pdf

Examples of q -space distance metrics include the q -space inverse variance (QIV) [5] and the mean squared displacement (MSD) [1], which are measures of diffusivity and only pertinent to planes. A useful distance metric defined on the hypersphere is the chordal distance [7], denoted χ . Consider the two q -space points $Q = (q_x, q_y, q_z)$ and $\widehat{Q} = (\widehat{q}_x, \widehat{q}_y, \widehat{q}_z)$ and their corresponding projections on the hypersphere $S = (s_1, s_2, s_3, s_4)$ and $\widehat{S} = (\widehat{s}_1, \widehat{s}_2, \widehat{s}_3, \widehat{s}_4)$, respectively. Then the Euclidean (4D) distance between the projections S and \widehat{S} on the hypersphere is given by the chordal distance, and it can be shown that

$$\chi = \chi(Q, \widehat{Q}) = \frac{2p_o^2 \sqrt{(q_x - \widehat{q}_x)^2 + (q_y - \widehat{q}_y)^2 + (q_z - \widehat{q}_z)^2}}{\sqrt{q^2 + p_o^2} \sqrt{\widehat{q}^2 + p_o^2}}, \tag{9}$$

where $\widehat{q} = \sqrt{\widehat{q}_x^2 + \widehat{q}_y^2 + \widehat{q}_z^2}$. If $\widehat{Q} = \mathbf{0}$, then the distance between the projection S and the south pole is $\chi_{sp} \equiv \chi(Q, 0) = \frac{2p_o q}{\sqrt{q^2 + p_o^2}}$. Likewise, if $\widehat{q} \rightarrow \infty$, then the distance between the projection S and the north pole is $\chi_{np} \equiv \chi(Q, \infty) = \frac{2p_o^2}{\sqrt{q^2 + p_o^2}}$. Geometrically, χ_{np} and χ_{sp} are the chords on the hypersphere that form the legs of a right triangle (by Thales' theorem), with the diameter of the hypersphere the hypotenuse. The ratio of these 2 chordal lengths illuminates the relationship between the hypersphere radius p_o and q -space: $p_o = q \frac{\chi_{np}(q)}{\chi_{sp}(q)}$ (= constant). It can be shown that $\cos \beta = \frac{\chi_{sp}^2 - \chi_{np}^2}{4p_o^2}$, which we define as the chordal squared difference (CSD). We then define the mean chordal squared difference (MCSD) as the CSD averaged over the surface of the signal-hypersphere i.e. $\langle \cos \beta \rangle$:

$$MCSD \equiv \langle \cos \beta \rangle = p_o^3 \int_{\Omega \in S^3} \cos \beta E_{p_o}(\Omega) d\Omega = \frac{\pi}{\sqrt{2}} p_o^3 C_{100}, \tag{10}$$

where we use the fact that $Z_{10}^0 = \sqrt{2} \cos \beta / \pi$. MCSD is an inherently hyperspherical metric, whereas Po is a native q -space metric.

2.6 Numerical Implementation and Estimation of ODF

Consider M diffusion signal measurements (including $b = 0$) spread across k shells in q -space. Denote \mathbf{G} as the $M \times 1$ vector representing the M measurements, \mathbf{C} the $W \times 1$ vector of unknown expansion coefficients C_{nlm} , and \mathbf{A} the $M \times W$ matrix constructed with the HSH basis. Thus, we have a simple linear model of the form $\mathbf{G} = \mathbf{A}\mathbf{C}$. This system of over-determined equations is solved via linear least squares (LLS) with Laplace-Beltrami regularization (LBR), yielding $\widehat{\mathbf{C}} = (\mathbf{A}^T \mathbf{A} + \lambda_l \mathbf{L}_{reg})^{-1} \mathbf{A}^T \mathbf{G}$, where \mathbf{L}_{reg} is the LBR diagonal matrix with entries $l^2(l + 2)^2$ along the diagonal. The regularization serves to reinforce the positivity constraint of the signal.

Using Eq. (4), we correct for the volume distortion induced in Po_{unc} by weighing each signal shell in q -space by $\left(\frac{q_i^2 + p_o^2}{2p_o}\right)^3$, where q_i is the radius of the i^{th}

shell, before signal fitting. The resulting “weighted” coefficients are then solely used for computing P_o via Eq. (8). The q -shell radii are listed in the next section. The volume weighting of each q -shell, in this case, can be viewed as a sampling density correction of the projected q -space points on the hypersphere.

Given the intricacy of Eq. (6), it is difficult to estimate the EAP analytically using the HSH framework. However, the zeroth-order diffusion ODF [8] can be numerically estimated. Lets construct a $11 \times 11 \times 11$ ($-q_{max} : \Delta q : q_{max}$) Cartesian lattice, which we map onto the 4D hypersphere via Eq. (3). Once we have computed the HSH expansion coefficients via LLS from the acquired data, Eq. (5) can then be used to estimate the signal at any location on the hypersphere, including the projected lattice points. Taking the fast Fourier transform (FFT) of the HSH-estimated signal for the lattice gives the EAP. The radial projection of the EAP then yields the ODF. Since the zeroth-order ODF is not inherently normalized, we min-max normalize it [8].

3 Experiments

The synthetic and *in vivo* datasets use a hybrid, non-Cartesian q -space sampling scheme (HYDI) [9], consisting of 7 baseline images acquired at $b = 0$ and 125 diffusion measurements spread across 5 shells in q -space. The number of encoding directions and b -value (in s/mm^2) for each shell are (6,300), (21,1200), (24,2700), (24,4800), and (50,7500); and $q_{min} = 15.79 \text{ mm}^{-1}$, $q_{max} = 78.95 \text{ mm}^{-1}$, and $\Delta q = 15.79 \text{ mm}^{-1}$.

Synthetic Data. Synthetic experiments were done the same way as in [5], with data generated via the bi-exponential mixture model. We look at two equally weighed fibers crossing at 45° , and set eigenvalues of each diffusion tensor to be $[1.6, 0.4, 0.4]e-3$, which gives $\text{FA} = 0.7071$. Monte Carlo noise simulations were performed to investigate the effect of SNR on the signal reconstruction. Five SNR levels ([10 20 30 40 80]) for the $b = 0$ image were simulated, 1000 times each, in a similar manner as in [5], and the quality of the HSH signal fit is assessed by computing the normalized mean squared error (NMSE), given by $\frac{\|S - \hat{S}\|^2}{\|S\|^2}$. The HSH parameters are $N = 4$, $p_o = 42$, and $\lambda_l = 10^{-6}$ and those of BFOR are taken from [5]. In the signal fitting, *HSH expends $W = 55$ parameters while BFOR uses 90.*

Fig. 1 displays the HSH signal fit for each shell and the corresponding ground truth in absence of noise, and shows that the HSH basis fits the diffusion signal nearly perfectly across all b -values. Fig. 2 displays the results of the noise simulations, with the NMSE plotted against SNR for each shell, and shows that the HSH and BFOR bases have a nearly identical robustness to noise. Fig. 3 shows the ground truth, HSH-estimated, and BFOR ODF profiles in absence of noise, indicating that the HSH basis successfully captures the geometry and orientation of the ODF profile. However, as with the Laplacian modeling framework in diffusion propagator imaging (DPI) [2], the HSH basis smoothens the ODF peaks. The BFOR-estimated ODF is not as accurate as that of HSH.

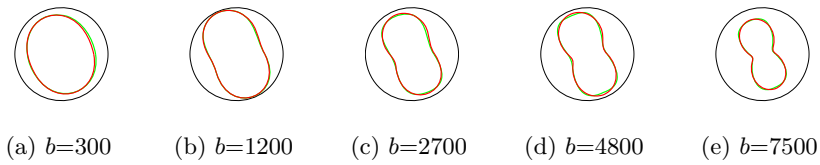


Fig. 1. The ground truth diffusion signal (green) and reconstructed signal (red) using HSH basis in absence of noise. Two equally weighted WM fibers were simulated crossing at 45° . Measurements from all 5 shells were used.

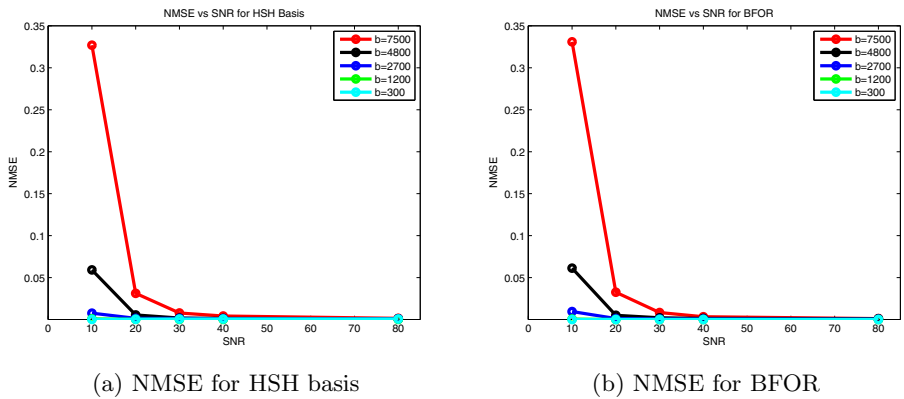


Fig. 2. The normalized mean squared error (NMSE) of the HSH signal fit, left, and BFOR signal fit, right, for each b -value plotted against SNR. 1000 noise trials were simulated for each SNR level for two equally weighted fibers crossing at 45° .

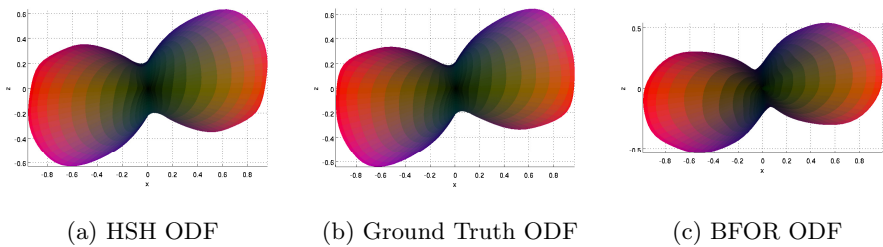


Fig. 3. The HSH-estimated ODF, ground truth ODF, and BFOR ODF in absence of noise for two equally weighted fibers crossing at 45° . The ODF is normalized to $[0\ 1]$.

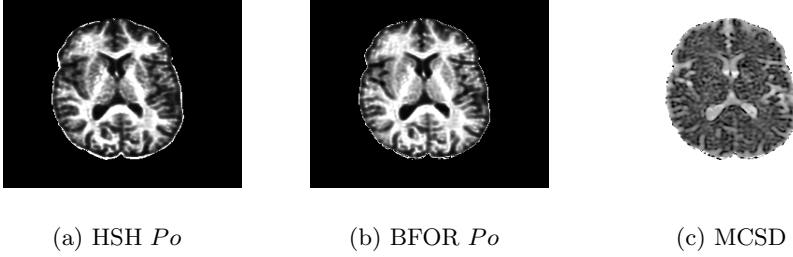


Fig. 4. Axial slices of P_o , computed via HSH and BFOR bases, and MCSD maps for a healthy, adult human

Real Data. HYDI was performed on a healthy, adult human using a 3 T GE-SIGNA whole body scanner. MR parameters were TE=102ms, TR=6500ms, FOV=24cm, matrix=96x96, voxel size=2.5x2.5mm², 43 slices with slice thickness=3mm, and scan time=15min. Diffusion parameters were $\delta = 37.86$ ms and $\Delta = 43.1$ ms.

Axial slices of P_o , computed via HSH and BFOR bases, and MCSD are shown in Fig. 4. The P_o map closely resembles BFOR's, exhibiting both tissue/CSF and WM/GM contrasts. The HSH P_o map, however, has sharper WM/GM contrast than BFOR's (compare the left and right putamen in both maps), which probably arises from the signal weighting operation. The MCSD map has tissue/CSF contrast but very little WM/GM contrast, and interestingly, closely resembles a mean squared displacement map. The MCSD can be viewed as a hyperspherical diffusivity measure, and specifically, an index of isotropic diffusion in neural tissue.

4 Discussion

We have introduced a new orthonormal basis to model the 3D q -space signal, and from which various metrics can be analytically derived. 4D HSH signal modeling allows for the capture of the radial and angular contents of the diffusion profile by a single basis function, and the basis' orthonormality provides robust numerical stability. The HSH basis' ability to give as good a signal reconstruction as BFOR and better ODF reconstruction, but with less fitting parameters, implies that it may be better suited to sparser sampling schemes than BFOR. Major drawbacks of the hyperspherical interpretation of q -space, however, are the destruction of the q -space signal's symmetry by stereographic projection and the difficulty in estimating the EAP via Eq. 6. Future work for HSH signal expansion includes implementing it on sparser q -space sampling schemes, estimating the EAP, and imposing a symmetry constraint.

References

1. Assaf, Y., Mayk, A., Cohen, Y.: Displacement imaging of spinal cord using q -space diffusion-weighted MRI. *Magn. Reson. Med.* 44, 713–722 (2000)
2. Descoteaux, M., Deriche, R., LeBihan, D., Mangin, J.F., Poupon, C.: Multiple q -shell diffusion propagator imaging. *Med. Image Anal.* 15, 603–621 (2011)
3. Domokos, G.: Four-dimensional symmetry. *Physical Review* 159, 1387–1403 (1967)
4. Fock, V.: Zur theorie des wasserstoffatoms. *Z. Physik* 98, 145–154 (1935)
5. Hosseinbor, A.P., Chung, M.K., Wu, Y.C., Alexander, A.L.: Bessel Fourier orientation reconstruction (BFOR): an analytical diffusion propagator reconstruction for hybrid diffusion imaging and computation of q -space indices. *NeuroImage* 64, 650–670 (2013)
6. Koay, C.G., Ozarslan, E., Basser, P.J.: A signal transformational framework for breaking the noise floor and its applications in MRI. *J. Magn. Reson.* 197, 108–119 (2009)
7. Saff, E.B., Snider, A.D.: *Fundamentals of Complex Analysis with Applications to Engineering and Science*, 3rd edn. Pearson Education, Inc. (2003)
8. Tuch, D.S.: Q-ball imaging. *Magn. Reson. Med.* 52, 1358–1372 (2004)
9. Wu, Y.C., Alexander, A.L.: Hybrid diffusion imaging. *NeuroImage* 36, 617–629 (2007)

Diffusion Propagator Estimation from Sparse Measurements in a Tractography Framework

Yogesh Rathi¹, Borjan Gagoski¹, Kawin Setsompop¹, Oleg Michailovich², P. Ellen Grant¹, and Carl-Fredrik Westin¹

¹ Harvard Medical School, Boston,

² Department of Electrical Engg., University of Waterloo, Canada

Abstract. Estimation of the diffusion propagator from a sparse set of diffusion MRI (dMRI) measurements is a field of active research. Sparse reconstruction methods propose to reduce scan time and are particularly suitable for scanning un-cooperative patients. Recent work on reconstructing the diffusion signal from very few measurements using compressed sensing based techniques has focussed on propagator (or signal) estimation at each voxel independently. However, the goal of many neuroscience studies is to use tractography to study the pathology in white matter fiber tracts. Thus, in this work, we propose a joint framework for robust estimation of the diffusion propagator from *sparse measurements* while simultaneously tracing the white matter tracts. We propose to use a novel multi-tensor model of diffusion which incorporates the bi-exponential radial decay of the signal. Our preliminary results on in-vivo data show that the proposed method produces consistent and reliable fiber tracts from very few gradient directions while simultaneously estimating the bi-exponential decay of the diffusion propagator.

1 Introduction

To obtain accurate information about the neural architecture, diffusion spectrum imaging (DSI) was proposed by [1]. However, this technique requires many measurements, making it impractical to use in clinical settings. Consequently, other imaging and analysis schemes, which use fewer measurements have been proposed in [2–6]. These techniques acquire important information about the neural tissue, which is missed by HARDI methods, yet, only a few of these are used in clinical studies. Traditional methods that compute the entire propagator incorporating the non-monoexponential decay of the signal, require many measurements at high b-values (greater than 3000 s/mm^2) [7, 8], making the scan time too long for un-cooperative patients. Thus, reducing the number of measurements is an important step towards making these techniques clinically viable.

Several methods have used the concept of compressed sensing to dramatically reduce the number of measurements [5, 9–12]. However, all of these methods estimate the fiber orientation distribution function (fODF) or the ensemble average propagator (EAP) at each voxel independently. Tractography is done as a post-processing step, making it susceptible to errors in estimation of the principal diffusion direction. On the other hand, most neuroscience studies require

tractography to analyze white matter fiber paths. Thus, we propose a joint framework for tractography and EAP estimation using a causal filter - the unscented Kalman filter (UKF), allowing for incorporating the correlation in water diffusion along the fiber tracts. With experiments on in-vivo data, we show that the proposed framework is rather robust to the number of measurements required, giving very similar results for sparse as well as dense set of measurements spread over two b-value shells (1000 and 4000 s/mm^2). Thus, we propose a model based framework which is an alternative to the compressed sensing based techniques in terms of the number of measurements required to accurately represent the bi-exponential decay of the diffusion signal.

2 Our Contribution

The proposed work has several novel contributions and extends the work of [13] in a significant way; *(i)* The method in [13] uses a multi-tensor model with the assumption of a mono-exponential signal decay, making it inaccurate for use with high b-values. Note that, several studies have established the bi-exponential decay of diffusion signal at high b-values [8]. Thus, we propose a novel multi-tensor bi-exponential model of diffusion to represent the signal in the entire q-domain, which also has an analytical form for computing the diffusion propagator (EAP). Further, this representation is not limited to the spherical sampling scheme as required by methods based on spherical functions. *(ii)* The proposed UKF based method is robust enough for estimation of the EAP with very few measurements, thus allowing for reducing the number of measurements required. *(iii)* Most sparsity based methods report error in estimation of the EAP at an individual voxel level, making it difficult to assess its effect on estimation of long and short range fiber tracts. To the best of our knowledge, for the first time, we compare the effect of using different number of measurements on tracing several different fiber bundles on a very high b-value data set. We also report error in estimation of return-to-origin probability (RTOP), a measure derived from EAP, for each fiber bundle. *(iv)* Further, we propose a novel probabilistic overlap metric to compute fiber bundle overlap, which is less sensitive to noise.

3 Methods

The diffusion signal $S(\mathbf{q}) : \mathbb{R}^3 \rightarrow \mathbb{R}^+$ is a real-valued function, which determines the value of S at location \mathbf{q} in q-space. Alternatively, S can also be written as a function of b and a unit vector \mathbf{u} , such that $S(b, \mathbf{u}) : \mathbb{R} \times \mathbb{S}^2 \rightarrow \mathbb{R}^+$, where $b = (2\pi q)^2 (\Delta - \frac{\delta}{3}) = (2\pi q)^2 \tau$ with δ being the duration of the gradient pulse, Δ is the time between the gradients and $q = \|\mathbf{q}\|$.

Signal Model: At low b-values, the signal decay as a function of b can be approximated by a Gaussian, however at higher b-values, the signal decay is markedly bi-exponential in nature [7, 14]. Signal from high b-values can provide subtle information about the tissue, such as, fast and slow diffusion fractions,

which cannot be obtained by assuming a mono-exponential decay. Further, as shown in [15], high b-value data is more sensitive to minor changes in the underlying tissue. Consequently, we propose a novel multi-tensor bi-exponential model to represent the diffusion signal in the entire q-space as follows:

$$S(b, \mathbf{u}) = \frac{1}{n} \sum_{i=1}^n w \exp(-b\mathbf{u}^T D_i \mathbf{u}) + (1 - w) \exp(-b\mathbf{u}^T \bar{D}_i \mathbf{u}), \quad (1)$$

where w is the weight fraction of the fast diffusing component, $D_i = \lambda_1 \mathbf{m}\mathbf{m}^T + \lambda_2 (\mathbf{p}\mathbf{p}^T + \mathbf{v}\mathbf{v}^T)$ and $\bar{D}_i = \lambda_3 \mathbf{m}\mathbf{m}^T + \lambda_4 (\mathbf{p}\mathbf{p}^T + \mathbf{v}\mathbf{v}^T)$ are the diffusion tensors, where we have assumed a cylindrical shape for the diffusion tensor as in [13]. The corresponding diffusion propagator (EAP) has the following analytical form:

$$P(\mathbf{r}) = \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{w}{(4\pi\tau)^3 |D_i|}} \exp\left(\frac{\mathbf{r}^T D_i^{-1} \mathbf{r}}{-4\tau}\right) + \sqrt{\frac{(1-w)}{(4\pi\tau)^3 |\bar{D}_i|}} \exp\left(\frac{\mathbf{r}^T \bar{D}_i^{-1} \mathbf{r}}{-4\tau}\right). \quad (2)$$

Several analytical formulae can be derived from this expression, for example, the return-to-origin probability (RTOP) $P(0)$, is simply given by:

$$P(0) = \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{w}{(4\pi\tau)^3 |D_i|}} + \sqrt{\frac{(1-w)}{(4\pi\tau)^3 |\bar{D}_i|}}. \quad (3)$$

Modeling Assumptions: From reported experimental data [14] and some *a-priori* knowledge about the human anatomy, we make certain assumptions about the model in (1), which allows us to reduce the number of unknowns while allowing for a robust estimation of the parameters. First, we assume that the fast and slow diffusing tensors D_i and \bar{D}_i have the same orientation but different shapes, i.e. different eigenvalues. This is a reasonable assumption, since the fast and slow diffusing components essentially sample the same biological tissue, albeit at different diffusion times (given by the low and high b-value data). Thus, they have the same orientation and differ only in the ‘‘amount’’ of diffusion in different directions, which can be easily represented using the two eigenvalues λ_3, λ_4 of \bar{D}_i . Similar to the work in [13, 16], in this preliminary work, we restrict our experiments to the case where $n = 2$ assuming at-most two fiber crossings. The proposed framework is however general enough to be used for three fiber crossings as well. This could be done by adaptively adding a third component if one of the two tensor components becomes planar, indicating a three-fiber crossing. Experiments done in [14] have shown that throughout the brain, the fraction of fast diffusing component is around 0.7, while the slow fraction is 0.3. We thus fix $w = 0.7$ in our work. This, however does not preclude us from modeling other diffusion fractions as can be seen by adjusting the eigenvalues: $\exp(-b\mathbf{u}^T D \mathbf{u}) = \exp(-b\mathbf{u}^T (D_{iso} + D_2) \mathbf{u}) = \alpha \exp(-b\mathbf{u}^T D_2 \mathbf{u})$, where we have decomposed a diffusion tensor D into a sum of isotropic tensor D_{iso} and an anisotropic part D_2 leading to $\alpha = \exp(-bD_{iso})$. Thus, changing the eigenvalues is equivalent to changing the volume fraction of the diffusion tensor. This also proves that bi-exponential models cannot have a unique solution. In our case, the UKF acts as a regularizer resulting in smooth parameter estimation.

Thus, the number of unknowns per tensor compartment is 7 (3 for the eigenvector \mathbf{m} , 2 eigenvalues of tensor D_i and 2 eigenvalues of tensor \bar{D}_i). For a two-fiber model, the total number of unknowns is 14. Note that, the proposed model is different from that of [7], where the restricted compartment is modeled in terms of the axonal diameter. On the other hand, the model of [14] is a general model with a separate bi-exponential fit done for each gradient direction. This model does not account for consistency in the spherical domain, making the estimation quite susceptible to noise (which is significant at high b-values). The NODDI model [17] utilizes Watson functions for representing the intra-cellular components and a customized acquisition sequence for optimal performance.

Unscented Kalman Filter: We use the unscented Kalman filter (UKF), as described in [13], for robust estimation of the parameters in (1). The UKF framework has the advantage of estimating the model parameters and performing tractography simultaneously, resulting in an inherent regularization of the model parameters and the tracts themselves. To use this state-space filter for estimating the model parameters, we define the following four filter components: (1) The system state (\mathbf{x}): the model parameters, (2) The state transition function (f): how the model changes as we trace the fiber, (3) The observation function (h): how the signal appears given a particular model state, (4) The measurement (\mathbf{y}): the actual signal obtained from the scanner.

The state \mathbf{x} of the system consists of the parameters of the model given by: $\mathbf{x} = [\mathbf{m}_1 \lambda_{11} \lambda_{12} \lambda_{13} \lambda_{14} \mathbf{m}_2 \lambda_{21} \lambda_{22} \lambda_{23} \lambda_{24}]$. As in [13], we assume the state transition function to be an identity matrix, since the diffusion signal does not change much when stepping from one location to the next within a voxel (we take very small step size). The observation function $h(\mathbf{x})$ is given by equation (1), which “predicts” the signal based on the model parameters \mathbf{x} . Finally, \mathbf{y} is the signal measured by the scanner (we perform interpolation as in [13] to obtain the signal at sub-voxel locations). The UKF allows for recursive estimation of the parameters \mathbf{x} and the confidence in the estimation of \mathbf{x} via a covariance matrix P . This is extremely important, since model parameters (and the corresponding tracts) estimated with low confidence can essentially be discarded.

To initialize the model parameters \mathbf{x} , we first estimate a single tensor using the low b-value data to obtain the principal diffusion direction \mathbf{m} . Next, the UKF is run for several iterations (typically 20) at the same location to obtain a reasonable estimate of the initial set of parameters for the bi-exponential model (1). Next, we start the tractography process, by alternately estimating the parameters \mathbf{x} and taking a small step in the direction of the principal diffusion direction, until a suitable termination criteria is reached.

4 Experiments

We performed several experiments on in-vivo human data to demonstrate the efficacy of the proposed method in computing the EAP and tracing the fiber tracts. Our data set involved 4 different acquisitions of one subject done during the same scanning session. The scans were acquired with $N = \{32, 40, 60, 120\}$

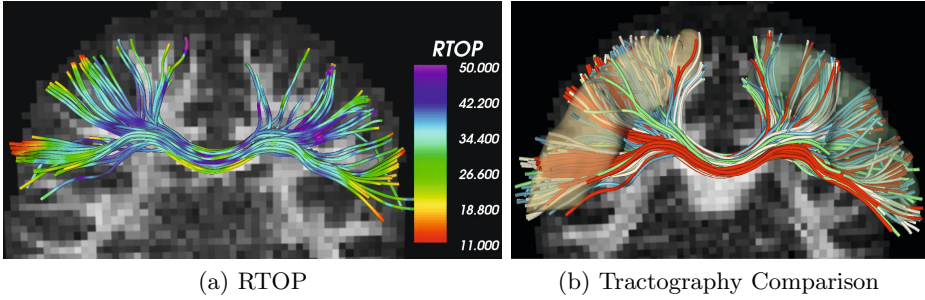


Fig. 1. (a) Tracts obtained with $N = 120$ measurements and colored using the estimated RTOP. Notice the low values (red) in the gray matter and high values (blue) in white matter, as expected. (b) Tracts generated using different number of measurements: red($N = 120$), green($N = 60$), blue($N = 40$) and white($N = 32$). Note the significant overlap between the fiber bundles. Overlap measure $B_{60} = 0.97$, $B_{40} = 0.94$, $B_{32} = 0.92$.

gradient directions at the following b-value shells: $b = \{1000, 4000\} s/mm^2$ and spatial resolution of $2.5mm^3$. We acquired different sparsely sampled scans to avoid interpolating the densely sampled data and sub-sampling it, which could potentially introduce errors and change the actual SNR (due to smoothing in the interpolation process). In addition a T1 and T2-weighted images with $1mm^3$ resolution were also acquired with a total scan time of 61 minutes. All diffusion images (of all scans) were spatially normalized and corrected for motion and eddy current distortions using FLIRT [18]. T1-image parcellation was performed with CMTK (www.nitrc.org/projects/cmtk) using the SRI24 atlas [19] and it was subsequently registered to the diffusion space.

We performed whole brain tractography on each of the four data sets, using the proposed method, by seeding 10 times (random) per voxel. Several fiber bundles connecting two regions in the SRI24 atlas were extracted for subsequent analysis. Figure 1 shows tracts connecting the precentral areas in the left and right hemisphere. On the left, we show the tracts obtained with dense sampling of 120 measurements (60 gradient directions per shell) and color coded with the return-to-origin probability (RTOP). As expected, RTOP is low in gray matter (yellow-red) and high white matter (green-blue). Figure 1 (right) also shows the precentral fiber bundles obtained using all the four sampling schemes; red($N = 120$), green($N = 60$), blue($N = 40$) and white($N = 32$). Note that, all the other fiber bundles overlap significantly with the one obtained using dense sampling (red). In Figure 2, we show two views of a part of the cortico-spinal tract (CST) obtained by selecting fibers that pass through the internal capsule. For the sake of clarity, we show results for $N = 120$ (red) and $N = 32$ (white).

Next, we provide some quantitative results on how well the proposed method traces various fiber bundles (from different scans) relative to the dense sampling method (gold standard) with $N = 120$. To quantify the overlap between two fiber bundles, we propose to use the Bhattacharyya metric B on probability distributions [20]. We compute probability distribution for each of the spatial

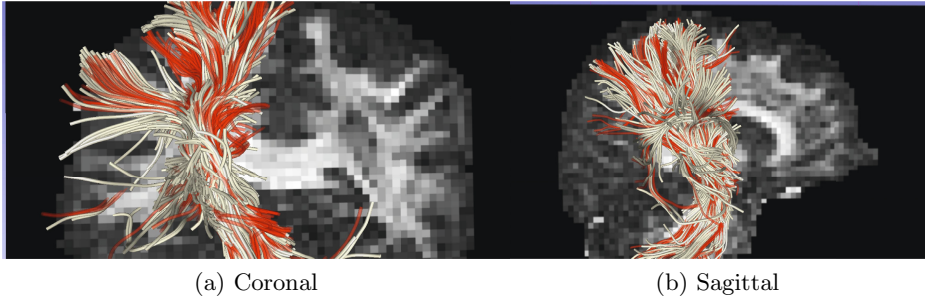


Fig. 2. Two views of the CST. Red - fibers obtained with dense sampling ($N = 120$) and white - fibers with sparse sampling ($N = 32$). Note that, despite the sparse sampling the later method traces similar regions of the brain. Overlap measure $B_{32} = 0.9$.

co-ordinates (x, y, z) of a fiber bundle (denote the probability distribution function (pdf) on the x -coordinate of the gold standard as $p_g(x)$). Then, the Bhattacharyya metric B_x is given by: $B_x = \int \sqrt{p_g(x)p(x)}dx$, where $p(x)$ is the pdf of a fiber bundle to be compared. To compute the distance between two fiber bundles, we simply take an equally-weighted combination in each co-ordinate:

$$B = \frac{1}{3} \left(\int \sqrt{p_g(x)p(x)} dx + \int \sqrt{p_g(y)p(y)} dy + \int \sqrt{p_g(z)p(z)} dz \right).$$

This metric has several advantages: 1). The values of B are bounded between 0 and 1. Thus, B will be 1 for a perfect match between two fiber bundles and 0 for no overlap at all. 2). Since the probability distributions are smooth, the metric accounts for minor deviations in tracts due to noise (as opposed to discretizing the fibers to obtain a label map for computing the dice coefficient).

We selected 66 different cortical regions from the SRI24 atlas and computed B for each of the fiber bundles that connected these regions. Fiber bundles obtained using $N = 120$ measurements were considered as the “gold standard” and distance (B) between fiber bundles obtained with sparser set of measurements were computed. Figure 3 shows the connectivity network for the three different acquisitions color coded with the distance B between the fiber bundles. Note that, B varies between 0.8 and 0.9 for all the acquisitions indicating a good overlap of the traced fibers. Further, as expected, the acquisition with $N = 120$ is closest to the gold standard, while the one with $N = 32$ has only a few fiber bundles with overlap close to 0.8.

We performed a similar analysis on the network, by computing the normalized mean error E in estimation of RTOP for each of the fiber bundles ($E = \frac{|m_g - m|}{m_g}$, where m_g and m is the mean RTOP for the gold standard and the other fiber bundle respectively). Figure 4 shows result for all the three acquisitions (with the dense one being the gold standard). In general, the percentage error in estimation of RTOP is small (around 2-3%), which is close to the variability between two different scans of the same subject. However, there are a few fiber bundles for which the percentage is around 10%, specially for the $N = 32$ acquisition data.

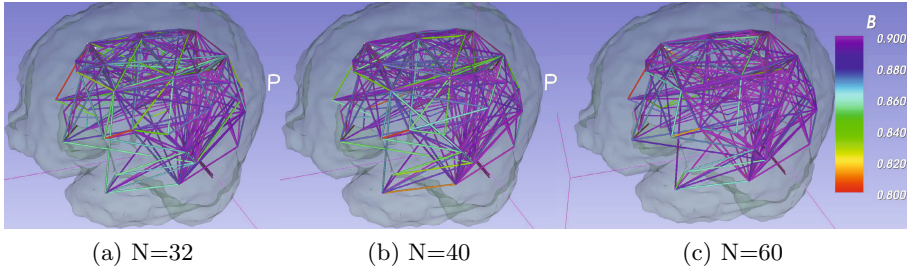


Fig. 3. Fiber bundle overlap measure B between the “gold standard” ($N = 120$) and the corresponding fiber bundles obtained from sparser acquisitions. B is greater than 0.8 in all cases (all acquisitions), and close to 0.9 in most cases.

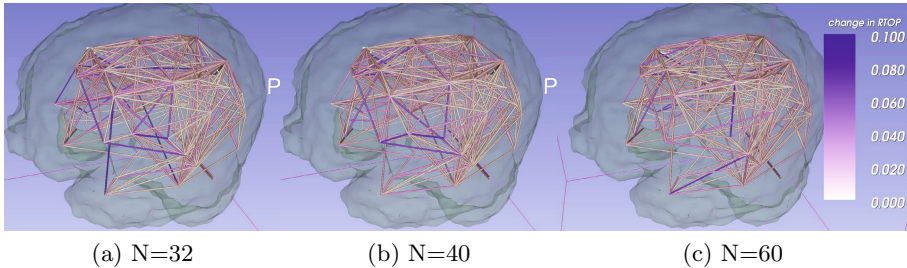


Fig. 4. Color coding reflects normalized mean error in estimation of RTOP compared to the “gold standard” in all fiber bundles traced for different acquisitions

5 Conclusion and limitations

In this work, we proposed a novel bi-exponential multi-tensor tractography framework for consistent estimation of the diffusion propagator and fiber tracts. We investigated our technique on four different data sets from the same subject but with different number of gradient directions. Our quantitative results showed that the tracts traced with sparser set of samples ($N = 32, 40, 60$) showed good overlap with those traced using dense sampling of $N = 120$. Further, the estimated diffusion measure of return-to-origin probability (RTOP) for sparser data sets was quite consistent with that obtained from the dense one. We should however note that, this work is quite preliminary in nature and essentially a proof-of-concept. Our future work entails comprehensive comparison on several subjects as well as quantifying error in different diffusion measures.

Acknowledgements. This work has been supported by NIH grants: R01MH097979, R01MH074794, P41RR013218, P41EB015902 and VR grant 2012-3682.

References

1. Wedeen, V., Hagmann, P., Tseng, W., Reese, T., Weisskoff, R.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine* 54(6), 1377–1386 (2005)

2. Wu, Y., Alexander, A.: Hybrid diffusion imaging. *NeuroImage* 36(3), 617–629 (2007)
3. Jensen, J., Helpert, J., Ramani, A., Lu, H., Kaczynski, K.: Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine* 53(6), 1432–1440 (2005)
4. Asselmlal, H.E., Tschumperlé, D., Brun, L., Siddiqi, K.: Recent advances in diffusion MRI modeling: Angular and radial reconstruction. *Medical Image Analysis* 15(4), 369–396 (2011)
5. Merlet, S., Caruyer, E., Deriche, R.: Parametric dictionary learning for modeling EAP and ODF in diffusion MRI. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III*. LNCS, vol. 7512, pp. 10–17. Springer, Heidelberg (2012)
6. Barmpoutis, A., Vemuri, B., Forder, J.: Fast displacement probability profile approximation from hardi using 4th-order tensors. In: *ISBI*, pp. 911–914 (2008)
7. Assaf, Y., Freidlin, R., Rohde, G., Basser, P.: New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine* 52(5), 965–978 (2004)
8. Mulkern, R.V., Vajapeyam, S., Robertson, R.L., Caruso, P.A., Rivkin, M.J., Maier, S.E.: Biexponential apparent diffusion coefficient parametrization in adult vs newborn brain. *Magnetic Resonance Imaging* 19(5), 659–668 (2001)
9. Landman, B.A., Bogovic, J.A., Wan, H., ElShahaby, F.E.Z., Bazin, P.L., Prince, J.L.: Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. *NeuroImage* 53, 2175–2186 (2012)
10. Gramfort, A., Poupon, C., Descoteaux, M.: Sparse DSI: Learning DSI structure for denoising and fast imaging. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II*. LNCS, vol. 7511, pp. 288–296. Springer, Heidelberg (2012)
11. Michailovich, O., Rathi, Y., Dolui, S.: Spatially regularized compressed sensing for high angular resolution diffusion imaging. *TMI* 30(5), 1100–1115 (2011)
12. Rathi, Y., Michailovich, O., Setsompop, K., Bouix, S., Shenton, M.E., Westin, C.-F.: Sparse multi-shell diffusion imaging. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part II*. LNCS, vol. 6892, pp. 58–65. Springer, Heidelberg (2011)
13. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Filtered multi-tensor tractography. *IEEE Trans. on Medical Imaging* 29, 1664–1675 (2010)
14. Clark, C.A., Le Bihan, D.: Water diffusion compartmentation and anisotropy at high b values in the human brain. *MRM* 44(6), 852–859 (2000)
15. Assaf, Y., Ben-Bashat, D., Chapman, J., Peled, S., et al.: High b-value q-space analyzed diffusion-weighted MRI: Application to multiple sclerosis. *MRM* 47(1), 115–126 (2002)
16. Behrens, T., Berg, H., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34(1), 144–155 (2007)
17. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C.: NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage* (2012)
18. Jenkinson, M., Smith, S., et al.: A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 5(2), 143–156 (2001)
19. Rohlfing, T., Zahr, N.M., Sullivan, E.V., Pfefferbaum, A.: The sri24 multichannel atlas of normal adult human brain structure. *HBM* 31(5), 798–819 (2009)
20. Kailath, T.: The divergence and bhattacharyya distance measures in signal selection. *IEEE Tran. Communication Technology* 15(1), 52–60 (1967)

Characterizing the DIstribution of Anisotropic MicrO-structural eNvironments with Diffusion-Weighted Imaging (DIAMOND)*

Benoit Scherrer¹, Armin Schwartzman², Maxime Taquet¹, Sanjay P. Prabhu¹,
Mustafa Sahin¹, Alireza Akhondi-Asl¹, and Simon K. Warfield¹

¹ Boston Children's Hospital, 300 Longwood Avenue, Boston, MA, 02115, USA

² Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA, 02115, USA

Abstract. Diffusion-weighted imaging (DWI) enables investigation of the brain microstructure by probing natural barriers to diffusion in tissues. In this work, we propose a novel generative model of the DW signal based on considerations of the tissue microstructure that gives rise to the diffusion attenuation. We consider that the DW signal can be described as the sum of a large number of individual *homogeneous spin packets*, each of them undergoing local 3-D Gaussian diffusion represented by a diffusion tensor. We consider that each voxel contains a number of large scale microstructural environments and describe each of them via a matrix-variate Gamma distribution of spin packets. Our novel model of DIstribution of Anisotropic MicrOstructural eNvironments in DWI (DIAMOND) is derived from first principles. It enables characterization of the extra-cellular space, of each individual white matter fascicle in each voxel and provides a novel measure of the *microstructure heterogeneity*. We determine the number of fascicles at each voxel with a novel model selection framework based upon the minimization of the generalization error. We evaluate our approach with numerous in-vivo experiments, with cross-testing and with pathological DW-MRI. We show that DIAMOND may provide novel biomarkers that captures the tissue integrity.

1 Introduction

Diffusion-weighted imaging (DWI) enables investigation of the brain microstructure by probing natural barriers to diffusion in tissues. Because the DWI spatial resolution is typically on the order of $6-27\text{mm}^3$, the measured DW signal in each voxel combines the signal arising from a variety of heterogeneous microstructural environments including multiple cell types, sizes, geometries and orientations and extra-cellular space. This is well known to give rise to an overall observed non-monoexponential decay [9,1,7,10]. Multiple models have been proposed to

* This work was supported in part by NIH grants 1U01NS082320, R01 NS079788-01A1, R01 EB008015, R01 LM010033, R01 EB013248, P30 HD018655, BCH TRP, R42 MH086984, UL1 TR000170 and R21 EB012177. MT was supported by F.R.S-FNRS and B.A.E.F.

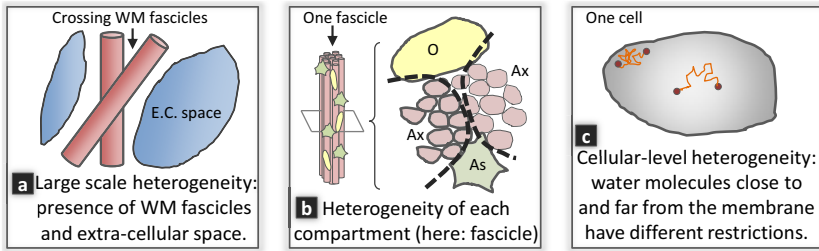


Fig. 1. Multiple scales of intra-voxel heterogeneity are responsible for the observed non-monoexponential decay (Ax: axons with various degrees of myelination; As: Astrocyte; O: Oligodendrocyte). (a): Large scale heterogeneity includes the mixing of large scale microstructural environments (LSME) such as the mixing of multiple WM fascicles with extra-cellular space. (b): Furthermore, each LSME may contain a complex varying microstructure such as axons with varying radii and degrees of myelination. (c): At an even smaller scale, other biophysical mechanisms such as intracellular heterogeneities and the proximity of cell membranes that locally restricts motion may contribute to the signal decay behavior. For example, Sehy *et al.* [8] observed a non-monoexponential decay within the intracellular space of a *single cell*, the frog oocyte.

account for the observed non-monoexponential decay. Among them, generative models focus on modeling the biophysical mechanisms underlying the MR signal formation and are of great interest to characterize the white-matter (WM) microstructure. In this context, Assaf *et al.* [1] proposed in CHARMED to represent the intra-axonal diffusion with a model inspired by the analytic diffusion in impermeable cylinders, which however required b-values up to $10000\text{s}/\text{mm}^2$ to distinguish between multiple fascicles. Zhang *et al.* [10] proposed in NODDI to represent it with a spherical Watson distribution of sticks. The appropriate model for representing each compartment, however, remains an open question.

The solution may lie in considering a more detailed model of the tissue microstructure that gives rise to the diffusion attenuation. Particularly, it is likely that the observed non-monoexponential decay arises from both *large scale* and *small scale* intra-voxel heterogeneity (see Fig.1). In [9], Yablonskiy *et al.* proposed a *statistical distribution model* of the apparent diffusion coefficient (ADC) that intrinsically reflects the presence of heterogeneous micro-structural environments in each voxel. They assumed that the DW signal in a voxel can be described as a sum of signals from a large number of individual spin packets, each of them undergoing local isotropic Gaussian diffusion described by an ADC D . Originally mono-directional, this model was extended to the multi-directional case by estimation of one ADC per direction. This model, however, does not capture the anisotropic diffusion observed in the brain. It cannot characterize the restricted diffusion such as occurs in dense WM fascicles. A generalization of [9] may be achieved by representing each spin packet with a full diffusion tensor \mathbf{D} . This, however, is analytically challenging because it implies the integration of a matrix-variate distribution of probability defined over the set of symmetric

positive-definite (SPD) matrices. Basser *et al.* [2] proposed a normal distribution for symmetric matrices that is however not restricted to SPD matrices.

In contrast, a natural distribution for SPD matrices is the matrix-variate Gamma distribution, which generalizes the Wishart distribution by allowing a non-integer number of degrees of freedom. In [5], a mixture of Wishart distributions with prespecified degree of freedom was used to discretize the manifold of the fascicle orientation distribution in a spherical deconvolution (SD) approach, and was shown to successfully capture the fascicle orientation. SD, however, relies on the definition of a prespecified convolution kernel that is assumed constant for all the brain. Therefore, variations of the fascicles microstructure (Fig.1b) are conflated with variations of the estimated mixing proportions, and SD cannot provide an indicator of the WM microstructure. Additionally, SD relies on an acquisition with a single non-zero b-value, and water molecules with very different restrictions such as water molecules in the extra-cellular space and in the intra-axonal space cannot be distinguished.

In contrast, a *generative model* based upon the 3-D generalization of the approach in [9] together with the acquisition of multiple non-zero b-values will enable characterization of both the WM structure and microstructure. However, unlike [5], this requires the identification of the appropriate model complexity, which is a challenging model order selection problem. In the literature, most approaches such as the Bayesian Information Criterion (BIC), the F-Test or the Bayesian Automatic Relevance Determination (ARD) focus on assessing the fitting error of each model while penalizing complex models to avoid overfitting. However, the choice of a penalization strategy and the trade-off between penalization and quality of fit are rather arbitrary and produce highly variable results. In contrast, generative models are *predictive models*, and a natural measure to identify the appropriate model complexity is the *generalization error* (GE). It describes how well a model can predict *new* data not included in the estimation. Typically, a model not complex enough to represent a dataset will have a large GE, and so will a too complex model so that it overfits the data. The GE, however, cannot be computed directly and must be approximated. Leave-one-out cross-validation provides an estimate with low bias but large variance, leading to high root mean squared errors [3]. *K*-fold cross-validation provides an estimator with lower variance but increased bias. Instead, the .632 bootstrap approach of [3] has been shown to provide low bias and low variance.

In this work, we propose a statistical distribution model of the diffusion in which we model the signal arising from each spin-packet with a 3-D diffusion tensor and the presence of multiple large scale microstructural environments in each voxel with a mixture of peak-shaped matrix-variate Gamma distribution of spin-packets. This has analytical solution and enables us to derive a novel generative model that describes the Distribution of Anisotropic Microstructural Environments with DWI (DIAMOND). Our model is derived from first principles and allows for the representation of both unrestricted diffusion and multiple fascicles with heterogeneous orientations, while providing a novel measure of heterogeneity of the microstructure. We determine the number of fascicles at

each voxel with a novel model selection framework based upon the minimization of the generalization error estimated with the bootstrap .632 approach [3,6]. We evaluate our approach with numerous in-vivo experiments, with cross-testing and with pathological DW-MRI. Importantly, we show that it may provide a novel biomarker that reflects the WM microstructure integrity.

2 Theory and Methods

A Generative Model of the Diffusion Signal. Following the ADC approach of [9], we consider that the measured signal can be described by a sum of signals arising from a large number of individual spin packets within the voxel. In contrast to [9], we consider that each spin packet undergoes *homogeneous 3-D* Gaussian diffusion represented by a diffusion tensor \mathbf{D} , whose contribution for a diffusion gradient \mathbf{g}_k is : $S_0 \exp(-b_k \mathbf{g}_k^T \mathbf{D} \mathbf{g}_k) d\mathbf{D}$. The fraction of spin packets described by a same \mathbf{D} in the voxel is given by a *matrix-variate* distribution $P(\mathbf{D})$, leading to the signal generation model :

$$S_k = S_0 \int_{\mathbf{D} \in \mathcal{S}_3^+} P(\mathbf{D}) \exp(-b_k \mathbf{g}_k^T \mathbf{D} \mathbf{g}_k) d\mathbf{D}, \tag{1}$$

where \mathcal{S}_3^+ is the set of 3×3 SPD matrices. If a voxel was composed of exactly a *single homogeneous* microstructural environment (ME) characterized by exactly \mathbf{D}_0 , $P(\mathbf{D})$ could be modeled by a matrix Dirac delta function $P(\mathbf{D}) = \delta(\mathbf{D} - \mathbf{D}_0)$ and our model is equivalent to DTI. If it were to contain several exactly identifiable ME, a mixture of delta functions could be used. However, it is more realistic to consider that a voxel contains multiple large-scale microstructural environments (LSME) (Fig. 1), each of them having some degree of heterogeneity.

We consider that a voxel contains N LSMEs and we model the composition of each LSME j with a matrix-variate Gamma probability distribution $P_{p^j, \Sigma^j}(\mathbf{D})$ of spin packets. Specifically, a random matrix $\mathbf{D} \in \mathcal{S}_3^+$ has a matrix-variate Gamma distribution with shape parameters $p^j > 1$ and $\Sigma^j \in \mathcal{S}_3^+$ if it has density:

$$P_{p^j, \Sigma^j}(\mathbf{D}) = \frac{|\mathbf{D}|^{p^j-2}}{|\Sigma^j|^{p^j} \Gamma_3(p^j)} \exp(-\text{trace}(\Sigma^{j-1} \mathbf{D})), \tag{2}$$

where Γ_3 is the 3-variate gamma function and $|\cdot|$ the matrix determinant. The distribution P_{p^j, Σ^j} is a peak-shaped distribution. Its expected value is $\mathbf{D}_0^j = p^j \Sigma^j$ and describes here *the average diffusivity* of the LMSE j . The shape parameter p^j determines the concentration of the distribution, the density (2) becoming more concentrated about \mathbf{D}_0^j as p^j increases. This captures the *microstructural heterogeneity* of each LMSE j . We consider that the LSMEs are in slow exchange by considering $P(\mathbf{D}) = \sum_{j=0}^N f_j P_{p^j, \Sigma^j}(\mathbf{D})$ where $f_j \in [0, 1]$ are the volume fractions of occupancy and sum to one, leading to:

$$S_k = S_0 \sum_{j=0}^N f_j \int_{\mathbf{D} \in \mathcal{S}_3^+} P_{p^j, \Sigma^j}(\mathbf{D}) \exp(-b_k \mathbf{g}_k^T \mathbf{D} \mathbf{g}_k) d\mathbf{D} \tag{3}$$

The integrals in the right-hand side of (3) are Laplace transforms of $P_{p^j, \Sigma^j}(\mathbf{D})$, which have a known analytical expression [4]. This leads to the generative model:

$$S_k = S_0 \sum_{j=0}^N f_j \exp \left(-p_j \log \left(1 + \frac{b_k \mathbf{g}_k^T \mathbf{D}_0^j \mathbf{g}_k}{p_j} \right) \right) \quad (4)$$

Using the Taylor expansion $-p \log(1 + \frac{u}{p}) = \sum_{l=1}^{\infty} \frac{(-1)^l}{l} \frac{u^l}{p^{l-1}}$ about $u = 0$ it follows that: $S_k = S_0 \sum_{j=0}^N f_j \exp \left(-b_k \mathbf{g}_k^T \mathbf{D}_0^j \mathbf{g}_k + \frac{1}{2p_j} (b_k \mathbf{g}_k^T \mathbf{D}_0^j \mathbf{g}_k)^2 - \frac{1}{3p_j^2} (b_k \mathbf{g}_k^T \mathbf{D}_0^j \mathbf{g}_k)^3 + \dots \right)$.

It shows that when $p^j \rightarrow \infty$ for all j , which corresponds to infinitely narrow $P_{p^j, \Sigma^j}(\mathbf{D})$'s, our model is equivalent to the multi-tensor model. In contrast, finite values of p^j captures the heterogeneity of each LMSE. Note that the decay rate decreases as the b-value increases, modeling a non-monoexponential decay.

Model Order Selection for Generative Models. We present our novel model order selection approach based on the minimization of the generalization error (GE). The model (4) is a generative model that relates input parameters x_k (the diffusion sensitization direction and strength) to output measurements y_k (the diffusion attenuation). Denoting by $\mathbf{z} = \{z_1, \dots, z_n\}$ with $z_i = (x_i, y_i)$ the set of n training data, by $\mathcal{G}_{\mathbf{z}}(x)$ the model whose parameters were estimated with \mathbf{z} , and by $z_0 = (x_0, y_0)$ a new hypothetical data point, the GE conditional on the observed data is :

$$E_g | \mathbf{z} = \mathbb{E}_{z_0 \sim F} [|y_0 - \mathcal{G}_{\mathbf{z}}(x_0)|^2 | \mathbf{z}], \quad (5)$$

where $\mathbb{E}[\cdot]$ is the statistical expectation and $z_0 \sim F$ indicates that the expectation is taken over the new data point that follows some distribution F . To account for the variability of the observed data points, the unconditional GE can be defined as the expectation of (5) over all \mathbf{z} : $E_g = \mathbb{E}_{z_i \stackrel{\text{iid}}{\sim} F} \{E_g | \mathbf{z}\}$. We propose to estimate E_g with the .632 bootstrap approach [3]. It counter-balances the positive bias of the leave-one-out bootstrap estimate \hat{E}_g^{BS} by the negative bias of the fitting error estimate \hat{E}_g^{fit} , by assessing: $\hat{E}_g^{.632} = 0.368 \hat{E}_g^{\text{fit}} + 0.632 \hat{E}_g^{\text{BS}}$. The 0.632 coefficient comes from that, on average, \hat{E}_g^{BS} uses $[1 - (1 - \frac{1}{n})^n]n$ data point at each bootstrap iteration, which is approximately equal to 0.632 for large n . We refer to [3] for details of the expressions of \hat{E}_g^{fit} and \hat{E}_g^{BS} . As in [6], we first consider a model with a single compartment and then progressively increase the model complexity as long as it provides a statistically significant decrease in GE.

Methods. At each voxel, we considered one matrix-variate Gamma distribution with isotropic $\mathbf{D}_0^{\text{iso}} = \text{diag}(3 \times 10^{-3}) \text{mm}^2/\text{s}$ to model the diffusion of unrestricted water and up to 3 matrix-variate Gamma distributions with tensor \mathbf{D}_0^j to represent up to three fascicles. The .632 bootstrap model order selection was performed with $B = 30$ bootstrap iterations. Similarly to [7], the model parameters were estimated using a maximum *a posteriori* approach by considering a diffusion model with gradually increasing complexity, from the ball-and-stick model to the full DIAMOND model.

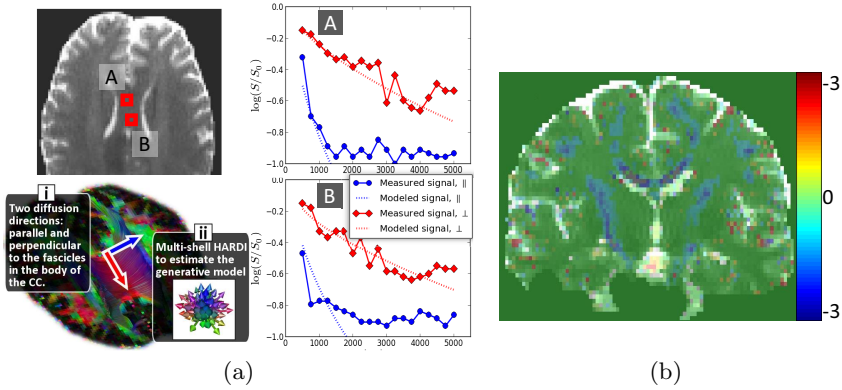


Fig. 2. (a) Plots of $\log(S_k/S_0)$. Our model captures the non-monoexponential decay observed in a region of a single fascicle direction. Note that the plotted data points were not used for the model estimation. (b) Cross-testing evaluation: difference between the mean-square prediction error of DIAMOND and MTM (lower is better).

Evaluation of the benefits of DIAMOND with actual MR measurements is challenging because we cannot rely on any ground truth providing the distribution of MEs in each voxel. First, we performed an experiment to illustrate that our model captures the non-monoexponential decay. In vivo imaging was carried out on a healthy volunteer using a Siemens 3T Trio scanner with a 32 channel head coil and the following parameters : FOV=220mm, 68 slices, matrix= 128×128 , resolution= $1.72 \times 1.7 \times 2 \text{ mm}^3$. We focused on imaging the body of the corpus callosum (see Fig.2), a region known to contain a single fascicle orientation. We measured the diffusion attenuation in both the parallel and perpendicular directions with respect to the fascicles (Fig.2i), with various b-values from 500 to 5000 by increments of 250. The number of repetition for each b-value was determined to ensure uniform SNR across b-values, resulting in a total of 548 DW images. We also imaged a multi-shell (Fig.2ii) with 95 DW-images (5 b=0, 30 b=1000 and 15 images at each of b=1500, 2000, 2500, 3000). The multi-shell HARDI was utilized to estimate the parameters of our model. We then compared the diffusion decay predicted by DIAMOND to the actual measured diffusion decay.

To further characterize DIAMOND, we performed a cross-testing analysis. This procedure consists in repeatedly splitting the set of DW images into a random *estimation set* and *testing set*, estimating the parameters with the former and evaluating the performance on the latter. This measures the prediction performance and objectively characterizes how well a model captures a phenomenon. This, however, requires a large number of measurements. We performed a multi-shell acquisition with 395 images (5b = 0 and 15 shells of 26 directions with $b \in [200, 3000]$ by increments of 200). We repeated the estimation-testing process 100 times, using at each iteration 70% of the data for estimation and 30% for testing. We computing the mean-square prediction error at each voxel across the iterations. We compared DIAMOND to the multi-tensor model (MTM), which corresponds to using infinitely narrow distributions ($p^j = \infty$).

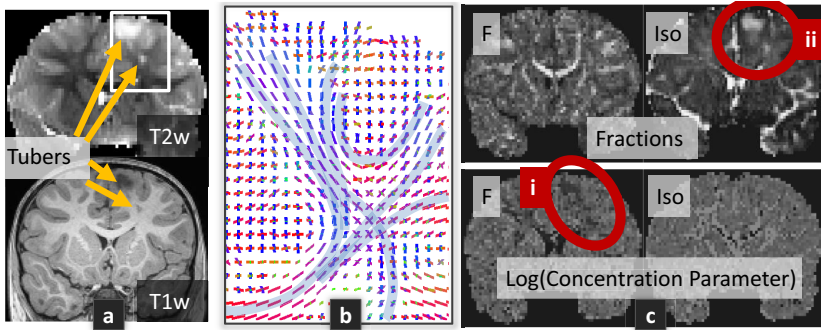


Fig. 3. TSC patient with CUSP65 imaging. Particularly, it shows that the orientation of the estimated fascicles (b) and the fractions of occupancy (c) correctly matches the known anatomy, while only 65 DW-images were acquired.

Finally, a great potential of assessing the distribution of MEs in the brain is the potential derivation of novel bio-markers that reflect the tissues integrity. We imaged a patient with Tuberosus Sclerosis Complex (TSC), a genetic disorder characterized by the presence of benign tumors in the brain called cortical tubers. 65 DW-images were acquired with a CUSP65 (CUbe and SPHERE) gradient encoding set [7], which achieves multiple b-values and directions with short echo time and high SNR. The data acquisition protocol was approved by the IRB.

3 Results

Fig 2a shows that DIAMOND successfully captures the non-monoexponential decay observed in the body of the corpus callosum. Fig 2b demonstrates that the cross-testing error is qualitatively lower with DIAMOND than with MTM. Quantitatively, a paired t-test on the differences between the testing errors at each voxel shows that DIAMOND is significantly better than MTM ($p < 10^{-8}$) with a mean error decreased by over 8%. Finally, Fig 3 reports DIAMOND imaging of a TSC patient. It shows decreased concentration parameter p_j (i) and increased fraction of unrestricted diffusion (ii) in the region of the tuber.

4 Discussion

We proposed a generative model motivated by biophysical considerations of the microstructure that gives rise to the DW signal. Inspired by the approach of [9], we considered that the signal in a voxel is the sum of the signal arising from a large number of homogeneous spin packets within each voxel. In contrast to [9], we considered that each spin packet locally undergoes 3-D Gaussian diffusion described by a diffusion tensor, capturing the 3-D geometrical structure of the local restrictions to water diffusion. We formulated the DIAMOND generative

model (4) which describes each large-scale microstructural environment (LSME) in the voxel with a matrix-variate Gamma distribution of spin packets. The concentration of each distribution was estimated, providing a novel measure of the microstructural homogeneity. Interestingly, DIAMOND is equivalent to the multi-tensor model when the distributions are infinitely concentrated. Unlike [5,10], our model does not rely on a convolution kernel with prespecified diffusivity. In contrast to [10], we have considered multiple fascicles per voxel (up to 3). We employed a novel model order selection approach based on the minimization of the generalization error. Using moderate b-values ≤ 3000 s/mm² (unlike [1]), we showed that both the estimated number of fascicles and fascicle orientations matches the known anatomy, even with a moderate number of DW images (Fig 3b). We showed that DIAMOND captures the non-monoexponential decay (Fig 2a) and better captures the underlying biophysical mechanisms underlying the DW signal formation compared to the MTM (Fig 2b). Interestingly, DIAMOND imaging in a patient with TSC showed that, in the region of the tuber, the estimated fraction of unrestricted diffusion is increased (Fig 3c.ii). This might reflect an increased extra-cellular space, the presence of perivascular spaces, or the presence of giant cells typically observed in TSC brain specimens. Importantly, we observed a reduction in the concentration parameter for the fascicle located in the tuber (Fig 3c.i), indicating an increased *anisotropic* heterogeneity *consistent with the orientation of the fascicle*. In contrast, there was no significant heterogeneity consistent with unrestricted diffusion. We speculate that this may reflect heterogeneous myelination or heterogeneous mixture of glial cells as observed in mice models of TSC. In future work we will compare DIAMOND to NODDI and CHARMED with cross-testing, and investigate the possibility of characterizing different types of tubers in TSC. DIAMOND imaging may enable novel investigations in both normal development and in clinical practice.

References

1. Assaf, Y., Basser, P.J.: Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *Neuroimage* 27(1), 48–58 (2005)
2. Basser, P.J., Pajevic, S.: A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. *IEEE T. Med Imaging* 22(7), 785–794 (2003)
3. Efron, B., Tibshirani, R.: Improvements on cross-validation: The .632 + bootstrap method. *Journal of the American Statistical Association* 92(438), 548–560 (1997)
4. Gupta, A.K., Nagar, D.K.: *Matrix Variate Distributions*. Chapman & Hall/CRC, Boca Raton (2000)
5. Jian, B., Vemuri, B.C., Ozarslan, E., Carney, P.R., Mareci, T.H.: A novel tensor distribution model for the diffusion-weighted MR signal. *Neuroimage* 37(1), 164–176 (2007)
6. Scherrer, B., Taquet, M., Warfield, S.K.: Reliable Selection of the Number of Fascicles in Diffusion Images by Estimation of the Generalization Error. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) *IPMI 2013*. LNCS, vol. 7917, pp. 742–753. Springer, Heidelberg (2013)
7. Scherrer, B., Warfield, S.K.: Parametric Representation of Multiple White Matter Fascicles from Cube and Sphere Diffusion MRI. *PLoS ONE* 7(11) (2012)

8. Sehy, J.V., Ackerman, J.J., Neil, J.J.: Evidence that both fast and slow water ADC components arise from intracellular space. *Magn. Reson. Med.* 48, 765–770 (2004)
9. Yablonskiy, D.A., Bretthorst, G.L., Ackerman, J.J.: Statistical model for diffusion attenuated MR signal. *Magn. Reson. Med.* 50(4), 664–669 (2003)
10. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C.: NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* 61(4), 1000–1016 (2012)

A Generative Model for Resolution Enhancement of Diffusion MRI Data

Pew-Thian Yap*, Hongyu An, Yasheng Chen, and Dinggang Shen

Department of Radiology and Biomedical Research Imaging Center (BRIC)
The University of North Carolina at Chapel Hill, U.S.A.
{ptyap,hongyu_an,yasheng_chen,dgshen}@med.unc.edu

Abstract. The advent of diffusion magnetic resonance imaging (DMRI) presents unique opportunities for the exploration of white matter connectivity *in vivo* and non-invasively. However, DMRI suffers from insufficient spatial resolution, often limiting its utility to the studying of only major white matter structures. Many image enhancement techniques rely on expensive scanner upgrades and complex time-consuming sequences. We will instead take a post-processing approach in this paper for resolution enhancement of DMRI data. This will allow the enhancement of existing data without re-acquisition. Our method uses a generative model that reflects the image generation process and, after the parameters of the model have been estimated, we can effectively sample high-resolution images from this model. More specifically, we assume that the diffusion-weighted signal at each voxel is an agglomeration of signals from an ensemble of fiber segments that can be oriented and located freely within the voxel. Our model for each voxel therefore consists of an arbitrary number of signal generating fiber segments, and the model parameters that need to be determined are the locations and orientations of these fiber segments. Solving for these parameters is an ill-posed problem. However, by borrowing information from neighboring voxels, we show that this can be solved by using Markov chain Monte Carlo (MCMC) methods such as the Metropolis-Hastings algorithm. Preliminary results indicate that our method substantially increases structural visibility in both subcortical and cortical regions.

1 Introduction

Diffusion magnetic resonance imaging (DMRI) [3] is a key imaging technique for the investigation and characterization of white matter pathways in the brain. It probes water diffusion in various directions and at various diffusion scales to characterize micro-structural compartments that are much smaller than the voxel size. However, limited by today's imaging technique, the typical $(2\text{mm})^3$ resolution of DMRI is too coarse to sufficiently capture the subtlety of neuronal axons, diameters of which range from $1\ \mu\text{m}$ to $30\ \mu\text{m}$ [3, 9, 11]. This causes significant partial volume effect since the signal collected at each voxel is likely to be

* Corresponding author.

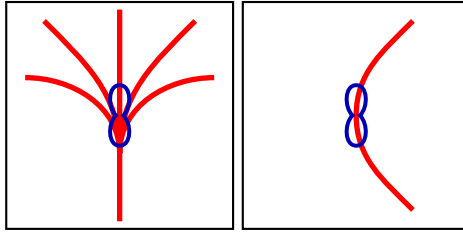


Fig. 1. Ambiguity Due to Partial Volume Effect. Radically different subvoxel fiber configurations (red) can result in similar fiber ODF shapes (blue).

due to multiple fascicles that concurrently traverse the voxel. Acquiring images with resolution higher than the typical $(2\text{mm})^3$, however, is extremely difficult without incurring unrealistic scan times and causing very low SNR due to reduced voxel size [9]. The impact of noise is aggravated in high angular resolution diffusion imaging (HARDI), which often requires prolonged echo time (TE) to achieve relatively high diffusion weighting.

In this paper, we propose to harness the rich connectivity information afforded by DMRI for estimating a generative model that best explains the observed data. By sampling from this model using a resolution that is higher than the acquisition resolution, high-resolution images can then be generated. More specifically, we assume that the diffusion-weighted signal at each voxel is an agglomeration of signals from an ensemble of fiber segments that can be oriented and located freely within the voxel. Our model for each voxel therefore consists of an arbitrary number of signal generating fiber segments, and the model parameters that need to be determined are the locations and orientations of these fiber segments.

Solving these parameters is an ill-posed problem. For example, Fig. 1 illustrates that, due to the symmetrical nature of diffusion-weighted MR measurements, the resulting fiber orientation distribution functions (ODFs) are symmetric and do not distinguish between curving and fanning fiber configurations. Put differently, even though the fiber segments that form the fibers traversing this voxel are located and oriented in radically different configurations, the signal observed within this voxel cannot be used to disambiguate between the configurations, let alone be used to estimate the configurations of the fiber segments. One viable solution to this is to gather information from neighboring, anatomically connected voxels to regularize the problem. The continuous nature of the fiber trajectories provides subvoxel information that can help super-resolve the voxel. This provides a powerful mechanism that allows us to collapse measurements across multiple voxels to estimate micro-structural properties with spatial resolution that is finer than the voxel dimensions. We will show that the associated problem can be solved by using the Metropolis-Hastings algorithm [2, 5], a Markov chain Monte Carlo (MCMC) method that is well suited for solving high-dimensional problems.

2 Approach

We assume that the signal at each voxel is an agglomeration of signals from constituent fiber segments that reside within the voxel. Formally, denoting the signal at location \mathbf{x} and gradient direction \mathbf{g} as $E(\mathbf{x}, \mathbf{g}) = S(\mathbf{x}, \mathbf{g})/S_0(\mathbf{x})$, we define the generative model of this signal in the form of spatio-angular decomposition:

$$E(\mathbf{x}, \mathbf{g}) = \sum_i G(\mathbf{x}, \mathbf{x}_i)E_i(\mathbf{g}) + \varepsilon(\mathbf{x}, \mathbf{g}) \quad \text{s.t. } \mathcal{M}. \quad (1)$$

Our task is to solve for the parameters of model \mathcal{M} by minimizing error $\varepsilon(\mathbf{x}, \mathbf{g})$ and at the same time imposing some form of regularity on the solution, i.e.,

$$\min_{\mathcal{M}} [\Theta(\mathcal{M})] = \min_{\mathcal{M}} \left\{ \int_{\Omega_{\mathbf{x}} \times \Omega_{\mathbf{g}}} |E(\mathbf{x}, \mathbf{g}) - \sum_i G(\mathbf{x}, \mathbf{x}_i)E_i(\mathbf{g})|^2 d\mathbf{x}d\mathbf{g} + \gamma\Phi(\mathcal{M}) \right\}. \quad (2)$$

Here $G(\mathbf{x}, \mathbf{x}_i) = e^{-|\mathbf{x}-\mathbf{x}_i|^2/\sigma^2}$ models the smoothing effect owing to signal averaging and $\Phi(\cdot)$ is a regularization term that enforces a certain degree of cross-voxel smoothness in the model \mathcal{M} , which encodes the fiber configuration. The predicted signal is generated via an ensemble of constituent signals $\{E_i(\cdot)\}$, each assumed to be generated by the i -th fiber segment located at \mathbf{x}_i . γ is a tuning parameter that balances the two terms. The L_2 -norm is evaluated over all locations (denoted by set $\Omega_{\mathbf{x}} \subset \mathbb{R}^3$) and gradient directions (denoted by set $\Omega_{\mathbf{g}} = \mathbb{S}^2$), i.e., the position-orientation space (POS). The goal here is to determine the parameters (i.e., locations and orientations) of the fiber segments that will result in a configuration that can best explain the observed signal $E(\mathbf{x}, \mathbf{g})$. Unlike approaches such as spherical deconvolution [10] that seek to decompose the signal at each voxel on the \mathbb{S}^2 domain; our framework seeks a $\mathbb{R}^3 \times \mathbb{S}^2$ decomposition. The additional spatial component provides sub-voxel information that is important for resolution enhancement. Note that since we are in practice only concerned with white matter, the observed and predicted signals in (2) are first centralized by removing their means so that isotropic diffusion will not affect the outcome; a similar approach was used in [4, 7].

2.1 Signal Generating Fiber Segments

Our method assumes that an arbitrary number of fiber segments can reside within a voxel space. Each fiber segment is represented by a cylinder that contributes a signal typical of parallel fibers within the voxel it resides. Each cylinder (see Figure 2) is defined by the tuple $h_i = (\mathbf{x}_i, \mathbf{v}_i, l_i, d_i)$. The three-dimensional vector \mathbf{x}_i specifies the center of the cylinder, and \mathbf{v}_i is a unit vector that defines its orientation. The length l_i and diameter

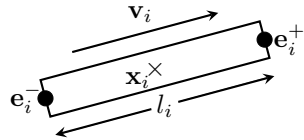


Fig. 2. A Fiber Segment. The signal at each voxel is assumed to be an accumulation of signal generated by an ensemble of fiber segments.

d_i are predefined and are identical for all cylinders (i.e., $l_i = l$, $d_i = d$, $\forall i$). The two ends \mathbf{e}_i^+ and \mathbf{e}_i^- of the cylinder h_i are determined by $\mathbf{e}_i^+ = \mathbf{x}_i + \frac{l_i}{2}\mathbf{v}_i$ and $\mathbf{e}_i^- = \mathbf{x}_i - \frac{l_i}{2}\mathbf{v}_i$. The cylinders can be connected at the ends, e.g., $\mathbf{e}_i^+ = \mathbf{e}_{i+1}^-$ and $\mathbf{e}_i^- = \mathbf{e}_{i-1}^+$. Each fiber segment i is assumed to generate signal based on a single-tensor model, i.e., $E_i(\mathbf{g}) = we^{-b\mathbf{g}^T\mathbf{D}_i\mathbf{g}}$, where b is the diffusion weighting, $\mathbf{D}_i = \lambda_1\mathbf{v}_i\mathbf{v}_i^T + \lambda_2\mathbf{I}$ denotes a diffusion tensor with diffusivities λ_1 and λ_2 and principal diffusion direction \mathbf{v}_i , which is equivalent to the orientation of the fiber segment. Constant $w \in \mathbb{R}^+$ controls the amount of signal contribution from each fiber segment.

2.2 Fiber Continuity as Regularization

The regularization term $\Phi(\mathcal{M})$ is important to ensure that the otherwise ill-posed minimization of $\Theta(\mathcal{M})$ in (2) is tractable. The model \mathcal{M} here consists of a set of fiber segments $\mathcal{H} = \{h_i\}$ as well as their connections $\mathcal{E} = \{(\mathbf{e}_i^{\alpha_i}, \mathbf{e}_j^{\alpha_j})\}$ with $\alpha \in \{+, -\}$. Smooth transition in both location and orientation is expected from fiber segments that are connected. Similar to [8], this is enforced by defining

$$\Phi(\mathcal{M}) = \sum_{(\mathbf{e}_i^{\alpha_i}, \mathbf{e}_j^{\alpha_j}) \in \mathcal{E}} \frac{1}{l^2} (\|\mathbf{e}_i^{\alpha_i} - \bar{\mathbf{x}}_{ij}\|^2 + \|\mathbf{e}_j^{\alpha_j} - \bar{\mathbf{x}}_{ij}\|^2) - L, \quad (3)$$

where $\bar{\mathbf{x}}_{ij} = \frac{\mathbf{x}_i + \mathbf{x}_j}{2}$ is the midpoint of the line connecting the centers of the i -th and j -th fiber segments. Parameter L controls the likeliness of connections. A large L causes two segments to be connected with higher likeliness. The first two terms of the above equation encourages the fiber segments to be close to each other, but not closer than the length of the segments. The fiber segments are also encouraged to be aligned by penalizing the misalignment of \mathbf{v}_i and \mathbf{v}_j .

2.3 The Metropolis-Hastings Algorithm

We minimize $\Theta(\mathcal{M})$ in (2) by utilizing the Metropolis-Hastings algorithm [2, 5]. This is equivalent to maximizing the posterior distribution $P(\mathcal{M}|\mathcal{D})$. That is, we need to determine the most probable \mathcal{M} given the observed data \mathcal{D} . The core idea of the Metropolis-Hastings algorithm is to update the model from \mathcal{M} to \mathcal{M}' based on a proposal distribution $Q(\mathcal{M}'|\mathcal{M})$ and accept the model update with probability $\min(1, R)$, where R is the so-called Green's ratio

$$R = \left(\frac{P(\mathcal{M}'|\mathcal{D})}{P(\mathcal{M}|\mathcal{D})} \right)^{\frac{1}{T}} \frac{Q(\mathcal{M}|\mathcal{M}')}{Q(\mathcal{M}'|\mathcal{M})}. \quad (4)$$

Temperature T is progressively lowered in a manner similar to simulated annealing so that the estimated posterior distribution can progressively become sharper and more defined. The algorithm works best if the proposal density $Q(\mathcal{M}'|\mathcal{M})$ matches the shape of the target distribution $P(\mathcal{M}'|\mathcal{D})$ from which direct sampling is difficult, that is $Q(\mathcal{M}'|\mathcal{M}) \approx P(\mathcal{M}'|\mathcal{D})$. We follow the approach outlined

in [8] to construct a proposal distribution. The model \mathcal{M} is allowed to be modified by creation/deletion, connection/disconnection, and shifting of fiber segments. For more details, please refer to [8].

2.4 Generating High-Resolution Images

Once the optimal model \mathcal{M}^* has been determined, high-resolution data can be sampled from the model using a grid with resolution (e.g., $(1 \text{ mm})^3$) that is higher than the acquisition resolution. More formally, the sampled signal $E'(\mathbf{x}, \mathbf{g})$ is obtained as

$$E'(\mathbf{x}, \mathbf{g}) = \sum_i G'(\mathbf{x}, \mathbf{x}_i) E_i(\mathbf{g}) \quad \text{s.t. } \mathcal{M}^*, \quad (5)$$

where $G'(\mathbf{x}, \mathbf{x}_i)$ is the same function as $G(\mathbf{x}, \mathbf{x}_i)$, but with the bandwidth σ reduced according to the up-sampling factor to reflect the reduced voxel size and the fact that the sampled signal should now come from a smaller neighborhood, in line with the actual MR acquisition mechanism. This in effect reduces the blurring effect associated with the larger voxel size and hence helps produce a super-resolved version of the data.

A scalar image indicating the anisotropy at each location can be generated with the help of (5). This can be achieved by considering the anisotropic energy of the fiber ODFs [10], i.e.,

$$A(\mathbf{x}) = \int_{\Omega_{\mathbf{g}}} E'(\mathbf{x}, \mathbf{g}) \otimes H(\mathbf{g}) d\mathbf{g} = \sum_i G'(\mathbf{x}, \mathbf{x}_i) \int_{\Omega_{\mathbf{g}}} E_i(\mathbf{g}) \otimes H(\mathbf{g}) d\mathbf{g}, \quad (6)$$

where \otimes denotes the spherical deconvolution operator, and $H(\mathbf{g})$ is the response function of a directionally coherent fiber bundle. If we let $H(\mathbf{g}) = w e^{-b\mathbf{g}^T \mathbf{D} \mathbf{g}}$, where $\mathbf{D} = \lambda_1 \mathbf{v} \mathbf{v}^T + \lambda_2 \mathbf{I}$ with an arbitrary \mathbf{v} , then (6) gives $A(\mathbf{x}) \propto \sum_i G'(\mathbf{x}, \mathbf{x}_i)$. Essentially, this implies that the anisotropic energy $A(\mathbf{x})$ at each location \mathbf{x} can be evaluated by a weighted count of fiber segments in the vicinity of \mathbf{x} . On the surface, this approach resembles track-density imaging (TDI), as reported in [1]; however, in TDI no signal model is attached to the fiber segments and an arbitrarily huge or tiny number of fibers are allowed to transverse each voxel. This makes interpretation of the fiber count based image contrast generated by TDI very difficult.

3 Experimental Results

We report here preliminary results from our evaluation of the proposed technique using two different *in vivo* datasets, one acquired at the common $(2 \text{ mm})^3$ resolution, the other at $(1 \text{ mm})^3$ resolution.

3.1 Materials

Diffusion-weighted images for an adult subject were acquired at the typical $(2 \text{ mm})^3$ resolution using a Siemens 3T TIM Trio MR scanner. Diffusion gradients were applied in 120 non-collinear directions with diffusion weighting

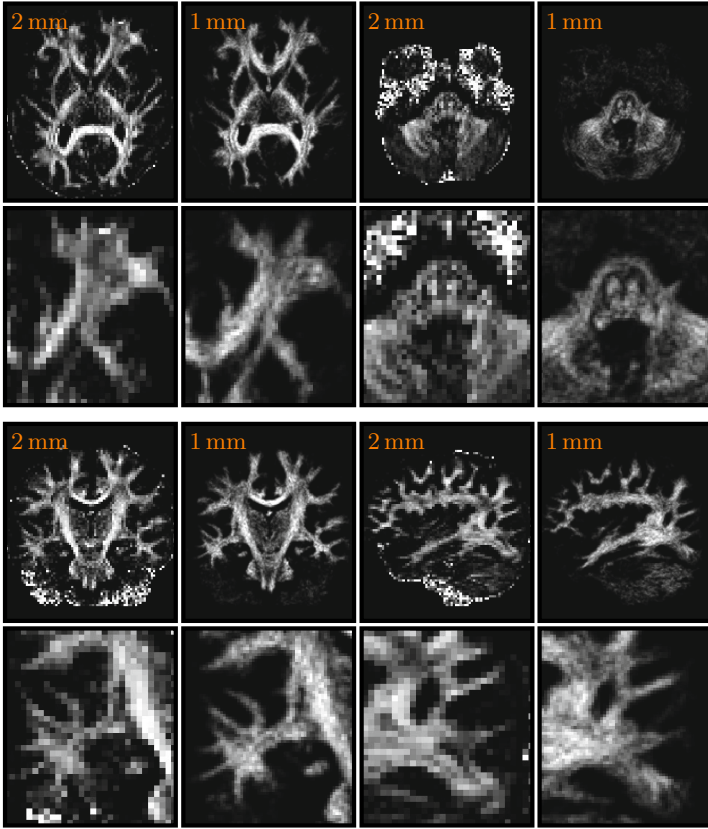


Fig. 3. Results for $(2\text{ mm})^3$ Data. Using our technique, the $(2\text{ mm})^3$ data are enhanced to $(1\text{ mm})^3$. The fractional anisotropy images are shown for the original data; the anisotropic energy images are shown for the enhanced data. The images in the second and fourth rows are closeup views of the images in first and third rows, respectively.

$b = 2,000\text{ s/mm}^2$. An additional set of high-resolution $(1\text{ mm})^3$ diffusion-weighted images were acquired from a different adult subject using the same scanner with the acquisition technique reported in [6]. Diffusion gradients were applied in 42 non-collinear directions with diffusion weighting $b = 1,000\text{ s/mm}^2$.

3.2 Parameters

Setting the parameters of our algorithm to the following values was found to yield reasonable results. Regularization tuning parameter: $\gamma = 1$; tensor model parameters: $b\lambda_1 = 1$, $b\lambda_2 = 0$; the weight of each fiber segment: $w = 0.0018$; the smoothing bandwidth: $\sigma = d = 1\text{ mm}$; the length of each fiber segment: $l = 3\text{ mm}$; the connection likelihood parameter: $L = 10$; and the initial temperature: $T = 0.1$, which was decreased to the final temperature $T = 0.001$ in 5×10^7 iterations. More details on how to set these parameters can be found in [7, 8].

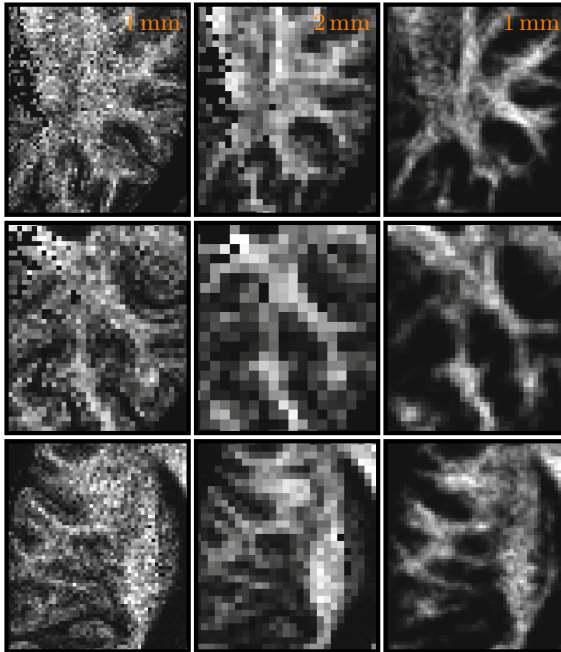


Fig. 4. Results for Simulated $(2\text{ mm})^3$ Data. The high resolution $(1\text{ mm})^3$ data (left) were down-sampled by averaging every 8 adjacent voxels to simulate the $(2\text{ mm})^3$ data (middle), which were then enhanced to the resolution of $(1\text{ mm})^3$ (right) by using our technique.

3.3 Results

For preliminary evaluation, we applied our technique to the $(2\text{ mm})^3$ data and enhanced the resolution to $(1\text{ mm})^3$. The results, shown in Fig. 3, indicate that structural visibility can be significantly improved by using the proposed technique. Structures not visible in the low resolution becomes visible after enhancement.

For better evaluation of the proposed technique, we down-sampled the $(1\text{ mm})^3$ data by averaging every 8 adjacent voxels to simulate a $(2\text{ mm})^3$ version of the data. This simulated low-resolution data were then enhanced using our technique to become $(1\text{ mm})^3$. The results, shown in Fig. 4, indicate that the resolution enhanced data retains most of the structures in the original high-resolution data. It can also be observed that the enhanced data are generally less noisy.

4 Conclusion

We have presented a method to learn a generative model for producing high-resolution DMRI data. The model consists of a set of fiber segments that are configured in a way that best explains the observed data. The high-resolution

data are generated by sampling from this learned model using a grid with resolution that is higher than the acquisition resolution. Even though the results reported were preliminary, they do demonstrate that the proposed method is effective and produces reasonable results.

Acknowledgment. This work was supported in part by a UNC start-up fund and NIH grants (EB006733, EB008374, EB009634, MH088520, AG041721, and MH100217).

References

1. Calamante, F., Tournier, J., Jackson, G., Connelly, A.: Track-density imaging (TDI): Super-resolution white matter imaging using whole-brain track-density mapping. *Neuroimage* 53(4), 1233–1243 (2010)
2. Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1), 97–109 (1970)
3. Johansen-Berg, H., Behrens, T.E. (eds.): *Diffusion MRI — From Quantitative Measurement to In vivo Neuroanatomy*. Elsevier (2009)
4. Kreher, B., Mader, I., Kiselev, V.: Gibbs tracking: A novel approach for the reconstruction of neuronal pathways. *Magnetic Resonance in Medicine* 60(4), 953–963 (2008)
5. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6), 1087–1092 (1953)
6. Porter, D.A., Heidemann, R.M.: High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magnetic Resonance in Medicine* 62(2), 468–475 (2009)
7. Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V.: Global fiber reconstruction becomes practical. *NeuroImage* 54(2), 955–962 (2011)
8. Reisert, M., Mader, I., Kiselev, V.: Global reconstruction of neuronal fibres. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI), Diffusion Modelling Workshop* (2009)
9. Scherrer, B., Gholipour, A., Warfield, S.: Super-resolution reconstruction to increase the spatial resolution of diffusion weighted images from orthogonal anisotropic acquisitions. *Medical Image Analysis* 16(7), 1465–1476 (2012)
10. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23(3), 1176–1185 (2004)
11. Yap, P.-T., Shen, D.: Resolution enhancement of diffusion-weighted images by local fiber profiling. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part III. LNCS, vol. 7512*, pp. 18–25. Springer, Heidelberg (2012)

Multi-atlas Segmentation without Registration: A Supervoxel-Based Approach

Hongzhi Wang and Paul A. Yushkevich*

Department of Radiology, University of Pennsylvania

Abstract. Multi-atlas segmentation is a powerful segmentation technique. It has two components: label transfer that transfers segmentation labels from prelabeled atlases to a novel image and label fusion that combines the label transfer results. For reliable label transfer, most methods assume that the structure of interest to be segmented have localized spatial support across different subjects. Although the technique has been successful for many applications, the strong assumption also limits its applicability. For example, multi-atlas segmentation has not been applied for tumor segmentation because it is difficult to derive reliable label transfer for such applications due to the substantial variation in tumor locations. To address this limitation, we propose a label transfer technique for multi-atlas segmentation. Inspired by the Superparsing work [13], we approach this problem in two steps. Our method first oversegments images into homogeneous regions, called supervoxels. For a voxel in a novel image, to find its correspondence in atlases for label transfer, we first locate supervoxels in atlases that are most similar to the supervoxel the target voxel belongs to. Then, voxel-wise correspondence is found through searching for voxels that have most similar patches to the target voxel within the selected atlas supervoxels. We apply this technique for brain tumor segmentation and show promising results.

1 Introduction

Multi-atlas segmentation has been widely applied in medical image analysis, e.g. [10,8,9,4,11,15]. This technique has two main components: label transfer and label fusion. In the label transfer step, it computes voxel-wise correspondence between pre-labeled images, called atlases, and a novel target image, from which the atlas segmentation labels are transferred to the target image. Label fusion is then applied to derive a consensus segmentation to reduce segmentation errors produced by label transfer.

So far, multi-atlas segmentation has been mostly applied to problems, where deformable image registration can establish reliable voxel-wise correspondences for label transfer. The advantage of image registration is that it enforces spatial regularization, such as smoothness on the correspondence map, to improve the reliability of voxel-wise correspondences obtained from local appearance matching. For applications such as brain parcellation where the structures of interest

* This work was supported by NIH awards AG037376, EB014346.

have stable spatial structures, registration usually can provide high quality label transfer. However, for problems such as tumor and lesion segmentation where the assumption of localized spatial support does not hold, it is not straightforward to apply registration for label transfer across different subjects. Note that although non-local mean label fusion [4,11] does not require non-rigid registration for label transfer, it still requires the structure of interest to have localized spatial support across subjects to make the technique practical due to its high computational cost for non-local averaging. In order to extend multi-atlas segmentation to problems where the structure of interest does not have localized spatial support, *our contribution is to develop a supervoxel-based label transfer scheme for multi-atlas segmentation.*

Our work is inspired by the Superparsing work [13], which was developed for semantic natural scene segmentation. Given pre-labeled training images and a novel target image, this technique first generates oversegmentation for all images using bottom-up segmentation techniques. Each segmented region is treated as a superpixel (or supervoxel in 3D). Typically, an image is represented by hundreds or thousands of superpixels. To segment a novel image, segmentation labels of superpixels in training images are transferred to superpixels in the novel image. For this task, each superpixel is described by a feature vector, which may include intensity, texture histograms and shape features extracted from all pixels within the superpixel. For each superpixel in the target image, the most similar superpixels in the training images are selected based on feature matching for label transfer. After label fusion, the consensus label produced for a superpixel in the target image is assigned to all pixels within the superpixel.

The advantages of employing superpixel for label transfer are three fold: 1) it can be applied to structures that do not have localized spatial support, 2) it significantly reduces the computational burden, making label transfer based multi-atlas segmentation practical for large datasets, and 3) feature matching between superpixels allows more reliable image statistics to be used, which gives more accurate matching than voxel-wise patch matching. The technique also has limitations. First, bottom up oversegmentation may contain errors, i.e. mixing pixels from different label classes into a single superpixel. Hence, its performance is limited by the performance of bottom-up oversegmentation. Second, as extensive studies in the multi-atlas segmentation literature have verified, label fusion should be performed in a way that spatially varies the relative contribution of atlases to accommodate spatially varying performance of label transfer [1,12]. Performing label fusion at the superpixel level can be considered as a semi-local approach, which may not fully capture spatial variations for optimal label fusion.

To address these limitations, instead of label transfer and fusion at the superpixel level, *we propose to use superpixel matching as an additional matching constraint for establishing voxel-wise label transfer and fusion.* For each voxel in a novel target image, we first find supervoxels in the atlases that are most similar to the supervoxel it belongs to. Then voxel-wise correspondences of the target voxel are located by searching the most similar voxels within the selected atlas supervoxels by appearance patch matching. This technique maintains the

computational efficiency achieved by employing supervoxels and also improves the accuracy of voxel-wise correspondences produced by local patch matching. For validation, we apply our multi-atlas segmentation technique to multimodal brain tumor segmentation and show promising results.

2 Method

2.1 Supervoxel Generation and Feature Extraction

To generate supervoxel representations, we apply the efficient graph-based segmentation technique [6] to oversegment images into homogeneous regions. [6] groups neighboring voxels based on their intensity differences, such that similar voxels are more likely to be grouped together. Since for the tumor segmentation problem addressed in our experiments, multi-modality magnetic resonance (MR) images, including T1, contrast-enhanced T1, T2, and FLAIR, are available, we define intensity difference between two neighboring voxels as the maximal absolute intensity difference between them in all modality channels. In addition, we specify the minimal region size in the resulting oversegmentation to be 100 voxels. These parameters were chosen so that about 1000~2000 supervoxels are produced for each brain image (see Fig. 1 for examples of produced oversegmentations). With such specifications, an image can be segmented within a few seconds on a single 2GHZ CPU.

Before extracting features, we apply image histogram equalization implemented by the *histeq* function in Matlab with default parameters to reduce the intensity scale variations across different subjects. After histogram equalization, the image intensities are normalized into $[0, 1]$ and all processed images have similar intensity histogram profiles. We include the following features to represent each supervoxel: the mean and standard deviation of voxel intensity and gradient within each supervoxel from each modality channel, the intensity and gradient histogram from each modality channel. The histograms are computed with 41 bins. Hence, each supervoxel is described by 344 features in total.

2.2 Supervoxel-Based Voxel-Wise Label Transfer

Our goal is to find voxel-wise correspondence between a target image and all atlases for label fusion. For this task, we apply image patch based appearance matching. However, directly searching voxel-wise correspondences based on local appearance similarity has limitations. First, local image appearance similarity is not always a reliable indicator for correspondence. Additional regularization constraints, such as smoothness on the correspondence map used in image registration, are often required to make local appearance matching based correspondence searching more reliable. Second, the computation cost for global voxel-wise correspondence searching is too high. Hence, the assumption of localized spatial support for the structure of interest across subjects is necessary for limiting the searching area [4,11].

One insight provided by the Superparsing work [13] is that regions obtained from bottom up segmentation provide meaningful information for correspondence matching. Hence, we propose to use supervoxel-based region matching as an additional constraint to guide voxel-wise correspondence searching. For each supervoxel in a novel target image, we find K most similar supervoxels in the atlases based on their feature vectors using $L1$ norm. Based on the selected corresponding supervoxels, we apply label fusion to derive the consensus label for the target supervoxel, as described below. If the target supervoxel is labeled as the structure of interest, then for each voxel within the target supervoxel, we search its corresponding voxels within the K selected atlas supervoxels. Again, $L1$ norm on appearance patches extracted for the voxels from all modality channels are used for voxel matching. N most similar voxels are selected as the correspondences of the target voxel. Their segmentation labels are transferred and fused into a consensus label for the target voxel.

Limiting the voxel-wise correspondence searching within selected corresponding atlas supervoxels significantly reduces potential matches into a small set of regions that are most likely to contain the correct correspondence and enforces weak spatial regularization on the voxel-wise correspondences. The above mentioned drawbacks in directly using local appearance matching for searching voxel-wise correspondence are properly addressed.

2.3 Label Fusion

For label fusion, we apply local weighted voting. The fused consensus label probability for a novel target image T_F is obtained by:

$$p(l|x, T_F) = \sum_{j=1}^N w_{x(j)} p(l|x(j), A), \quad (1)$$

where x indexes through all voxels in the target image and l indexes through all possible labels. A is the set of all atlases. $x(j)$ is the j_{th} selected corresponding voxel for x from A using the method described in section 2.2. $p(l|x(j), A)$ is the probability that $x(j)$ votes for label l , which is 1 if $x(j)$ has label l and 0 otherwise. $w_{x(j)}$ is the voting weights for $x(j)$, with $\sum_{j=1}^N w_{x(j)} = 1$. The consensus segmentation is obtained by selecting the label with the maximal probability for each voxel in the target image. To compute the voting weights, we apply the joint label fusion technique [15], and the solutions are given as:

$$\mathbf{w}_x = \frac{M_x^{-1} \mathbf{1}_N}{\mathbf{1}_N^t M_x^{-1} \mathbf{1}_N}, \quad (2)$$

where $\mathbf{1}_N = [1; 1; \dots; 1]$ is a vector of size N . M_x is a dependency matrix capturing the pairwise dependencies of the N selected corresponding atlas voxels voting for wrong labels for the target voxel x , which is computed as:

$$M_x(j, k) \sim \sum_{m=1}^{L_M} (|A_F^m(\mathcal{N}(x(j))) - T_F^m(\mathcal{N}(x))|, |A_F^m(\mathcal{N}(x(k))) - T_F^m(\mathcal{N}(x))|), \quad (3)$$

where m indexes through all modality channels and $|A_F^m(\mathcal{N}(x(j))) - T_F^m(\mathcal{N}(x))|$ is the vector of absolute intensity difference between a selected atlas image and the target image over local patches \mathcal{N} centered at voxel $x(j)$ and voxel x , respectively. In our experiments, we applied a patch size $5 \times 5 \times 5$. $\langle \cdot, \cdot \rangle$ is the dot product. L_M is the total number of modality channels.

Label fusion at the supervoxel level. In our experiments, we compare our voxel-wise label transfer and fusion approach with the standard supervoxel-wise label transfer and fusion approach [13]. For supervoxel-wise label fusion, we compute the label distribution of all voxels within an atlas supervoxel to represent its label votes. As in [13], we apply majority voting to fuse the votes from the K selected corresponding supervoxels. The label that has the greatest consensus vote is assigned to all voxels within the target supervoxel.

Machine learning based error correction. Multi-atlas segmentation may produce systematic segmentation errors with respect to manual segmentation. We apply the technique described in [14] to train AdaBoost classifiers [7] to automatically correct systematic errors produced by our multi-atlas segmentation method on a voxel by voxel basis.

3 Experiments

We evaluate our method on brain tumor segmentation. Tumor segmentation is an ideal application for testing our method because the classical deformable registration based multi-atlas segmentation technique and non-local mean methods cannot be directly applied due to the substantial variation in tumor locations.

3.1 Data and Experiment Setup

To facilitate comparisons with state-of-the-art brain tumor segmentation techniques, we evaluate our method on the data from MICCAI 2012 Multimodal Brain Tumor Segmentation (BRATS) challenge. This dataset contains both real patient data and simulated data. The images are skull-stripped multimodal MR images, including T1, contrast enhanced T1, T2, and FLAIR. Both high grade and low grade tumor data are available for real patient data and simulated data. Overall, there are 20 real patient data with high grade tumors, 10 real patient data with low grade tumors. For simulated data, both high grade and low grade tumor data contain 25 image sets. For evaluation, we conduct leave-one-out cross-validation for each of the four datasets separately, i.e. each image is segmented by the remaining images of its kind.

Since empirical studies have shown that the performance of multi-atlas segmentation usually reaches saturated levels when 20 or more atlases are used, e.g. [2], we fixed the parameter K , the number of selected candidate corresponding atlas supervoxels, and N , the number of selected corresponding voxels for label transfer, to be 20 in our experiments.

3.2 Results

Table 1 summarizes the performance for edema and tumor segmentation in terms of Dice Similarity Coefficient (DSC) produced by each method for each of the four datasets. Overall, automatic segmentation algorithms all produced more accurate results for simulated data, possibly due to the fact that simulated data have better image quality. Our supervoxel matching based voxel-wise label fusion approach produced substantial improvements over the supervoxel-wise label fusion approach. Machine learning based error correction (EC) produced substantial improvement over our voxel-wise label fusion approach for real patient data, but did not produce much improvement for the simulated data, indicating that our method did not produce prominent systematic errors for simulated data. In terms of computational cost, supervoxel-wise label fusion segments an image within a few minutes using ten or twenty atlases. Supervoxel matching based voxel-wise label fusion usually takes about 30 minutes to segment one image, which is significantly faster than deformable registration-based label fusion.

Table 1. The performance of edema and tumor segmentation in terms of Dice similarity coefficient ($2|A \cap B|/|A| + |B|$) produced by each method

Real patient	edema(High)	tumor (High)	edema(Low)	tumor (Low)
supervoxel	0.58±0.20	0.62±0.29	0.32±0.14	0.41±0.27
voxel-wise	0.64±0.17	0.63±0.28	0.37±0.16	0.54±0.22
voxel-wise + EC	0.66±0.15	0.67±0.28	0.41±0.15	0.60±0.25
[16]	0.70±0.09	0.71±0.24	0.44±0.18	0.62±0.27
[3]	0.61±0.15	0.62±0.27	0.35±0.18	0.49±0.26
Simulated	edema(High)	tumor (High)	edema(Low)	tumor (Low)
supervoxel	0.55±0.29	0.86±0.05	0.39±0.24	0.74±0.07
voxel-wise	0.65±0.29	0.92±0.04	0.61±0.25	0.82±0.06
voxel-wise + EC	0.67±0.28	0.92±0.04	0.59±0.26	0.83±0.06
[16]	0.65±0.27	0.90±0.05	0.55±0.23	0.71±0.20
[3]	0.68±0.26	0.90±0.06	0.57±0.24	0.74±0.10

Overall, our method compares well with the state-of-the-art algorithms [16] and [3], which finished in the top two places in the BRATS challenge.¹ Both methods apply decision forest based classification technique [5]. Our results are consistently better than those reported in [3], except for edema segmentation for high grade real patient data. In comparison with [16], our results on real patient data are worse but our results on simulated data are better. Note that classification-based approach implicitly chooses training samples for classifying new data. The key difference from our method is that no spatial regularization is used for locating training samples for classifying target voxels, which may compromise the reliability of the located training samples.

¹ Ranking is based on on-site evaluation, for which the data is not available for testing.

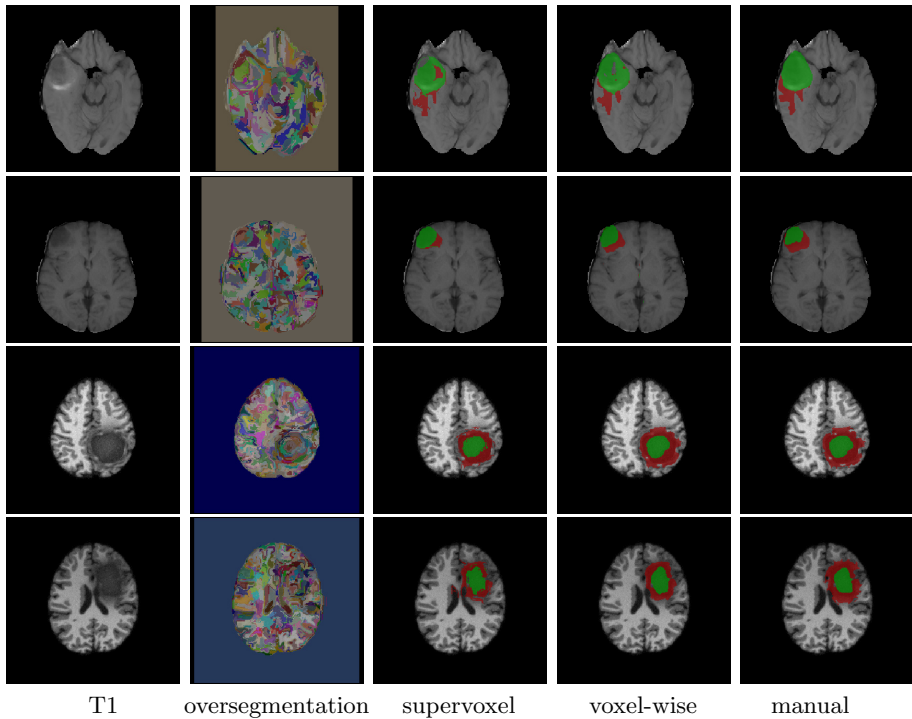


Fig. 1. Edema (red) and tumor (green) segmentation produced by supervoxel label fusion and our voxel-wise label fusion. From first row to fourth row: high/low grade real data, high/low grade simulated data.

Fig. 1 shows a few segmentation results produced by supervoxel label fusion and our voxel-wise label fusion. Note that since the supervoxels produced by bottom up segmentation have irregular shapes, the results produced by supervoxel label fusion also have irregular shapes. This effect is greatly reduced when voxel-wise label fusion is applied.

4 Conclusions

We introduced a novel region matching based voxel-wise label transfer scheme for multi-atlas segmentation. This technique is much faster than registration-based label transfer and can be easily applied on large datasets. As a proof of concept, we showed promising performance on a brain tumor segmentation problem.

In our current experiments, a fairly crude set of criteria and simple features were used to create supervoxel representations. There is room left for improvement by specially tuning the supervoxel representation for multimodal MRI tumor segmentation. We will also explore the performance of the supervoxel-based label transfer scheme for segmenting structures with less distinctive appearance patterns than tumors.

References

1. Artaechevarria, X., Munoz-Barrutia, A., de Solorzano, C.O.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE TMI* 28(8), 1266–1277 (2009)
2. Awate, S.P., Zhu, P., Whitaker, R.T.: How many templates does it take for a good segmentation?: Error analysis in multiatlas segmentation as a function of database size. In: Yap, P.-T., Liu, T., Shen, D., Westin, C.-F., Shen, L. (eds.) *MBIA 2012*. LNCS, vol. 7509, pp. 103–114. Springer, Heidelberg (2012)
3. Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes, M.: Segmentation of brain tumor images based on integrated hierarchical classification and regularization. In: *MICCAI BraTS workshop*, Nice (2012)
4. Coupe, P., Manjon, J., Fonov, V., Pruessner, J., Robles, N., Collins, D.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
5. Criminisi, A.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends® in Computer Graphics and Vision* 7(2-3), 81–227 (2011)
6. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. *IJCV* 59(2), 167–181 (2004)
7. Freund, Y., Schapire, R.: A decision-theoretic generalization of on-line learning and an application to boosting. In: *Proceedings of the 2nd European Conf. on Computational Learning Theory*, pp. 23–27 (1995)
8. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33, 115–126 (2006)
9. 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 Trans. on MI* 28(7), 1000–1010 (2009)
10. 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)
11. Rousseau, F., Habas, P.A., Studholme, C.: A supervised patch-based approach for human brain labeling. *IEEE TMI* 30(10), 1852–1862 (2011)
12. Sabuncu, M., Yeo, B., Leemput, K.V., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE TMI* 29(10), 1714–1720 (2010)
13. Tighe, J., Lazebnik, S.: Superparsing. *International Journal of Computer Vision* 101(2), 329–349 (2013)
14. Wang, H., Das, S., Suh, J.W., Altinay, M., Pluta, J., Craige, C., Avants, B., Yushkevich, P.: A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. *NeuroImage* 55(3), 968–985 (2011)
15. Wang, H., Suh, J.W., Das, S., Pluta, J., Craige, C., Yushkevich, P.: Multi-atlas segmentation with joint label fusion. *IEEE Trans. on PAMI* 35(3), 611–623 (2013)
16. Zikic, D., Glocker, B., Konukoglu, E., Shotton, J., Criminisi, A., Ye, D., Demiralp, C., Thomas, O., Das, T., Jena, R., Prince, S.: Context-sensitive classification forests for segmentation of brain tumor tissues. In: *MICCAI BraTS workshop*, Nice

Adaptive Voxel, Texture and Temporal Conditional Random Fields for Detection of Gad-Enhancing Multiple Sclerosis Lesions in Brain MRI

Zahra Karimaghloo¹, Hassan Rivaz², Douglas L. Arnold³
D. Louis Collins², and Tal Arbel^{1,*}

¹ Centre for Intelligent Machines, McGill University, Canada

² Montreal Neurological Institute, McGill University, Canada

³ NeuroRx Research, Montreal, Canada

Abstract. The detection of Gad-enhancing lesions in brain MRI of Multiple Sclerosis (MS) patients is of great clinical interest since they are important markers of disease activity. However, many of the enhancing voxels are associated with normal structures (i.e. blood vessels) or noise in the MRI, making the detection of Gad-enhancing lesions a challenging task. Furthermore, these lesions are typically small and in close proximity to vessels. In this paper, we present a probabilistic Adaptive Multi-level Conditional Random Field (AMCRF) framework, capable of leveraging spatial and temporal information, for detection of MS Gad-enhancing lesions. In the first level, a voxel based CRF with cliques of up to size three, is used to identify candidate lesions. In the second level, higher order potentials are incorporated leveraging robust textural features which are invariant to rotation and local intensity distortions. Furthermore, we show how to exploit temporal and longitudinal images, should they be available, into the AMCRF model. The proposed algorithm is tested on 120 multimodal clinical datasets acquired from Relapsing-Remitting MS patients during multi-center clinical trials. Results show a sensitivity of 93%, a positive predictive value of 70% and average False Positive (FP) counts of 0.77. Moreover, the temporal AMCRF results show the same sensitivity as the AMCRF model while decreasing the FP counts by 22%.

1 Introduction

Multiple Sclerosis (MS) is one of the most common neurological disease in young adults. Conventional Magnetic Resonance Imaging (MRI) techniques, such as T2-weighted (T2) and Gadolinium-enhanced T1-weighted (T1) sequence are sensitive in detecting its white matter (WM) plaques known as lesions. Specifically, due to their ability to reflect areas of blood-brain barrier disruption and acute inflammations, Gad-enhancing lesions¹ lesions serve as a measure of disease activity. At present, the number of Gad lesions is a widely used MRI outcome

* This work was supported by an NSERC CRD grant (CRDPJ 411455-10).

¹ We refer to Gad-enhancing lesions simply as Gad lesions hereafter.

parameter in MS clinical trials. Gad lesions are generally segmented manually, a laborious task which is subject to intra- and inter-rater variability and very expensive for clinical trials that involve enormous amounts of data from multiple centers. It is desirable to have an automatic segmentation method that is robust to data variability due to different scanners and protocols. Moreover, it is necessary for any automatic technique to have high sensitivity and low False Positive (FP) rate to be clinically relevant. Unfortunately, there exists huge variability in the size (as small as 3 voxels), texture, intensity and location of Gad lesions making the detection task very challenging. Furthermore, the presence of numerous non-lesional enhancements (e.g. blood vessels, MRI noise and partial volume effects) renders maintaining low FP rate a challenging task. Most of the existing methods for Gad segmentation described in the literature are either not fully automatic [1,2], or depend on non-conventional MRI acquisition sequences [2,3], or require prior segmentation of T2 lesions in order to remove the FPs [3,4]. In [5,6,7], conditional random field models are proposed for addressing this problem, which were shown to outperform standard MRF, SVM and linear regression models. However, these models incorporate mainly local, voxel-level features and FPs still remain. As Gad lesions are typically very small and noisy, higher order features could be integrated in order to express more complex patterns. However, computing such features for all enhancing voxels is computationally prohibitive. Also, since MS is a longitudinal disease, clinical trials often consist of multiple scans of each patient over time which can provide additional information to the manual raters by observing persistence of enhancements in scans acquired at least six months apart (Gad lesions are generally enhance for less than six months). No automated methods have explored how additional temporal information (if available) can be leveraged in Gad lesion segmentation for further removal of the possible FPs.

In this work, we present an Adaptive Multi Level Conditional Random Field (AMCRF) classifier which incorporates both local voxel level and robust higher order textural patterns into the model. Specifically, at the voxel level, a local CRF (with cliques of up to size 3) is developed to infer binary labels at each voxel (i.e. lesions/non-lesion). At this level, the classifier is tuned to be highly sensitive at the expense of additional FP detections. At the second level, voxels with the same label are grouped together to form lesion candidates. Each candidate is further analyzed by considering new textural patterns, derived from a larger neighborhood, along with its voxel-wise observations to differentiate true and false lesion detections. To this end, SPIN image and RIFT features [8] are explored, two texture descriptors that are invariant to rotation and local intensity distortions. In addition to removing false lesions, the AMCRF also refines the boundaries of lesions at the second level and as such, is adaptive. We also show effective ways to exploit the temporal information from a past or future time point (should it be available) into our AMCRF model to further improve our results. The temporal AMCRF outperforms other methods with a sensitivity of 93%, a positive predictive value of 75% and average False Positive (FP) counts of 0.60.

2 Method

2.1 Adaptive Multi-level CRF

Our single timepoint AMCRF model infers the posterior distribution of the labels in two levels: the first voxel-based level, and the second level which incorporates higher order texture information. The first level is similar to the model presented in [7], except that in addition to the unary and pairwise interactions, here triplet cliques are considered as well. We first describe these two levels at a single time-point, and then present algorithms for exploiting temporal information, should it be available.

Let $\mathbf{x}_i \in \mathbb{R}^d$ denote the observation vector (e.g. intensity values) at voxel i in the image and $y_i \in \{1, 0\}$ be a binary random variable indicating its label (e.g. lesion vs. non-lesion). Given a test image, X , the goal of a probabilistic classifier is to infer the posterior distribution of the labels given the observations, i.e $p(Y|X)$ where $X = \{\mathbf{x}_i\}_1^n$, $Y = \{y_i\}_1^n$, and n is the total number of voxels in the image. At the first level, we introduce a voxel-based CRF with cliques of size up to 3 to formulate p^v , the posterior of labels given the observations:

$$\begin{aligned}
 p^v(Y|X, \boldsymbol{\lambda}^v) &= \frac{1}{Z} \exp[-(\sum_{i=1}^n \boldsymbol{\lambda}_\phi^v \phi(y_i|\mathbf{x}_i) + \sum_{i,j \in N_i} \boldsymbol{\lambda}_\varphi^v \varphi(y_i, y_j|\mathbf{x}_i, \mathbf{x}_j) \\
 &+ \sum_{i,j \in N_i} \boldsymbol{\lambda}_\delta^v \delta_2(y_i, y_j) + \sum_{i,(j,k) \in N_i} \boldsymbol{\lambda}_\psi^v \psi(y_i, y_j, y_k|\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k))] \quad (1)
 \end{aligned}$$

where Z is the partition function. ϕ , φ and ψ represent the voxel level potentials for the unary, pairwise and triplet cliques respectively. The smoothing constant $\delta_2(\cdot)$ is for penalizing discrepancies in the labels of neighbouring pairs. It is zero if the two labels are equal and is one otherwise. N_i represents the first order neighborhood of voxel i . The voxel level parameters $\boldsymbol{\lambda}^v$, modulate the effect of each term in the final decision and are learned at the training stage (Sec.3.1). ϕ , φ and ψ are modeled as:

$$\phi(y_i|\mathbf{x}_i) = -\log p(y_i|\mathbf{x}_i), \quad (2)$$

$$\varphi(y_i, y_j|\mathbf{x}_i, \mathbf{x}_j) = -\log p(y_i, y_j|\mathbf{x}_i, \mathbf{x}_j) \quad (3)$$

$$\psi(y_i, y_j, y_k|\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = -\log p(y_i, y_j, y_k|\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) \quad (4)$$

Random Forest [RF] [9] can be used to model the probabilities in Eq.(2) to (4). A set of labeled voxels are inferred as the result of the voxel level analysis. At this stage, the goal is to capture all of the lesions at the expense of additional FPs.

So far only voxel-wise interactions are considered. However, the pure intensity at each voxel might be distorted due to the presence of noise or other artifacts such as partial volume. Hence, higher order textural patterns that are robust to local intensity distortions are incorporated in to the model to remove the possible FPs. To that end, at the second level, voxels with the same label are grouped together to form a set of lesion candidates. A bounding box (BB) is

considered around each candidate with a 2 voxel margin from each side, and a new CRF is constructed for the voxels inside the BB modeling voxel-wise interactions together with higher order textural patterns:

$$p(Y_{BB}|X_{BB}, \boldsymbol{\lambda}^l, \boldsymbol{\lambda}_\Omega^l) = p^v(Y_{BB}|X_{BB}, \boldsymbol{\lambda}^l) \exp\left(-\sum_i \boldsymbol{\lambda}_\Omega^l \Omega(y_i|H(X_{BB}))\right) \quad (5)$$

where X_{BB} and Y_{BB} indicate the observations and labels inside BB. $p^v(Y_{BB}|X_{BB}, \boldsymbol{\lambda}^l)$ represents a set of voxel-wised cliques similar to Eq.(1) with a new set of modulating parameters $\boldsymbol{\lambda}^l$ which along with $\boldsymbol{\lambda}_\Omega^l$ are learned for the second level. $H(X_{BB})$ is the textural pattern derived from the region inside the BB. In principal, any textural feature can be used to represent $H(X_{BB})$. In this work, inspired by [8], we chose two novel descriptors: *SPIN image* and *RIFT* which are 2D histograms encoding the appearance pattern inside each BB based on its intensity and gradient orientation distributions. $\Omega(y_i|H(X_{BB})) = -\log(p(\textit{lesion}_{BB}|H(X_{BB})))$ which represents the likelihood of detecting a Gad lesion inside BB given $H(X_{BB})$. It should be noted that the SPIN image and RIFT descriptors are computationally expensive, but the proposed hierarchical framework computes them only at the second level where we are left with only a few candidates. During training, we apply the voxel level model to a subset of the training data to obtain a set of lesion candidates. Spin image and RIFT features are computed for each candidate and are saved as a *textural pattern dictionary* according to whether it is a true or false detection. At test time, we use a KNN classifier (e.g. K=100) to find the K closest match between the textural patterns of the test candidate and the ones in the dictionary. Specifically, Earth Movers Distance (EMD) [10] can be used to find the distance between textural patterns. The probability of $\textit{lesion}_{BB} = 1$ is proportional to the number of true detections among the K nearest neighbours.

2.2 Leveraging Temporal Information

In clinical practice, temporal information can be available to help the rater detect Gad lesions. Let X^t and $X^{t\pm m}$, respectively, denote the image at the current time point and the one acquired m months before or after. In this paper, we focus on the context where the temporal interval, m , is large enough such that, if a Gad lesion is enhanced in X^t , it is most likely not enhanced in $X^{t\pm m}$. In clinical practice, this typically translates to scanning intervals of 6 months or more (i.e. $m=6$). In order to incorporate this temporal information, at the voxel level of the AMCRF model, we use the voxel intensities of both X^t and $X^{t\pm m}$ for all cliques. At the second level, in addition to comparing the textural pattern of the detected region at X^t with those in the dictionary, we also compare it with the textural pattern at the same location at $X^{t\pm m}$. Hence, the second level is modeled as:

$$p(Y_{BB}|X_{BB}^t, X_{BB}^{t\pm m}, \boldsymbol{\lambda}^{l'}, \boldsymbol{\lambda}_\Omega^{l'}, \boldsymbol{\lambda}_\Gamma^{l'}) = p^v(Y_{BB}|X_{BB}^t, X_{BB}^{t\pm m}, \boldsymbol{\lambda}^{l'}) \exp\left(-\sum_i \boldsymbol{\lambda}_\Omega^{l'} \Omega(y_i|H(X_{BB}^t))\right) \exp\left(-\sum_i \boldsymbol{\lambda}_\Gamma^{l'} \Gamma(y_i|G(X_{BB}^t, X_{BB}^{t\pm m}))\right) \quad (6)$$

where $\Gamma(y_i|G(X_{BB}^t, X_{BB}^{t\pm m})) = -\log(p(\text{Lesion}_{BB}|G(X_{BB}^t, X_{BB}^{t\pm m})))$ and $G(X_{BB}^t, X_{BB}^{t\pm m})$ is the EMD between textural patterns of X^t and $X^{t\pm m}$ at the same location. A RF classifier is designed to model this term. As before, λ' , λ'_{Ω} , and λ'_r are modulating parameters learned in training. For a non-lesional enhancement at X^t , textural patterns are similar to the ones extracted from the same location at $X^{t\pm m}$ (compare Fig.1(e)-(f) to Fig.1(g)-(h)). This is typical for enhancing structures such as blood vessels, for example. However they look different for a lesional enhancement (compare Fig.1(m)-(n) to Fig.1(o)-(p)).

3 Experiments and Results

3.1 Parameter Learning and Inference

There are two sets of parameters in our model: the RF parameters used in ϕ , φ , ψ and Γ and the modulating parameters. RF parameters are learned separately for each clique. However, due to the complexity of the partition function, exact learning of the modulating parameters is intractable. In this work we used an iterative approach proposed by Taskar *et al.* [11] in order to find the modulating parameters.

In the inference stage, considering the CRF model at each level and its learned parameters, the most probable labeling is found. Graph Cuts are chosen to solve this optimization problem primarily because of their ability to find globally optimal solutions for binary classifications [12].

3.2 Data Pre-processing

The training and test data was acquired from multi center clinical trials with RRMS patients with varying numbers of Gad lesions, each located in different areas of the brain WM. Each acquisition was composed of five sequences: pre- and post-contrast T1, T2, PD and FLAIR. Therefore, our voxel-wise observation vector, \mathbf{x} , consists of the intensities of the above five modalities, WM and partial volume tissue priors² and spatial locations of each voxel. For the particular data set that we had access to, the “silver standard” manual labels were determined using a protocol where two trained experts label the data separately, followed by consensus agreement. Prior to classification, pre-processing steps including bias-field inhomogeneity correction as well as removal of non-brain regions from the MRI are performed [13]. Furthermore, all intra-subject sequences are registered to a common coordinate space and the intensity histogram of all sequences is normalized [14]. The training data consists of 1760 scans (880 pairs of two time points) from 160 different centers and testing is based on 120 scans (60 pairs of two time points) from a *separate* clinical trial consisting of 24 centers in order

² The WM prior is built based on statistical tissue frequencies of 152 manually labelled brains (ICBM 152). The PV atlas was built based on overlapping locations between Grey Matter and Cerebrospinal Fluid (CSF) atlases and WM and CSF atlases.

to examine the robustness of the method to different multi center trials. Before computing the statistics, any detected region with size 1 or 2 is deleted according to clinical protocol that requires Gad lesions to consists of at least 3 voxels. If at least three voxels of a lesion are classified correctly, it is counted as a TP, otherwise it is a False Negative (FN). Any candidate with size greater than two that does not correspond to an enhancing lesion is counted as an FP. Sensitivity ($\frac{TP}{TP+FN}$), Positive Predictive Value ($\frac{TP}{TP+FP}$) and average number of FPs are reported.

3.3 Single Timepoint Results

Figure 1 compares the higher order textural descriptors for a non-lesional (first row) and lesional (second row) enhancement. As it is observed, the proposed textural patterns look very different for false (Fig.1(e)-(f)) and true (Fig.1(m)-(n)) detections and hence when compared to the dictionary of textural patterns (by computing the EMD), they can be distinguished from each other.

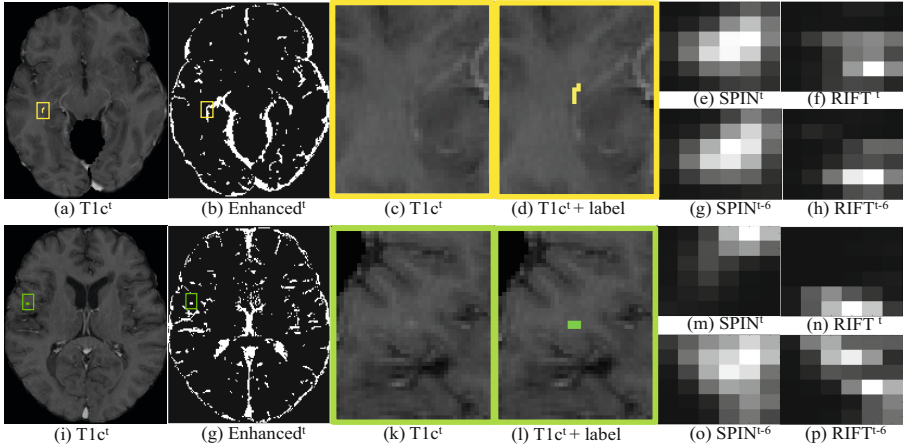


Fig. 1. Comparisons of the higher order textural features for a non-lesional and lesional enhancement. Post contrast T1 images are shown in the first column. Zoomed views are shown without and with labels in the third and fourth columns. The second column shows all the high enhancement voxels. The textural features are shown for the detected regions at the current timepoint image and a previous timepoint.

In Table 1(a), we quantitatively compare the performance of the AMCRF model with the HCRF model proposed in [7] and an MRF. Here, a conventional MRF model is considered consisting of a unary clique and a non data dependant smoothing term (i.e. Eq.(1) without φ and ψ). For this experiment, a mask outlining the “new” enhancing voxels at the current time point is made available to all three methods. The results show that the proposed model has the highest sensitivity and PPV rate over all methods. Lower sensitivity in the MRF stems from the lack of observations in the smoothing term resulting in over smoothing

the small lesions. In HCRF, the high level features used in the second level were not robust enough to capture all lesions. We also show the overall performance of the AMCRF model along with its break down based on the size of the detected regions in Table 1(b). The AMCRF model achieves overall sensitivity of 0.93, PPV of 0.70 and average FP count of 0.77. As the size of the detected regions get larger both sensitivity and PPV values increase. Furthermore, Fig.2(a) shows that majority of the false detections are very small (i.e. less than 5 voxels).

Table 1. (a) Quantitative comparison of the performance of the AMCRF, HCRF and MRF models. (b) The performance of the AMCRF based on the voxel size.

	(a)			(b)							
	AMCRF	HCRF [7]	MRF	overall	1-5	6-10	11-20	21-50	51-100	101 ⁺	
Sens	0.93	0.86	0.78	#Les	231	64	44	35	53	20	15
FPS	0.77	0.76	0.80	Sens	0.93	0.89	0.93	0.94	1	1	1
PPV	0.70	0.68	0.66	PPV	0.70	0.37	0.55	0.82	0.91	1	1

3.4 Temporal Model Results

Fig.1 shows qualitative results depicting the improvement caused by the higher order textural patterns when longitudinal data is available. The higher order textural features of a false detection at time t are very similar to those of the same location at time $t - 6$ (compare Fig.1(e)-(f) to Fig.1(g)-(h)). On the other hand, the higher order textural features of a true detection at time t are very different from those of the same location at time $t - 6$ (compare Fig.1(m)-(n) to Fig.1(o)-(p)). The classifier designed to capture these differences (by computing the EMD between the two textures) can appropriately distinguish true and false detections.

Fig.2(c) shows the quantitative comparisons of the AMCRF model with the temporal AMCRF. As it is observed, incorporation of the temporal data has increased the PPV value by 5% without changing the sensitivity. Also, the average number of false detections has been reduced by 22%. The histogram of the voxel size of the FPs is also shown in Fig.2(b). Once again we see that majority of the FP counts are small.

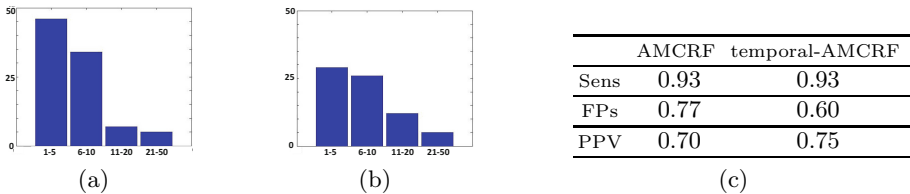


Fig. 2. (a) and (b) show the histograms of the voxel wise size of the total FP detections for the AMCRF and temporal AMCRF models respectively. (c) is the quantitative comparison of the performance of the AMCRF and temporal AMCRF.

4 Discussion

In this paper, we propose a new Adaptive Multi-level CRF (AMCRF) model to detect Gad lesions in brain MRI that embeds contextual information at multiple levels. At the first level, a local voxel-based CRF is used to identify candidate lesions. In the second level, a CRF model is designed to further examine the lesion candidates. We also proposed exploiting temporal data into our model. The temporal AMCRF outperforms other methods with a sensitivity of 93%, a positive predictive value of 75% and average False Positive (FP) counts of 0.60.

References

1. Miki, Y., et al.: Computer-assisted quantitation of enhancing lesions in multiple sclerosis: correlation with clinical classification. *Am. J. Neur* 18, 705–710 (1997)
2. Bedell, B., Narayana, P.: Automatic segmentation of Gadolinium-enhanced multiple sclerosis lesions. *Magn. Reson. Med.* 39, 935–940 (1998)
3. He, R., Narayana, P.: Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. *Med. Phys.* 29, 1536–1546 (2002)
4. Datta, S., et al.: Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis. *J. Magn. Reson. Imag.* 25, 932–937 (2007)
5. Karimaghloo, Z., Shah, M., Francis, S.J., Arnold, D.L., Collins, D.L., Arbel, T.: Detection of gad-enhancing lesions in multiple sclerosis using conditional random fields. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 41–48. Springer, Heidelberg (2010)
6. Karimaghloo, Z., et al.: Automatic detection of Gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields. *TMI* (2012)
7. Karimaghloo, Z., Arnold, D.L., Collins, D.L., Arbel, T.: Hierarchical conditional random fields for detection of gad-enhancing lesions in multiple sclerosis. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 379–386. Springer, Heidelberg (2012)
8. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. *IEEE PAMI* 27, 1265–1278 (2005)
9. Leo, B.: Random forests. *Machine Learning*, 5–32 (2001)
10. Rubner, Y., Tomasi, C., Guibas, L.: The earth mover’s distance as a metric for image retrieval. *IEEE PAMI*, 99–121 (2000)
11. Taskar, B., Chatalbashev, V., Koller, D., Guestrin, C.: Learning structured prediction models: a large margin approach. In: *ICML*, pp. 896–903 (2005)
12. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 1124–1137 (2004)
13. Smity, S.: Fast robust automated brain extraction. *Hum. Brain Mapp.*, 143–155 (2002)
14. Nyul, L., Udupa, J.: On standardizing the MR image intensity scale. *Comp. Assisted Tomography* 42, 1072–1081 (1999)

Minimizing Joint Risk of Mislabeling for Iterative Patch-Based Label Fusion

Guorong Wu¹, Qian Wang¹, Shu Liao¹, Daoqiang Zhang²,
Feiping Nie³, and Dinggang Shen¹

¹Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA

²Department of Computer Science, Nanjing University of Aeronautics and Astronautics, China

³Department of Computer Science, University of Texas Arlington, USA

Abstract. Automated labeling of anatomical structures in medical images is very important in many neuroscience studies. Recently, patch-based labeling in the non-local manner has been widely investigated to alleviate the possible misalignment when registering atlases to the target image. However, the weights used for label fusion from the registered atlases in conventional methods are generally computed independently and thus lack the capability of preventing the ambiguous atlas patches from contributing to the label fusion. More critically, these weights are often calculated based only on the simple patch similarity, thus not necessarily providing optimal solution for label fusion. To address these issues, we present a novel patch-based label fusion method in multi-atlas scenario, for the goal of labeling each voxel in the target image by the best representative atlas patches that also have the lowest joint risk of mislabeling. Specifically, sparse coding is used to select a small number of atlas patches which best represent the underlying patch at each point of the target image, thus minimizing the chance of including the misleading atlas patches for labeling. Furthermore, we examine the joint risk of any pair of atlas patches in making similar labeling error, by analyzing the correlation of their morphological error patterns and also the labeling consensus among atlases. This joint risk will be further recursively updated based on the latest labeling results to correct the possible labeling errors. To demonstrate the performance of our proposed method, we have evaluated it on both whole brain parcellation and hippocampus segmentation, and achieved promising labeling results, compared with the state-of-the-art methods.

1 Introduction

With the advent of modern imaging techniques, image analysis plays an important role in quantitatively measuring the structural differences between either individuals or groups. In many neuroscience and clinical studies, certain regions of interest (ROIs), e.g., hippocampus, in the human brain are widely investigated due to their close relation to brain diseases, such as Alzheimer's disease (AD). Consequently, automatic and accurate labeling of anatomical structures becomes critical in those studies that often have large datasets. However, automatic labeling still remains as a

challenging problem because of complicated brain anatomy and high inter-subject variability.

Recently, patch-based labeling methods have emerged as a popular direction for multi-atlas based segmentation [1-4]. The basic assumption of these methods is that, *if two image patches are similar in appearance, they should have the same anatomical label* [1]. Most patch-based labeling methods perform label fusion from registered atlases in a non-local manner [1, 2, 5]. Specifically, to label a patch in the target image, all possible candidate patches from different atlases are considered, with their contributions weighted by their similarities w.r.t. the target patch. In this way, these non-local based labeling methods can alleviate the labeling errors due to possible registration errors.

Although patch-based labeling methods are effective, they also have several limitations. First, all candidate patches from atlases contribute to label fusion, as long as they are similar to the target patch. However, even patches with high appearance similarities from atlases could bear the wrong labels, thus affecting the final labeling result. Second, weights assigned to candidate patches during label fusion are often estimated independently. Thus, if majority of candidate patches have wrong labels, they will dominate the label fusion process and lead to incorrect labeling result [4]. Third, the weights calculated from patch appearance similarity are often directly applied for label fusion. However, although these weights are optimal for patch representation, i.e., making the combination of candidate patches close to the target patch, they are not necessarily optimal for label fusion.

To address these limitations, we aim to label the target patch by the best representative candidate patches that have the lowest joint risk of mislabeling. Specifically, sparse coding [6] is used to select only a small number of candidate patches that can best represent the target patch. In this way, the chance of including the misleading candidate patches for label fusion can be minimized. Moreover, we examine the joint risk of mislabeling by atlas patches in two ways. First, we measure the pairwise correlation of morphological error patterns for any pair of candidate patches, in order to reject those candidate patches with repeated incorrect labels. Second, we further examine whether the labeling result for the target patch achieves the largest labeling consensus among the candidate patches. In this way, our label fusion method is able to iteratively improve the label fusion result by gradually refining the estimation of the joint risk of mislabeling among the candidate atlas patches. We have applied our proposed label fusion method to both whole brain parcellation and hippocampus segmentation. Promising labeling results have been achieved with comparison of the state-of-the-art labeling methods, which indicates the high applicability of our patch-based labeling method in neuroscience and clinical scenario.

2 Methods

Let S be the set of N atlas images $I = \{I_k | k = 1, \dots, N\}$ and their corresponding label maps $L = \{L_k | k = 1, \dots, N\}$, already registered to the target (to-be-labeled) image T . For each point $v \in \Omega_{I_k}$ at the domain of deformed atlas I_k , $\vec{L}_k(v)$ is its binary vector of $\{0,1\}^M$, representing the particular label, where M is the total number of labels.

Thus, the goal of label fusion for labeling the target image T is to propagate the labels from the registered atlases to the target image. For each point $u \in \Omega_T$ in the target image T , its label $\vec{L}_T(u)$ will be estimated through the interaction between the target patch $\mathcal{P}_T(u)$ centered at point u and all possible candidate atlas patches $\mathcal{P}_k(v)$ from the registered atlas images I_k . In order to reduce the computation burden, the search range is usually confined to a relatively small neighborhood $n(u) \subset \Omega_T$. Given the weight $w_k(u, v)$ associated with $\mathcal{P}_T(u)$ and $\mathcal{P}_k(v)$, we can calculate the assignment vector $\theta(u)$ for voxel u as:

$$\vec{\theta}(u) = \frac{\sum_{k=1}^N \sum_{v \in n(u)} w_k(u, v) \vec{L}_k(v)}{\sum_{k=1}^N \sum_{v \in n(u)} w_k(u, v)}. \quad (1)$$

It is worth noting that $\vec{\theta}(u) = [\theta^1(u), \dots, \theta^m(u), \dots, \theta^M(u)]$ is a vector of likelihood for each label at point u after label fusion. Then, the final label at point u can be determined by binarizing the vector $\vec{\theta}(u)$ to a binary label vector $\vec{L}_T(u) = [l^1(u), \dots, l^m(u), \dots, l^M(u)]$, where

$$l^m(u) = \begin{cases} 1, & \text{if } \theta^m(u) \text{ has the highest value} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

2.1 Conventional Patch-Based Labeling Method by Non-local Voting

For each point $u \in \Omega_T$ in the target image, we vectorize the local patch $\mathcal{P}_T(u)$ (red box in Fig. 1(a)) into the column vector \vec{y} . In order to account for the registration uncertainty, a set of candidate patches are included in a search neighborhood $n(u)$ (blue boxes in Fig. 1(a)) from different atlas images. For clarity, we arrange each candidate patch $\mathcal{P}_k(v)$ (pink boxes in Fig. 1(a)) into a column vector \vec{a}_j and then assemble them into a matrix $A = [\vec{a}_j]_{j=1, \dots, Q}$, where $j = (k, v)$ is a bivariate index of particular candidate patch $\mathcal{P}_k(v)$ and $Q = N \cdot |n(u)|$ denotes the total number of candidate patches. With the same order, we assemble the label vector $\vec{L}_k(v)$ of each candidate patch into the label matrix, denoted as $\Lambda = [\vec{\lambda}_1, \vec{\lambda}_2, \dots, \vec{\lambda}_Q]$. In non-local voting, each candidate patch \vec{a}_j contributes to label fusion, with the non-local weight w_j in the column vector $\vec{w} = [w_1, w_2, \dots, w_Q]'$ calculated by:

$$w_j = w_k(u, v) = e^{-\frac{\|\vec{y} - \vec{a}_j\|^2}{\sigma^2}} \quad (3)$$

where σ is named as the decay parameter in [1] to control the strictness on distance penalty. Given the weighting vector \vec{w} for the point u , we are able to predict the label $\vec{\eta} = \vec{L}_T(u)$ by following Eqs. (1) and (2). The procedure of non-local voting is demonstrated in Fig. 1(b), where each one-end arrow indicates the independent calculation of weight w_j for each candidate patch \vec{a}_j .

2.2 Joint Label Fusion with Sparse Patch Representation

In order to alleviate the issue of ambiguous patches in label fusion, sparsity constraint is imposed so that only a small number of candidate patches will contribute to label

fusion. Specifically, by regarding the set of candidate patches \mathbf{A} as the local dictionary for the point u , the weights for label fusion can be obtained by solving the following LASSO problem [7]:

$$\widehat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\tilde{\mathbf{y}} - \mathbf{A}\mathbf{w}\|_2^2 + \rho \|\mathbf{w}\|_1, \quad s. t. w_j \geq 0, \forall j, \quad (4)$$

where ρ is the scalar which controls the strength of sparsity. Eventually, most elements in the weighting vector \mathbf{w} approach to zero after imposing sparsity constraint. As demonstrated in Fig. 1(c), only a few candidate patches (connected to the red box with solid curves) will contribute to the final label fusion, which is different from the conventional non-local based methods (Fig. 1(b)).

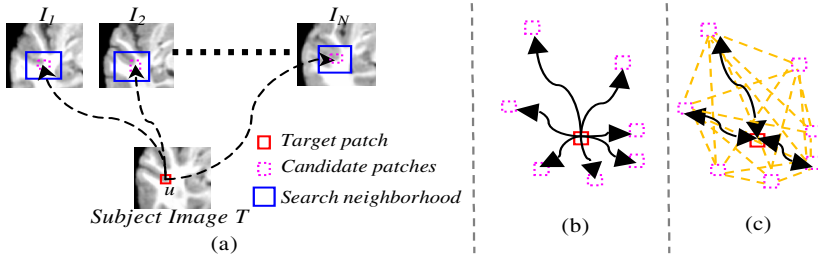


Fig. 1. The overview of patch-based labeling method in multi-atlas scenario. As shown in (a), the target patch (red box) seeks for the linear combination from all possible candidate atlas patches (pink boxes) in a search neighborhood (blue box). (b) and (c) demonstrate the labeling procedure on the particular point by non-local voting and our method, respectively.

It is apparent that Eq. 4 does not account for the dependency among candidate patches in label fusion. In order to make the estimation tractable, we define ϕ_{ij} as the joint risk of mislabeling that two candidates patches \vec{a}_i and \vec{a}_j simultaneously suffer from labeling error. Thus, besides seeking for the optimal patch representation by Eq. 4, the weight vector \mathbf{w} is also required to minimize the joint risk of labeling error by any pair of patches as

$$\widehat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{i=1}^Q \sum_{j=1}^Q w_i \phi_{ij} w_j = \arg \min_{\mathbf{w}} \mathbf{w} \Phi \mathbf{w}', \quad (5)$$

where $\Phi = [\phi_{ij}]_{Q \times Q}$ is a symmetric matrix.

In general, we can learn the joint risk ϕ_{ij} in two ways: (1) correlation of morphological error patterns (w.r.t. \vec{y}) between each pair of a_i and a_j in the dictionary \mathbf{A} , and (2) the labeling consensus between a pair of λ_i and λ_j in \mathbf{A} and the tentative label fusion result $\vec{\eta}$. For the first criterion, we assume that two candidate patches \vec{a}_i and \vec{a}_j have high chance to produce similar labeling error only if their error patterns, i.e., $(\vec{a}_i - \vec{y})$ and $(\vec{a}_j - \vec{y})$, are highly correlated. Thus, we define ϕ_{ij}^A as the correlation of two patches in terms of error pattern as ¹:

$$\phi_{ij}^A = \frac{(\vec{a}_i - \vec{y}) \cdot (\vec{a}_j - \vec{y})}{\|\vec{a}_i - \vec{y}\| \cdot \|\vec{a}_j - \vec{y}\|}. \quad (6)$$

¹ We will further normalize ϕ_{ij}^A to the range from 0 to 1.

Given the tentatively estimated weighting vector $\widehat{\mathbf{w}}$, we can determine the labeling result $\widehat{\eta}$ by Eq. 1~2. Here, we go one step further to examine whether the estimation $\widehat{\eta}$ achieves the largest labeling consensus between any pair of label λ_i on patch \vec{a}_i and λ_j on patch \vec{a}_j . We use ϕ_{ij}^Λ to measure the labeling consensus between λ_i and λ_j as:

$$\phi_{ij}^\Lambda = 1 - \frac{[\delta(\lambda_i - \widehat{\eta}) + \delta(\lambda_j - \widehat{\eta})]}{2}, \quad (7)$$

where $\delta(\cdot)$ is the Dirac pulse function at zero, i.e., $\delta(0) = 1$. It is apparent that $\phi_{ij}^\Lambda = 1$ only if both candidate patches \vec{a}_i and \vec{a}_j bear different labels against $\widehat{\eta}$, for which our method will panelize the weights w_i and w_j to have high values. In this way, our method is able to not only prevent the repeated labeling error from two correlated patches, but also provide our label fusion method a chance to iteratively correct the possible mislabeling by refining the labeling consensus ϕ_{ij}^Λ . Finally, the joint risk of labeling error ϕ_{ij} is defined as:

$$\phi_{ij} = (1 - r) \cdot \phi_{ij}^\Lambda + r \cdot \phi_{ij}^\Delta \quad (8)$$

where $0 \leq r \leq 1$ is the scalar balancing the two measurements. $r = 0$ in the beginning since there is no estimation of $\widehat{\eta}$ at that moment. We increase r from 0 to 0.5 linearly and iteratively. As we will show later, the labeling performance keeps improving as iteration progresses.

By integrating Eq. 8 into Eq. 4, the new energy function can be formulated as:

$$(\widehat{\mathbf{w}}, \widehat{\eta}) = \arg \min_{(\mathbf{w}, \eta)} \|\vec{y} - \mathbf{A}\vec{w}\|_2^2 + \beta \cdot \vec{w} \Phi \vec{w}' + \rho \cdot \|\vec{w}\|_1, \quad s. t. w_j \geq 0, \forall j, \quad (9)$$

where β controls the strength of joint labeling risk. Fig. 1(c) demonstrates the improvements of our method: (1) a small number of candidate patches are eventually involved in label fusion (i.e., few dash boxes are connected with red box by solid curves in Fig. 1(c)); (2) the dependencies of candidate patches are clearly described (as indicated by the dash lines in Fig. 1(c)); (3) the label fusion result $\widehat{\eta}$ is recursively used to refine the labeling result by updating the joint labeling risk Φ (as indicated by the two-end arrows in Fig. 1(c)). It is worth noting that our method will degrade to the sparse patch-based labeling method [8] if $\hat{\mathbf{a}} = \mathbf{0}$.

To solve this problem, we alternatively repeat two sub-steps, i.e., (1) optimizing \vec{w} by fixing $\widehat{\eta}$, and (2) updating Φ with the latest $\widehat{\eta}$ by Eq. 1~2. For step (1), we use coordinate descent method [9] to efficiently find \vec{w} . The idea is to go through each w_j in \vec{w} and minimize the energy function in Eq. 9 along one w_j at a time. Specifically, for each w_j , we discard all terms in Eq. 9 that are not related with w_j and turn Eq. 9 into:

$$\widehat{w}_j = \arg \min_{w_j} \|\vec{\xi} - w_j \vec{a}_j\|_2^2 + \beta \phi_{jj} w_j^2 + 2\beta b w_j + \rho |w_j|, \quad (10)$$

where $\vec{\xi} = \vec{y} - \sum_{i=1, i \neq j}^Q w_i \vec{a}_i$ and $b = \sum_{i=1, i \neq j}^Q \phi_{ij} w_i$.

By letting $\mathfrak{l} = (\vec{\xi}^T \vec{a}_j - \beta b) / (\vec{a}_j^T \vec{a}_j + \beta \phi_{jj})$ and $\tau = \rho / (\vec{a}_j^T \vec{a}_j + \beta \phi_{jj})$, we further rewrite Eq. 10 as:

$$\widehat{w}_j = \arg \min_{w_j} (w_j - \mu)^2 + \tau |w_j|, \quad s. t. w_j \geq 0, \quad (11)$$

which turns to the classic l_2 regression problem [9]. Thus, by requiring the directional derivative along the coordinate direction of w_j coincide with the ordinary partial derivative $\partial(w_j - \mu)^2 / \partial w_j$, the optimal solution to particular \hat{w}_j is given as:

$$\hat{w}_j = \begin{cases} \mu - \frac{\tau}{2} & \mu > \frac{\tau}{2} \\ 0 & \mu \leq \frac{\tau}{2} \end{cases} \quad (12)$$

In our implementation, we randomly set the order to visit each element w_j and fix the iteration number as 200 in optimizing \bar{w} for each point u in the target image.

3 Experiments

We apply our joint patch-based labeling method (Joint-PBM) on both ADNI and NIREP-NA40 [10] datasets to evaluate the labeling performance. For comparison, we also deploy the conventional patch-based method by non-local weighting (Nonlocal-PBM) and recently proposed sparse patch-based method [8] (Sparse-PBM) on the same datasets. For each subject, we first use FLIRT in FSL package (<http://fsl.fmrib.ox.ac.uk/>) and then diffeomorphic Demons²[11] to deform all atlases to the underlying subject space.

In the following experiment, we fix the patch size as $7 \times 7 \times 7$ and the search range as $9 \times 9 \times 9$ for all label fusion methods. We follow the patch pre-selection strategy and local adaptive selection of decay parameter in [1] for Nonlocal-PBM. Also, according to [8], we set the parameter for l_1 -norm strength as 0.1 for sparse-PBM. For our Joint-PBM method, we fix $\beta = 1.0$ and $\rho = 0.1$ for all experiments. Particularly, we repeat the steps of estimating \bar{w} and updating labels $\bar{\eta}$ for 5 times.

3.1 Experiment Result on Hippocampus Labeling in ADNI Dataset

In this experiment, we randomly select 61 NC (normal control) subjects, 96 MCI (Mild Cognitive Impairment) subjects, and 41 AD (Alzheimer’s disease) subjects from the ADNI dataset. For each subject image, the skull is stripped and the intensity range has been normalized by histogram matching. The segmentation ground truth of hippocampus for each image is also provided in the ADNI dataset.

The overall Dice ratio of left/right hippocampus by three label fusion methods are shown in Table 1, where our Joint-PBM method has achieved 4.7% and 2.2% improvements over Nonlocal-PBM and Sparse-PBM, respectively. Moreover, we show the average and standard deviation of Dice ratio in NC, MCI and AD groups by three label fusion methods in Table 2. Again, our Joint-PBM method has the best labeling performance in each group. It is worth noting that the highest Dice ratio of hippocampus is 0.893 in [4]. However, only 57 NC and 82 MCI subjects are included in [4].

² The parameters for diffeomorphic Demons are: 15, 15, and 15 iterations in low, middle, and high resolutions. The kernel size for deformation smoothing is set to 1.8. The option of histogram matching is turned on.

As shown in Table 2, labeling hippocampus of AD subjects are more challenging than MCI and NC groups. Discarding the AD subject, the overall Dice ratio of NC and MCI groups by our method is able to reach 0.896. As our labeling method iteratively refines the labeling result by increasing r (Eq. 8), the Dice ratio increases consistently from 0.882 to 0.887 in the end of labeling.

Table 1. The mean and standard deviation of Dice ratio on left/right hippocampus by Nonlocal-PBM, Sparse-PBM, and Joint-PBM

Method	Left Hippocampus	Right Hippocampus	Overall
Nonlocal-PBM	0.854 \pm 0.040	0.849 \pm 0.043	0.852 \pm 0.042
Sparse-PBM	0.877 \pm 0.032	0.869 \pm 0.036	0.873 \pm 0.034
Joint-PBM	0.890 \pm 0.022	0.884 \pm 0.023	0.887 \pm 0.022

Table 2. The mean and standard deviation of Dice ratio in hippocampus labeling for three groups (NC, MCI, and AD) by Nonlocal-PBM, Sparse-PBM, and Joint-PBM

Method	NC	MCI	AD
Nonlocal-PBM	0.866 \pm 0.034	0.859 \pm 0.039	0.831 \pm 0.046
Sparse-PBM	0.882 \pm 0.030	0.873 \pm 0.036	0.864 \pm 0.041
Joint-PBM	0.899 \pm 0.014	0.893 \pm 0.019	0.870 \pm 0.032

3.2 Experiment Result on Whole Brain Labeling in NIREP-NA40 Dataset

The NIREP-NA40 dataset consists of 16 MR images of 8 normal male adults and 8 normal female adults, each with 32 manually-delineated ROIs. The image size is $256 \times 300 \times 256$ and the voxel resolution is $0.7 \times 0.7 \times 0.7mm^3$.

Table 3. The average Dice ratio in NIREP-NA40 dataset by three label fusion methods

ROI (left+right)	Nonlocal-PBM	Sparse-PBM	Joint-PBM
Occipital Lobe	0.801 / 0.813	0.815 / 0.820	0.833 / 0.859
Cingulate Gyrus	0.815 / 0.812	0.811 / 0.814	0.819 / 0.852
Insula Gyrus	0.851 / 0.873	0.855 / 0.878	0.862 / 0.890
Temporal Pole	0.837 / 0.829	0.838 / 0.841	0.842 / 0.875
Superior Temporal Gyrus	0.779 / 0.777	0.781 / 0.784	0.801 / 0.811
Infero Temporal Region	0.832 / 0.833	0.848 / 0.832	0.867 / 0.871
Parahippocampal Gyrus	0.829 / 0.843	0.831 / 0.851	0.842 / 0.864
Frontal Pole	0.820 / 0.804	0.824 / 0.820	0.849 / 0.852
Superior Frontal Gyrus	0.807 / 0.785	0.805 / 0.800	0.822 / 0.837
Middle Frontal Gyrus	0.791 / 0.753	0.809 / 0.763	0.819 / 0.805
Inferior Gyrus	0.755 / 0.751	0.758 / 0.775	0.785 / 0.790
Orbital Frontal Gyrus	0.833 / 0.831	0.841 / 0.835	0.863 / 0.860
Precentral Parietal Lobule	0.762 / 0.744	0.785 / 0.762	0.801 / 0.789
Superior Parietal Lobule	0.742 / 0.739	0.756 / 0.780	0.802 / 0.805
Inferior Parietal Lobule	0.759 / 0.738	0.771 / 0.802	0.800 / 0.812
Postcentral Gyrus	0.707 / 0.694	0.713 / 0.711	0.756 / 0.738

The leave-one-out strategy is used in this experiment by alternatively taking 16 images as the target image in each leave-one-out case. Table 3 shows the average Dice ratio in each brain structure (left and right combined) across 16 leave-one-out cases. The overall Dice ratios across 32 ROIs are 0.792 by Nonlocal-PBM, 0.803 by Sparse-PBM, and 0.827 by Joint-PBM. Our Joint-PBM method achieves the best labeling accuracy over the other two methods, by 3.5% and 2.4%, respectively.

4 Conclusion

In this paper, we present a novel patch-based label fusion method for multi-atlas segmentation. Sparsity constraint is introduced in our method to suppress the misguidance from ambiguous patches. Furthermore, our method explicitly describes the dependency of patches from different atlases or same atlas to guide the label fusion, which is iteratively updated based on the latest label fusion result. In this way, our method is able to gradually improve the labeling accuracy by reducing the chance of making the repeated labeling errors. Our experiments on hippocampus segmentation and whole brain parcellation show the promising labeling results, indicating its applicability in neuroscience and clinical studies.

References

1. Coupe, P., et al.: Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954 (2011)
2. Rousseau, F., Habas, P.A., Studholme, C.: A Supervised Patch-Based Approach for Human Brain Labeling. *IEEE Trans. Medical Imaging* 30(10), 1852–1862 (2011)
3. Tong, T., et al.: Segmentation of Brain Images via Sparse Patch Representaion. In: STML, Nice, France (2012)
4. Wang, H., et al.: Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* (2012)
5. Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: *Computer Vision and Pattern Recognition*, Beijing, China (2005)
6. Vinje, W.E., Gallant, J.L.: Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287(5456), 1273–1276 (2000)
7. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58(1), 267–288 (1996)
8. Zhang, D., Guo, Q., Wu, G., Shen, D.: Sparse Patch-Based Label Fusion for Multi-Atlas Segmentation. In: Yap, P.-T., Liu, T., Shen, D., Westin, C.-F., Shen, L. (eds.) *MBIA 2012. LNCS*, vol. 7509, pp. 94–102. Springer, Heidelberg (2012)
9. Wu, T.T., Lange, K.: Coordinate descent algorithms for lasso penalized regression. *The Annals of Applied Statistics* 2(1), 224–244 (2008)
10. Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the non-rigid image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) *WBIR 2006. LNCS*, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
11. Vercauteren, T., et al.: Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage* 45(suppl. 1), S61–S72 (2009)

Lateral Ventricle Segmentation of 3D Pre-term Neonates US Using Convex Optimization

Wu Qiu, Jing Yuan, Jessica Kishimoto, Eranga Ukwatta, and Aaron Fenster

Robarts Research Institute, University of Western Ontario, London, ON, Canada

Abstract. Intraventricular hemorrhage (IVH) is a common disease among preterm infants with an occurrence of 12-20% in those born at less than 35 weeks gestational age. Neonates at risk of IVH are monitored by conventional 2D ultrasound (US) for hemorrhage and potential ventricular dilation. Compared to 2D US relying on linear measurements from a single slice and visually estimates to determine ventricular dilation, 3D US can provide volumetric ventricle measurements, more sensitive to longitudinal changes in ventricular volume. In this work, we propose a global optimization-based surface evolution approach to the segmentation of the lateral ventricles in preterm neonates with IVH. The proposed segmentation approach makes use of convex optimization technique in combination with a subject-specific shape model. We show that the introduced challenging combinatorial optimization problem can be solved globally by means of convex relaxation. In this regard, we propose a coupled continuous max-flow model, which derives a new and efficient dual based algorithm, that can be implemented on GPUs to achieve a high-performance in numerics. Experiments demonstrate the advantages of our approach in both accuracy and efficiency. To the best of our knowledge, this paper reports the first study on semi-automatic segmentation of lateral ventricles in neonates with IVH from 3D US images.

Keywords: lateral ventricle segmentation, convex optimization, pre-term neonate, 3D ultrasound imaging.

1 Introduction

Preterm neonates that are born of low birth weight ($< 1500g$) are at an increased risk of intraventricular hemorrhage (IVH) [1], bleeding in and around the ventricles. Diagnosis of IVH is done using 2D clinical ultrasound (US). The hemorrhage is ranked using the Papile grading system [2] that qualitatively describes the location and approximate volume of ventricle affected. By this system, grade I and II indicate a minor hemorrhage and tends towards favourable patient outcomes, and grades III and IV indicate severe hemorrhage often with following ventricular dilation (ventriculomegaly) and poor patient outcomes. Patients often have ventriculomegaly that spontaneously resolves, but some fraction of those will progress to hydrocephalus, and may require interventional therapy to reduce the amount of cerebral spinal fluid (CSF) accumulated in the brain. 3D US can monitor the ventricular system in neonates [3,4], and can be done at the

bed side, however; to incorporate 3D US into clinical setting, a segmentation algorithm would be required to adequately reduce the time required to obtain the ventricle volume from the 3D image.

Previous cerebral ventricle segmentation algorithms have been exclusively for CT [5] and MRI images [6], and mainly for adult populations. While studies have quantified 3D US ventricle volumes in neonates [3,4], all have used expert manually contoured regions in lieu of a semi-automatic or automatic approach. Unlike in a healthy neonate, IVH patients ventricles provide an increased difficulty to segment due to both the bleed itself as well as the irregular manner in which the ventricle deforms. The bleeding creates a hyperechoic region around and inside the relatively hypoechoic region of the CSF filled ventricles making most threshold or boundary based segmentation algorithms not useful for this application. Atlas based segmentations fail as dilation and subsequent reduction in conjunction with brain growth cause deformations to the structure that cannot be accounted for and vary drastically from patient to patient. Manual segmentation, though done in previous studies to quantify 3D US ventricle volumes [3,4], is too arduous and time consuming to be clinically feasible. Moreover, manual segmentation of such a structure is challenging due to indistinct boundaries and irregular shape deformation, even for an expert. Thus, accurate and efficient automatic or semi-automatic ventricle segmentation is highly desired in clinical practice.

Contributions: In this study, we propose convex optimization based approach for delineating lateral ventricles of pre-term neonates from 3D US images, which incorporates a subject-specific model as the shape prior. We show that the introduced challenging combinatorial optimization problem can be globally optimized by means of convex relaxation. Moreover, we introduce a coupled continuous max-flow model which is equivalent to the formulated convex relaxed optimization problem. With the help of the coupled continuous max-flow formulation, we derive an efficient multiplier augmented algorithm, which avoid directly tackling the original non-smooth convex energy functional and can be readily implemented on a GPU to achieve a substantial speed-up in computation. To the best of our knowledge, this paper reports the first study on semi-automatic segmentation of lateral ventricles of premature newborn brains from 3D US images.

2 Method

2.1 Segmentation Pipeline

A specific subject model is used to facilitate the segmentation task considering great individual shape variations and inhomogeneous context. A 3D US image at baseline is first segmented manually by an expert. Then, two segmented lateral ventricle surfaces (left and right lateral ventricles) as subject-specific models are rigidly registered to the following repeat images (from the second time-point) through six chosen landmarks. The registered models are used as an initial guess of lateral ventricles for repeat images, estimating the prior probability density

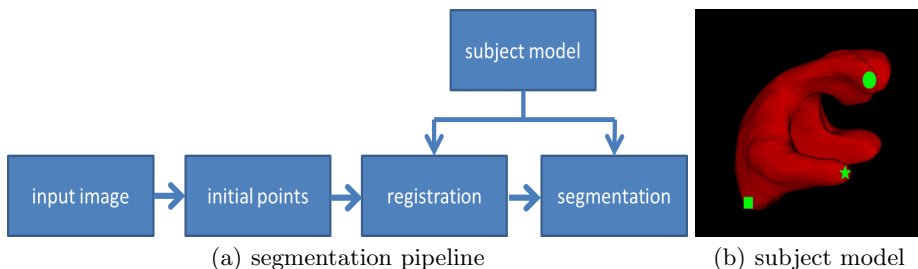


Fig. 1. Schematic diagram of segmentation pipeline; (a) segmentation pipeline, (b) example of a 3D rendered ventricle model with landmarks

functions (PDF) for the background and foreground. The subject-specified models also served as shape priors to assist the subsequent convex optimization based segmentation algorithm. Figure 1(a) shows the segmentation pipeline.

2.2 Rigid Registration

Three landmarks, locating at anterior, posterior, and inferior horn of right lateral ventricle, respectively (green circle, square, and star in Fig. 1(b)), are manually chosen from each ventricle in a given subject-specified model at baseline, resulting in six landmarks for the two ventricles. Six corresponding initial landmarks are also manually chosen from the repeat images. An affine transformation is then calculated from those six corresponding point pairs, which are used to rigidly register the subject model to the following repeat images. It should be noted that a rigid registration can not guarantee a good segmentation for repeat images due to the nonlinear deformation caused by neonate brain growing, ventricle deformation and bleeding. Thus, a more accurate segmentation algorithm is needed to refine the rough registration initial result.

2.3 Convex Optimization Based Surface Evolution with Shape Constraint

In this work, a convex optimization based method with shape constraint was used to refine the registered surface. We denote \mathcal{R}_i , $i = L, R$, and B as the left ventricle, right ventricle, and background pixels, respectively; $u_i \in \{0, 1\}$, $i = L, R$, and B are denoted as the corresponding indicator labeling functions of \mathcal{R}_L , \mathcal{R}_R , and \mathcal{R}_B , respectively. Since there is no overlap between left and right ventricle in \mathcal{R} , we have

$$u_L \cap u_R = \emptyset, \quad \forall x \in \mathcal{R}. \quad (1)$$

Multiple-Region Segmentation: The segmentation of \mathcal{R} into three regions of \mathcal{R}_B , \mathcal{R}_L and \mathcal{R}_R can be formulated as a coupled continuous min-cut problem, which minimizes the following energy function:

$$\min_{u_{L,R}(x) \in \{0,1\}} \sum_{i=L,R} \left\{ E_{\text{matching}}(u_i) + \int_{\mathcal{R}} g_i(x) |\nabla u_i(x)| dx \right\} \quad (2)$$

where $E_{\text{matching}}(u)$ is Bhattacharyya distance [7,8] used to measure the distance between the estimated PDFs for the two estimated regions and their corresponding PDF models(see Section 2.1); the weight function $g(x)$ in (2) is positive and is given by

$$g(x) = \lambda_1 + \lambda_2 \exp(-\lambda_3 |\nabla I(x)|), \quad \lambda_{1,2,3} \geq 0. \tag{3}$$

Note that, $I(x) \in \mathcal{Z}$ be a given repeat image, where \mathcal{Z} is the set of image intensity values, and the values of $g(x)$ fall within the range $[\lambda_1, \lambda_1 + \lambda_2]$. We can estimate the PDF $p_i(u, z)$, where $z \in \mathcal{Z}$ and $i = L, R, B$, for the estimated region \mathcal{R} by the Parzen method [9], such that

$$p(u, z) = \frac{\int_{\mathcal{R}} K(z - I(x)) u \, dx}{\int u \, dx},$$

where $K(\cdot)$ is the Gaussian kernel function such that $K(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/2\sigma^2)$. $q(z)$ is defined as the intensity PDF model of the region \mathcal{R} , where $z \in \mathcal{Z}$. Thus, $E_{\text{matching}}(u)$ is defined as:

$$E_{\text{matching}}(u) = - \sum_{z \in \mathcal{Z}} \sqrt{p(u, z) q(z)} \tag{4}$$

Convex optimization with shape constraint: In addition, we penalize the difference between the evolved ventricle surface and the registered subject-specified model so as to enforce that the resulting surface is not far from the initial guess. Let \mathcal{C}^* be the registered surface model, and $u^*(x) \in \{0, 1\}$ the indicator function of \mathcal{C}^* , which is used to assist the segmentation task of \mathcal{C} [10]. The indicator function $u(x) \in \{0, 1\}$ of the ventricle regions within \mathcal{C} is optimized over the following energy function:

$$\min_{u_{L,R}(x) \in \{0,1\}} \sum_{i=L,R} \left\{ E_{\text{matching}}(u_i) + \omega_1 \int_{\mathcal{R}} g_i(x) |\nabla u_i(x)| \, dx + \omega_2 \int_{\mathcal{R}} |u_i - u_i^*| \, dx \right\} \tag{5}$$

where $E_{\text{matching}}(u)$ formulates the statistical intensity distribution matching energy inside and outside two ventricles, the second weighted total-variation function gives the surface smoothness term and $\omega_{1,2} > 0$ are the positive penalty parameters ($\omega_1 = 0.3$ and $\omega_2 = 0.05$ were used in our experiments). The last energy term of (5) encodes and penalizes the difference between the evolved ventricle surface and the registered model.

To optimize the energy function (5), which is often highly nonlinear, we introduce the convex optimization based evolution approach [11], which can efficiently move the given surface to the object of interest. In contrast to the traditional evolution methods, e.g. active contour or level-set, the introduced surface evolution approach provides a fully time implicit scheme in numerics, for which a large time step-size is allowed to significantly speed up the surface evolution process and, during each discrete evolution time frame, the surface can be moved to its

globally optimal position by globally solving the following continuous min-cut problem:

$$\min_{u_{L,R}(x) \in \{0,1\}} \sum_{i=L,R} \left\{ \langle 1 - u_i, C_s^i \rangle + \langle u_i, C_t^i \rangle + \int_{\mathcal{R}} g_i(x) |\nabla u_i| dx \right\} \quad (6)$$

where the functions $C_{s,t}^i(x)$ are set up w.r.t. the current surface[11]. The continuous min-cut problem (6) can be solved globally and efficiently by means of the continuous max-flow method[12] and implemented on the modern parallel computing platforms (GPUs) to obtain the high-performance in numerics. The proof of the duality between the continuous max-flow model and the convex relaxed optimization model (6) is omitted due to the limited space.

3 Experiments

Image acquisition: A motorized 3D US system developed for cranial US scanning of pre-term neonates was used to acquire the images. Following the routine cranial US exam, the 2D US transducer (Phillips C8-5 broadband curved array) is placed into the motorized housing and the 3D US image is acquired. The US technician locates the midline of lateral ventricles through the anterior fontanelle. Images are then acquired while the motor housing is held firmly while the device rotates the transducer at an axis at the probe tip. Scans were performed with a 60-72 degree scan angle, a step size of 0.3 degrees at a frame rate of 25 frames/s with scan times between 8-12 seconds [13]. The image sizes ranged from $300 \times 300 \times 300$ to $450 \times 450 \times 450$ voxels at the same voxel spacing of $0.22 \times 0.22 \times 0.22 \text{mm}^3$. Scans were performed 1-2 times per week for the first month enrolled in the study and 1-4 per month for the duration of the patients stay in the neonatal intensive care unit.

Evaluation Metrics: Ventricles were manually segmented in parallel slices of the 3D US image with 0.75 mm between slices. Our segmentation method was then evaluated by comparing the algorithm to manual segmentation results using *volume-based metrics*: Dice similarity coefficient (DSC)[14]; and *distance-based metrics*: the mean absolute surface distance (MAD) and maximum absolute surface distance (MAXD)[15]. In addition, each image was segmented three times by the same observer for assessing the intra-observer variability. The mean run time of three repeated segmentations for each 3D US image was considered as the segmentation time to assess the algorithm's efficiency.

4 Results

Accuracy: Figure. 2 shows two algorithm segmented lateral ventricles (green contours) and manual delineations (red contours) of one patient, demonstrating good agreement. Table. 1 shows the mean quantitative segmentation results for 20 patient images using the proposed method. Our approach obtained a mean

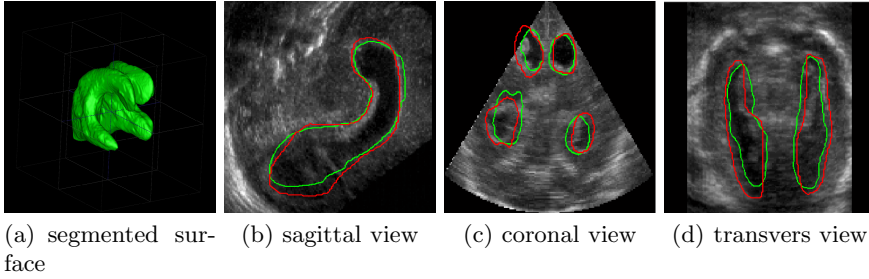


Fig. 2. Segmented ventricles (green contour) overlapped with manual segmentations (red contour)

Table 1. Segmentation results of twenty 3D US images of four patients (five time points each) in terms of DSC, MAD, and MAXD (p1-p4: patient ID)

	p1	p2	p3	p4	Mean
DSC (%)	71.2 ± 1.5	74.3 ± 2.0	70.5 ± 2.1	70.5 ± 2.1	72.4 ± 2.5
MAD (mm)	0.5 ± 0.2	0.6 ± 0.1	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1
MAXD (mm)	3.8 ± 1.2	2.7 ± 1.6	3.1 ± 2.2	3.4 ± 1.5	3.5 ± 1.6

DSC of $72.4 \pm 2.5\%$, a MAD of $0.7 \pm 0.1mm$, and a MAXD of $3.5 \pm 1.6mm$ for the patients' two ventricles. Considering that values of DSC above 70% are usually regarded as a satisfactory level of agreement between two segmentations[14,16], the segmentation accuracy generated by our method is useful for clinic use.

Reliability: The result of the intra-observer variability test showed that the proposed method yielded a DSC of $72.5 \pm 2.1\%$, $73.4 \pm 2.6\%$ and $73.2 \pm 2.4\%$ for three segmentations from the same observer, respectively. ANOVA analysis with a single factor failed to demonstrate a statistically significant difference between these three segmentations ($p = 0.80$, $F = 0.75$). The approximately similar DSCs yielded by algorithm in all three repetitions suggest a high reproducibility of our approach.

Computational Time: The proposed approach was implemented using parallel computing architecture (CUDA, NVIDIA Corp., Santa Clara, CA) and the user interface in Matlab (Natick, MA). The experiments were conducted on a Windows desktop with an Intel i7-2600 CPU (3.4 GHz) and a GPU of NVIDIA Geforce 5800X. The segmentation time including registration and convex optimization was calculated as the mean run time of three repeated segmentations for each 3D US image. The mean segmentation time of our method was $15 \pm 2s$ in addition to $50 \pm 5s$ for initialization, resulting in a total segmentation time of less than **1.5 minutes** for a single 3D lateral ventricle US image, significantly less than 30 minutes for manual segmentation.

5 Discussion and Conclusion

This paper proposes an accurate, reproducible and numerically efficient solution to a challenging segmentation problem of the lateral ventricles of pre-term neonates from 3D US images, which makes use of the latest development of convex optimization technique combined with a subject-specific shape model. The experimental results show that the proposed method is reliable to the initial segmentation of the first image, where the rough manual segmentation is registered to the other images as the shape prior, in combination with the local image features, to assist the segmentation of the following images. In practice, the local image appearance can derive the correct region boundaries even when the shape prior has errors. In addition, the intra-observer experiments show that the variability introduced by the user-selected shape prior is small in terms of DSC, which demonstrates that the proposed method is reliable at least for a single expert. Although the current algorithm is semi-automatic, and still requires manual segmentation for the first image, the 1.5 min computational time is still promising for repeat images in clinical trials compared to half an hour for a currently used manual segmentation. The algorithm performance suggests that it has the potential to be used for measuring the volume of lateral ventricles of pre-term neonates.

Many investigators are focusing on segmentation of brain anatomic structures in 3D MRI, such as lateral ventricle and hippocampus. However, these are found to be application-dependent, which can not be directly applied in our 3D US application due to the difference of modality and segmentation object. We tested the classic active contour [17], level set [18], and graph cut [19] algorithms in our dataset. Visual inspection show that all of these failed to generate favourable results for all testing images. It should be noted that our algorithm cannot handle the situation when bleeding extends outside ventricles (grade IV haemorrhage), which greatly concerns obstetricians and neonatologist along with ventricle volume.

Acknowledgments. The authors are grateful for the funding support from the Canadian Institutes of Health Research (CIHR), the Ontario Institute of Cancer Research (OICR), the Canada Research Chairs (CRC) Program, and Academic Medical Organization of Southwestern Ontario (AMOSO).

References

1. Wilson-Costello, D., Friedman, H., Minich, N., Fanaroff, A.A., Hack, M.: Improved survival rates with increased neurodevelopmental disability for extremely low birth weight infants in the 1990s. *Pediatrics* 115(4), 997–1003 (2005)
2. Papile, L.A., Burstein, J., Burstein, R., Koffler, H.: Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. *The Journal of Pediatrics* 92(4), 529–534 (1978)
3. Abdul-Khaliq, H., Vogel, M., Lange, P.: Feasibility of brain volumetric analysis and reconstruction of images by transfontanel three-dimensional ultrasound. *Journal of NeuroImaging* 10(3), 147–150 (2000)

4. McLean, G., Coombs, P., Sehgal, A., Paul, E., Zamani, L., Gilbertson, T., Ptasznik, R.: Measurement of the lateral ventricles in the neonatal head: Comparison of 2-d and 3-d techniques. *Ultrasound in Medicine & Biology* (2012)
5. Liu, J., Huang, S., Ihar, V., Ambrosius, W., Lee, L.C., Nowinski, W.L.: Automatic model-guided segmentation of the human brain ventricular system from ct images. *Academic Radiology* 17(6), 718–726 (2010)
6. Liu, J., Huang, S., Nowinski, W.L.: Automatic segmentation of the human brain ventricles from mr images by knowledge-based region growing and trimming. *Neuroinformatics* 7(2), 131–146 (2009)
7. Michailovich, O., Rathi, Y., Tannenbaum, A.: Image segmentation using active contours driven by the bhattacharyya gradient flow. *IEEE Transactions on Image Processing* 16(11), 2787–2801 (2007)
8. Ukwatta, E., Yuan, J., Rajchl, M., Qiu, W., Tessier, D., Fenster, A.: 3d carotid multi-region MRI segmentation by globally optimal evolution of coupled surfaces. *IEEE Tran. Med. Imag.* 32(4), 770–785 (2013)
9. Parzen, E.: On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3), 1065–1076 (1962)
10. Qiu, W., Yuan, J., Ukwatta, E., Tessier, D., Fenster, A.: Prostate segmentation in 3d TURS using convex optimization with shape constraint. In: *SPIE, Medical Imaging* (2013)
11. Yuan, J., Ukwatta, E., Tai, X.C., Fenster, A., Schnoerr, C.: A fast global optimization-based approach to evolving contours with generic shape prior. Technical report CAM-12-38, UCLA (2012)
12. Yuan, J., Bae, E., Tai, X.: A study on continuous max-flow and min-cut approaches. In: *CVPR 2010* (2010)
13. Kishimoto, J., Lee, D., Lawrence, K.S., Romano, W., Fenster, A., de Ribaupierre, S.: Development of a 3d ultrasound system to investigate post-hemorrhagic hydrocephalus in pre-term neonates. In: *SPIE Medical Imaging*, p. 86751M (2013)
14. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C.: Morphometric analysis of white matter lesions in mr images: method and validation. *IEEE Transactions on Medical Imaging* 13(4), 716–724 (1994)
15. Garnier, C., Bellanger, J.J., Wu, K., Shu, H., Costet, N., Mathieu, R., de Crevoisier, R., Coatrieux, J.L.: Prostate segmentation in HIFU therapy. *IEEE Trans. Med. Imag.* 30(3), 792–803 (2011)
16. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A.D., Rueckert, D., Hajnal, J.V.: Automatic segmentation and reconstruction of the cortex from neonatal mri. *Neuroimage* 38(3), 461–477 (2007)
17. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
18. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Img. Process.* 10(2), 266–277 (2001)
19. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(11), 1222–1239 (2001)

Semi-automatic Brain Tumor Segmentation by Constrained MRFs Using Structural Trajectories^{*}

Liang Zhao¹, Wei Wu², and Jason J. Corso¹

¹ Computer Science and Engineering, SUNY at Buffalo, Buffalo, NY, USA

² Wuhan University of Science and Technology, Wuhan, Hubei, China
{lzhao6, jcorso}@buffalo.edu

Abstract. Quantifying volume and growth of a brain tumor is a primary prognostic measure and hence has received much attention in the medical imaging community. Most methods have sought a fully automatic segmentation, but the variability in shape and appearance of brain tumor has limited their success and further adoption in the clinic. In reaction, we present a semi-automatic brain tumor segmentation framework for multi-channel magnetic resonance (MR) images. This framework does not require prior model construction and only requires manual labels on one automatically selected slice. All other slices are labeled by an iterative multi-label Markov random field optimization with hard constraints. Structural trajectories—the medical image analog to optical flow—and 3D image over-segmentation are used to capture pixel correspondences between consecutive slices for pixel labeling. We show robustness and effectiveness through an evaluation on the 2012 MICCAI BRATS Challenge Dataset; our results indicate superior performance to baselines and demonstrate the utility of the constrained MRF formulation.

1 Introduction

Magnetic resonance imaging provides detailed information of the human brain and is an essential tool for the neuro-oncologist. Quantifying the volume of a brain tumor is the key prognostic measurement of tumor progression [1, 2]. Yet, manually labeling a brain tumor in 3D MRI is a time-consuming and error-prone task. The medical imaging community has hence invested a significant amount of effort in methods for automatic brain tumor segmentation [3–7]. However, despite these efforts, there has been limited success in translation to the clinical environment; the current performance of automatic methods does not meet the prognostic needs [1] (e.g., the best performer in BRATS 2012 has a Jaccard score of 0.5 for high-grade tumor). This may be due to the underlying variability of tumor shape and appearance, or due to assumptions of the approaches.

Semi-automatic methods that require some input from the user are a plausible alternative. They have received comparatively little attention for brain tumor

^{*} This work was partially supported by the Chinese National Science Foundation (61273241) and the NSF CAREER grant IIS-0845282.

imaging, e.g., [2, 8, 9]. In contrast, 2D image interactive labeling has been hotly studied, e.g., GrabCut [10], LazySnapping [11]. Similar ideas have been used on 2D brain tumor segmentation, such as [8]. However, generalizing these ideas to 3D is non-trivial. The interaction mechanisms in these papers, such as drawing lines to specify foreground and background, moving a square brush or lasso, or giving a bounding box, are no longer trivial tasks in a 3D volume for complex objects, tumor and edema. Recently, some approaches for object segmentation in 3D medical images by iterative energy optimization based on shape model were proposed [12, 13]. However, compared to some organs, such as the liver, the edema and tumor shape are comparatively harder to model.

We propose an alternative mechanism for incorporating human input in semi-automatic brain tumor segmentation. Our main idea is based on the assumption that pixels with the same label in consecutive slices will have a similar feature distribution and strong spatial correspondence. Hence, we require the human to manually label only one slice (using standard 2D annotation techniques) and then all other slices are sequentially labeled based on a constrained Markov random field model. The constraints in the model are created based on a 3D notion of optical flow, which we call *structural trajectories*, and on over-segmentation. The initial slice to be labeled is automatically selected based on an asymmetry heuristic. Our thorough experiments on the BRATS 2012 data set [3] demonstrate the potential of our approach with limited input of only one manually labeled slice (our Jaccard score for high-grade tumor is 0.75, which is a 50% improvement over the best fully automatic method). To the best of our knowledge, this is the first time an optical flow-like calculation has been used to provide a global 3D structural consistency measure for brain tumor segmentation.

2 Methods

Our proposed semi-automatic segmentation approach labels the volume slice-by-slice using a constrained Markov random field (MRF) energy minimization on neighboring slices. In this procedure, the structural correspondence between adjacent slices is detected by optical flow estimation, which we call *structural trajectories* and helps in the iterative pixel labeling as hard constraints.

Given a sequence of slices, $S = \{s_i, i = 1 : n\}$, the whole procedure of the proposed approach is as follows:

1. Compute structural trajectories $TR = \{tr_j, j = 1 \dots M\}$.
2. Select the most asymmetric slice $s_{i_{max}}$.
3. Label pixels in $s_{i_{max}}$ as tumor, edema, or background manually.
4. In $s_{i_{max}+1}$ to s_n and $s_{i_{max}-1}$ to s_1 , label pixels slice-by-slice with Alg. 1.

2.1 Slice Selection by Asymmetric Area Detection

Consider a 3D MRI brain image, Img . We first flip it sagittally, yielding Img_{mirror} and then refine it with a non-rigid 3D registration [14] to the original Img . We then compare the refined flipped image Img_r to Img to look for regions of high

Algorithm 1. Sequential Optimization on Three Consecutive Slices**Input:** Three consecutive slices, $\{s_i, s_k, s_{2k-i}\} (k = i \pm 1)$. Labels of pixels in s_i .**Output:** Labels of pixels in s_k

- 1: Estimating the intensity distributions on s_i with GMM
- 2: Selecting some pixels in s_k to label by estimating the constraints between s_i and s_k . Calling the set of these pixels $Cons_k$ (Sec. 2.3, 2.4)
- 3: Labeling pixels in $s_k \setminus Cons_k$ by MRF optimization on $J_{i+1} = \{s_i, s_k, s_{2k-i}\}$ with hard constraints on the labeled pixels. (Sec. 2.4)

asymmetry. Concretely, for pixel $p \in Img(x, y, z)$, let $I(p)$ be its feature vector (4D MRI). To define $Asym(p)$, we measure local asymmetry of p :

$$Asym(p) = \min_{q \in N(p)} \|I(p) - Img_r(q)\| \quad (1)$$

where, $N(p)$ is the neighborhood of p . We use Otsu's Method [15] on $Asym(p)$, $p \in I$, resulting in the threshold θ_{Asym} and define the most asymmetric slice: $i_{max} = \arg \min_i \sum_{p \in s_i} \mathbf{1}(Asym(p) \geq \theta_{Asym})$, where $\mathbf{1}(\cdot)$ returns 1 if the argument is true and 0 otherwise.

2.2 Annotation

For manual annotation of the selected slide, we have developed an interface that allows the human to make labels (*Tumor*, *Edema* and *Background*) in all four channels. The interface supports drawing rectangles and curves with the mouse. When a pixel is labeled in one channel, it will be labeled in all the other three channels in the same time in the interface. This approach has been a minimal burden on the user as the asymmetric slice typically finds large tumor and edema regions and at the same time, our method is robust to minor errors in initial labeling (see Sec. 2.4 for details). A novice annotator with no background on medicine spends about 2–4 minutes (used in this paper).

2.3 Structural Trajectories

To exploit the 3D structural consistency of the MRI volume, we develop a method to constrain the ultimate pixel label problem based on tracking voxels through the volume, which we call structural trajectories. Ultimately, these will form constraints for our MRF formulation (i.e., linking a tumor pixel in one slice to the corresponding tumor pixel in the next slice, if it exists, and analogously for the other types of tissue). We use an optical flow algorithm [16] to track points between neighboring slices by successive registrations, resulting in spatially dense trajectories, capturing the global correspondence of pixels over the whole volume.

First, we register neighboring slices by minimizing intensity and gradient pixel matching scores [17]. Given a sequence of slices $S = \{s_i, i = 1 \dots n\}$, define a structural trajectory to be a sequence of pixels: $tr_j = \{p_t^j | p_t^j \in s_t, t = t_0 \dots t_1, 1 \leq t_0 \leq$

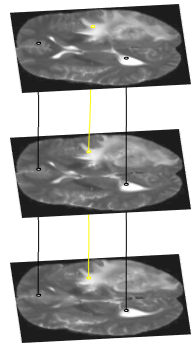


Fig. 1. Example trajectories

Algorithm 2. Over-Segmentation Based Local Consistency**Input:** $J_k = s_i, s_k, s_{2k-i}$

-
- 1: **for** $label1 \in \{tumor, edema, background\}$ **do**
 - 2: Defining s_i^{label1} , such that $I(s_i^{label1}(x, y)) = I(s_i(x, y)) * \delta(f_{(x,y)} == label1)$
 - 3: Defining $J_k^{label1} = \{s_i^{label1}, s_k, s_{2k-i}\}$
 - 4: Making over-segmentation on J_k^{label1} , resulting in a set of segments, $SEG^{label1} = \{Seg_j^{label1}, j = 1 \dots T\}$ using the method of [18] with the implementation of [19]
 - 5: Defining $L_k^{label1} = \{p \in s_k | \exists [Seg1 \in SEG^{label1}, q \in s_i], s.t. f_q == label1, p \in Seg1, q \in Seg1\}$
 - 6: **end for**
 - 7: $Cons1_k^{tumor} \leftarrow L_k^{tumor} \setminus (L_k^{edema} \cup L_k^{background})$
 - 8: $Cons1_k^{edema} \leftarrow L_k^{edema} \setminus (L_k^{tumor} \cup L_k^{background})$
 - 9: $Cons1_k^{background} \leftarrow L_k^{background} \setminus (L_k^{edema} \cup L_k^{tumor})$
-

$t_1 \leq n\}$. Trajectories hence capture slice-to-slice pairwise pixel correspondence (Fig. 1), even if the pair of pixels are not neighboring. In a brain MRI 3D image, a point trajectory refers to a sequence of physical points in brain which intensities are successively similar. We use the point tracking with forward-backward checking method to compute the trajectories. To the best of our knowledge, this is the first time an optical flow-like computation has been used to enforce structural consistency in 3D MRI brain tumor segmentation.

2.4 Constrained MRF Pixel Labeling

Given the label of pixels of s_i , we define a traditional Markov random field on pixels of $J_k = \{s_i, s_k, s_{2k-i}\}$. ($k = i \pm 1$). In this pixel labeling problem we have a 3D image J_k with a 6-connected neighborhood system N and a set of labels $L = \{tumor, edema, background\}$. A labeling f assigns a label $f_p \in L$. The feature vector of p is defined as $I(p)$. The goal is to find a labeling minimizing an energy function of the form,

$$E(f) = \sum_{p \in J_k} D_p(f_p) + \sum_{(p,q) \in N} V_{pq}(f_p, f_q) \quad (2)$$

where $D_p(f_p) = -\log(P(I(p)|f_p))$ is the data term and smoothness is

$$V_{pq}(f_p, f_q) = Smoothness(f_p, f_q) \times \exp(-\alpha * Dis(I(p), I(q), \Sigma)) \quad (3)$$

where $Smoothness$ is a 3×3 non-negative matrix, α is a positive scalar, $Dis(\cdot)$ is Mahalanobis distance, $Dis(x, y, \Sigma) = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$. Σ is computed with the training data.

Estimation of Intensity Distribution: Using the labeled pixels in s_i as training data, we model the node-class likelihoods $P(I(p)|f_p)$ as a Gaussian mixture (GMM). Let Σ be the covariance matrix of $\{I(p)|p \in s_i, f_p == background\}$.

Estimation of Constraints: With the constraints between s_i and s_k , we use two different methods to select two sets of pixels in s_j to label; here a constraint

Algorithm 3. Local Modification of Structural Trajectories

```

1: for  $q \in s_i$  do
2:    $NEAR(q) = \arg \min_{p \in s_k} (\|I(p) - I(q)\| | pre(p) \in N(q))$ 
3: end for
4: for  $p_0 \in s_k$  do
5:    $voter\_1(p_0) = \{q \in s_i | p_0 = NEAR(q)\}$ 
6:    $cr(p_0) = \arg \min_{q \in voter\_1(p_0)} (\|I(p) - I(q)\|)$ 
7:    $discr(p_0) = \|I(p_0) - I(cr(p_0))\|$ 
8:    $voter(p_0) = \{q \in voter\_1(p_0)\}$ 
9: end for
10: for  $label1 \in \{tumor, edema, background\}$  do
11:    $Cons2_k^{label1} = \{p \in s_k | \forall [q \in voter(p)], f_p = label1\}$ 
12: end for

```

means that we will fix the label of the pixel during inference use it to propagate the labels to neighboring unconstrained pixels (see Sec. 2.5 for details).

Erosion-like Processing with Over-Segmentation: The basic idea is that if a pixel is grouped (in the over-segmentation) with other pixels of different labels on s_i , then its labeling has a high uncertainty and it should hence not be selected as a hard-constraint, which makes our method robust to minor errors in the labeling. The estimation process is performed according to Alg. 2.

Local Modification of Structural Trajectories: Structural-trajectories cross slices and there is strong correspondence between the pixels in the same trajectory. We make a local modification of these trajectories to select some pixels in s_k to label. For a pixel $p_0 \in s_k$, let $pre(p_0)$ be the pixel in s_i and in the same trajectory with p_0 . Obviously, p_0 is likely to have the same label with $pre(p_0)$. Let $voter(p_0)$ be the set of pixels in s_i that have the same label with p_0 . Define $Cons2_k^{label1}$ as the set of pixels in s_k that we select to label as $label1$. Alg. 3 shows how to compute $Cons2_k^{label1}$ and Fig. 2 gives a visualization. We then intersect the two constraint sets, $Cons_s_k^{label1} = Cons1_k^{label1} \cap Cons2_k^{label1}$ where $label1 \in \{tumor, edema, background\}$ and $Cons_k = \cup Cons_k^{label1}$, as a conservative step since these will be used as hard constraints in the MRF optimization (all pixels in $Cons_s_k^{label1}$ are given label $label1$), which we discuss next.

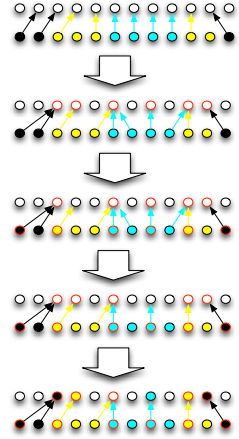


Fig. 2. Local modification of structural trajectories

2.5 Optimizing MRF with Hard Constraints

In $J_k = \{s_i, s_k, s_{2k-i}\}$, we have pixels in $s_i \cup Cons_k$ labeled. Hence, the Markov random field on J_k is constrained by the labeled pixels. We implement these hard constraints through with the following MRF conversion. We define $H_k =$

$s_i \cup Cons_k$, $J_k^H = J_k \setminus H_k$. We can define a new MRF on J_k^H with the same neighborhood system N , which has no hard constraints. To make the two MRFs equivalent, we define a new energy function based on the original one in Eq. (2):

$$E^H(f^H) = \sum_{p \in J_k^H} D_p^H(f_p^H) + \sum_{(p,q) \in N} V_{pq}^H(f_p^H, f_q^H). \quad (4)$$

$$D_p^H(f_p^H) = D_p(f_p^H) + \sum_{(p,q) \in N, q \in H} V_{pq}(f_p^H, f_q^H) \quad (5)$$

$$V_{pq}^H(f_p^H, f_q^H) = V_{pq}(f_p^H, f_q^H) \quad (6)$$

It is easy to prove, if $\forall p \in J_k^H$, $f_p = f_p^H$, then $E^H(f^H) = E(f)$. Hence, we can optimize $E(f)$ on J_k by optimizing $E^H(f^H)$ on J_k^H . Our MRF is a conventional multi-class MRF, we hence optimize with the alpha-expansion graph cut [20].

3 Experiments and Results

Data Set: We evaluate our method on 2012 BRATS Training Data [3], which is now the standard benchmark for brain tumor segmentation research. This dataset contains 30 cases from patients with high- and low-grade gliomas. The data has four channels, T1, T2, T1C and FLAIR; the feature $I(p)$ in our experiment is hence a 4D vector. The typical volume size is about $130 \times 170 \times 170$. We also evaluate our method on 2012 BRATS Challenge Data [3], which contains 11 high- and 4 low-grade gliomas from patients, and compare the result with a state-of-the-art semi-automatic method, Tumor-cut [9]. The gold standard labels are not given and DICE scores of Tumor and Complete Tumor (Tumor+Edema) are computed by an online evaluation infrastructure (hence we avoid a bias by the human annotator, who has never seen the gold standard). The DICE Scores of [9] are given by 2012 BRATS [3].

Parameters: The MRF parameter α is set to 0.001 and *Smoothness* is given by Table 1 based on empirical experiments.

Baselines and Metrics: We use the Jaccard and DICE metrics in our evaluation. j^* refers to our method. We define a set of appropriate baselines: to evaluate the hard constraints from the structural tra-

jectories and the over-segmentation, denote the baseline without any hard constraints as j^B . We also assess the value of the structural trajectories, which are global through the whole volume, with respect to the local over-segmentation constraints; j^L uses only the local over-segmentation. We follow the exact evaluation regime specified in [3].

Results: We show the quantitative results in Table 2. As a point of comparison to the current state of the art fully automatic method, we also compare our results to supervised automatic method [21] on the same dataset (denoting it j^A), which won the first prize in the BRATS Challenge 2012 [3]. Although this

Table 1. Smoothness term

<i>Smoothness</i>	bg	tumor	edema
bg	0	2	1
tumor	2	0	1
edema	1	1	0

Table 2. Comparative Jaccard scores on BRATS Training Set and DICE scores on BRATS Challenge Set: our method (j^*), no hard constraints (j^B), no structural trajectories (j^L), the best fully automatic method [21] from [3] (j^A), and the state-of-the-art semi-automatic Tumor-cut(j^T) [9]

	BRATS Training Set								BRATS Challenge Set		
	Tumor				Edema				Tumor		Complete tumor
	j^B	j^L	j^*	j^A	j^B	j^L	j^*	j^A	j^*	j^T	j^*
High-grade	0.121	0.653	0.750	0.500	0.113	0.451	0.629	0.450	0.683±0.242	0.694	0.835±0.089
Low-grade	0.072	0.564	0.657	0.360	0.050	0.144	0.215	0.230	0.563±0.345	0.324	0.848±0.087

comparison will clearly favor our semi-automatic method, we show it to get a sense of what value our approach yields. Note that not all of the semi-automatic methods in Table 2 outscore the automatic counterparts. The results clearly demonstrate that the proposed semi-automatic method, which requires only one slice to be labeled by a human, outperforms the best fully automatic method [21] by a significant margin on all cases except the low-grade edema class, which is known to be challenging. The results also indicate the value of both the local over-segmentation constraints and the global structural trajectories. The average runtime of one case is about 5 minutes in Matlab. Fig. 3 shows an example.

Failure Modes: Mode 1: In some cases, especially low-grade cases, tumor or edema is not successive slice-by-slice. In these cases, because we cannot estimate the feature distribution accurately and no pixel correspondence is available, our results may just contain some connected components of tumor or edema. Mode 2: Let A be a small homogeneous region contained in slice s_k . If its feature distribution is very different from any of the feature distribution of tumor, edema or background estimated by the previous labeled slice s_i , and it is strongly adjacent to both of two different tissues (e.g. tumor and edema), then the pixel labeling of A by MRF optimization will not be accurate.

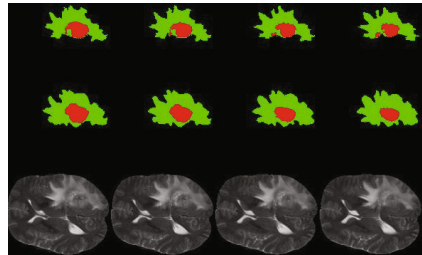


Fig. 3. Example results. Top: results of our method (Tumor: Red; Edema: Green). Middle: gold standard. Bottom: T1C-MRI.

4 Conclusion

In this work, we have proposed a novel semi-automatic brain tumor segmentation method. We just need to ask for a manual labeling on a single slice, which is selected automatically based on asymmetry. All the other slices are labeled iteratively with the local intensity distribution and both global and local constraints in a constrained MRF framework. The constraints are extracted automatically

based on an optical flow-like measure, which is the first time such a structural global measure has been used for brain tumor segmentation, to the best of our knowledge. Our method does not depend on the assumption that training data and testing data have similar intensity distribution or similar shape prior, unlike many existing methods. Our experiments demonstrate that the proposed approach yields significant improvements over fully automatic methods (e.g., 50%) in most cases, as expected, and the novel structural constraints greatly improve the MRF optimization over the conventional MRF labeling.

References

1. Patel, M.R., Tse, V.: Diagnosis and staging of brain tumors. *Seminars in Roentgenology* 39(3), 347–360 (2004)
2. Liu, J., Udupa, J.K., Odhner, D., Hackney, D., Moonis, G.: A system for brain tumor volume estimation via mr imaging and fuzzy connectedness. *CMIG* 29(1), 21–34 (2005)
3. Menze, B., Jakab, A., Bauer, S., Reyes, M., Leemput, K.V.: Proceedings of the miccai 2012 challenge on multimodal brain tumor segmentation (2012)
4. Corso, J.J., Sharon, E., Dube, S., El-Saden, S., Sinha, U., Yuille, A.: Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Trans. on Medical Imaging* 27(5), 629–640 (2008)
5. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. *MEDIA* 8, 275–283 (2004)
6. Lee, C.-H., Schmidt, M., Murtha, A., Bistriz, A., Sander, J., Greiner, R.: Segmenting brain tumors with conditional random fields and support vector machines. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) *CVBIA 2005*. LNCS, vol. 3765, pp. 469–478. Springer, Heidelberg (2005)
7. Geremia, E., Zikic, D., Clatz, O., et al.: Classification forests for semantic segmentation of brain lesions in multi-channel mri. *Decision Forests for Computer Vision and Medical Image Analysis* (2013)
8. Birkbeck, N., Cobzas, D., Jägersand, M., Murtha, A., Kesztyues, T.: An interactive graph cut method for brain tumor segmentation. In: *WACV* (2009)
9. Hamamci, A., Kucuk, N., Karaman, K., Engin, K., Ünal, G.B.: Tumor-cut: Segmentation of brain tumors on contrast enhanced mr images for radiosurgery applications. *IEEE Trans. Med. Imaging* 31(3), 790–804 (2012)
10. Rother, C., Kolmogorov, V., Blake, A.: Grabcut -interactive foreground extraction using iterated graph cuts. In: *SIGGRAPH* (2004)
11. Li, Y., Sun, J., Tang, C.K., Shum, H.Y.: Lazy snapping. In: *SIGGRAPH* (2004)
12. Chen, X., Yao, J., Zhuge, Y., Bagci, U.: 2d automatic anatomy segmentation based on graph cut-oriented active appearance models. In: *ICIP* (2010)
13. Afifi, A., Nakaguchi, T.: Liver segmentation approach using graph cuts and iteratively estimated shape and intensity constrains. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II*. LNCS, vol. 7511, pp. 395–403. Springer, Heidelberg (2012)
14. Ourselin, S., Stefanescu, R., Pennec, X.: Robust registration of multi-modal images: Towards real-time clinical applications. In: Dohi, T., Kikinis, R. (eds.) *MICCAI 2002, Part II*. LNCS, vol. 2489, pp. 140–147. Springer, Heidelberg (2002)
15. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics* (1979)

16. Sundaram, N., Brox, T., Keutzer, K.: Dense point trajectories by GPU-accelerated large displacement optical flow. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part I. LNCS, vol. 6311, pp. 438–451. Springer, Heidelberg (2010)
17. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004, Part IV. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
18. Grundmann, M., Kwatra, V., Han, M., Essa, I.A.: Efficient hierarchical graph-based video segmentation. In: CVPR (2010)
19. Xu, C., Corso, J.J.: Evaluation of super-voxel methods for early video processing. In: CVPR (2012)
20. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(11), 1222–1239 (2001)
21. Bauer, S., Fejes, T., Slotboom, J., Wiest, R., Nolte, L.P., Reyes1, M.: Segmentation of brain tumor images based on integrated hierarchical classification and regularization. In: MICCAI BRATS Challenge (2012)

A Probabilistic, Non-parametric Framework for Inter-modality Label Fusion

Juan Eugenio Iglesias¹, Mert Rory Sabuncu¹, and Koen Van Leemput^{1,2,3}

¹ Martinos Center for Biomedical Imaging, MGH, Harvard Medical School, USA

² Department of Applied Mathematics and Computer Science, DTU, Denmark

³ Departments of Information and Computer Science and of Biomedical Engineering and Computational Science, Aalto University, Finland

Abstract. Multi-atlas techniques are commonplace in medical image segmentation due to their high performance and ease of implementation. Locally weighting the contributions from the different atlases in the label fusion process can improve the quality of the segmentation. However, how to define these weights in a principled way in inter-modality scenarios remains an open problem. Here we propose a label fusion scheme that does not require voxel intensity consistency between the atlases and the target image to segment. The method is based on a generative model of image data in which each intensity in the atlases has an associated conditional distribution of corresponding intensities in the target. The segmentation is computed using variational expectation maximization (VEM) in a Bayesian framework. The method was evaluated with a dataset of eight proton density weighted brain MRI scans with nine labeled structures of interest. The results show that the algorithm outperforms majority voting and a recently published inter-modality label fusion algorithm.

1 Introduction

Automated segmentation of brain MRI scans is a key step in most neuroimaging pipelines. Manual delineation of structures of interest is time consuming and rater dependent, making automated approaches desirable. Some of the most popular segmentation methods in the recent literature are based on the multi-atlas paradigm, in which a set of training images with manual annotations (henceforth “atlases”) are deformed to the image to analyze. The deformations are used to propagate the annotations to *target* space, where they are finally merged into an estimate of the segmentation; this step is known as label fusion. Multi-atlas techniques overcome the main limitation of using a single atlas: the fact that a single template can seldom cover all the anatomical variability within a population.

The simplest form of label fusion is majority voting, in which the most frequent label is assigned to each voxel [1]. Better results can be achieved by locally weighting the contribution of the atlases by their similarity to the target scan after registration. In [2], Isgum et al. define weights by inverting the absolute difference in image intensities. A more principled framework based on a generative model was proposed by Sabuncu et al. [3]: a smooth, discrete, latent membership

field determines from which deformed atlas the intensity and label are taken at each spatial location; Bayesian inference is used to compute the segmentation.

These methods rely on the consistency of voxel intensities across the atlases and the target scan. This assumption falters in inter-modality scenarios, including MRI when the atlases and the target have been acquired with different hardware or pulse sequences. This is often the case when analyzing clinical or legacy data. Another example of application that could benefit from inter-modality label fusion is the analysis of infant brain MRI, in which the intensities are very different than in scans from adults due to ongoing myelination.

While the image registration literature has managed inter-modality scenarios using metrics based on mutual information [4], label fusion across modalities remains an open problem. One possible approach is to arbitrarily define weights based on the mutual information or cross-correlation computed in a window around each voxel [5], yet the optimality of such an approach is unclear.

A principled way of carrying out label fusion across modalities was presented in [6]. This method is based on a generative model in which the intensity of the voxels corresponding to each label follows a Gaussian distribution. The parameters of the Gaussian are estimated from the data, making the fusion robust against changes in modality or MRI contrast. While this approach outperforms heuristic schemes based on local cross-correlations [7], it still has two disadvantages. First, the performance is poor when the Gaussian assumption is violated, such as in the thalamus or the putamen in brain MRI scans. And second, since the fusion only considers the deformed labels from the atlases, it ignores potentially valuable information from their intensities.

Here we propose a generalization of Sabuncu et al.’s model to inter-modality scenarios. The generative model is essentially the same; however, we do not hypothesize a Gaussian relation between the intensities of the atlases and the target scan. Instead, we assume a more flexible model based on a semi-parametric or non-parametric conditional distribution of the intensities of the target given the intensities of the atlases. Using a Bayesian framework, we first estimate this conditional distribution and also a multiplicative bias field. Then, the segmentation is computed as the most likely labels given these estimates and the input image.

2 Methods

2.1 Generative Model

The proposed generative model is shown in Fig. 1a, and the corresponding equations in Fig. 1b. We assume that registration is a preprocessing step: the intensities $\{I_n\}$ and corresponding labels $\{L_n\}$ of the N_{atl} deformed atlases are constant during the fusion. $M(\mathbf{x}) \in \{1, \dots, N_{atl}\}$ is a discrete, latent membership field that indexes which atlas generated the label and intensity of the voxel at spatial location \mathbf{x} . $M(\mathbf{x})$ is smooth thanks to a Markov random field (MRF) prior (Eq. 1 in Fig. 1b, where \mathcal{V}_x is the 6-neighborhood of \mathbf{x} and $\delta(\cdot)$ is Kronecker’s delta).

Given $M(\mathbf{x})$, the label of a voxel $L(\mathbf{x})$ is sampled from a categorical distribution given by a logOdds model [8] defined by the warped labels of atlas $M(\mathbf{x})$

(Eq. 2 in Fig. 1b, where ρ is the slope of the model and D_n^l is the signed distance transform for atlas n and label $l \in \{1, \dots, \mathcal{L}\}$). This prior has been shown to outperform taking the label $L_{M(\mathbf{x})}(\mathbf{x})$ directly [3]. Given $L(\mathbf{x})$, the uncorrupted (i.e., bias field corrected) intensity of the voxel $\tilde{I}(\mathbf{x})$ is sampled from the conditional distribution $p(\tilde{I}(\mathbf{x})|I_{M(\mathbf{x})}(\mathbf{x}), \Theta)$, which depends on the model parameters Θ (Eq. 3 in Fig. 1b). Here we have made two assumptions. First, that the conditional PDF is stationary in space. Second, we assume that the image intensities are consistent across atlases, which is fair when the atlases are from the same modality. This allows us to represent the conditional probability density function (PDF) with a single, atlas-independent PDF.

For the conditional PDF of intensities, we consider two different models. First, a semi-parametric (SP) model which describes $\tilde{I}(\mathbf{x})$ with a Gaussian PDF with mean and variance depending on $I_{M(\mathbf{x})}(\mathbf{x})$. And second, a non-parametric (NP) model based on a collection of conditional 1D histograms (h) of $\tilde{I}(\mathbf{x})$ depending on $I_{M(\mathbf{x})}(\mathbf{x})$. These two forms of the algorithm allow us to isolate and understand the effects of the two contributions of this model with respect to [6]: the departure from the Gaussian model (NP) and using the intensities of the atlases in the fusion (NP and SP). The models are given by:

$$\text{Semi-parametric (SP): } p(\tilde{I}(\mathbf{x})|I_{M(\mathbf{x})}(\mathbf{x}), \Theta) = \mathcal{N}(\tilde{I}(\mathbf{x}); \mu_{Q[I_{M(\mathbf{x})}]}, \sigma_{Q[I_{M(\mathbf{x})}]}^2). \quad (1)$$

$$\text{Non-parametric (NP): } p(\tilde{I}(\mathbf{x})|I_{M(\mathbf{x})}(\mathbf{x}), \Theta) = h(Q[\tilde{I}(\mathbf{x})]; Q[I_{M(\mathbf{x})}(\mathbf{x})]), \quad (2)$$

where $Q[\cdot]$ is a nearest neighbor interpolator that quantizes the intensities of the target scan and the atlases into discrete sets A and B , respectively.

Finally, $\tilde{I}(\mathbf{x})$ is corrupted by a low-frequency, multiplicative bias field to yield the observed intensities $I(\mathbf{x})$ (Eq. 4 in Fig. 1). The bias field is modeled as the exponential (to ensure non-negativity) of a linear combination of smooth basis functions $\{\psi_k\}$. The linear coefficients $\mathbf{b} = \{b_k\}$ are grouped with the parameters of the conditional PDF of intensities into the model parameters $\theta = (\mathbf{b}, \{h(a; b)\})$ (NP) or $\theta = (\mathbf{b}, \{\mu_b, \sigma_b^2\})$ (SP). A flat prior distribution $p(\theta) \propto 1$ completes the model. Note that the denominator in Eq. 5 in Fig. 1b ensures integration to one.

2.2 Inference

The segmentation \hat{L} of image I is estimated by maximizing the posterior probability: $p(L|I, \{I_n\}, \{L_n\})$. This leads to an intractable integral for which we can use the approximation that the posterior distribution of the model parameters is a Dirac's delta, i.e., $p(\theta|I, \{I_n\}) \approx \delta(\theta = \hat{\theta})$:

$$\hat{L} = \underset{L}{\operatorname{argmax}} \int p(L|\theta, I, \{I_n\}, \{L_n\}) p(\theta|I, \{I_n\}) d\theta \approx \underset{L}{\operatorname{argmax}} p(L|\hat{\theta}, I, \{I_n\}, \{L_n\}), \quad (3)$$

Thus, we first need to estimate $\hat{\theta}$ and then use it to compute \hat{L} with Eq. 3.

Estimating $\hat{\theta}$: the problem here is

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} p(\theta|I, \{I_n\}) = \underset{\theta}{\operatorname{argmax}} \log p(I|\{I_n\}, \theta). \quad (4)$$

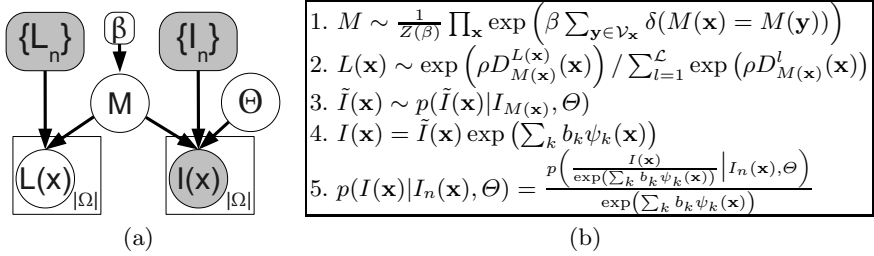


Fig. 1. Graphical model (a) and corresponding equations (b). Random variables are in circles, constants in boxes, observed variables shaded and plates indicate replication.

Note that θ does not depend on $\{L_n\}$. Optimizing Eq. 4 requires marginalizing over M , which is intractable due to the MRF. Therefore, we use VEM to compute an approximate solution by optimizing a lower bound J instead:

$$\log p(I|\{I_n\}, \theta) \geq J(q(M), \theta) = \log p(I|\{I_n\}, \theta) - KL[q(M)||p(M|I, \theta, \{I_n\})] \quad (5)$$

$$= H[q] + \sum_M q(M) \log p(M, I|\theta, \{I_n\}), \quad (6)$$

where $H[\cdot]$ is the entropy of a random variable, KL represents the (non-negative) Kullback-Leibler divergence and q is a distribution over M which is restricted to having a simpler form than $p(M|I, \theta, \{I_n\})$. We alternately optimize J with respect to q (E step) and Θ (M step).

In the **E step**, we work with Eq. 5. Maximizing J amounts to minimizing the KL divergence. The standard “mean field” approximation assumes that q factorizes as $q(M) = \prod_{\mathbf{x}} q_{\mathbf{x}}(M(\mathbf{x}))$, where $q_{\mathbf{x}}(m)$ is a categorical distribution over the atlas indices $m = 1, \dots, N_{atl}$, and $\sum_m q_{\mathbf{x}}(m) = 1$. This yields:

$$\operatorname{argmin}_q \sum_{\mathbf{x}} \sum_m q_{\mathbf{x}}(m) \log \frac{q_{\mathbf{x}}(m)}{p(I(\mathbf{x})|I_m(\mathbf{x}), \theta)} - \beta \sum_{\mathbf{x}} E_{q_{\mathbf{x}}} \left[\sum_{y \in \mathcal{V}_{\mathbf{x}}} q_y(M(\mathbf{x})) \right]. \quad (7)$$

Building the Lagrangian and setting derivatives to zero gives:

$$q_{\mathbf{x}}(m) \propto p(I(\mathbf{x})|I_m(\mathbf{x}), \theta) \exp[\beta \sum_{y \in \mathcal{V}_{\mathbf{x}}} q_y(m)], \quad (8)$$

such that $\sum_m q_{\mathbf{x}}(m) = 1$. We can solve this equation with fixed point iterations.

In the **M step**, it is more convenient to work with Eq. 6: since we are optimizing for θ , we can disregard $H(q)$. Because of the structure of q , we have:

$$\operatorname{argmax}_{\Theta} \sum_{\mathbf{x}} \sum_m q_{\mathbf{x}}(m) \left[\log p(\tilde{I}(\mathbf{x})|I_m(\mathbf{x}), \Theta) - \sum_k b_k \psi_k(\mathbf{x}) \right], \quad (9)$$

with $\tilde{I}(\mathbf{x}) = I(\mathbf{x})e^{-\sum_k b_k \psi_k(\mathbf{x})}$. The solution depends on whether we consider the SP or the NP model. In the first case, replacing $p(\tilde{I}(\mathbf{x})|I_m(\mathbf{x}), \Theta)$ by the

corresponding Gaussian (Eq. 1), taking derivatives with respect to the means and variances, and setting them to zero gives:

$$\mu_b = \sum_{\mathbf{x}} \sum_m w_x^b(m) I(\mathbf{x}) e^{-\sum_k b_k \psi_k(\mathbf{x})} / \sum_{\mathbf{x}} \sum_m w_x^b(m), \quad (10)$$

$$\sigma_b^2 = \sum_{\mathbf{x}} \sum_m w_x^b(m) (I(\mathbf{x}) e^{-\sum_k b_k \psi_k(\mathbf{x})} - \mu_b)^2 / \sum_{\mathbf{x}} \sum_m w_x^b(m), \quad (11)$$

where $w_x^b(m) = q_x(m) \delta(Q[I_m(\mathbf{x})] = b)$. Eqns. 10 and 11 are weighted means and variances depending on the (approximate) membership posteriors $q_x(m)$.

For the NP model, we substitute Eq. 2 into Eq. 9 and build a Lagrangian to ensure integration to one. Taking derivatives and setting them to zero yields:

$$h(a; b) \propto \sum_{\mathbf{x}} \sum_m w_x^b(m) \delta(Q[I(\mathbf{x}) e^{-\sum_k b_k \psi_k(\mathbf{x})}] = a). \quad (12)$$

such that $\sum_{a \in A} h(a; b) = 1/\Delta$, where Δ is the quantization interval. Again, Eq. 12 is simply a weighted histogram.

Finally, we use a quasi Newton solver with an explicit line search to optimize Eq. 9 for the bias field coefficients, both in the SP and BP case.

Computing the Final Segmentation: given θ , estimating the final segmentation with Eq. 3 still requires an intractable sum over M . However, since $q(M)$ minimizes the KL divergence with $p(M|\hat{\theta}, I, \{I_n\})$, we approximate:

$$\begin{aligned} \hat{L} &= \operatorname{argmax}_L \sum_M p(L|M, \{L_n\}) p(M|\hat{\theta}, I, \{I_n\}) \approx \operatorname{argmax}_L \sum_M p(L|M, \{L_n\}) q(M) \\ &= \operatorname{argmax}_L \prod_{\mathbf{x}} \sum_m q_x(m) p(L(\mathbf{x})|L_m) \Rightarrow \hat{L}(\mathbf{x}) = \operatorname{argmax}_{L(\mathbf{x})} \sum_m q_x(m) p(L(\mathbf{x})|L_m) \end{aligned} \quad (13)$$

Summary of the Algorithm: we initialize the bias field coefficients $b_k = 0$, and the distribution $q_x(m) = 1/N_{atl}$. Next, we alternate the E and M steps until convergence. The E step updates q with fixed point iterations of Eq. 8. The M step first updates the bias field by numerically optimizing Eq. 9 with respect to $\{b_k\}$, and then the parameters of the conditional PDF with Eqns. 10, 11 (SP model) or Eq. 12 (NP). Upon convergence, the final segmentation is computed with Eq. 13. The method is illustrated with a simple example in Fig. 2.

3 Experiments and Results

We used 39 manually delineated (see protocol in [9]) T1 MRI scans as atlases to segment 36 brain structures in eight proton-density (PD) scans. The annotations of the PD scans were made on co-registered T1 data, allowing consistent annotations across the two datasets. FreeSurfer [10] was used to skull-strip the volumes and intensity-normalize the atlases, since consistent intensities are assumed.

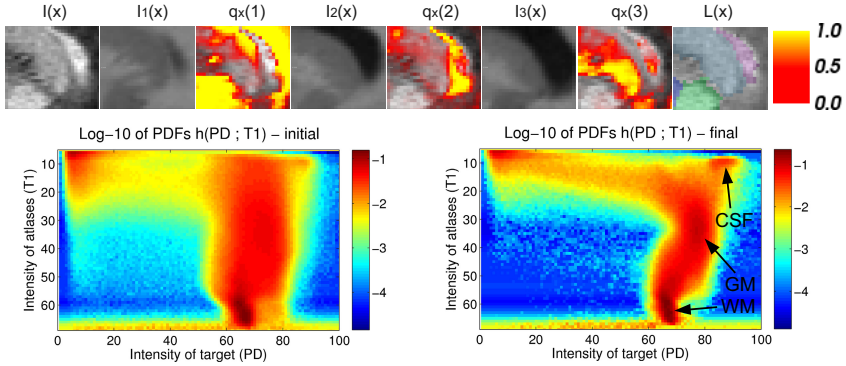


Fig. 2. Intermediate outputs for a sagittal patch of a PD scan segmented with three T1 atlases and the NP model. Top row: I , $\{I_n\}$, $q_x(m)$ overlaid on I , and segmentation L (green = putamen, purple = ventricle, blue = caudate); $q_x(m)$ highlights where each atlas contributed to generating I, L . Bottom row: initial and final estimates of the conditional PDFs $h(a; b)$; the latter are much sharper and emphasize three modes corresponding to gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF).

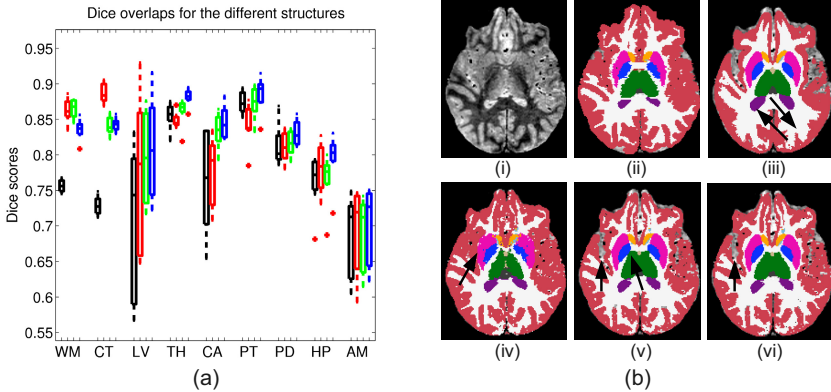


Fig. 3. (a) Box plot of Dice scores for majority voting (black), the Gaussian model [6] (red) and the proposed SP (green) and NP (blue) methods. Lines are at the three quartile values, whiskers extend to 1.5 times the interquartile range from the box, and dots mark outliers. (b-i) Axial slice of a PD scan, (ii) manual labels, (iii) segmentation from majority voting, (iv) the Gaussian model, (v) the SP model, (vi) the NP model. Arrows point at mistakes. The color code is: red = CT, white = WM, pink = PT, dark blue = PD, light blue = CA, green = TH, purple = LV, orange = accumbens.

We used Elastix [11] to register the T1 to the PD scans (b-spline transform, mutual information). We compared four label fusion methods: majority voting, the Gaussian model from [6] and the proposed approach (SP and NP). We set $\rho = 1$, $\beta = 0.75$ (as in [6]), $\{\psi_k\}$ to a fourth order polynomial and the number of bins $|A| = |B| = 64$. The Dice overlap between the manual and automatically generated labels was used as measure of performance. Statistical significance

Table 1. Mean Dice scores (in %, highest in bold) and p-values with respect to NP

Method	WM	CT	LV	TH	CA	PT	PD	HP	AM	All combined
Maj. Vot.	75.6	72.8	70.7	85.5	76.2	87.4	81.2	76.3	68.6	77.2
p-value	0.008	0.008	0.008	0.008	0.016	0.078	0.023	0.008	0.008	$6.8 \cdot 10^{-13}$
Gaussian	86.3	88.7	77.3	84.9	77.7	84.2	81.1	78.2	69.3	80.8
p-value	0.008	0.008	0.11	0.008	0.008	0.008	0.016	0.078	0.008	$8.2 \cdot 10^{-4}$
SP	86.5	84.1	79.6	86.2	83.4	87.4	81.7	76.6	68.9	81.6
p-value	0.008	0.46	0.016	0.008	0.055	0.016	0.023	0.008	0.008	$2.8 \cdot 10^{-5}$
NP	83.5	84.3	80.9	88.1	84.6	88.6	83.1	79.6	70.1	82.5

was assessed with paired Wilcoxon signed rank tests. For simpler presentation of results, we merged the label of each left structure with its right counterpart, and used a representative subset of structures in the evaluation (as in [3]): white matter (WM), cortex (CT), lateral ventricle (LV), thalamus (TH), caudate (CA), putamen (PT), pallidum (PA), hippocampus (HP) and amygdala (AM).

Box plots of the structure-wise Dice scores are shown in Fig. 3a, whereas segmentations for a sample axial slice are shown in Fig. 3b. The p-values for the statistical tests comparing the NP method (the top-performing algorithm) with the other competing approaches are shown in Tab. 1. Majority voting produces decent outputs for the subcortical structures, but fails to extract the convoluted white matter surface, which is very difficult to register (see Fig. 3b-iii). It also produces bad results for the ventricles, as illustrated in the same figure. The Gaussian model gives excellent results for the cortex, but falters when the normality assumption does not hold. This is often the case for the thalamus and the putamen. For instance, the latter leaks into the white matter in Fig. 3b-iv.

The NP version of the proposed approach significantly outperforms majority voting for every brain structure (Tab. 1). It also yields Dice scores significantly higher than those from the Gaussian model for all subcortical structures. Only in the cortex and the white matter the performance is inferior; see for instance the mistake marked by the arrow in Fig. 3b-vi. This is because the registration is poor for these convoluted structures, making a simple Gaussian intensity model more suitable. Overall, the mean improvement in Dice score is $\sim 2\%$ over the Gaussian model and $\sim 5\%$ over majority voting. The SP version also beats majority voting and the Gaussian model. However, it performs slightly worse than the NP method, likely due to its inability to describe multimodal shapes in the conditional intensity PDF (e.g., see atlas intensity range 10-20 in Fig. 2).

4 Discussion

We presented a cross-modality label fusion method based on a generative model that describes the relationship between image intensities in a SP or NP manner. The algorithm often converges in less than 15 iterations (about 20 minutes on a modern PC). The results show that using the intensities of the atlases in the fusion allows the SP algorithm to outperform previously proposed inter-modality

label fusion techniques. Moreover, departure from the Gaussian model allows the NP model to further improve the results. Exploring more flexible SP models (such as those based on mixtures of Gaussians), incorporating the registration step into the framework and using more accurate approximations than nearest neighbors when estimating the conditional intensity PDF remain as future work.

Acknowledgements. This research was supported by NIH NCCR (P41-RR14075), NIBIB (R01EB006758, R01EB013565, 1K25EB013649-01), NINDS (R01NS052585), NIH 1KL2RR025757-01, Academy of Finland (133611), TEKES (ComBrain), Harvard Catalyst, and financial contributions from Harvard and affiliations.

References

1. 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)
2. Isgum, I., Staring, M., Ruttan, 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 Trans. Med. Im.* 28(7), 1000–1010 (2009)
3. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Im.* 29, 1714–1729 (2010)
4. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Im.* 16(2), 187–198 (1997)
5. Yushkevich, P.A., Wang, H., Pluta, J., Das, S.R., Craige, C., Avants, B.B., Weiner, M.W., Mueller, S.: Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. *NeuroImage* 53(4), 1208–1224 (2010)
6. Iglesias, J., Sabuncu, M., Van Leemput, K.: A generative model for multi-atlas segmentation across modalities. In: 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 888–891 (2012)
7. Iglesias, J.E., Sabuncu, M.R., Van Leemput, K.: A generative model for probabilistic label fusion of multimodal data. In: Yap, P.-T., Liu, T., Shen, D., Westin, C.-F., Shen, L. (eds.) MBIA 2012. LNCS, vol. 7509, pp. 115–133. Springer, Heidelberg (2012)
8. Pohl, K.M., Fisher, J., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Logarithm odds maps for shape representation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 955–963. Springer, Heidelberg (2006)
9. Caviness Jr., V., Filipek, P., Kennedy, D.: Magnetic resonance technology in human brain science: blueprint for a program based upon morphometry. *Brain Dev.* 11(1), 1–13 (1989)
10. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355 (2002)
11. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Im.* 29(1), 196–205 (2010)

Weighted Functional Boxplot with Application to Statistical Atlas Construction

Yi Hong¹, Brad Davis³, J.S. Marron¹, Roland Kwitt³, and Marc Niethammer^{1,2}

¹ University of North Carolina (UNC) at Chapel Hill, NC

² Biomedical Research Imaging Center, UNC-Chapel Hill, NC

³ Kitware, Inc., Carrboro, NC

Abstract. Atlas-building from population data is widely used in medical imaging. However, the emphasis of atlas-building approaches is typically to compute a mean / median shape or image based on population data. In this work, we focus on the statistical characterization of the population data, once spatial alignment has been achieved. We introduce and propose the use of the weighted functional boxplot. This allows the generalization of concepts such as the median, percentiles, or outliers to spaces where the data objects are functions, shapes, or images, and allows spatio-temporal atlas-building based on kernel regression. In our experiments, we demonstrate the utility of the approach to construct statistical atlases for pediatric upper airways and corpora callosa revealing their growth patterns. Furthermore, we show how such atlas information can be used to assess the effect of airway surgery in children.

1 Introduction

Atlas-building from population data has become an important task in medical imaging to provide templates for data analysis. Numerous methods for atlas-building exist, ranging from methods designed for cross-sectional, longitudinal, and random design data. These approaches typically estimate a representative data object (e.g., shape, surface, image) for the population; e.g., a population mean [7] or median [3] with respect to spatial deformations and appearance. This is a restrictive representation, as much of the population data is discarded. In the literature, this has been acknowledged, e.g., by multi-atlas approaches [1] or manifold learning approaches [5] which retain population information by using sets of representative objects or by identifying a low-dimensional data representation.

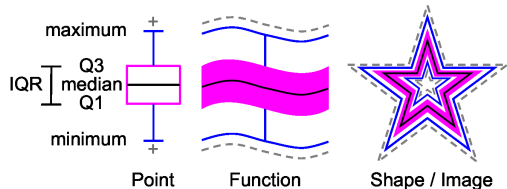


Fig. 1. Illustration of boxplots for points, functions, shapes and images. Median (middle black line), confidence region (magenta) and the maximum non-outlying envelope (two outward blue lines). The gray dash lines are the outliers.

An alternative strategy to retain population information is to represent additional aspects of the full data distribution, such as percentiles, the robust minimum and maximum, variance, confidence regions and outliers as captured by a boxplot for scalar-valued data. The functional boxplot [12] allows just this for functions. Similarly, we can use it to treat shapes and images (see Fig. 1) and therefore as a simple method to augment atlases with additional population information while avoiding restrictive point-wise analyses of data-objects. Note that we focus in this paper on augmenting atlases with statistical information and assume a given spatial alignment of data objects. However, the method could be extended to build order statistics from low-dimensional manifold embeddings where point-wise analysis becomes meaningful as each point then represent a full data object.

As subject data typically has associated individual characteristics (e.g., age, weight, gender) we want to be able to compute the statistical information continuously parameterized by these characteristics. For example, given a subject at a particular age we want to compute subject age-specific confidence regions to assess similarity with respect to the full data population.

We make the following contributions in this paper:

- *We develop a weighted variant of the functional boxplot* in Sec. 2. This allows us for example to use kernel-regression to build spatio-temporal atlases.
- *We show the effectiveness of the method in comparison to point-wise analysis* in Sec. 3 highlighting the importance of object-oriented data analysis.
- *We show applicability of the method to functions, shapes, and images* in Sec. 4 and demonstrate how an atlas can robustly be augmented with statistical data for two applications: capturing changes in pediatric airway development and changes of the corpus callosum over time. We also briefly sketch how our method could be used to build order-statistics on manifolds.
- *We show the use of our method for airway surgery assessment in children* in Sec. 5, where an age-adapted atlas can be used to quantify how “normal” a child suffering from airway obstruction is before and after surgery.

2 Weighted Functional Boxplots and Atlas-Building

The population of data-objects for atlas building could be functions, shapes, and images with associated to subjects characteristics. As an example, we consider subject age and demonstrate spatio-temporal atlas-building as a combination of weighted functional boxplots and kernel smoothing.

2.1 Atlas Building with Kernel Regression

Given spatially aligned data objects we want to capture population changes for example with respect to age. This can be achieved through kernel regression which essentially assigns weights to data-objects with respect to the regressor (say a desired age \bar{a}). We can use for example a Gaussian weighting function

$w_i(a_i; \sigma, \bar{a}) = ce^{(a_i - \bar{a})^2 / 2\sigma^2}$, where a_i is the age for the observation i , σ is the standard deviation for the Gaussian distribution and c a normalization constant to assure that the weights sum up to one. For scalar-valued data the weights can simply be used to define a weighted mean. When deformations are of concern they can be used as weights in an atlas-building procedure for images [2]. Here, we are interested in augmenting an atlas with functional statistical information and hence need to develop a weighted functional boxplot to obtain a regressed median (which is an actual data-object from the population), α central region, maximum non-outlying envelop, and outliers.

2.2 Weighted Functional Boxplots

To define a weighted functional boxplot consistent with the functional boxplot introduced by Sun et. al. [12] requires the definition of a consistent *weighted band depth* for functional data. This imposes an ordering of the weighted observations (data-objects) with respect to the (to be determined) central data-object.

Weighted Band-Depth. The functional boxplot is defined through the concept of band-depth [9, 12]. Since each observation has a different weight, we need to define a weighted band-depth. Such a definition immediately defines the weighted functional boxplot. To motivate our choice, assume we want to compute a standard weighted median of scalar values, which is given by $\mu^* = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^n w_i |x_i - \mu|$, where μ is the sought-for median, $\{x_i\}$ are the measurements, and $w_i > 0$ are weights for the individual measurements. Assume that all weights are natural numbers, i.e., $w_i \in \mathbb{N}^+$. This can be achieved exactly for arbitrary rational w_i and approximately in general by multiplying the energy with a suitable constant and does not change the minimizer. Hence, we replace the weighted problem with the equivalent unweighted minimization problem $\mu^* = \underset{\mu}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{m_i} |x_i - \mu|$, where the individual measurements are simply repeated based on their multiplicities, $m_i = w_i$. Similarly, repeating observations (according to weight), the sampled band-depth can be written as

$$BD_{\bar{n}}^{(j)}(y) = \frac{1}{C} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq \bar{n}} I\{G(y) \subseteq B(\bar{y}_{i_1}, \dots, \bar{y}_{i_j})\}, \tag{1}$$

$$\text{s.t. } \{\bar{y}_{i_1}, \dots, \bar{y}_{i_j}\} \text{ contains unique observations.} \tag{2}$$

where C is a normalization constant (i.e., contains the number of admissible permutations), I denotes the indicator function, $G(y)$ is the graph of the function $y(\mathbf{x})$, and B is the band delimited by the observations given as its arguments. We made use of the fact that, according to our definition, we only want to consider unique observations for the depth measure; the $\{\bar{y}_i\}$ contain the original observations $\{y_i\}$, but according to their respective multiplicity given by the weights. Rewriting the sampled band-depth as

$$WBD_{\bar{n}}^{(j)}(y) = \frac{\sum_{1 \leq i_1 < i_2 < \dots < i_j \leq \bar{n}} w_{i_1} w_{i_2} \dots w_{i_j} I\{G(y) \subseteq B(y_{i_1}, \dots, y_{i_j})\}}{\sum_{1 \leq i_1 < i_2 < \dots < i_j \leq \bar{n}} w_{i_1} w_{i_2} \dots w_{i_j}} \tag{3}$$

defines the weighted band-depth and generalizes to non-natural-numbered weights $w_i \in \mathbb{R}^+$. In fact, this is a “natural” way to define a weighted band-depth and, in further consequence, a weighted functional boxplot. Computing the weighted band-depth in this way is intuitive, as only bands with large weights for all its individual observations have a large impact. Furthermore, this weighted version can be also adapted to the modified band-depth proposed in [12], i.e.,

$$WMBD_n^{(j)}(y) = \frac{\sum_{1 \leq i_1 < i_2 < \dots < i_j \leq n} w_{i_1} w_{i_2} \dots w_{i_j} \lambda_m \{A(y; y_{i_1}, \dots, y_{i_j})\}}{\sum_{1 \leq i_1 < i_2 < \dots < i_j \leq n} w_{i_1} w_{i_2} \dots w_{i_j}} \quad (4)$$

where $A_j(y) \equiv A(y; y_{i_1}, \dots, y_{i_j}) \equiv \{\mathbf{x} \in \mathbb{R}^m : \min_{r=i_1, \dots, i_j} y_r(\mathbf{x}) \leq y(\mathbf{x}) \leq \max_{r=i_1, \dots, i_j} y_r(\mathbf{x})\}$, m is the observation’s dimension, $\lambda_m(y) = \lambda(A_j(y)) / \lambda(\mathbb{R}^m)$ and λ is the Lebesgue measure on \mathbb{R}^m .

With the above definitions, the band depths of all the sampled observations can be calculated and ranked in descending order, $y_{[1]}(\mathbf{x}) \geq \dots \geq y_{[n]}(\mathbf{x})$. $y_{[1]}(\mathbf{x})$ is the deepest observation and regarded as the median of the population, whereas $y_{[n]}(\mathbf{x})$ is the most outlying observation which is a potential outlier.

α Central Region. The concept of central region was introduced in [8]. We define the α central region for the weighted functional boxplot based on the weights of observations. The band of the α central region is delimited by the α proportion of all weights, i.e., the accumulated weights of the first p deepest observations

$$WCR_\alpha = \{(\mathbf{x}, y(\mathbf{x})) : \min_{r=1, \dots, p} y_{[r]}(\mathbf{x}) \leq y(\mathbf{x}) \leq \max_{r=1, \dots, p} y_{[r]}(\mathbf{x}), \left(\sum_{r=1, \dots, p-1} w_{[r]} < \alpha \right) \cap \left(\sum_{r=1, \dots, p} w_{[r]} \geq \alpha \right), 0 \leq \alpha \leq 1\}, \quad (5)$$

where $w_{[r]}$ corresponds to the weight for the r -th deepest observation. When $\alpha = 0.5$, (5) corresponds to the 50% central region $WCR_{0.5}$. In practice, the 50% central region is commonly chosen as the confidence region for analysis because it 1) is a robust range for interpretation and 2) enables visualization of the data spread which is less affected by outliers or extreme-values.

Outlier Detection. In classical boxplots, the outliers can be detected by the 1.5 *IQR* (interquartile range). This is comparable to 1.5 times the height of the 50% central region for the weighted functional boxplot. Besides, the weights of the observations also need to be taken into consideration during outlier detection. According to the probability density function for a boxplot based on a normal distribution, the *IQR* is equal to the 50% distribution and the 1.5 *IQR* covers the 99.3% distribution. Hence, we define fences by combining the one of the 1.5 *IQR* with the accumulated weights consistent with the 1.5 *IQR* of the normal distribution, and any objects outside the fences will be flagged as outliers:

$$C_{fences} = \{(\mathbf{x}, y(\mathbf{x})) : \max(\min_{r=1, \dots, q} y_{[r]}(\mathbf{x}), \min(WCR_\alpha) - 1.5 * IQR) \cup \min(\max_{r=1, \dots, q} y_{[r]}(\mathbf{x}), \max(WCR_\alpha) + 1.5 * IQR), \left(\sum_{r=1, \dots, q-1} w_{[r]} < \beta \right) \cap \left(\sum_{r=1, \dots, q} w_{[r]} \geq \beta \right), \beta = 0.993\} \quad (6)$$

3 Comparisons of Boxplots for Analysis

We compare atlases built by 1) weighted point-wise boxplots and 2) functional boxplots, using synthetic observations defined by

$$y_i(x) = 500 * (1 + \sin(2\pi x + 0.1\pi i)) + 2 * age_i,$$

where $x \in [0, 1]$, i is the curve index and age_i its age. Fig. 2 shows the curves colored by age.

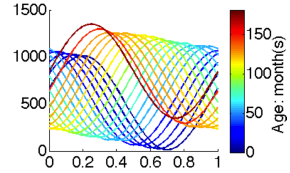


Fig. 2. Observations

Fig. 3(a) shows an atlas built with the weighted point-wise boxplot including four typical percentiles and the point-wise median. While the median curve follows the overall population trend, it is not close to any of the observations because weighted boxplots applied in a point-wise manner to a population of functions disregard the spatial aspect of the functional data. In contrast, our method 1) provides a median curve which corresponds to a curve in the data set, and 2) allows for the computation of *functional* outliers (gray dashed lines) which results in a more robust statistical description for the atlas.

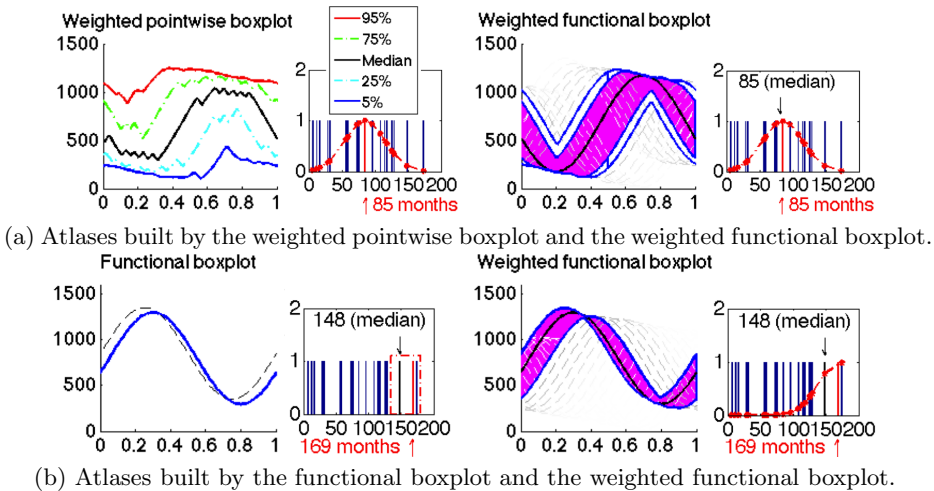


Fig. 3. Comparisons of boxplots on the synthetic data

To construct an atlas at a particular age using standard functional boxplots, we use a uniform window to pick curves centered around the age of interest. As shown in Fig. 3(b), only two curves are available in the uniform window for atlas-building with functional boxplots, and one of them is flagged as an outlier. This atlas includes little information about the population. The atlas built using the weighted functional boxplot (with a Gaussian window size that is comparable to the uniform one according to [10]) captures the population data much better as it does not suffer from the local data sparsity and makes use of all the data.

4 Applications

4.1 Data

The data objects for the weighted functional boxplot can for example be functions, shapes and images (with shapes and images converted to long vectors).

Functions: Our first application is the construction of a pediatric airway atlas for normal subjects to assess airway malformations (subglottic stenosis (SGS)). The observations are a population of 1D functions describing airway cross-sectional areas parameterized along the centerline of the airway. Functions are generated from 3D CT data for 44 normal subjects using the approach in [6] followed by landmark based spatial alignment [11]. We focus the analysis on the region between the true vocal cord and the trachea carina, where SGS locates.

Shapes: The second application is to build a corpus callosum atlas and to explore shape changes with age. The observations are a collection of 32 corpus callosum shapes of varying ages from [4]. Each shape is represented by 64 2D boundary points. We perform affine alignment before atlas constructions.

Images: The third application is to understand age-related changes of the corpus callosum using binary images of the corpus callosum segmentations. The images are converted from the aligned corpus callosum shapes.

4.2 Comparison with Point-Wise Boxplots

We compare the functional boxplot to the point-wise approach on above real datasets to further demonstrate the advantages of our method. Fig. 4 shows the median (the black curve) and the confidence region (the 50% central region, magenta) for both point-wise and functional boxplots. We count the number of data objects inside the confidence region: for the point-wise boxplots only 5 (of 44) functions and none of the shapes or images are fully within the confidence region. However, the functional boxplots by construction achieves a confidence region containing 50% of the data objects. Hence it is a more intuitive representation of true data-object variation. To construct the point-wise confidence regions for shapes we locally compute distances with respect to the median point which establishes an (unsigned) ordering. The confidence region is then the convex hull of the closest half of the points. This strategy would extend to constructing approximate confidence regions with respect to manifold embedding coordinates.

4.3 Atlas Construction with Weighted Functional Boxplots

The weighted functional boxplot is used to build a pediatric airway atlas with variance $\sigma = 30$ months for the weighting function, Fig. 5(a), and the corpus callosum shape/image atlases with $\sigma = 10$ years, Fig. 5(b). The pediatric airway atlases capture increases in cross-sectional airway area with age which is consistent with the growth pattern for pediatric airways and indicates the necessity of building an age-adapted atlas as a reference. The corpus callosum atlases reveal

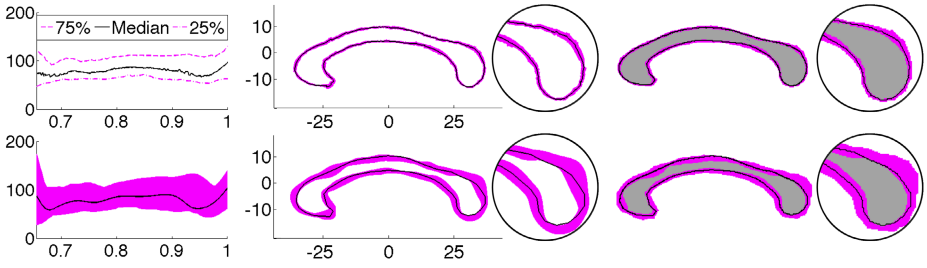


Fig. 4. Comparison between point-wise (top) and functional (bottom) boxplots on functions, shapes and images (from left to right)

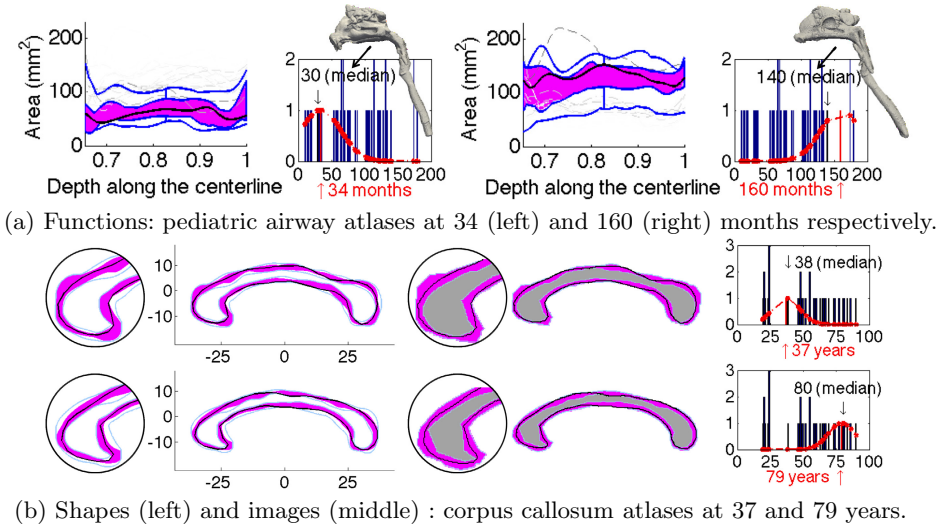


Fig. 5. Age-adapted atlases for functions, shapes, and images

the thinning trend in the shape and the decreasing volume in the image with age, especially at the anterior and posterior parts consistent with [4].

5 Assessment with Statistical Atlas

To test the utility of the statistical atlas built by weighted functional boxplots we show (Fig. 6) airway changes of a SGS subject before (at 9 months) and after (at 20 months) surgery compared to the age-matched normal control airway atlas. Before treatment, there is a constricted region outside the atlas; after treatment, the airway size increases and the corresponding curve is almost entirely within the maximal non-outlying envelope, indicating a successful surgery.

6 Discussion and Conclusions

We proposed a method to compute weighted functional boxplots and use it for spatio-temporal atlas building. We applied it to construct a pediatric airway atlas to assess children with subglottic stenosis and a corpus callosum atlas capturing aging. The proposed method is general, easy to compute, and allows robust statistical description of functional, shape, and image data.

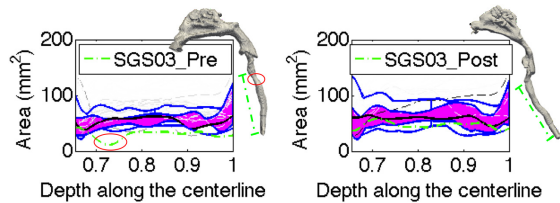


Fig. 6. Airway changes for a subject pre- and post-surgery (green lines) compared to the age-matched atlas. The stenosis of the airway is marked by the red ellipse on the pre-surgery geometry and no stenosis exists in the post-surgery geometry.

Acknowledgement. This publication was supported by NIH 5P41EB002025-28, NIH 1R01HL105241-01, NSF EECS-1148870, NSF EECS-0925875, and NIH 2P41EB002025-26A1.

References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46, 726–739 (2009)
2. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. *International Journal of Computer Vision* 90(2), 255–266 (2010)
3. Fletcher, P., Venkatasubramanian, S., Joshi, S.: The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage* 45(suppl. 1), S143–S152 (2009)
4. Fletcher, T.: Geodesic regression on Riemannian manifolds. In: 3rd MICCAI Workshop on Mathematical Foundations of Computational Anatomy, pp. 75–86 (2011)
5. Gerber, S., Tasdizen, T., Fletcher, P.T., Joshi, S., Whitaker, R.: Manifold modeling for brain population analysis. *Medical Image Analysis* 14(5), 643–653 (2010)
6. Hong, Y., Niethammer, M., Andruejol, J., Kimbel, J., Pitkin, E., Superfine, R., Davis, S., Zdanski, C., Davis, B.: A pediatric airway atlas and its application in subglottic stenosis. In: *International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1194–1197 (2013)
7. Joshi, S., Davis, B., Jomier, M.: Unbiased diffeomorphic atlas construction for computational anatomy. *Neuroimage* 23(suppl. 1), S151–S160 (2004)
8. Liu, R., Parelius, J., Singh, K.: Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics* 27, 783–858 (1999)
9. López-Pintado, S., Romo, J.: On the concept of depth for functional data. *Journal of the American Statistical Association* 104, 718–734 (2009)
10. Marron, J., Nolan, D.: Canonical kernels for density estimation. *Statistics and Probability Letters* 7, 195–199 (1988)
11. Ramsay, J., Silverman, B.: *Functional Data Analysis*. Springer (2005)
12. Sun, Y., Genton, M.: Functional boxplots. *Journal of Computational and Graphical Statistics* 20, 316–334 (2011)

Bayesian Estimation of Probabilistic Atlas for Anatomically-Informed Functional MRI Group Analyses

Hao Xu¹, Bertrand Thirion^{2,3}, and Stéphanie Allassonnière¹

¹ CMAP Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France

² Parietal Team, INRIA Saclay-Île-de-France

³ CEA, DSV, I²BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France

Abstract. Traditional analyses of Functional Magnetic Resonance Imaging (fMRI) use little anatomical information. The registration of the images to a template is based on the individual anatomy and ignores functional information; subsequently detected activations are not confined to gray matter (GM). In this paper, we propose a statistical model to estimate a probabilistic atlas from functional and T1 MRIs that summarizes both anatomical and functional information and the geometric variability of the population. Registration and Segmentation are performed jointly along the atlas estimation and the functional activity is constrained to the GM, increasing the accuracy of the atlas.

Keywords: Probabilistic atlas, geometric variability, joint registration segmentation, atlas-based segmentation, multi-modal, T1 MRI and fMRI.

1 Introduction

Brain atlas is a useful tool in medical image analysis for both segmentation and registration. Probabilistic atlases yield a useful summary of a given dataset [6,7], as they take into account the uncertainty on the underlying tissue type, which is related to partial volume effect (PVE) or to perfectible registration. In [3], a probabilistic framework was proposed for joint nonlinear registration, intensity normalization and segmentation of a single image, from which it infers tissue probability maps. In [10], a probabilistic model was proposed to segment a heterogeneous data set of brain MRIs simultaneously while constructing the probabilistic atlases. In spite of its convincing results, this model is not consistent as the deformations are considered as parameters (whereas segmentation is an unobserved random variable). In [13], the model proposed in [3] was generalized in order to provide estimates of individual segmentation as well as the probabilistic atlas from a set of anatomical images. This approach handles both the segmentation and registration as hidden variables, leading to a coherent convergent statistical estimator. However, this model is limited to scalar images. Here, we generalize it to create a probabilistic atlas that provides the probabilistic

templates of each tissue as well as the degree of activation on GM voxels and the geometric variability.

Functional Magnetic Resonance Imaging of the brain is used to localize functional areas in the cortex and deep nuclei by measuring MRI signal changes associated with neural activity. It is a tool of choice for cognitive studies that aim at identifying specific regions of the brain that are activated in perceptual, cognitive or motor tasks. The most popular type of analysis is Statistical Parametric Mapping (SPM) [5], an approach that estimates the probability that some activation can be due to chance alone and provides p-value maps. Group analysis is then used to detect regions that show a positive mean activation across subjects [4,12]. Accurate realignment of individual scans is most often obtained by normalizing individual anatomical images to a T1 MRI template. These processing steps are done without considering the complementarity of the anatomical and functional information available in each subject. Therefore, detected activations are not confined to gray matter. Few fMRI segmentation methods have been proposed to take into account multi-modal data, such as T1 and functional MRI. An implementation of cortical-based analysis of fMRI data was proposed in [2]. The fMRI data is mapped to the cortical surface, then activations are detected on the surface. It has been shown to achieve anatomically accurate activation detection. In [8], Markov Random Fields (MRF) were used as a spatial regularization in fMRI detection and anatomical information was incorporated into the MRF-based detection framework. In [11], both anatomical and functional data are used to improve the group-wise registrations. Anatomical information appears helpful in fMRI detection; however, the approaches so far do not incorporate a group model into the analysis. In this paper, we process multi-modal data jointly to ensure that the detected active areas are conditioned to gray matter while registration is informed by functional information. More specifically, group analysis first performs the realignment of individual images to a T1 MRI template and then segments active regions by thresholding. However, performing registration and segmentation jointly is generally more effective than performing them sequentially [13,14]. In this paper, we take advantage of such coupling.

To deal with all the issues described above, we propose an atlas estimation procedure that can improve the template image estimation and the detection of the active areas. We generalize the model proposed in [13]. The input is now multivariate, as it encodes multi-modal patient observations (gray level T1 and functional MRIs). The estimated active areas are conditioned to GM segmentation. We perform the estimation by coupling the segmentation and registration steps. We estimate a probabilistic atlas that accounts for the variability of active areas in the population. We also learn the geometry as the metric on the space of deformations which drives the coupled segmentation. We use a stochastic algorithm with known guarantees on the convergence in the estimation procedure. The output of the algorithm is the probabilistic atlas, the individual active areas and the means and variances of each tissue type in each modality.

The rest of this paper is organized as follows. In Section 2, we present the model, the estimation procedure, the algorithm. Section 3 yields experimental results on simulated and real data.

2 Methods

Statistical Model. Let us consider n pairs of T1- and f- MRIs $(y_{1,i}, y_{2,i})_{1 \leq i \leq n}$ from n patients. Each image is observed on a grid of voxels Λ embedded in a continuous domain $D \subset \mathbb{R}^3$. We denote $x_j \in D$ the location of voxel j . We consider that each T1 MRI is composed of voxels belonging to one of the four classes, corresponding to four tissue types: gray matter, white matter (WM), CSF and background (BG). Each fMRI is composed of voxels belonging to one class among $3 + K$, corresponding to WM, CSF and BG, where no activation is expected to occur, and K different levels of activation in gray matter. We assume that the signal in the $3 + K$ classes is normally distributed with class dependent means $(\mu_{1,f(k)}, \mu_{2,k})_{k \in \{WM, CSF, BG, GM_1, \dots, GM_K\}}$ and variances $(\sigma_{1,f(k)}^2, \sigma_{2,k}^2)_{k \in \{WM, CSF, BG, GM_1, \dots, GM_K\}}$, where $f(k) = k$ if $k \in \{WM, CSF, BG\}$, GM otherwise. The whole set of parameters is denoted by Θ .

As mentioned previously, we are working with gray level images which have not been pre-segmented. The unknown class of each voxel is supposed to be the discretization on Λ of a random deformation of probability maps $(P_k)_{1 \leq k \leq K+3}$. These probability maps correspond to the probability of each voxel to belong to each class in the template domain. They form the *probabilistic template* of the population. The random deformations from this template to each subject are also unobserved as the images are not pre-registered. We define them through a random field $z : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that for $j \in \Lambda$ the prior probability of a voxel j from subject i to be in the k^{th} class is given by:

$$\mathbb{P}(c_i^j = k) = P_k(x_j - z(x_j)). \tag{1}$$

We define the deformation field as a finite linear combination of a given kernel K_g centered at some fixed equi-distributed control points in the domain D , $(x_g)_{1 \leq g \leq k_g}$, with parameter $\beta \in (\mathbb{R}^3)^{k_g}$

$$\forall x \in D, z_\beta(x) = \sum_{g=1}^{k_g} K_g(x, x_g)\beta(g), \tag{2}$$

where K_g is chosen as a radial Gaussian Kernel. Note that we expect the tissue-specific information to be found in all the brain volume, hence the whole volume has to be covered with control points. As for the deformation model, the probability template maps $P_k : \mathbb{R}^3 \rightarrow]0, 1[$, $\forall k \in \llbracket 1, K + 3 \rrbracket$ are parametrized by the coefficients $\alpha_k \in]0, 1[^{k_p}$ which satisfy $\forall l \in \llbracket 1, k_p \rrbracket, \sum_{k=1}^{K+3} \alpha_k^l = 1$. Let $(p_l)_{1 \leq l \leq k_p}$ be some control points :

$$\forall x \in D, P_k(x) = \sum_{l=1}^{k_p} K_p(x, p_l)\alpha_k^l, \tag{3}$$

where $K_p(x, p_l) = 1$ if p_l is the nearest neighbor of x among $(p_j)_j$, 0 otherwise.

The previous hypothesis provides a generative statistical model for a sample of pairs of gray level images. The random variables are the deformation vector β , the class of each voxel c and the observed gray levels of the images. We assume that the deformation vector follows a normal distribution with mean zero and non-diagonal covariance matrix Γ_g . The hierarchical model is given by:

$$\left\{ \begin{array}{l} \beta_i \sim \mathcal{N}(0, \Gamma_g) | \Gamma_g; \quad c_i^j \sim \sum_{k=1}^{K+3} \delta_k P_k(x_j - z_{\beta_i}(x_j)) | \beta_i, \\ \left(\begin{array}{c} y_{1,i}^j \\ y_{2,i}^j \end{array} \right) \sim \mathcal{N} \left(\left(\begin{array}{c} \mu_{1,f(k)} \\ \mu_{2,k} \end{array} \right), \left(\begin{array}{cc} \sigma_{1,f(k)}^2 & 0 \\ 0 & \sigma_{2,k}^2 \end{array} \right) \right) \Big| c_i^j = k, \left(\begin{array}{c} \mu_{1,f(k)} \\ \mu_{2,k} \end{array} \right), \left(\begin{array}{c} \sigma_{1,f(k)}^2 \\ \sigma_{2,k}^2 \end{array} \right) \end{array} \right. \quad (4)$$

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ is the normal density with mean μ and variance σ^2 and δ_k is a Dirac function. The covariance matrix Γ_g is not assumed to have any particular pattern of zeros. This makes it possible to model local and global correlations between control point moves, in particular, very correlated displacements can be captured such as translation of a large area of the images.

The parameters to estimate are the covariance matrix Γ_g of the deformation distribution (Eq. (2)), $(\alpha_k)_{1 \leq k \leq K+3}$ the coefficients that define the template maps (Eq. (3)), $(\mu_{1,f(k)}, \mu_{2,k})_{1 \leq k \leq K+3}$ and $(\sigma_{1,f(k)}^2, \sigma_{2,k}^2)_{1 \leq k \leq K+3}$ the class dependent means and variances. As medical images are high-dimensional but usually come in small samples, we work in a Bayesian framework. We use the standard conjugate priors for the covariance matrix, the class dependent means and variances with fixed hyper-parameters. All priors are assumed independent.

Estimation Algorithm. A maximum a posteriori (MAP) approach yields estimates of the model parameters: $\hat{\theta}_n = \arg \max_{\theta \in \Theta} q_B(\theta | (y_{1,1}, y_{2,1}), \dots, (y_{1,n}, y_{2,n}))$, where q_B denotes the posterior distribution of the parameters given the n observations $(y_{1,1}, y_{2,1}), \dots, (y_{1,n}, y_{2,n})$. As we are in an incomplete-data setting, we choose the Stochastic Approximation Expectation-Maximization (SAEM) algorithm coupled with a Markov Chain Monte Carlo method to take advantage of its theoretical and numerical properties [1,13]. The SAEM algorithm is an iterative procedure that consists of three steps. First, we simulate the missing data using a Metropolis-Hastings algorithm within Gibbs sampler. Then a stochastic approximation is done on the sufficient statistics using the simulated value of the missing data. Last, we maximize the expected log-likelihood with respect to the model parameters.

3 Experiments and Results

We test our algorithm on both simulated data and real data. As the SAEM algorithm is an iterative procedure, we run 250 iterations which was checked to reach convergence. We initialize $\beta_0 = 0$ and a random classification c_0 .

Simulated Data. We use a pair of $64 \times 64 \times 8$ images as the reference images. We consider here $K = 3$, i.e. three different levels of activation in GM and 6

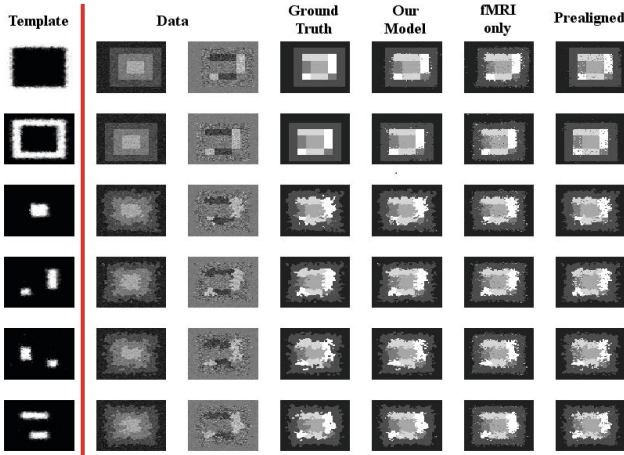


Fig. 1. Experiments on simulated data. The first column displays the first slice of the probabilistic template, each row corresponding to a class and white/black colors to high/low probability. The second and third columns show one slice of six pairs of data images. The fourth to seventh columns correspond to the ground truth and the estimated segmentation for different models.

classes in total. We define the means and the standard deviations as follows (taking values that are observed in real fMRI for the standard deviations):

$$\begin{pmatrix} \mu_{1,1:6} \\ \mu_{2,1:6} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 & 3 & 3 & 3 \\ 0 & 0 & 0 & 2.5 & 0 & -2.5 \end{pmatrix}, \quad \begin{pmatrix} \sigma_{1,1:6} \\ \sigma_{2,1:6} \end{pmatrix} = \begin{pmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.24 & 1.22 & 0.91 & 0.78 & 0.71 & 0.83 \end{pmatrix}$$

The training data is composed of 20 pairs of images with random deformations of our template following Eq. (4) with previous parameters. We take 64 fixed control points for the deformation model given in Eq. (2), i.e. one control point in each $4 \times 4 \times 4$ cube. We take all the points in the image as landmarks for the template model given in Eq. (3).

The most important output of our estimation procedure is the probabilistic template. The estimated probabilistic maps are shown in the first column in Fig.1, each row corresponding to one class. The white/dark colors represent high/low probability of the tissues. Our probabilistic maps are sharp, as most voxels in each class have a probability larger than 0.9. Only voxels at the boundary between two classes are fuzzy which takes into account both the accuracy of deformation and the level of noise.

As mentioned previously, our model uses both the T1- and f- MRIs because we want the active areas to be conditioned to GM. We compare our model with the segmentation model in [13] using fMRI only. The result is shown in Fig.1. The second and third columns correspond to one slice of six pairs of data images. The ground truth and the final estimated segmentation of different methods (our model, the model using fMRI only) are shown from the fourth to sixth columns. From the fourth and fifth columns, we can see that the segmentation obtained with our atlas estimation is accurate. From the fifth and sixth columns, we see

Table 1. Experiments on synthetic data. Jaccard Index for the different methods averaged across all data.

	BG	CSF	WM	GM_1	GM_2	GM_3
Our model	98.5%	92.3%	88.4%	88.0%	73.1%	90.4%
fMRI only	89.0%	75.0%	69.7%	78.3%	48.2%	81.4%
Pre-Aligned	96.6%	91.1%	87.0%	86.6%	67.8%	86.6%

the improvement using the information provided from the T1 MRI. We calculate the Jaccard index for each class as a quantitative validation (Table 1) for each method. Our model yields an accurate segmentation, as only few voxels are misclassified. For the model using fMRI only, we are only interested in GM_1 and GM_3 which correspond to the active areas. As the other classes are non-active, the means of these classes are close to zero, therefore they are difficult to segment without the MRI tissue type information, leading to lower values for these classes. Moreover, the Jaccard indexes for the active area are lower than those obtained with our procedure, which shows that the coupling of both images information increases the accuracy of detection.

In our model, the registration and segmentation are done jointly, which avoids any pre-registration. In the preprocessing, each fMRI is pre-aligned to its corresponding MR image. However, the inter-subject non-rigid registration is not done, as it would require a template and would not take into account the fMRI observation to drive this preprocessing step. We compare our model with the pre-aligned model which does the registration and the segmentation sequentially. First, we use the segmentation model in [13] using the T1 MRIs, we get the deformation vector and individual tissue segmentation as our output. Then we apply the same deformation to the fMRI and detect the activation only in GM. The estimated segmentation of the pre-aligned model is shown in the seventh columns in Fig.1. Comparing the fifth and seventh columns, the segmentations look similar which makes it difficult to say which method gives the better result. However our model gives less isolated points. Moreover, looking at the Jaccard indexes (Table 1), we see that our model outperforms the pre-aligned model. This shows the improvement of doing registration and segmentation jointly.

In-Vivo Data. The proposed method was also tested on a real MRI and fMRI dataset described in [9]. Both anatomical and functional data were subject to standard preprocessing using SPM8, including spatial normalization and General Linear Model analysis. Images are sampled at 3mm resolution, yielding volumes of shape $46 \times 53 \times 63$. We select a contrast from the fMRI that yields differential effect of a computation task versus a simple instruction reading/listening. We have $K = 3$ levels of activation in the GM, hence 6 classes overall.

We take 792 fixed control points for the deformation model given in Eq. (2), corresponding to one control points in each $6 \times 6 \times 6$ cube and $23 \times 27 \times 32$ points in the image as the landmarks for the template model given in Eq. (3), corresponding to one landmark in each $2 \times 2 \times 2$ cube.

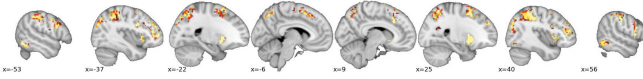


Fig. 2. The estimated template on real data. The yellow/red colors correspond to high/low probability of the activation for the computation task.

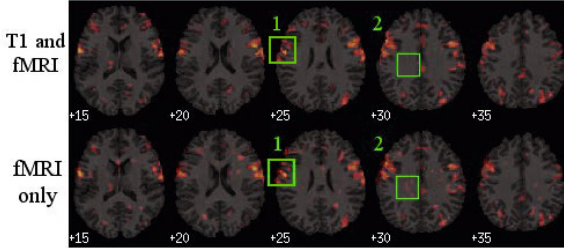


Fig. 3. Experiments on real data showing the detected active areas $p > .95$ for the computation task. The first row for our method using both T1- and f- MRI and the second row for the standard method using fMRI only. Each column corresponds to one slice of the same patient.

The estimated probabilistic maps, thresholded at the $p > .95$ level, are shown in Fig.2. The yellow/red colors correspond to high/low probability of the computation task activation. Our probabilistic maps are sharp. The detected areas are well conditioned to GM and fits the known active areas for the computation task. For example, in the slice $x = 25mm$, we find well the Putamen.

We compared our model with the standard method that thresholds the group-level mean activation. We represent the active areas in the computation task overlaid on T1 images. The results of one patient are shown in Fig.3. The first row for our method uses both T1- and f- MRI and the second row for the method uses fMRI only. Each column corresponds to one slice of the same patient. In zone 1, we see that the areas detected as active by our method are limited to the GM. However, a part of the detected active areas by the non-anatomically aware method are outside of the brain. In zone 2, the standard method detects some active areas in WM, while our method does not. These show that we reach our goal, i.e. the detected active areas are well conditioned to GM. The detected active areas by our method are similar to those by the standard method in GM, this shows that our segmentation is accurate.

4 Conclusion

In this study, we proposed a statistical model to detect the active areas in the brain using both T1 and functional MRI. We used a stochastic algorithm to perform registration, segmentation and to create a probabilistic atlas simultaneously. Our model has several advantages. First, the probabilistic atlas contains both the templates and the geometric variability of the population. Second, we

do not need any pre-registration to perform the segmentation which is automatically obtained as an output. Third, the detected active areas are confined to GM with the information provided from the MRI data. Our experiments show that we get better results with our algorithm than the standard method. The detected active areas are well conditioned to GM and the atlas is sharp.

Acknowledgment. We thank Digiteo for funding MMoVNI.

References

1. Allasonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via stochastic approximation algorithm: A convergence study. *Bernoulli Journal* 16(3), 641–678 (2010)
2. Andrade, A., Kherif, F., Mangin, J.F., Worsley, K.J., Paradis, A.L., Simon, O., Dehaene, S., Bihan, D.L., Poline, J.B.: Detection of fMRI activation using cortical surface mapping. *Human Brain Mapping* 12, 79–93 (2001)
3. Ashburner, J., Friston, K.J.: Unified segmentation. *NeuroImage* 26(3), 839–851 (2005)
4. Friston, K.J., Holmes, A.P., Price, C.J., Büchel, C., Worsley, K.J.: Multisubject fMRI studies and conjunction analyses. *NeuroImage* 10(4), 385–396 (1999)
5. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Firth, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 2(4), 189–210 (1995)
6. Gouttard, S., Styner, M., Joshi, S., Smith, R.G., Cody, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. In: *SPIE* (2007)
7. Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE TMI* 20(8), 677–688 (2001)
8. Ou, W., Wells III, W.M., Golland, P.: Combining spatial priors and anatomical information for fMRI detection. *Medical Image Analysis* 14(3), 318–331 (2010)
9. Pinel, P., Thirion, B., Mériaux, S., Jobert, A., Serres, J., Bihan, D.L., Poline, J.B., Dehaene, S.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neuroscience* 8 (2007)
10. Ribbens, A., Hermans, J., Maes, F., Vandermeulen, D., Suetens, P.: Sparc: Unified framework for automatic segmentation, probabilistic atlas construction, registration and clustering of brain mr images. In: *IEEE ISBI*, pp. 856–859 (2010)
11. Sabuncu, M.R., Singer, B.D., Conroy, B., Bryan, R.E., Ramadge, P.J., Haxby, J.V.: Function-based intersubject alignment of human cortical anatomy. *Cerebral Cortex* 20, 130–140 (2010)
12. Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.B.: Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage* 35, 105–120 (2007)
13. Xu, H., Thirion, B., Allasonnière, S.: Probabilistic atlas and geometric variability estimation to drive tissue segmentation. *Statistics in Medicine* (submitted)
14. Yeo, B., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical demons: Fast diffeomorphic landmark-free surface registration. *IEEE TMI* 29(3), 650–668 (2010)

Exhaustive Search of the SNP-SNP Interactome Identifies Epistatic Effects on Brain Volume in Two Cohorts

Derrek P. Hibar¹, Jason L. Stein¹, Neda Jahanshad¹, Omid Kohannim¹,
Arthur W. Toga¹, Katie L. McMahon², Greig I. de Zubicaray³,
Grant W. Montgomery⁴, Nicholas G. Martin⁴, Margaret J. Wright⁴,
Michael W. Weiner^{5,6}, and Paul M. Thompson¹

¹ Imaging Genetics Center, Laboratory of Neuro Imaging,
UCLA School of Medicine, Los Angeles, USA

² Center for Magnetic Resonance, School of Psychology,
University of Queensland, Brisbane, Australia

³ Functional Magnetic Resonance Imaging Laboratory, School of Psychology,
University of Queensland, Brisbane, Australia

⁴ Genetic Epidemiology Laboratory,

Queensland Institute of Medical Research, Brisbane, Australia

⁵ Departments of Radiology, Medicine, Psychiatry, UC San Francisco,
San Francisco, USA

⁶ Department of Veterans Affairs Medical Center, San Francisco, USA

Abstract. The SNP-SNP interactome has rarely been explored in the context of neuroimaging genetics mainly due to the complexity of conducting $\sim 10^{11}$ pair-wise statistical tests. However, recent advances in machine learning, specifically the iterative sure independence screening (SIS) method, have enabled the analysis of datasets where the number of predictors is much larger than the number of observations. Using an implementation of the SIS algorithm (called EPISIS), we used exhaustive search of the genome-wide, SNP-SNP interactome to identify and prioritize SNPs for interaction analysis. We identified a significant SNP pair, rs1345203 and rs1213205, associated with temporal lobe volume. We further examined the full-brain, voxelwise effects of the interaction in the ADNI dataset and separately in an independent dataset of healthy twins (QTIM). We found that each additional loading in the epistatic effect was associated with $\sim 5\%$ greater brain regional brain volume (a protective effect) in both the ADNI and QTIM samples.

Keywords: epistasis, interaction, genome, sure independence, tensor-based morphometry.

1 Introduction

Traditional univariate methods can test the association of common genetic variants with complex quantitative traits, but they only consider the marginal effect of a single locus and potentially miss variance explained by synergistic or interacting effects of

pairs or sets of SNPs¹ [1]. For many complex traits, the similarity of family members drops faster than would be expected as relatedness decreases [2]. This implies that there are non-additive (epistatic) interactions involved in the etiology of many complex traits. Statistical interactions have been demonstrated to be plausible representations of the complex interactions of genes in biological pathways [3-4].

Some prior studies have examined second-order interactive effects of SNPs on brain structure [5-7]. However, none of these studies has considered genome-wide genotype data; the closest conceptually related study tested for SNP effects on diffusion imaging measures, and aggregated all SNPs with correlated effects into a network [8]. The concept here is different, and aims to assess gene pairs that influence each other's effects on the brain. Prior studies tested interaction effects only for a limited number of popular candidate genes. Any approach based on pre-selecting a pair of genes will overlook a vast search space of potential interactions among SNPs in the genome that have no obvious prior connection. Also, a large main effect is not necessary to be able to detect significant second-order interactions [9]. Given this, prior hypotheses focusing on SNPs with large individual effects may also overlook large second-order effects. Importantly, power estimates for detecting interactive effects are comparable to those for single SNP tests [1]. In simulation studies, the inclusion of interaction terms can boost the power to detect main effects, at least for certain genetic tests [10]. Here we examined the genome-wide, SNP-SNP interactome² to test genetic associations with a quantitative biomarker of Alzheimer's disease (temporal lobe volume) in the public Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. We further examine the whole-brain effects of interaction pairs in statistical parametric maps generated with tensor-based morphometry (TBM); we also replicate our tests in an independent, non-overlapping dataset of young healthy twins from the Queensland Twin Imaging (QTIM) study [11].

2 Methods

2.1 Imaging Parameters and Study Information

We downloaded the full baseline set of 818 high-resolution, T1-weighted structural MRI brain scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is a multi-site, longitudinal study of patients with Alzheimer's disease (AD), mild cognitive impairment (MCI) and healthy elderly controls (HC). Subjects were scanned with a standardized protocol to maximize consistency across sites. We used the baseline 1.5 Tesla MRI scans, i.e., the T1-weighted 3D MP-RAGE scans, with TR/TE = 2400/1000 ms, flip angle = 8°, slice thickness = 1.2 mm, and a final voxel resolution = 0.9375 x 0.9375 x 1.2 mm³. Raw MRI scans were pre-processed to

¹ **SNP** (=single nucleotide polymorphism): a single-letter variant in the genome; these variations are common, even in healthy human populations, and their effects on brain measures can be assessed using association testing, at one SNP or up to a million genotyped SNPs.

² **Interactome**: The study of interactions between genetic variants or sets of variants in terms of their effects on traits such as brain measures.

remove signal inhomogeneity, non-brain tissue, and affine registered to the MNI template (using 9 parameters).

Additionally, we obtained 753 high-resolution, T1-weighted structural MRI brain scans from the Queensland Twin Imaging (QTIM) study. QTIM is a longitudinal neuroimaging and genetic study of young, healthy twins and their family members. All structural MRI scans were acquired on a single 4-Tesla scanner (Bruker Medspec): T1-weighted images, inversion recovery rapid gradient echo sequence, TR/TE = 1500/3.35 ms, flip angle = 8° , slice thickness = 0.9 mm, 256×256 acquisition matrix, with a final voxel resolution = $0.9375 \times 0.9375 \times 0.9 \text{ mm}^3$. Raw MRI scans were pre-processed to remove signal inhomogeneity, non-brain tissue, and affine registered to the ICBM template (using 9 parameters).

2.2 Genotype Pre-processing and Study Demographics

Genome-wide genotyping data were available for the full set of ADNI subjects. We performed standard quality control procedures to ascertain the largest homogenous genetic sub-population in the dataset, using multi-dimensional scaling (MDS) compared to a dataset of subjects of known genetic identity (HapMap III; <http://hapmap.ncbi.nlm.nih.gov/>). The largest subset contained 737 subjects from the CEU population (Caucasians). We therefore removed the remaining 81 subjects from our analysis to limit the effects of genetic stratification on our statistical analyses [12]. Additionally, we applied filter rules to the genotype data to remove rare SNPs (minor allele frequency < 0.01), violations of Hardy-Weinberg Equilibrium (HWE $p < 5.7 \times 10^{-7}$), and poor call rate (<95%). Data were further “phased” to impute any missing individual genotypes after filtering using the MaCH program [13] following the ENIGMA imputation protocol [14]. After filtering and phasing, 534,033 SNPs remained.

All QTIM subjects were ascertained for genetic similarity, so no subjects were removed before analysis. All 753 subjects in the QTIM dataset clustered with the CEU population, in the MDS analysis. The same genotype filter rules from the ADNI dataset were applied to the QTIM sample’s genetic data. After filtering and phasing, 521,232 SNPs remained.

After all rounds of genotype pre-processing, the ADNI sample contained 737 subjects (mean age \pm sd: 75.5 \pm 6.8 yrs; 436 males) comprised of 173 patients diagnosed with Alzheimer’s disease, 358 subjects with mild cognitive impairment, and 206 healthy elderly controls. The QTIM sample contained 753 subjects (mean age \pm sd: 23.1 \pm 3.0 yrs; 286 males) and consisted of 110 monozygotic twin pairs, 147 dizygotic twin pairs, 3 dizygotic twin trios, 143 singletons, and 87 siblings from 438 families.

2.3 Tensor-Based Morphometric Differences in the Full Brain

We calculated information on regional brain morphometry using an elastic, nonlinear registration algorithm (3DMI) [15] applied to the entire brain. Voxelwise volumetric differences were stored, using the Jacobian value of the deformation matrix obtained by nonlinearly registering a subject’s scan to a study-specific minimum deformation

template (MDT). Scans from the ADNI and QTIM datasets were processed and analyzed separately (using separate study templates). The MDT for the ADNI sample is a nonlinear average of 40 age-and-sex matched healthy elderly controls [16]. The MDT for the QTIM is a nonlinear average of 32 age- and sex- matched, unrelated subjects [17]. Nonlinear registration with 3DMI yields a $110 \times 110 \times 110$ voxel statistical parametric map, where the Jacobian value at each voxel represents the expansion required to match the same voxel in the study-specific MDT.

2.4 Genome-Wide, Gene-Gene Interaction Testing

The EPISIS software is an implementation of the machine-learning algorithm called *sure independence screening* (SIS) developed by Fan and Lv [18]. The SIS algorithm is a correlation learning method that can be applied to ultra-high dimensional datasets where the number of predictors p is much greater than the number of observations n . Despite the development of robust methods for cases where $p > n$ (e.g., the Dantzig selector of Candès and Tao [19]) the properties of the selector fail when $p \gg n$. Fan and Lv [18] developed the SIS algorithm to reduce the ultra-high dimension of p to a moderately-sized subset, while guaranteeing that the subset still explains the maximum amount of variance explained by the full set of predictors.

We conducted an exhaustive search of association tests of genome-wide SNP-SNP interactions with temporal lobe volume (computed by integrating the Jacobian over an temporal lobe ROI on the MDT) [20] in the ADNI dataset using the EPISIS software. EPISIS utilizes the massively parallel processing available in GPGPU (General-purpose computing on graphics processing units) framework to test $p(p-1)/2$ SNP-SNP interactions in the ADNI dataset in a feasible timeframe. We used the SIS algorithm with cell-wise dummy coding (CDC) [21] to reduce the full predictor space into a subset d of $n/\log(n)$ interaction terms [18]. After screening the full set of possible two-way SNP-SNP interactions, we applied ridge regression [22-23] to the subset of interaction terms (the multiplicative loading of each SNP-SNP pair) and selected significant SNP-SNP interaction terms using the extended Bayesian Information Criterion (EBIC) [24] with $\gamma = 0.5$. The choice of the parameter γ was chosen based on simulations [21]. The EPISIS software is implemented in CUDA and optimized for parallel processing across multiple NVIDIA GPU cards as detailed elsewhere [21]. A single exhaustive search of the genome-wide, SNP-SNP interactome with EPISIS was completed in 7 hours (using one NVIDIA Tesla C2050 GPU card).

2.5 Voxelwise Interaction Analysis and Replication

We tested the significant SNP-SNP interaction pair selected by ridge regression for association with voxelwise, regional volume differences (V) at each point, i , in the full brain. The association test at each voxel in the ADNI dataset followed the multiplicative interaction model in multiple linear regression:

$$V_i \sim \beta_0 + \beta_{\text{age}}X_{\text{age}} + \beta_{\text{sex}}X_{\text{sex}} + \beta_{\text{snp1}}X_{\text{snp1}} + \beta_{\text{snp2}}X_{\text{snp2}} + \beta_{\text{snp1,2}}X_{\text{snp1}}*X_{\text{snp2}} + \varepsilon \quad (1)$$

Additionally, we used QTIM as an independent replication sample of the top SNP-SNP interaction pair identified by ridge regression after EPISIS. The voxelwise association tests assume the multiplicative interaction model, detailed previously. Due to the family design of the QTIM sample, we tested association using mixed-effects modeling as implemented in the R package *kinship* (version 1.3) in order to account for relatedness.

3 Results

After screening the full set of SNP-SNP interaction pairs for association with temporal lobe volume in the ADNI dataset, we obtained a subset d of SNP-SNP interaction pairs such that $d = n/\log(n)$. The subset is chosen by ranking the marginal correlation coefficients of each interaction pair and selecting the top d SNP-SNP pairs (correlation learning) [18], in this case $d = 111$ pairs. Next, we applied ridge regression to the pruned subset of SNP-SNP interaction pairs. Using the extended BIC ($\gamma = 0.5$) [21] to estimate significance in our ridge regression, we identified a significant interaction between rs1345203 and rs1213205. The distribution of alleles for each SNP and their interaction is given in Table 1.

Table 1. The distribution of alleles for the significant SNPs and the number of subjects with each genotype by study. For rs1345203 the minor allele is G and the major allele is A in both studies. The minor allele is A and the major allele is G for rs1213205. The association testing assumes an additive model (each subject is assigned a value 0,1,2 based on the number of minor alleles they have at a given SNP). The interaction column gives the number of subjects in each category after multiplying together the counts of each of the alleles.

Study	rs1345203	rs1213205	Interaction
ADNI (n=737)	G/G: 27	A/A: 93	0 loadings: 612
	A/G: 223	G/A: 297	1 loadings: 79
	A/A: 487	G/G: 347	2 loadings: 46
QTIM (n=753)	G/G: 5	A/A: 78	0 loadings: 664
	A/G: 193	G/A: 300	1 loadings: 70
	A/A: 555	G/G: 375	2 loadings: 19

We further examined the significant SNP pair, rs1345203 and rs1213205, for whole-brain effects in the statistical parametric maps generated using tensor-based morphometry (TBM). In the ADNI dataset, we found broad effects bilaterally in the temporal and occipital lobes (Fig. 1) after correcting for multiple tests at a 5% false discovery rate (FDR) using the searchlight FDR method [25].

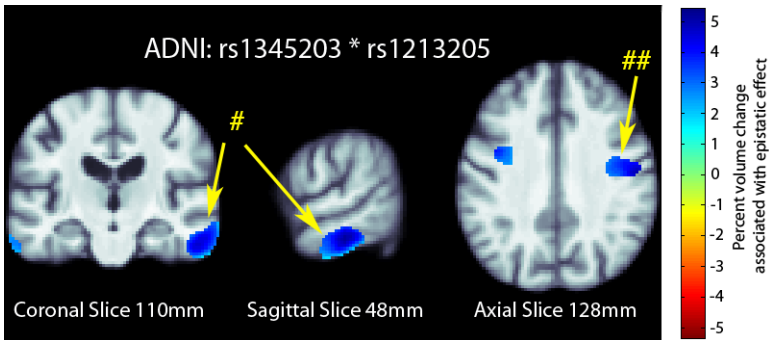


Fig. 1. 3D maps of percent tissue change for each additional genetic variant in the interaction in ADNI. Only significant regions are shown after correcting for multiple comparisons with searchlight FDR [25] at a 5% false discovery rate. Images follow radiological orientation. The origin is placed at the Posterior-Right-Inferior corner. Cooler colors over the tissue represent tissue expansion (larger regional brain volume) compared to an average template. There is a clear protective effect of the epistatic loadings bilaterally in the temporal (# in the figure) and occipital lobes (## in the figure): as the number of alleles a subject has increases, the amount of local brain tissue they have is also increased on average.

We examined the whole-brain effects of the SNP pair on voxelwise, regional brain volume in the statistical parametric maps in an independent dataset (QTIM). The distribution of alleles for each SNP and their interaction in the QTIM sample is given in Table 1. In the QTIM, we identified significant effects in the left temporal lobe and along the border of the left frontal and occipital lobes (Fig. 2) after correction for multiple tests at 5% false discovery rate (FDR) using the searchlight FDR method.

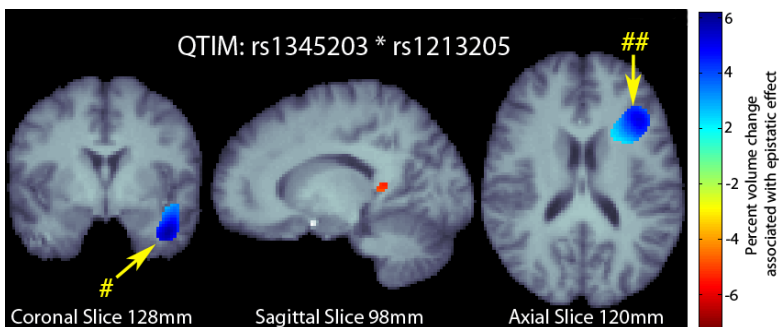


Fig. 2. 3D maps of percent tissue change for each additional genetic variant in the interaction in QTIM. Only significant regions are shown after correcting for multiple comparisons with searchlight FDR [25] at a 5% false discovery rate. Images follow radiological orientation. The origin is placed at the Posterior-Right-Inferior corner. Cooler colors over the tissue represent tissue expansion (larger regional brain volume) compared to an average template. There is a clear protective effect of the epistatic loadings in the left temporal (# in the figure) and along the boundary of the frontal and occipital lobes (## in the figure): as the number of alleles a subject has increases, the amount of local brain tissue they have is also increased on average.

4 Discussion

The genome is incredibly complex and statistical epistasis has been suggested as an appropriate model for the biological interactions among genes and protein products in related pathways [3-4]. Here we examined the multiplicative effect of SNP-SNP pairs on brain volume differences. Significant interaction terms explain additional variance in brain volume beyond what is already explained by the additive SNP terms. In our primary tests of associations with temporal lobe volume in the ADNI dataset, we screened 10^{11} possible SNP-SNP interaction pairs using the GPU acceleration implemented in the EPISIS software. The top 111 interaction pairs were selected after ranking the marginal effect of each SNP-SNP pair on temporal lobe volume, using an implementation of the sure independence screening (SIS) algorithm [18]. We used ridge regression and the extended BIC [24] to identify a significant interaction between rs1345203 and rs1213205. The functional relevance of the two SNPs is as yet unknown. However, data obtained from the ENCODE dataset (<http://genome.ucsc.edu/>) show that rs1345203 is located in a transcription factor gene (ELF1/CEBPB) that demonstrates regulatory influence on the DNA structure. The SNP rs1213205 is located in a region of hypersensitivity to cleavage by DNase regulatory elements. It is worth noting that the parameter choices made in the interaction analysis may influence the results, however, parameters were chosen based on the recommended values for EPISIS [21] and SIS [18]. Additional work is still required to identify precisely how these two SNPs might affect brain structure, and to further replicate their interaction. Specifically, we need to identify how changes at a given SNP are related to changes in activity in gene transcription or translation into protein products involved in similar biological pathways.

References

1. Marchini, J., et al.: Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics* 37(4), 413–417 (2005)
2. Wray, N.R., et al.: Multi-locus models of genetic risk of disease. *Genome Med.* 2(10) (2010)
3. Moore, J.H., Williams, S.M.: Epistasis and its implications for personal genetics. *American Journal of Human Genetics* 85(3), 309 (2009)
4. Stich, B., et al.: Power to detect higher-order epistatic interactions in a metabolic pathway using a new mapping strategy. *Genetics* 176(1), 563–570 (2007)
5. Pezawas, L., et al.: Evidence of biologic epistasis between BDNF and SLC6A4 and implications for depression. *Molecular Psychiatry* 13(7), 709–716 (2008)
6. Tan, H.-Y., et al.: Epistasis between catechol-O-methyltransferase and type II metabotropic glutamate receptor 3 genes on working memory brain function. *PNAS* 104(30), 12536–12541 (2007)
7. Wang, Y., et al.: Evidence of Epistasis Between the Catechol-O-Methyltransferase and Aldehyde Dehydrogenase 3B1 Genes in Paranoid Schizophrenia. *Biological Psychiatry* 65(12), 1048–1054 (2009)
8. Chiang, M.-C., et al.: Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *Journal of Neuroscience* 32(25), 8732–8745 (2012)

9. Cordell, H.J.: Detecting gene–gene interactions that underlie human diseases. *Nature Reviews Genetics* 10(6), 392–404 (2009)
10. Cordell, H.J., et al.: Statistical modeling of interlocus interactions in a complex disease: rejection of the multiplicative model of epistasis in type 1 diabetes. *Genetics* 158(1), 357–367 (2001)
11. de Zubicaray, G.I., et al.: Meeting the challenges of neuroimaging genetics. *Brain Imaging Behavior* 2, 258–263 (2008)
12. Lander, E.S., Schork, N.J.: Genetic dissection of complex traits. *Science* 265(5181), 037–2048 (1994)
13. Abecasis, G.R., et al.: MaCH: Using Sequence and Genotype Data to Estimate Haplotypes and Unobserved Genotypes. *Genetic Epidemiology* 34(8), 816–834 (2010)
14. ENIGMA2 Genetics Support Team. ENIGMA2 1KGP cookbook (v3) (Online). The Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) consortium (accessed July 27, 2012)
15. Leow, A., et al.: Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties. *Inf. Process. Med. Imaging* 19, 493–503 (2005)
16. Hua, X., et al.: Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer’s disease clinical trials. *Neuroimage* (2012) (EPUB)
17. Jahanshad, N., et al.: Brain structure in healthy adults is related to serum transferrin and the H63D polymorphism in the HFE gene. *Proc. Natl. Acad. Sci.* 109(14), E851–E859 (2012)
18. Fan, J., Lv, J.: Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(5), 849–911 (2008)
19. Candès, E., Tao, T.: The Dantzig selector: Statistical estimation when p is much larger than n . *The Annals of Statistics* 35(6), 2313–2351 (2007)
20. Stein, J.L., et al.: Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer’s disease. *Neuroimage* 51(2), 542–554 (2010)
21. Ueki, M., Tamiya, G.: Ultrahigh-dimensional variable selection method for whole-genome gene-gene interaction analysis. *BMC Bioinformatics* 13(1), 72 (2012)
22. Hoerl, A.E.: Application of ridge analysis to regression problems. *Chemical Engineering Progress* 58, 54–59 (1962)
23. Kohannim, O., et al.: Boosting power to detect genetic associations in imaging using multi-locus, genome-wide scans and ridge regression. *Biomedical Imaging: From Nano to Macro IEEE 2011* (2011) (EPUB)
24. Chen, J., Chen, Z.: Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95(3), 759–771 (2008)
25. Langers, D.R., et al.: Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate. *NeuroImage* 38(1), 43–56 (2007)

Sparse Representation of Group-Wise fMRI Signals

Jinglei Lv^{1,2}, Xiang Li², Dajiang Zhu², Xi Jiang², Xin Zhang^{1,2}, Xintao Hu¹,
Tuo Zhang^{1,2}, Lei Guo^{1,2}, and Tianming Liu²

¹ School of Automation, Northwestern Polytechnical University, Xi'an, China,

² Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA, USA

Abstract. The human brain function involves complex processes with population codes of neuronal activities. Neuroscience research has demonstrated that when representing neuronal activities, sparsity is an important characterizing property. Inspired by this finding, significant amount of efforts from the scientific communities have been recently devoted to sparse representations of signals and patterns, and promising achievements have been made. However, sparse representation of fMRI signals, particularly at the population level of a group of different brains, has been rarely explored yet. In this paper, we present a novel group-wise sparse representation of task-based fMRI signals from multiple subjects via dictionary learning methods. Specifically, we extract and pool task-based fMRI signals for a set of cortical landmarks, each of which possesses intrinsic anatomical correspondence, from a group of subjects. Then an effective online dictionary learning algorithm is employed to learn an over-complete dictionary from the pooled population of fMRI signals based on optimally determined dictionary size. Our experiments have identified meaningful Atoms of Interests (AOI) in the learned dictionary, which correspond to consistent and meaningful functional responses of the brain to external stimulus. Our work demonstrated that sparse representation of group-wise fMRI signals is naturally suitable and effective in recovering population codes of neuronal signals conveyed in fMRI data.

Keywords: DTI, Task-based fMRI, Sparse coding.

1 Introduction

The human brain function intrinsically involves complex processes with population codes of neuronal activities [1-2, 4]. In the neuroscience community, a large amount of research has supported that when determining neuronal activity, sparse population coding is an effective exploration [3]. For example, the primary visual cortex V1 receives image signals with a sparse set of sensory neurons [2], and similarly, the middle temporal lobe (MTL) neurons fire selectively to visual stimulus [4]. In other words, a sparse set of neurons encode specific concepts rather than responding to the input independently [3]. Inspired by these findings, significant amount of research efforts from the machine learning and pattern recognition fields has been recently devoted to sparse representations of signals and patterns, and remarkable achievements have been made [5].

In general, fMRI neuroimaging takes the advantage of the coupling between neuronal activities and hemodynamics in the human brain, and thus fMRI signals, in principle, represent the population codes of large-scale neuronal activities. Given the remarkable successes of sparse representations in the machine learning and pattern recognition fields [5], in which the achievements were originally inspired and motivated by brain sciences discoveries, it is natural and well-justified to explore sparse representation of fMRI signals and the associated brain activity patterns. In the literature, there have been several pioneering efforts along this direction. For instance, a data-driven sparse coding fMRI analysis approach with K-SVD method and general linear model were developed to extract more accurate individually adaptive activation patterns in [6]. In [7], the authors used the Fisher Discriminative Dictionary learning (FDDL) method to cluster and differentiate functional brain states in resting and under task performance based on resting state fMRI and task-based fMRI datasets [7]. However, there have been very few studies that aim to examine the sparse representation of fMRI signals at the population level and to investigate how atoms in the learned sparse dictionary correspond to meaningful functional brain responses.

To address the above questions, in this paper, we design and apply a novel group-wise sparse representation framework for task-based fMRI signals from multiple subjects via dictionary learning methods. Specifically, we employ an effective online dictionary learning method [8] to learn an over-complete dictionary for group-wise sparse representation of the fMRI signals pooled from large-scale corresponding cortical landmarks of a group of subjects. For each subject, we adopted the publicly available DICCCOL (Dense Individualized and Common Connectivity-Based Cortical Landmarks) system [9] to locate 358 consistent cortical landmarks based on DTI data. Since the 358 DICCCOLs have been shown to possess intrinsic structural and functional correspondence across subjects [9], we extract fMRI signals for each of them across a group of subjects and then the pooled fMRI signals are used to learn a dictionary for group-wise sparse representations based on optimally determined dictionary size. The major advantage of using DICCCOL is that the small set of samples reduce the computing consume for group analysis, and the anatomical correspondence enable statistical exploration to the sparse representation. Applications of this novel framework on a working memory task-based fMRI dataset [10] have identified meaningful Atoms of Interests (AOI) in the learned dictionary, corresponding to consistent and meaningful functional responses of the brain to block-based external stimulus.

2 Materials and Methods

2.1 Overview

Our novel computational framework of group-wise sparse representation of fMRI signals is summarized in Fig.1. First, 358 corresponding DICCCOL landmarks (Fig.1a) are localized and optimized on DTI datasets of all subjects via the methods in [9]. For each subject, after the linear intra-subject image registration of DTI and fMRI data, we extract task-based fMRI signal for each landmark. Then, the signals of 358

landmarks from k subjects are arranged into a signal matrix $S \in \mathbb{R}^{t \times n}$ (Fig.1b) in order, where n columns are representative fMRI signals from the landmarks of k subjects ($n_1 = n_2 = \dots = n_k = 358$, $n = n_1 + n_2 + \dots + n_k = k \times 358$), and t is the fMRI volume number. By using an effective online dictionary learning and sparse coding method [8], each fMRI signal vector in S will be modeled as the linear combination of atoms of a learned dictionary D (Fig.1c), i.e. $s_i = D \times \alpha_i$ and $S = D \times \alpha$, where α is the coefficient matrix for sparse representation and each column α_i is the corresponding coefficient vector for s_i . Finally, as shown in Fig.1d, we optimize the dictionary size, identify meaningful Atoms of Interests (AOI), and perform statistics and interpretations on the sparse α matrix in the context of the correspondences of DICCCOL landmarks.

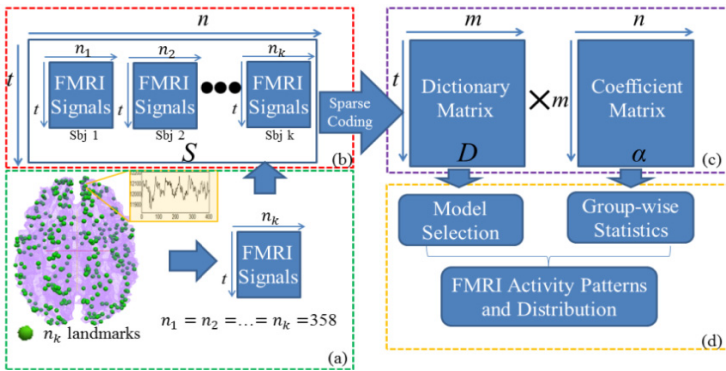


Fig. 1. The framework of sparse representation of group-wise fMRI signals

2.2 Data Acquisition and Pre-processing

In an IRB approved working memory task-based fMRI experiment [10], fMRI images of 19 subjects were scanned on a 3T GE Signa scanner. Briefly, acquisition parameters are as follows: fMRI: 64×64 matrix, 4mm slice thickness, 220mm FOV, 30 slices, $TR=1.5s$, $TE=25ms$, $ASSET=2$. Each participant performed a modified version of the OSPAN task (3 block types: OSPAN, Arithmetic, and Baseline) while fMRI data is acquired. DTI data was acquired with dimensionality $128 \times 128 \times 60$, spatial resolution $2mm \times 2mm \times 2mm$; parameters are $TR 15.5s$ and $TE 89.5ms$, with 30 DWI gradient directions and 3 B0 volumes acquired. Detailed task design and preprocessing steps are referred to [10]. For each subject, the 358 DICCCOL landmarks are localized and optimized via the computational framework in [9]. DTI images of each subject were registered into the fMRI space via FSL FLIRT. As landmarks were localized and optimized on the white matter cortical surfaces of DTI data, we first translate the landmark locations into voxels of DTI image, and with the registration of DTI and fMRI data, we then locate the landmarks on fMRI images and extract corresponding fMRI signal for each landmark. As our work majorly focused on the fluctuation of fMRI signals, we normalize each extracted signal to have zero mean and standard deviation of 1.

2.3 Dictionary Learning

In our approach, we aim to learn a meaningful and over-complete dictionary $D \in \mathbb{R}^{t \times m}$ ($m > t$, $m < n$) [8] for the sparse representation of S . For the signal set $S = [s_1, s_2, \dots, s_n] \in \mathbb{R}^{t \times n}$, the empirical cost function is summarized as Eq.(1) considering the average loss of regression to n signals.

$$f_n(D) \triangleq \frac{1}{n} \sum_{i=1}^n \ell(s_i, D), \tag{1}$$

With the aim of sparse representation using D , the loss function is defined in Eq.(2) with a ℓ_1 regularization that yields to a sparse resolution of α_i , and here λ is a regularization parameter to trade off the regression residual and sparsity level.

$$\ell(s_i, D) \triangleq \min_{\alpha_i \in \mathbb{R}^m} \frac{1}{2} \|s_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \tag{2}$$

As we majorly focus on the fluctuation shape of basis fMRI activity and also in order to prevent D from arbitrarily large values, columns d_1, d_2, \dots, d_m are constrained with Eq.(3).

$$C \triangleq \{D \in \mathbb{R}^{t \times m} \text{ s.t. } j = 1, \dots, m, \quad d_j^T d_j \leq 1\} \tag{3}$$

$$\min_{D \in C, \alpha \in \mathbb{R}^{m \times n}} \frac{1}{2} \|S - D\alpha\|_F^2 + \lambda \|\alpha\|_{1,1} \tag{4}$$

In summary, the whole problem can be rewritten as a matrix factorization problem in Eq.(4) [8], and we use the effective online dictionary learning method in [8] to learn the dictionary for sparse representation of group-wise fMRI signals on corresponding DICCCOL landmarks.

2.4 Sparse Representation and Group-Wise Analysis

With sparse representation, the most relevant bases of fMRI activity are selected and linearly combined to represent the original fMRI signals. The Orthogonal Matching Pursuit algorithm [11] is employed to represent a given signal using the dictionary learned D with the ℓ_0 regularization $\|\alpha_i\|_0 \leq L$ [11], as present in Eq.(5), where L is the largest number of atoms in D that are allowed to use in regression .

$$\min_{\alpha_i \in \mathbb{R}^m} \|s_i - D\alpha_i\|_2^2, \text{ s.t. } \|\alpha_i\|_0 \leq L \tag{5}$$

Afterwards, for each signal in S , we obtain a corresponding coefficient vector which is a column of α in Fig.1 and Fig.2. The matrix of α is decomposed into k sub-matrices for DICCCOL landmarks of each subject, as illustrated in Fig.2. The element (i,j) in each

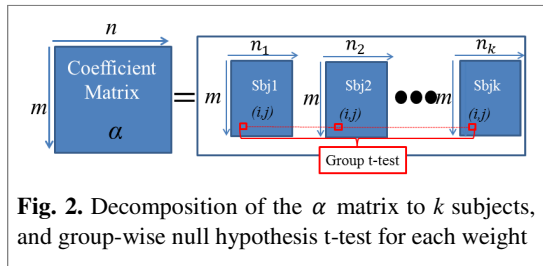


Fig. 2. Decomposition of the α matrix to k subjects, and group-wise null hypothesis t-test for each weight

sub-matrix stores the weight coefficient to the i^{th} atom in the dictionary of the j^{th} DICCCOL landmark (Fig.2). With a group-wise null hypothesis t -test, we determine the non-zero reference (i,j) as a group-wise effective landmark reference to the i^{th} atom of the j^{th} landmark. Then, with the statistics of global reference rates and

group-wise effective landmark reference rate, we select the most popular dictionary atoms and map them to the DICCCOL landmark atlases for further analysis.

In our proposed pipeline, we train dictionary with a weak l_1 -norm (λ in Eq.(4) is set with a small value) in order to minimize the representation residual and guarantee the atom signals' quality in D . While in the representation step, we choose a strong l_0 -norm (L in Eq.(5) is set with a small integer) with the purpose of discriminating ROIs using the same smallest number of explanatory atoms from the learned D . In this way, we trade off the reliability of D and discriminative capability of α .

3 Experimental Results

3.1 Dictionary Size Selection

With the online dictionary learning method [8], we learned the dictionary for sparse representation of our signal dataset. Here, the parameter λ is set with a small value of 0.15 in order to learn dictionary with relatively less representation residual [8]. Note that we tried different settings of λ from 0.5-2.0 with a fixed dictionary size, and the learned dictionaries are quite similar in term of corresponding atom signals' shape. There are no common criteria to determine the size of learned dictionary for the online dictionary learning method. But a good dictionary should be efficient while being not overfitting [8]. In term of efficiency, the representation residual should be minimized, and meanwhile the atoms should be less correlated to avoid overfitting. As discussed in Section 2.3, the appropriate dictionary size m would be $m > t, m < n$ ($t=270, n=358 \times 19=6802$). Here, to select the optimal dictionary for our problem, we repeated the training with the m ranging from 300 to 900 with an interval of 50.

For each learned dictionary, we define the average representation residual as

$$R = \frac{1}{n} \|S - DA\|_F^2 \tag{6}$$

where A is the regression coefficient matrix of S using D , calculated with the method described in Section 2.4, and here we set the sparsity constraint $L=m$ to minimize R in each repeat. As plotted in Fig.3a, it is obvious that R is monotonically decreasing with larger dictionary size.

As the DICCCOL landmarks possess intrinsic structural and functional correspondence, we measure the consistency of representation for signals of each corresponding landmark. For each cortical landmark i , we collect the corresponding coefficient vectors from k subjects in A , as $a_i = [a_{i1}, a_{i2}, \dots, a_{ik}]$, $a_{ij} \in \mathbb{R}^m$. And their consistency is measured by the Cronbach's α [12] as

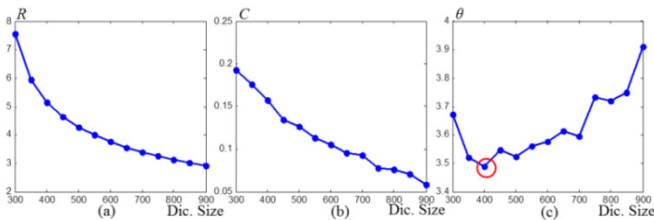


Fig. 3. The plotted values of R , C and θ with dictionary size vary from 300 to 900

$$c_i = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \sigma_{a_{ij}}^2}{\sigma_{a_i}^2} \right) \tag{7}$$

The average consistency of all DICCCOL landmarks is:

$$C = \frac{1}{n_k} \sum_{i=1}^{n_k} c_i, (n_k = 358) \tag{8}$$

As shown in Fig.3b, C also decreases with increasing of the dictionary size.

In our approach, considering the correspondence of DICCCOL landmarks as a group-wise regularization, the C is supposed to be higher, while R is supposed to be lower. In order to trade-off the two criteria, we define

$$\theta = \ln R - \ln C \tag{9}$$

to select optimal dictionary size with the lowest θ value. In Fig.3c, we select the best dictionary size $m=400$, with the lowest θ value highlighted with a red circle.

3.2 Sparse Representation and Evaluation

We represented each fMRI signal in S with the sparsity constraint of $L=20$ using the method in Section 2.4. Statistics was performed on the global reference rate for each atom in the learned D , which was the signal number in S that refers to the atom, as shown in Fig.4a. We also performed statistics on the group-wise effective landmark reference rate in Section 2.4, which was the landmark number that possesses group-wise reference to each atom in Fig.4b. It is striking that in both Fig.4a and Fig.4b, we found 4 corresponding peaks that have both high global reference rate and group-wise landmark reference rate. These atoms are thus selected as Atoms of Interest (AOI). Notably, we tried different sparsity constraints of $L=10, 20, 30$ and 40 , and it is interesting that the peaks AOIs are consistently the same, suggesting the robustness of

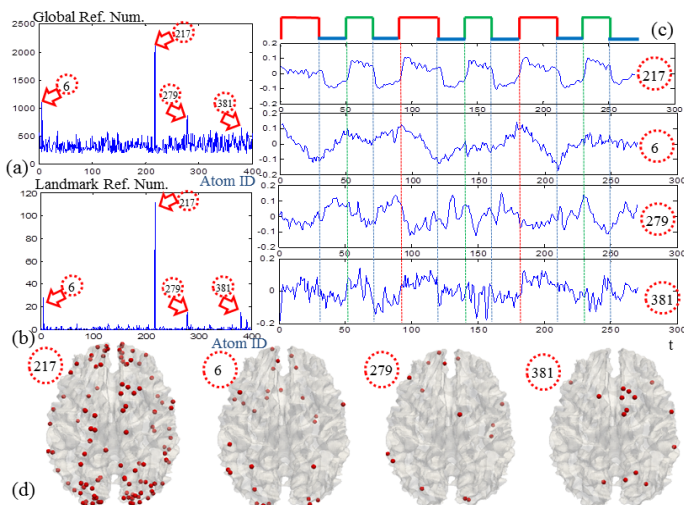


Fig. 4. (a) Global reference numbers of 400 atoms in the learned dictionary. (b) Group-wise landmark reference numbers of 400 atoms. (c) Atom shapes of 4 peaks in (a) and (b). (d) The landmark distributions with reference of the 4 AOIs respectively.

our methods. Also, the shapes of AOIs are compared with the task stimulus curve in Fig.4c. It is striking that all of the AOI shapes appear to be task-related response patterns. For instance, the AOIs #217 and #279 correspond to the task and anti-task responses; the AOI #6 exhibits half of the frequency of AOI #217; while AOI #381 shows similar global shape as AOI #217 but with much higher frequency fluctuations.

Afterwards, the DICCCOL landmarks that possess reference to each AOI in Fig. 4c are mapped to a cortical surface, respectively, in Fig.4d. These DICCCOLs were selected as landmarks of interest (LOI) that are involved in working memory task. Among all of the LOIs in Fig. 4d, the DICCCOLs, that possess reference to only one AOI, are visualized in Fig.5a, and the ones possessing references to multiple AOIs are shown in Fig.5b-5c. As a result, we separated the LOIs in Fig. 4d into 10 groups depending on the combinations of AOIs they refer to, with the number of each group shown in the 2nd row of Table 1 and spheres marked in Figs. 5a-5c with different colors. It is clear that the brain's responses to block task are complex, e.g., the landmarks in Fig.5a are only involved in one functional activity, but the landmarks in Fig.5b-5c are involved in multiple responses. These results not only support the importance of sparse representation of fMRI signals, but also demonstrate the multiple functional roles possibly played by the same cortical landmark.

For additional comparisons, we performed group-wise activation detection using the FSL FEAT based on the GLM and mixed-effect model [13] to the fMRI volumes of the same fMRI dataset. We mapped the detected activations (p -value=0.05, z -value>3.0) to the cortical surface as yellow regions in Figs. 5d-5f. The 10 groups of selected DICCCOLs in Figs. 5a-5c were overlaid on the activation maps of Figs. 5d-5f respectively for visual comparisons. It is evident that some DICCCOLs with the AOI #217 are within the activation maps, but many of them are outside of the activation maps, even though these landmarks possess a response component very similar to the stimulus curve. This result suggests that the red spheres referring to the AOI #217 might be a substantially more sensitive way to detect activated brain landmarks than the traditional FSL FEAT, considering different levels of noise.

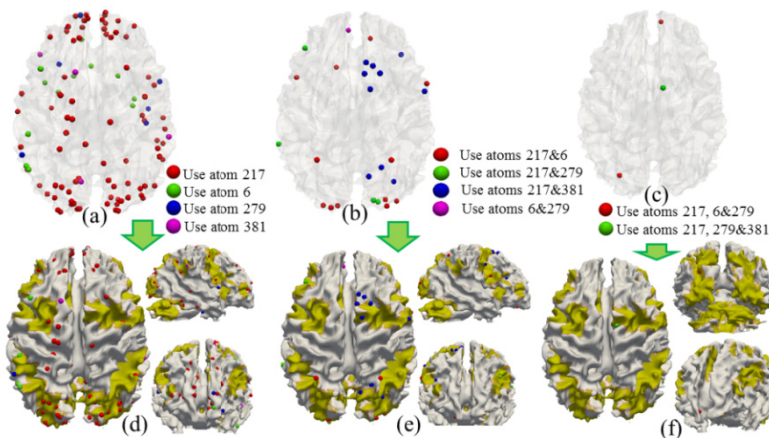


Fig. 5. (a) Landmarks with reference of only one of the 4 AOIs. (b) Landmarks with reference of two of the 4 AOIs. (c) Landmark with reference of three of the 4 AOIs. (d-f) are the activation maps detected by using group-wise activation detection (via FSL FEAT), overlaid with the landmarks in (a-c), respectively.

Table 1. The number of DICCCOLs of each group within the activation areas. The 2nd row is the total number of each DICCCOL group using different AOIs. The following four rows are the overlaid numbers with activations detected with different z-value thresholds.

Used AOIs	#217	#6	#279	#381	#217 & #6	#217 & #279	#217 & #381	#6 & #279	#217, #6 & #279	#217, #279 & #381
Total	83	13	7	4	12	4	10	1	2	1
$z=2.5$	37	1	3	2	9	3	6	0	1	1
$z=3.0$	33	1	2	2	9	3	6	0	1	1
$z=3.5$	23	1	1	2	7	3	5	0	1	0
$z=4.0$	14	0	1	1	6	3	5	0	1	0

For quantitative comparison, we selected activation foci with different z-value thresholds, and the DICCCOL numbers of the 10 groups that are within the yellow activation maps are presented in Table 1 (Rows 3-6). By comparing the total number of the 10 groups in the 2nd Row, we can see that even with a relatively low threshold (z-value=2.5), the traditional method by FSL FEAT can only detect a limited number of the selected DICCCOL landmarks that possess reference to AOI #217 with similar pattern as the stimulus curve. Also, it performs even poorer in detecting DICCCOLs that have other activity pattern (AOIs #6, #279 and #381). These results support that group-wise decoding of fMRI activity using sparse coding is much more robust to noise than GLM-based activation method, and is much more adaptive in decoding multiple task-related fMRI activity patterns.

4 Conclusion

We have described a novel group-wise sparse representation framework for task-based fMRI signals pooled via consistent DICCCOL landmarks, and demonstrated by extensive experiments that the framework can recover consistent and functionally meaningful atoms that represent population codes in task-based fMRI data. Our work demonstrated that sparse representation is effective in representing task-based fMRI signals and functional brain activity patterns. Thus, our work offers a promising general framework for representation and modeling of fMRI data.

References

1. Olshausen, B.A., Field, D.J.: Sparse coding of sensory inputs. *Current Opinion in Neurobiology* 14(4), 481–487 (2004)
2. Olshausen, B.A.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381(6583), 607–609 (1996)
3. Daubechies, I., et al.: Independent component analysis for brain fMRI does not select for independence. *Proceedings of the National Academy of Sciences* 106(26), 10415–10422 (2009)

4. Quian Quiroga, R., Kreiman, G., Koch, C., Fried, I.: Sparse but not ‘grandmother-cell’ coding in the medial temporal lobe. *Trends in Cognitive Sciences* 12(3), 87–91 (2008)
5. Wright, J., et al.: Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE* 98(6), 1031–1044 (2010)
6. Lee, K., Tak, S., Ye, J.C.: A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion. *IEEE Transactions on Medical Imaging* 30(5), 1076–1089 (2011)
7. Zhang, X., et al.: Characterization of Task-Free/Task-Performance Brain States. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 237–245. Springer, Heidelberg (2012)
8. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research* 11, 19–60 (2010)
9. Zhu, D., et al.: DICCCOL: Dense Individualized and Common Connectivity-Based Cortical Landmarks. *Cerebral cortex* (2012), <http://dicccol.cs.uga.edu/>
10. Faraco, C.C., et al.: Complex span tasks and hippocampal recruitment during working memory. *NeuroImage* 55(2), 773–787 (2011)
11. Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41(12), 3397–3415 (1993)
12. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. *Psychometrika* 16(3), 297–334 (1951)
13. Beckmann, C.F., Jenkinson, M., Smith, S.M.: General multilevel linear modeling for group analysis in FMRI. *Neuroimage* 20(2), 1052–1063 (2003)

Anatomy-Guided Discovery of Large-Scale Consistent Connectivity-Based Cortical Landmarks

Xi Jiang^{1,*}, Tuo Zhang^{2,1,*}, Dajiang Zhu¹, Kaiming Li³, Jinglei Lv^{2,1},
Lei Guo², and Tianming Liu¹

¹ Cortical Architecture Imaging and Discovery Lab, Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA, USA

² School of Automation, Northwestern Polytechnical University, Xi'an, China

³ Biomedical Imaging Technology Center,

Emory University/Georgia Institute of Technology, Atlanta, GA, USA

Abstract. Establishment of structural and functional correspondences across different brains is one of the most fundamental issues in the human brain mapping field. Recently, several multimodal DTI/fMRI studies have demonstrated that consistent white matter fiber connection patterns can predict brain function and represent common brain architectures across individuals and populations, and along this direction, several approaches have been proposed to discover large-scale cortical landmarks with common structural connection profiles. However, an important limitation of previous approaches is that the rich anatomical information such as gyral/sulcal folding patterns has not been incorporated into the landmark discovery procedure yet. In this paper, we present a novel anatomy-guided discovery framework that defines and optimizes a dense map of cortical landmarks that possess group-wise consistent anatomical and fiber connectional profiles. This framework effectively integrates reliable and rich anatomical, morphological, and fiber connectional information for landmark initialization, optimization and prediction, which are formulated and solved as an energy minimization problem. Validation results based on fMRI data demonstrate that the identified 555 cortical landmarks are producible, predictable and exhibit accurate structural and functional correspondences across individuals and populations, offering a universal and individualized brain reference system for neuroimaging research.

Keywords: DTI, fMRI, anatomical, connectivity, cortical landmarks.

1 Introduction

Establishment of structural and functional correspondences across different brains is one of the most fundamentally important issues in the brain imaging field. Current popular approaches to establishing the correspondences of brains regions across individuals can be broadly classified into three categories: image registration algorithms [1-3], cortical parcellation [4, 5], and manual/semi-automatic regions of interests

* These authors contributed equally to this work.

(ROI) analysis [6, 7]. Although these methods have their own advantages and have been successfully applied in different scenarios [1-7], they are limited due to the lack of quantitative representation of the regularity and variability of brain structure and function. In recognition of this limitation, recent literature studies have proposed to automatically define and discover common and consistent brain ROIs/landmarks with intrinsic structural/functional correspondences in a group of subjects or populations [8-11]. The underlying neuroscience basis is that consistent white matter fiber connection patterns can be used as common landmarks to predict brain function [8-12]. However, those approaches [8-11] also have limitations in that they do not consider rich and meaningful anatomic information of human brains and the accuracy of the discovered brain landmarks can be significantly improved.

In response to the above challenges, in this paper, we propose a novel framework for large-scale consistent connectivity-based cortical landmark discovery that defines and optimizes landmarks via integrating reliable and rich anatomical, morphological, and connectional information. These meaningful constraints have been used for landmark initialization, optimization and prediction based on multimodal DTI/fMRI datasets. The major novelty of our framework is that based on the predefined gyral/sulcal folding pattern homology, our landmark optimization and prediction simultaneously consider structural connection pattern similarity and homogeneity, and landmark spatial constraint, all of which are computationally formulated into a unified energy minimization problem. The proposed framework discovers 555 cortical landmarks that are consistent, reproducible, and predictable across individuals and populations, as demonstrated by extensive validations. Our results suggest that this set of 555 cortical landmarks effectively represents common cortical architectures and potentially provides opportunities for numerous applications in brain sciences.

2 Materials and Methods

2.1 Overview

The framework includes three major steps as shown in the red boxes in Fig.1. We perform landmark initialization and optimization under the guidance of four constraints (green boxes) for model brains, and conduct landmark prediction for testing brains. Details will be presented in the following sections.

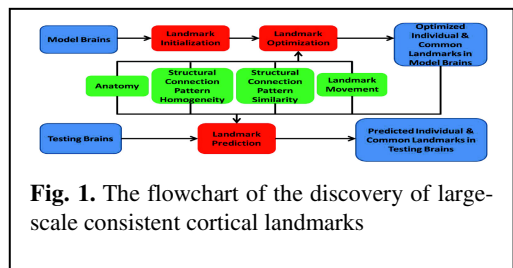


Fig. 1. The flowchart of the discovery of large-scale consistent cortical landmarks

2.2 Multimodal Data Acquisition, Preprocessing and Landmark Initialization

Two multimodal DTI/fMRI datasets were acquired and used in this work. In brief, dataset 1 includes DTI and four task-based fMRI scans (semantic decision making, emotion, empathy and fear networks) of eleven healthy young adults. Dataset 2

includes DTI and working memory task-based fMRI scans of twenty three healthy adult students. Imaging parameters and preprocessing steps of these datasets are referred to [10]. We randomly selected ten subjects from dataset 2 as the model brains. The other brains in datasets 1 and 2 were used as testing brains.

The landmark initialization procedure is as follows. First, we randomly selected one of the ten model brains as the template and other model brains were linearly registered to it via FSL FLIRT so that their global shape differences are removed and their cortical surfaces are in the same space for comparison. Second, for each corresponding major clearly identifiable gyrus/sulcus of each model brain according to the brain template used in the BrainVoyager Brain Tutor (<http://www.brainvoyager.com>), a certain number (ranging from 3-20) of landmarks were interactively labeled at cortical surface mesh vertices that are roughly distributed evenly along the gyral ridge/sulcal valley, and are sufficiently dense to ensure the full coverage of the whole gyral ridge/sulcal valley. In total, we manually labeled 594 landmarks for each model brain. It should be noted that during the following optimization step, the neighboring initialized landmarks that satisfy specific criteria are merged (Section 2.4). In this way, the number of initial

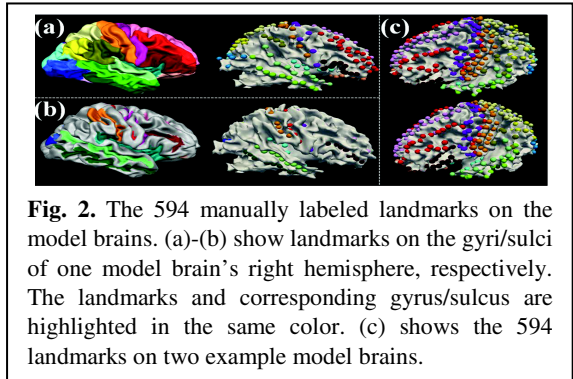


Fig. 2. The 594 manually labeled landmarks on the model brains. (a)-(b) show landmarks on the gyri/sulci of one model brain's right hemisphere, respectively. The landmarks and corresponding gyrus/sulcus are highlighted in the same color. (c) shows the 594 landmarks on two example model brains.

landmarks (594) is not a critical issue. Since the variability of folding pattern across subjects is huge, the manually initialization and homology determination of the 594 roughly corresponding cortical landmarks in the ten model brains based on the gyral/sulcal folding patterns effectively enforces the first anatomy constraint for landmark optimization. That is, corresponding cortical landmarks in different brains should be located on the same clearly identifiable gyrus or sulcus in order to preserve the same anatomical identity.

2.3 Structural Connection Pattern Similarity and Homogeneity Constraints

The second constraint for landmark optimization is that corresponding landmarks across ten model brains should possess similar structural connection patterns, which is represented by the DTI-derived fiber bundles emanating from the landmark. For each initialized landmark with extracted fiber bundles emanating around its neighbourhood, we used the trace-map model in [8], which is represented as a 144-dimensional vector, to quantitatively describe the structural connection pattern. Thus, the problem of quantitatively comparing the similarities across connections is converted to comparing the similarities across 144-dimensional trace-map vectors.

The third constraint for landmark optimization is that the corresponding landmarks across ten model brains should move toward to the location with local maximum of

structural connection pattern homogeneity, which is not considered in previous works [8-11]. The homogeneity is defined as the similarity between the trace-map representations of fiber bundles connected to the current ROI and its outside neighbouring ROIs, and calculated by the Kendall's coefficient of concordance [13]. Prior studies have demonstrated that the structural connection profile of a cortical ROI can be highly nonlinear, that is, a slight change to the location, size or shape to the ROI could significantly alters its fiber connection patterns [8, 11]. Therefore, this high nonlinearity can cause uncertainties and instabilities in the optimization and discovery of consistent and reproducible cortical landmarks.

In this paper, we examined the nonlinearity of the entire cortex of the ten model brains, and found that there are cortical areas with substantially less nonlinearity, or more homogeneity. By measuring the fiber pattern similarity via the trace-map model as mentioned above, we found that the fiber patterns extracted from the local maximum of homogeneity (red bubbles) across

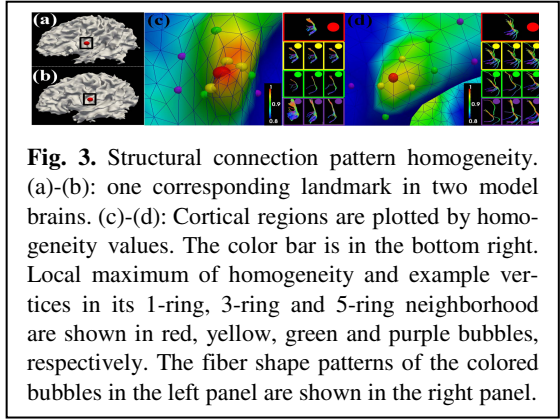


Fig. 3. Structural connection pattern homogeneity. (a)-(b): one corresponding landmark in two model brains. (c)-(d): Cortical regions are plotted by homogeneity values. The color bar is in the bottom right. Local maximum of homogeneity and example vertices in its 1-ring, 3-ring and 5-ring neighborhood are shown in red, yellow, green and purple bubbles, respectively. The fiber shape patterns of the colored bubbles in the left panel are shown in the right panel.

corresponding landmarks in ten model brains have higher similarity than those extracted from the vertices in outside neighbourhoods, such as those in the 1-ring, 3-ring, and 5-ring surface mesh neighbourhood (yellow, green and purple bubbles). Fig. 3 suggests that there are cortical regions with homogeneity peaks on the cortex, and importantly, these peaks exhibit quite consistent fiber connection patterns. Thus, the initialized landmarks across ten model brains should move toward to the cortical regions with homogeneity peaks within a neighborhood. The fourth constraint for landmark optimization is that the landmark should move within a neighborhood of the initial location with a predefined size.

2.4 Landmark Optimization

With the availability of initialized cortical landmarks in Section 2.2 and four meaningful constraints in Sections 2.2-2.3, we performed landmark optimization by searching all possible combinations of candidate landmark locations within their local morphological neighbourhoods in different model brains, and to seek the optimal solution with the optimal group-wise consistency and homogeneity. In this paper, we formulated and solved the landmark optimization problem by jointly modeling the four constraints. The goal is to minimize the group-wise variance of these jointly modeled profiles. Assume that there are M brains (i is i -th brain) and j is the current landmark that to be optimized. \tilde{k}_i^j is the initial location of j in brain i , and k_i^j is the candidate location in its neighborhood $C_{\tilde{k}_i^j}$ ($k_i^j \in C_{\tilde{k}_i^j}$). The maximum principal

curvature of k_i^j is represented by $p_{k_i^j} \begin{cases} \geq 0, \text{gyrus} \\ < 0, \text{sulcus} \end{cases}$ and it is used as the anatomical constraint in the following energy function. Mathematically, the group-wise variance is modeled as the energy E that we want to minimize as follows.

$$E(j) = \lambda_1 E_S(j) + \lambda_2 E_H(j) + \lambda_3 E_D(j), \quad k_i^j \in C_{\tilde{k}_i^j} \text{ and } p_{k_i^j} \begin{cases} \geq 0, \text{gyrus} \\ < 0, \text{sulcus} \end{cases} \quad (1)$$

where E_S is the structural connection pattern similarity constraint, E_H is the structural connection pattern homogeneity constraint, and E_D is the landmark movement constraint. Here, we have weights $\lambda_1 + \lambda_2 + \lambda_3 = 1$ ($0 \leq \lambda_1, \lambda_2, \lambda_3 \leq 1$).

First, $E_S(j)$ is defined to ensure that the corresponding landmark j across M brains have similar fiber bundle shape patterns as mentioned in Section 2.3.

$$E_S(j) = \text{var}(\{tr(k_1^j), tr(k_2^j), \dots, tr(k_i^j)\}) \quad (2)$$

where $tr(k_i^j)$ is a 144-dimension trace-map vector of k_i^j . $\text{var}(\cdot)$ is the variance.

Second, $E_H(j)$ is defined to ensure that the corresponding landmark j across M brains should move toward to the location with local maximum of homogeneity within $C_{\tilde{k}_i^j}$. We assume there are Q vertices in the candidate vertex's neighborhood $C_{\tilde{k}_i^j}$ and they are regarded as the objects to be ranked. Each of 144 dimensions of trace-map is considered as a judge [13] and the number of judges is denoted by P . Define object q is given the rank $r_{q,p}$ by judge p , t_k is the number of tied ranks in k -th of m groups of ties. The Kendall's coefficient of concordance [13] of landmark j of brain i is defined as

$$W_i^j(tr(k_i^j)) = \frac{12 \sum_{q=1}^Q (\sum_{p=1}^P r_{q,p})^2 - 3P^2Q(Q+1)^2}{P^2(Q^3 - Q) - P \sum_{k=1}^m (t_k^3 - t_k)} \quad (3)$$

$$E_H(j) = \sum_{i=1}^M (1 - W_i^j(tr(k_i^j))) \quad (4)$$

Third, $E_D(j)$ is defined to ensure the landmark j moves within $C_{\tilde{k}_i^j}$ where $p_{k_i^j} \begin{cases} \geq 0, \text{gyrus} \\ < 0, \text{sulcus} \end{cases}$ during the optimization procedure.

$$E_D(j) = \sum_{i=1}^M \text{dist}(k_i^j, \tilde{k}_i^j) \quad (5)$$

where $\text{dist}(\cdot)$ is the Euclidean distance between k_i^j and \tilde{k}_i^j on the cortical surface.

The energy minimization is solved as follows. For each iteration, by searching the whole-space of landmarks candidate locations in different model brains for one corresponding landmark, we can find an optimal combination of landmark locations that minimizes E (Eq. (1)). The convergence criterion is that the distance of landmark locations between two consecutive iterations is less than ε ($\varepsilon = 2$ mm, since the distance between two adjacent surface mesh vertices is about 2 mm). Notably, for each

iteration, if the distance between two neighboring landmarks to be optimized is less than or equal to a threshold t_d ($t_d=2$ mm, since the distance between two adjacent surface mesh vertices is about 2 mm) across $p\%$ (here $p=80$) of all model subjects, we label these two landmarks as ‘merged’ and only optimize one of them in the next iteration. In our implementation, we considered about 30 candidate locations (3-ring neighborhood) for each landmark. We tested different combinations of λ_1, λ_2 and λ_3 in Eq. (1) and chose the ones with best optimization results (for gyri, $\lambda_1 = 0.5, \lambda_2 = 0.4$ and $\lambda_3 = 0.1$ and for sulci, $\lambda_1 = 0.6, \lambda_2 = 0.3$ and $\lambda_3 = 0.1$). Other efficient approaches for energy minimization may be considered in the future.

2.5 Landmark Determination and Prediction

To examine and ensure the reproducibility of the discovered cortical landmarks, we divided the ten model brains into two groups and performed landmark optimization separately. As a result, two independent groups of optimized cortical landmarks were obtained. Then, for each optimized landmark in all of the ten brains in two groups, we evaluated the consistency of landmarks using both quantitative (trace-map similarity [8]) and qualitative (visual inspection) methods the same as in [10]. In brief, for each corresponding landmark, we calculated its mean trace-map distance [8], which measures the similarity of fiber shape patterns, between two groups and adopted the same criterion in [10] to verify if the landmark was similar across groups of brains [8, 10]. Furthermore, we used in-house batch visualization tool [10] to visually check the fiber patterns in all model brains of two groups. If the landmark in any of the ten model brains has substantially different trace-map distance value, and is confirmed to have different fiber shape patterns by visual inspection, this landmark is discarded. Finally, we retained 555 landmarks which exhibit consistent fiber connection patterns across all ten model brains.

Based on the 555 landmarks in the ten model brains, we predicted them in other testing brains (Section 2.2). Given a testing brain, we first mapped the 555 landmarks of one model brain to the testing brain via linear image registration (FSL FLIRT) to have rough locations. Then we optimized the locations of the 555 landmarks on the testing brain by minimizing the energy E (Eq. (1)) across ten model brains and the testing brain. The landmark prediction procedure is fast and efficient (about 15 minutes for one testing brain on a typical PC).

3 Experimental Results

3.1 Reproducibility and Predictability of 555 Cortical Landmarks

Figs. 4a-4b show the 555 landmarks (yellow bubbles) in three model brains and three testing brains, respectively. As an example, we randomly selected 5 landmarks and visualized their fiber connectational patterns in the three model and three testing brains in the left and right panel of Fig. 4, respectively. Quantitatively, the average trace-map distances of the corresponding landmarks across model and testing brains are 2.08, 2.18, 2.15, 2.20 and 2.24, respectively, which are considered as quite low [8, 10]. By visual inspection, there is also no much difference among the fiber patterns of

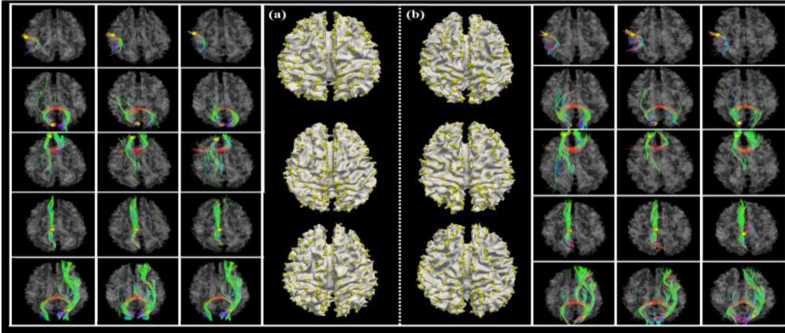


Fig. 4. The 555 landmarks (yellow bubbles) in (a) three model brains and (b) three testing brains, respectively. Five landmarks are randomly selected and their fiber connectional patterns in the model and testing brains are shown in the left and right panel, respectively.

the same corresponding landmark in all model and testing brains. Importantly, all of these 555 landmarks have been confirmed to possess the above-mentioned characteristics in all ten model brains and testing brains, indicating that the 555 landmarks represent a common structural brain architecture that is reproducible and predictable across different subjects.

3.2 Functional Annotation of Landmarks

We adopted the two fMRI datasets including five task-based scans in Section 2.2 to examine the functional roles and correspondences of the 555 DTI-derived cortical landmarks. The benchmark fMRI activation peaks were detected and selected using FSL FEAT as the functional locations. In total, we identified 69 functional locations from the five functional networks. For each functional location, we first identified five closest (Euclidean distance) cortical landmarks within each model brain as the candidates. Then, the cortical landmark with the most votes as the closest to the functional location across all ten model brains was annotated by the functional location. In total, 46 cortical landmarks were annotated. The same cortical landmark is truly in the specific functional network across all subjects,

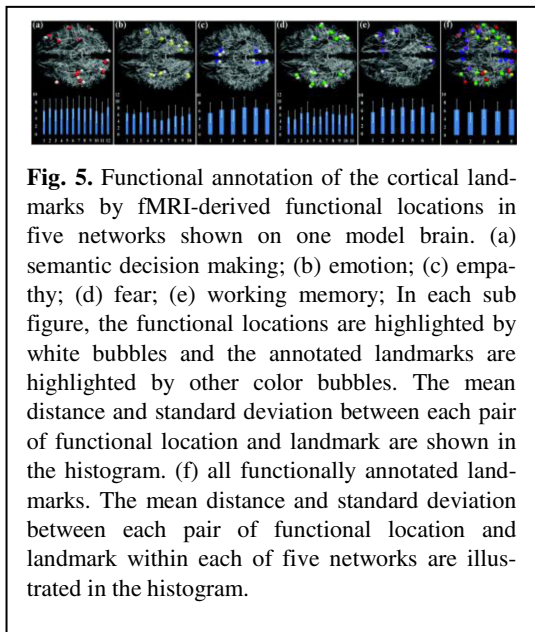


Fig. 5. Functional annotation of the cortical landmarks by fMRI-derived functional locations in five networks shown on one model brain. (a) semantic decision making; (b) emotion; (c) empathy; (d) fear; (e) working memory; In each sub figure, the functional locations are highlighted by white bubbles and the annotated landmarks are highlighted by other color bubbles. The mean distance and standard deviation between each pair of functional location and landmark are shown in the histogram. (f) all functionally annotated landmarks. The mean distance and standard deviation between each pair of functional location and landmark within each of five networks are illustrated in the histogram.

suggesting that these landmarks specifically align with regions of task-related processing. The mean distances for the five functional networks are 6.27 mm, 5.68 mm, 6.38 mm, 5.91 mm and 6.33 mm, respectively. On average, the distance is 6.11 mm. The results in Fig. 5 demonstrate the corresponding structural connectivity-based landmarks are consistently co-localized with the same functional regions, and reflect the common structural/functional cortical architectures that are reproducible across subjects. It should be noted that more cortical landmarks may be functionally annotated if more specific, large-scale fMRI tasks can be designed and performed in the future for the purpose of functional annotation of landmarks.

3.3 Comparisons with Image Registration Algorithms

Finally, we compared the functional annotation accuracies by our 555 landmarks and those by four different linear/nonlinear image registration algorithms (linear: FSL FLIRT; nonlinear: FSL FNIRT, ANTS, and HAMMER). The working memory functional locations in Fig. 5 were used as benchmarks. The average annotation errors by the five methods (our landmarks, FSL FLIRT, FSL FNIRT, ANTS and HAMMER) are 6.33 mm, 7.76 mm, 8.01 mm, 7.74 mm, and 7.73 mm respectively, indicating that our landmarks have superior functional annotation accuracy than those four image registration algorithms.

4 Conclusion

In this paper, we presented a novel anatomy-guided discovery framework to define and optimize a dense map of cortical landmarks that possess group-wise consistent anatomical/connectional profiles. Extensive experiments demonstrated its reproducibility and predictability. Furthermore, a validation study via task-based fMRI was provided for a subset of the discovered landmarks, suggesting the accurate structural and functional correspondences of these landmarks across individuals and populations. In the future, we will compare our 555 landmarks with the available DICCCOL system [10]. Other possible future studies will apply this dense map of 555 landmarks as a universal and individualized brain reference system on clinical datasets for connectivity analysis and mapping human brain connectomes.

References

1. Thompson, P., Toga, A.W.: A surface-based technique for 1336 warping 3-dimensional images of the brain. *IEEE Trans. Med. Imaging* 15(4), 402–417 (1996)
2. Fischl, B., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3), 341–355 (2002)
3. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21(11), 1421–1439 (2002)
4. Johansen-Berg, et al.: Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 101(36), 13335–13340 (2004)

5. Jbabdi, S., et al.: Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. *NeuroImage* 44(2), 373–384 (2009)
6. Poldrack, R.A.: The future of fMRI in cognitive neuroscience. *NeuroImage* (2011), doi:10.1016/j.neuroimage.2011.08.007
7. Liu, T.: A few thoughts on brain ROIs. *Brain Imaging and Behavior* (2011), doi:10.1007/s11682-011-9123-6
8. Zhu, D., et al.: Discovering Dense and Consistent Landmarks in the Brain. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011. LNCS*, vol. 6801, pp. 97–110. Springer, Heidelberg (2011)
9. Zhang, T., et al.: Predicting functional cortical landmarks via DTI-derived fiber shape models. *Cerebral Cortex* (2011)
10. Zhu, D., et al.: DICCCOL: Dense Individualized and Common Connectivity-based Cortical Landmarks. *Cerebral Cortex* (2012), doi:10.1093/cercor/bhs072
11. Li, K., et al.: Individualized ROI Optimization via Maximization of Group-wise Consistency of Structural and Functional Profiles. In: *Advances in Neural Information Processing Systems*, NIPS (2010)
12. Honey, C.J., et al.: Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 106(6), 2035–2040 (2009)
13. Siegel, S., Castellan Jr., N.J.: *Nonparametric Statistics for the Behavioral Sciences*, 2nd edn., p. 266. McGraw-Hill, New York (1988)

Sparse Representation of Higher-Order Functional Interaction Patterns in Task-Based fMRI Data

Shu Zhang^{1,2}, Xiang Li¹, Jinglei Lv², Xi Jiang¹, Dajiang Zhu¹, Hanbo Chen¹,
Tuo Zhang², Lei Guo², and Tianming Liu¹

¹ Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA, USA

² School of Automation, Northwestern Polytechnical University, Xi'an, China

Abstract. Traditional task-based fMRI activation detection methods, e.g., the widely used general linear model (GLM), assume that the brain's hemodynamic responses follow the block-based or event-related stimulus paradigm. Typically, these activation detections are performed voxel-wise independently, and then are usually followed by statistical corrections. Despite remarkable successes and wide adoption of these methods, it remains largely unknown how functional brain regions interact with each other within specific networks during task performance blocks and in the baseline. In this paper, we present a novel algorithmic pipeline to statistically infer and sparsely represent higher-order functional interaction patterns within the working memory network during task performance and in the baseline. Specifically, a collection of higher-order interactions are inferred via the greedy equivalence search (GES) algorithm for both task and baseline blocks. In the next stage, an effective online dictionary learning algorithm is utilized for sparse representation of the inferred higher-order interaction patterns. Application of this framework on a working memory task-based fMRI data reveals interesting and meaningful distributions of the learned sparse dictionary atoms in task and baseline blocks. In comparison with traditional voxel-wise activation detection and recent pair-wise functional connectivity analysis, our framework offers a new methodology for representation and exploration of higher-order functional activities in the brain.

Keywords: task-based fMRI, GES, sparse coding, dictionary learning, higher-order interaction.

1 Introduction

Voxel-based fMRI activation detection has been widely adopted in the functional brain mapping field. For instance, the general linear model (GLM) [1] is often used to detect activated voxels in task-based fMRI data, and followed by statistical corrections of the detected foci. However, voxel-wise activation detection methods have their limitations in terms of revealing the complex functional interaction patterns, since the brain often functions a network behavior. In recognition of this limitation, recently, several new studies have examined the functional connectivities during task performance in task fMRI data. The authors in [2] proposed a fiber-centered

activation detection method to find the activated connectivity patterns and the results demonstrated activated fiber-connected regions covered substantially wider brain areas than the traditional voxel-based activation methods. Another recent literature study in [3] examined the temporal dynamics of functional connectivity during task performance and found that the whole-brain's functional connectivity pattern well correlated with the block-based stimulus curve [3]. The results in [2, 3] suggested the feasibility and promise of examining functional connectivity patterns in task-based fMRI.

However, all the above-mentioned method [1, 2, 3] are still constrained in terms of the lack of quantitative representation of higher-order functional brain responses, and multivariate functional interaction patterns within brain networks are omitted in these methods. Essentially, both basic neuroscience research and computational modeling of neuroimaging data have proved that brain functions are typically realized via higher-order functional interactions among specific networks [4, 5]. In the literature, there are several other methods published to deal with higher-order interactions among multiple ROIs (regions of interests) such as independent component analysis [6], Granger causality modeling [7], dynamic causal modeling [8] and Bayesian graphical models [4, 5]. In particular, the Bayesian graphical causal models are proven to exhibit superior performance on estimating the network structure [4, 5] in both simulated and real data. Conceptually, Bayesian models are based on marginal and conditional probabilistic dependencies, which determines this method more suitable in estimating the network structure and less sensitive to the noises in fMRI signals [4]. However, only Bayesian models is not enough, the inferred interaction patterns in both baseline and task blocks among the whole-brain structural connectomes could be potentially overlapping with each other, their temporal transitions could be gradual.

In this paper, our novel algorithmic framework offers a new methodology to explore and represent higher-order functional activities in task-based fMRI data. Specifically, the greedy equivalence search (GES) algorithm [9] is employed to infer multivariate functional interactions for both task performance and baseline blocks. Then, an effective online dictionary learning algorithm [10] is utilized for sparse representation of the inferred higher-order functional interaction patterns obtained by GES. Both of the GES algorithm [9] and online dictionary learning algorithm [10] are not new to the world, but the novelty of this paper lies in its integration of both methods into an effective framework for modeling task-based fMRI data. Importantly, applying the proposed method on a working memory task-based fMRI data [11] has revealed distinguished and interesting of the learned sparse dictionary atoms in task and baseline blocks.

2 Materials and Methods

2.1 Overview

Based on the recently developed and publicly released DICCOL (Dense Individualized and Common Connectivity-based Cortical Landmarks) system [12], structural connectome is constructed (Fig.1a) and the higher-order functional interactions are

modeled on the relevant sub-networks (Fig.1b) of the connectome. Then, the fMRI signals from each subject split into blocks by following the task and baseline paradigm (Fig.1c). All these segmented blocks of fMRI signals then generate one functional interaction pattern inferred by GES (Fig.1d). Afterwards, all of these interaction patterns are arranged into one matrix, as the bases of online dictionary learning algorithm used for sparse coding (Fig.1e). Finally, the proposed methods are applied on an operational span (OSPAN) working memory task-based fMRI dataset [11]. The flow-chart is summarized in Fig. 1.

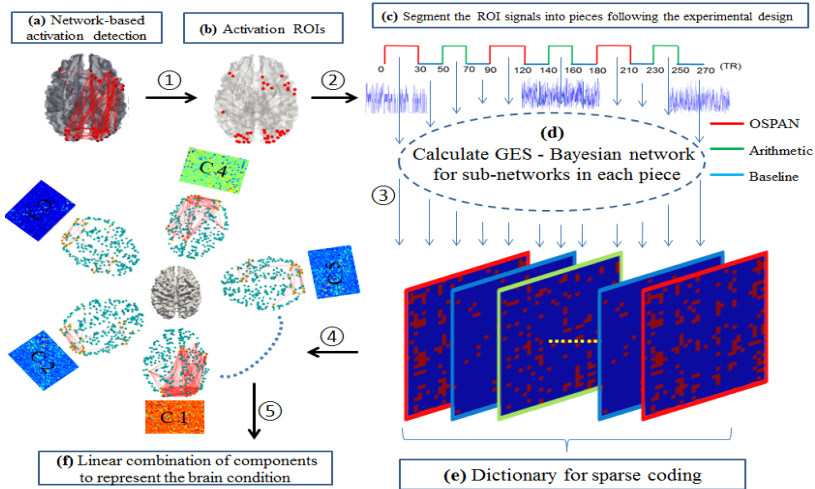


Fig. 1. The overview of sparse representation of higher-order functional interaction patterns in task-based fMRI data. The framework includes five main steps.

2.2 Data Acquisition and Pre-processing

In an operational span (OSPAN) working memory task-based fMRI experiment under IRB approval [11], 19 healthy young adult subjects were scanned and fMRI images were acquired on a 3T GE Sigma scanner. Briefly, acquisition parameters were taken as following: fMRI: 64×64 matrix, 4mm slice thickness, 220mm FOV, 30 slices, $TR=1.5s$, $TE=25ms$, $ASSET=2$. Each subject was performed by a modified version of the OSPAN task (3 block types: OSPAN, Arithmetic, and Baseline) while fMRI data were acquired [11]. DTI data was acquired with dimensionality $128 \times 128 \times 60$, spatial resolution $2mm \times 2mm \times 2mm$; parameters were $TR 15.5s$ and $TE 89.5ms$, with 30 DWI gradient directions and 3 B0 volumes acquired. More details about pre-processing can be referred to [11].

2.3 Bayesian Network Modeling and The GES Algorithm

Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a direct acyclic graph (DAG).

For instance, $D = (V, E)$ is a directed acyclic graph satisfying the causal condition, and thus Markov factorization of the distribution can be utilized. Here, V is a finite set of DAG nodes and E is a finite set of directed edges between the DAG nodes. As a measure of how well a DAG is, D represents the conditional independencies between the random variables, we can use the relative probability:

$$S(D) = p(D, d) = p(d|D)p(D) \quad (1)$$

where $S(D)$ refers to a network score [13], $p(d|D)$ calculates the probability for observing a given dataset d under a given graphical model D with un-weighted edges, where an “improved” model will have larger $p(d|D)$. Thus the direction obtained by GES is based on the fact that two graphical models with the same undirected connectivity but different directionality would have different $p(d|D)$, resulting in the different scores for $S(D)$. In this work, we adopt the GES algorithm [10] to infer the D from fMRI signals extracted for the relevant sub-networks of DICCCOL-based structural connectomes. Briefly, GES begins with an empty graph, and then each time it searches for one edge to add to the graph over the space of Markov equivalence, and it only stops when the graph is not further improved by adding any more edges. Then, GES starts to search backwards, and each time it removes one edge until no improvement occurs by removing any edge. Thus the converged network graph is obtained.

In this work, each fMRI signal is split into 12 temporal segments by following the task paradigm for the 19 subjects. It leads to 228 pieces of fMRI segments. Then, for each segment, GES is used to generate a functional interaction graph that can represent the current higher-order brain activities.

The GES toolbox we used is the publicly available TETRAD system [14], which is an effective tool for the GES algorithm. Until now, there are 228 GES-derived graphs obtained, among which 57 are for OSPAN blocks, 57 graphs for the Arithmetic blocks, and the last 114 graphs for the Baseline blocks. Two examples with different pattern of the GES graphs are shown in Fig. 2. Accordingly, higher-order functional interactions, instead of pair-wise connectivities, on the structural connectomes can be clearly appreciated.

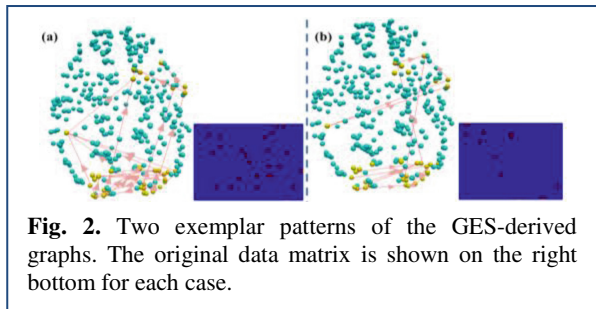


Fig. 2. Two exemplar patterns of the GES-derived graphs. The original data matrix is shown on the right bottom for each case.

2.4 Dictionary Learning and Sparse Coding

A large amount of recent studies in the machine learning field have demonstrated that sparse coding is superior in representing features and patterns. In this paper, we adopt the effective online dictionary learning algorithm [10] for sparse representation of the higher-order functional interaction patterns obtained in section 2.3. Specifically, for the problem interested, each GES graph matrix derived from section 2.3 with m_V rows and m_V columns is first reorganized into a vector x_i with the size

of $m = m_V \times m_V$, where m_V is the number of vertex in the graph. Then, given the intrinsically-established correspondences of DICCCOL-based connectomes across individual brains [12], we pool all the GES graphs from 19 subjects together to form a set of training samples $X = [x_1, x_2, \dots, x_i \dots x_n], x_i \in \mathbb{R}^m$ for dictionary learning. Thus, by using a learned dictionary $D \in \mathbb{R}^{m \times k}$, the aim is to represent each sample with a weight vector $\alpha_i \in \mathbb{R}^k$ that sparsely and linearly combine dictionary atoms, i.e., $x_i = D \times \alpha_i$. Here, the empirical cost function is defined:

$$f_n(D) \triangleq \frac{1}{n} \sum_{i=1}^n \ell(x_i, D) \quad (2)$$

in which the loss function to minimize the regression error of x_i with D is defined with a ℓ_1 regulation that seeks a sparse solution of α_i in Eq.(3). The parameter λ is a weight to control sparsity level. In our experiments, λ is all empirically set as 0.1 to minimize regression error residual.

$$\ell(x_i, D) \triangleq \min_{\alpha_i \in \mathbb{R}^k} \frac{1}{2} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \quad (3)$$

With the effective online dictionary learning method in [10], this problem can be solved using the publicly available SPAM-toolbox [10, 15]. Here, the learned dictionary D consists of k columns of components ($d_j \in \mathbb{R}^m$) are believed to be bases of GES graph patterns in the form of vectors. As our training dataset is not large with $n=228$, we empirically choose a relatively smaller dictionary size $k=10$. Further, in order to identify the most relative GES graph pattern atom for each sample x_i , a sparse coding approach is implemented via the fast Orthogonal Matching Pursuit algorithm [3] in the SPAM-toolbox. Specifically, the purpose is to minimize the representation error with a limited number of dictionary atoms, i.e., $\|\alpha_i\|_0 \leq L$, see Eq.(4), which used for sparse coding of the input GES graph with the constraint on the number of atoms used

$$\min_{\alpha_i \in \mathbb{R}^k} \|x_i - D\alpha_i\|_2^2, \quad s.t. \|\alpha_i\|_0 \leq L \quad (4)$$

In our experiments, the sparsity constrain L is set as 5, and in this way, each input GES graph pattern can be sparsely represented by 5 or less than 5 atoms in the learned dictionary. The sparse coding representation of the group-wise GES graph patterns can be used to discriminate different states of the brain during working memory task.

3 Experimental Results

3.1 Dictionary Learning Results

According to the algorithmic pipeline in Fig.1, first, group-wise activation detection is performed on 358 DICCCOLs via traditional GLM method, and 37 identified ROIs are most consistently activated in order to narrow down the number of ROIs concerned. Then, the 228 segments of fMRI signals within this sub-network of 37 ROIs are obtained via the online dictionary learning methods [10]. The size of the dictionary is empirically set as 10 in our application and the learned dictionary is visualized in Fig.3. It is interesting that the dictionary atom #2, #4 and #7 exhibit different levels of intense functional interactions within this sub-network, they are very important as

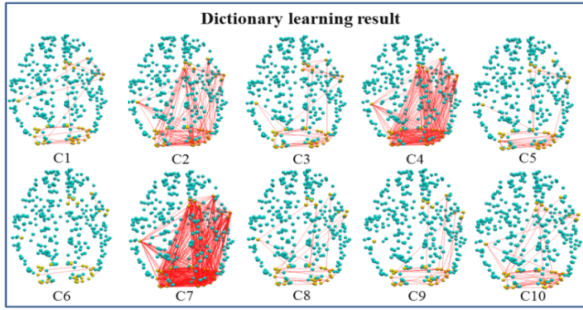


Fig. 3. The 10 learned dictionary atom components (C1-C10, respectively) are shown

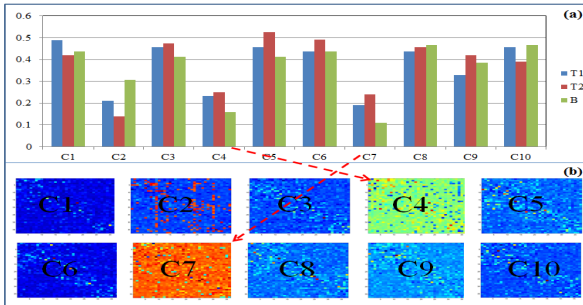


Fig. 4. (a) Frequency distributions of the 10 dictionary atoms in the three types of blocks. Here, T1 stands for OSPAN task, T2 stands for Arithmetic task, and B for Baseline. (b) Visualization of the 10 atoms (C1-C10) in the functional interaction matrices. The red color in the matrix represents higher functional interaction and the blue indicates lower functional interaction.

they have similar interaction patterns with the result from group-wise GLM, while other atoms show much less but variable functional interaction patterns. Moreover, they have similar patterns because the learning process needs more atoms to regress the dense (thus higher-weighted) interaction patterns. Though the frequencies of individual atoms are not high, their combined presence in task blocks is nearly 50%, which is really high in Fig.4a.

In addition, the 10 dictionary atoms are visualized in the matrix formats of functional interactions in Fig.4b and the statistical analysis of the distributions of these atoms is performed in three types of blocks as shown in Fig.4a. It is interesting that the two most active atoms #4 and #7 are substantially more frequent in both OSPAN and Arithmetic task blocks than in the Baseline block, as illustrated by the two red lines in Fig.4. This result is proven to be reasonable since task blocks exhibit more active interaction patterns. Also, for

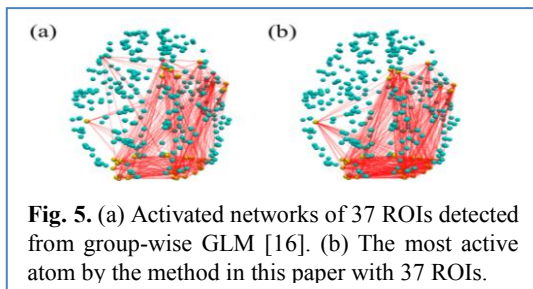


Fig. 5. (a) Activated networks of 37 ROIs detected from group-wise GLM [16]. (b) The most active atom by the method in this paper with 37 ROIs.

comparison purpose, the activated networks are visualized within the same set of 37 ROIs via our recently developed methods in [16] via group-wise GLM in Fig.5a. It is evident that the functional interaction pattern in atom #7 in Fig.5b exhibits similar activated pair-wise connections as those in Fig.5a, though there are more higher-order functional interactions in Fig.5b. This result not only suggests the validity of higher-order functional interaction modeling in this paper, but also demonstrates the superiority of GES-based inference of multivariate interaction patterns.

3.2 Reproducibility Study

As a reproducibility study of our methods, the z-value threshold was decreased when detecting the consistently activated DICCCOL landmarks via group-wise GLM, and thus 60 and 74 activated ROIs were obtained, respectively, to re-perform the experiments in section 3.1. Then we applied the methods in section 2 on these two experiments with different sizes of ROIs, and the results were shown in Fig.6. In both experiments, there were two active dictionary atoms, atoms #5 and #7 in Fig.6a and atoms #7 and #9 in Fig.6b. Again, these active atoms were substantially more frequent in both OSPAN and Arithmetic blocks than in the Baseline block. This result replicated the conclusion in section 3.1, suggesting the good reproducibility of our methods and results. Notably, for some inactive interaction patterns such as atom #10 in Fig.6a and atom #8 in Fig.6b, their frequencies were substantially higher in the Baseline block than in OSPAN and Arithmetic blocks, as illustrated by the green lines in Fig.6. This result further manifested the validity of our methods. What I want to emphasize is the atoms with same number in different experiment are independent.

For another comparison, we visualized the activated connections in the networks obtained by group-wise GLM methods and the interaction pattern of the most active atoms by our proposed methods with different numbers of selected ROIs in Fig.7. It was inspiring to observe more and more active connections from the left to the right in Fig.7, and the connections on the left side figure tended to be a sub-set of those on the right side figure. This result triggered us to re-think the limitations of traditional voxel-based [1] and connection-based [2, 3] activation detection methods. Hence, a large portion of the whole brain might be responsive to external tasks, such as the working memory in this paper, instead of a very small number of activated foci as detected by traditional voxel-based and connection-based methods [1-3].

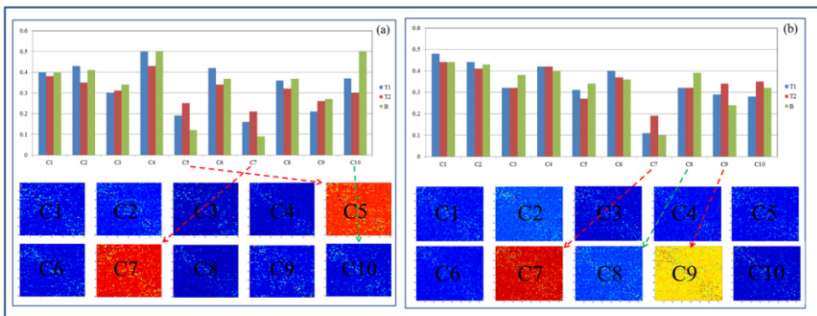


Fig. 6. Reproducibility study. In the charts, blue means OSPAN, red means Arithmetic and green means Baseline. (a) Results for 60 ROIs. (b) Results for 74 ROIs.

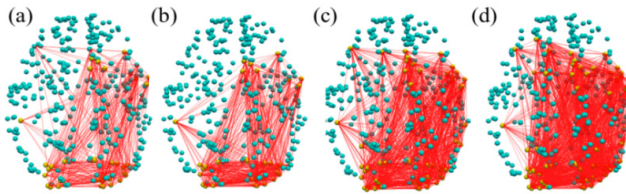


Fig. 7. Visualization of interaction patterns in (a) activated networks via group-wise GLM method, (b) most active atom with the ROI number of 37, (c) most active atom with the ROI number of 60, and (d) most active atom with the ROI number of 74

4 Conclusion

In this paper, in comparison with traditional voxel-wise activation detection and recent pair-wise functional connectivity analysis, a novel algorithmic framework was presented, which statistically inferred and sparsely represented higher-order functional interaction patterns within the working memory network during task performance and in the baseline. Experimental results on a working memory task-based fMRI data demonstrated meaningful, reproducible and interesting results. In the future, further evaluations and applying this framework on larger scale task-based fMRI datasets are to be planned.

References

1. Friston, K.J., et al.: Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* 2(4), 189–210 (1994)
2. Lv, J., Guo, L., Li, K., Hu, X., Zhu, D., Han, J., Liu, T.: Activated Fibers: Fiber-centered Activation Detection in Task-based FMRI. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011*. LNCS, vol. 6801, pp. 574–587. Springer, Heidelberg (2011)
3. Li, X., Lim, C., Li, K., Guo, L., Liu, T.: Detecting Brain State Changes via Fiber-Centered Functional Connectivity Analysis. *Neuroinformatics* (2012)
4. Sun, J., et al.: Inferring Consistent Functional Interaction Patterns from Natural Stimulus FMRI Data. *NeuroImage* (2012)
5. Ramsey, J., et al.: Six problems for causal inference from fMRI. *Neuroimage* 49(2), 1545–1558 (2010)
6. Calhoun, V.D., et al.: A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping* 14(3), 140–151 (2001)
7. Roebroeck, A., Formisano, E., Goebel, R.: Mapping directed influence over the brain using Granger Causality and fMRI. *NeuroImage* 25, 230–242 (2005)
8. Friston, K.J., et al.: Dynamic causal modelling. *NeuroImage* 19(3), 1273–1302 (2003)
9. Meek, C.: *Graphical Models: Selecting Causal and Statistical Models*. Carnegie Mellon University, Pittsburgh (1997)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: *Proceedings of the International Conference on Machine Learning, ICML (2009)*
11. Faraco, C.C., et al.: Complex span tasks and hippocampal recruitment during working memory. *NeuroImage* 55(2), 773–787 (2011)

12. Zhu, D., Li, K., Guo, L., Jiang, et al.: DICCCOL: Dense Individualized and Common Connectivity-Based Cortical Landmarks. *Cerebral Cortex* (2012)
13. Susanne, G., et al.: DEAL: A Package for Learning Bayesian Networks (2003)
14. Eberhardt, F., Hoyer, P.O., Scheines, R.: Combining Experiments to Discover Linear Cyclic Models with Latent Variables. *Journal of Machine Learning, Workshop and Conference Proceedings (AISTATS 2010)* 9, 185–192 (2010)
15. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research* 11, 19–60 (2010)
16. Shu, Z., et al.: Activated Cliques: Network-based Activation Detection in Task-based FMRI. In: *ISBI 2013* (2013)

Fusing Functional Signals by Sparse Canonical Correlation Analysis Improves Network Reproducibility

Jeffrey T. Duda¹, John A. Detre¹, Junghoon Kim²,
James C. Gee¹, and Brian B. Avants¹

¹ University of Pennsylvania

² Moss Rehabilitation Research Institute

Abstract. We contribute a novel multivariate strategy for computing the structure of functional networks in the brain from arterial spin labeling (ASL) MRI. Our method fuses and correlates multiple functional signals by employing an interpretable dimensionality reduction method, sparse canonical correlation analysis (SCCA). There are two key aspects of this contribution. First, we show how SCCA may be used to compute a multivariate correlation between different regions of interest (ROI). In contrast to averaging the signal over the ROI, this approach exploits the full information within the ROI. Second, we show how SCCA may simultaneously exploit both the ASL-BOLD and ASL-based cerebral blood flow (CBF) time series to produce network measurements. Our approach to fusing multiple time signals in network studies improves reproducibility over standard approaches while retaining the interpretability afforded by the classic ROI region-averaging methods. We show experimentally in test-retest data that our sparse CCA method extracts biologically plausible and stable functional network structures from ASL. We compare the ROI approach to the CCA approach while using CBF measurements alone. We then compare these results to the joint BOLD-CBF networks in a reproducibility study and in a study of functional network structure in traumatic brain injury (TBI). Our results show that the SCCA approach provides significantly more reproducible results compared to region-averaging, and in TBI the SCCA approach reveals connectivity differences not seen with the region averaging approach.

1 Introduction

Functional MRI (fMRI) is capable of measuring subject-specific and long-range correlations in brain activity (i.e. networks) as measured by changes in a direct or indirect time-series measurement of cerebral blood flow (CBF). EPI-BOLD is the standard protocol for studying network structure, however a second approach, arterial spin labeling (ASL) MRI, more directly measures CBF by tagging arterial blood and tracking changes in magnetization over time. ASL provides a quantitative measure of blood flow, which is believed to be more directly related to neuronal activity than the measure provided by EPI-BOLD [1]. One advantage

of ASL is greater signal stability and reproducibility when compared to EPI-BOLD especially over the range of resting state frequencies [2]. Additionally, the ASL acquisition contains images that exhibit BOLD contrast (ASL-BOLD) [1]. Although the temporal resolution of ASL is lower than EPI-BOLD, resting state fluctuations are thought to reside well within the range of frequencies that may be captured by ASL (0.01 Hz to 0.1 Hz).

While EPI-BOLD has been used extensively to examine functional brain connectivity in large-scale brain networks, only a small number of studies have examined functional connectivity in ASL [3,4]. Two studies have compared ASL-connectivity and BOLD connectivity measured with either EPI-BOLD [5], or ASL-BOLD [6]. To our knowledge no previous work has combined the CBF and BOLD components of the ASL signal to obtain a functional connectivity measure that exploits the full information provided by this modality. The scarcity of related work may be due, in part, to the fact that most ASL sequences collect relatively fewer time frames (impacting the stability of correlations) in a given amount of scan time as well as the lack of off-the-shelf methodology for computing ASL networks. There is no work that we are aware of employing ASL-connectivity in TBI.

In this work, we contribute a new multivariate method for ASL-based network analysis. We improve upon existing approaches in two ways. First, we extend standard region-based methods with a sparse dimensionality reduction method that optimally correlates two ROIs. This is achieved by formulating the correlation between ROIs as a sparse selection optimization algorithm that finds non-uniformly weighted sub-ROIs that are most related. Second, we show how this method may jointly find these sub-ROIs by using both ASL-BOLD and CBF time series signal. Both of these advances relax some of the assumptions of standard region-based approaches while retaining the interpretability afforded by these classic approaches. In short, our contributions are: (1) We detail a new multivariate network analysis method; (2) We show how it may be used to fuse simultaneous time series measurements from multiple signal sources to estimate correlation matrices; (3) We evaluate these approaches in terms of reproducibility and applicability to studying TBI; (4) The method is freely available in a public open source toolkit [7].

2 Methods

ROI Analysis for Network Construction. Denote the matrices that describe the ASL-BOLD or ASL-CBF time series within a whole-brain ROI as X and Y respectively. Additionally, for a given set of anatomical ROIs for which there are L regions, we denote the ASL-BOLD sub-matrix extracted from ROI ^{i} as X^i . Then Y^i will contain that same ROI's ASL-CBF measurements. The classic region-based analysis will compute $x^i = 1/n \sum_k x_k^i$ which denotes averaging the x_k^i columns of X^i and similarly for y^i . From these region-averaged time-signals, a correlation matrix, \mathcal{R} , of size $L \times L$ is calculated, where $\mathcal{R}(i, j) = Corr(x^i, x^j)$ with $Corr$ denoting the Pearson's correlation.

Fusing Functional Signals via SCCA. Canonical correlation analysis (CCA) is a method for elucidating the relationship between two sets of measurements taken across a population [8] and is thus well-suited to multivariate neuroimaging data. Here, we show how CCA generalizes the basic ROI-based approach to network analysis described above. CCA introduces new unknown vectors, u^i and u^j , that act as weighted averages of X^i, X^j . CCA will optimize:

$$\arg \max_{u^i, u^j} \text{Corr}(u^i X^i, u^j X^j). \tag{1}$$

This formulation allows for the inclusion of the full time signal in all voxels in an ROI and is “nice” in that it can be solved by singular value decomposition if the number of samples is larger than the minimum number of columns in X^i or X^j . Sparse CCA extends CCA with additional constraints that allow the problem to be solved even when the input matrices are “fat” i.e. the number of columns far exceeds the number of rows, as is typically the case in functional MRI data. The SCCA formulation optimizes:

$$\begin{aligned} &\arg \max_{u^i, u^j} u^i (X^i)^T X^j u^j \\ &\text{subject to} \\ &\sum_i \|u^i\|_1 \leq s, \quad u^i \geq 0, \quad \sum_i \|u^j\|_1 \leq t, \quad u^j \geq 0, \quad \|u^i X^i\| = \|u^j X^j\| = 1, \end{aligned}$$

where s, t determine sparseness. Due to the non-linear (even np-hard) nature of subset selection from a large matrix, optimizing for a single canonical variate, u_i , involves a nonlinear gradient descent on the objective function above. This is one disadvantage of these methods. However, one gains robustness and the ability to exploit the full information of the input data. An additional advantage is that the formulation shown above may easily incorporate both ASL-CBF and ASL-BOLD data for simultaneous analysis. In all experiments presented here, the sparsity parameters were fixed and equal ($s = t$). Additionally a positivity constraint was imposed upon u^i and u^j .

Recall that we represent a given ROI’s BOLD and ASL signal as X^i and Y^i where each matrix is $n \times p$ (rows by columns) with n the number of acquired time points and p the number of voxels in ROI^{*i*}. Since both ASL-CBF and ASL-BOLD derive from the same acquisition, X^i and Y^i will always have the same dimensions. To examine both time series measurements simultaneously, we can column-append the two matrices: $Z^i = [[X^i] [Y^i]]$ resulting in a $n \times 2p$ matrix. For clarity, X will be used in further equations with the knowledge that it could be replaced with Y or Z with no resulting changes to the algorithm.

Now note that in the standard approach to ROI-based network analysis, the correlation matrix is given by $\mathcal{R}_{ROI}(i, j) = \text{Corr}(a(X^i), a(X^j))$ where $a(\cdot)$ indicates averaging over the ROI. The SCCA solution is trivial to use in the same manner, producing $\mathcal{R}_{SCCA}(i, j) = \text{SCCA}(X^i, X^j) = \text{Corr}(u^i X^i, u^j X^j)$. Note that the key difference is that SCCA will optimize the canonical variates to specifically identify the sub-regions within ROI^{*i*} and ROI^{*j*} that are most mutually informative.

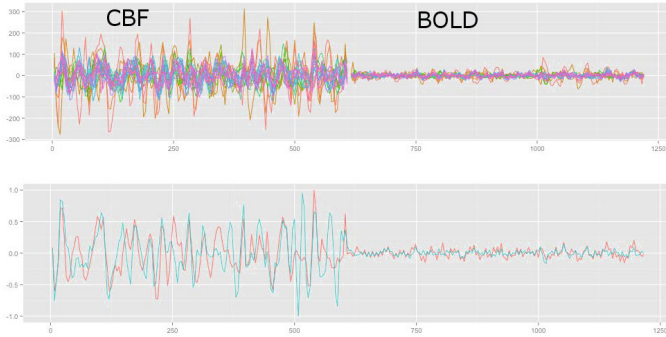


Fig. 1. ASL signals in the cingulate gyrus of a normal individual showing ASL-CBF and ASL-BOLD in all voxels in the ROI (top). The average ASL-CBF and ASL-BOLD are illustrated on the bottom in red and the ASL-CBF and ASL-BOLD that result from using SCCA to examine connectivity with the hippocampus is shown in blue. Both signals on the bottom have been intensity normalized for visualization

To obtain a connectivity matrix for an individual, the original ASL data and corresponding anatomical labels are required. For each labeled voxel, the ASL-CBF and ASL-BOLD time series are extracted. Motion-correction parameters are regressed out of the signals and a band pass filter is applied so that only frequencies between 0.01 Hz and 0.1 Hz are retained. This filtered time series data is then used to construct a matrix for each labeled region in the subject. For the region-averaged approach, the time-series are averaged to produce a single vector for each ROI while the SCCA method utilizes the full matrix to estimate connectivity.

3 Results

Our experimental design will establish the impact of SCCA-based network analysis on: (1) test-retest reliability of network correlation matrices; (2) how reliability changes with different signal (BOLD, CBF, concatenated BOLD-CBF) extracted from the input ASL time series; (3) the impact of the SCCA strategy on a population analysis of traumatic brain injury.

3.1 Evaluating Reproducibility via Test-Retest Data

Neuroimaging Data. The cohort consists of 12 healthy young adult participants (mean age 25.5 ± 4.5 , 7 female). For each subject, data was acquired at three time points. Two of these time points were acquired on the same day, in separate scanning sessions, while the third was acquired one week away from the same-day data. For each time point high resolution T1-weighted anatomic images were obtained using a 3D MPRAGE imaging sequence and the following

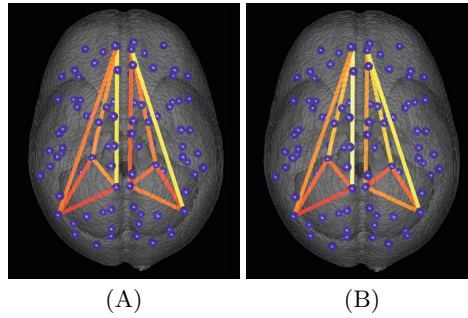


Fig. 2. Intrahemispheric connectivity in the default mode network of an individual subject is visualized using A) ROI averaged combined CBF and BOLD and B) SCCA based combined CBF and BOLD. The centroid of all labeled regions are illustrated with blue spheres, and the connections in the DMN are illustrated with line segments. The connectivity matrix values are used along with heat-mapping to color the line segments where higher connectivity values result in “hotter” colors.

acquisition parameters: TR = 1620 ms, TI = 950 ms, TE = 3 ms, flip angle = 15 degrees, 160 contiguous slices of 1.0 mm thickness, FOV = $192 \times 256 \text{ mm}^2$, matrix = 192×256 , 1NEX with a scan time of 6 min. The resulting voxel size was 1 mm^3 . Additionally, pulsed ASL (PASL) images were acquired with 80 alternating tag/control images and 2 M0 images all with 14 contiguous slices of 7.5mm thickness, FOV = $220 \times 220 \text{ mm}^2$, matrix = 64×64 . Additional acquisition parameters: TI1 = 700ms, TI2 = 1700ms.

Image Processing. The set of T1 images from each subjects first time point was used to construct a template using ANTs [9]. This template was brain masked and labeled with the AAL dataset [10]. A three-tissue segmentation was performed to allow the labels to be partially masked so only cortex and deep gray structures were labeled. For each time point, the T1 image was registered to the template image. Additionally, registration was used to find an intra-subject mapping between the T1 image and the M0 image that is acquired as a reference for the PASL acquisition. These transforms were composed to map the cortical labels into ASL native space for each time point. For PASL images, the M0 image served as a reference for motion-correction of all time-point volumes. Sinc interpolation was used to estimate the full time-series for both the control and tag data. The difference between control and tag was used along with relevant acquisition parameters to calculate the ASL-CBF over time, while the average of the two signals was calculated for ASL-BOLD [1].

Reproducibility Testing. To examine reproducibility, functional connectivity matrices are calculated for each time point using the classic region-averaged approach and the SCCA method on: ASL-CBF, ASL-BOLD, and combined ASL-CBF and ASL-BOLD. Graph correlation [11] is used for the comparison of connectivity matrices in order to examine reproducibility between images acquired

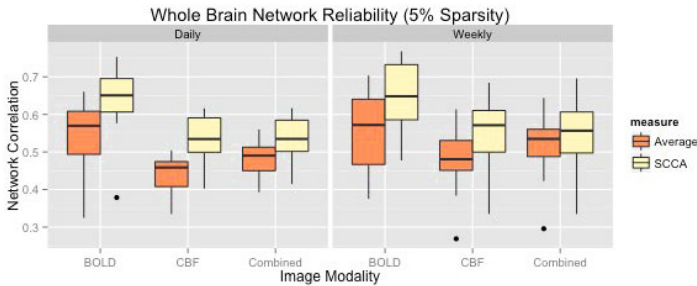


Fig. 3. For each metric, using both region averaging (orange) and SCCA (yellow), connectivity matrices were calculated from ASL data acquired in separate acquisitions in the same day and for data acquired one week apart. Whole network correlations were then calculated to examine reliability for the daily (left) and weekly (right) data for each subject. Here we illustrate results using sparsity values of $s = t = 0.05$. A range of sparsity values ($s = t$) up to 0.25 were examined and these higher values did not produce qualitatively different results.

on the same day, and images acquired one week apart. The results are illustrated in figure 3. A paired, one-sided Student's t-test was used to determine when the SCCA based method produced significantly more reproducible results than the corresponding region-averaged metric. Here we list the percent improvement provided by the SCCA method along with the p-value from the t-test.

Daily. ASL-BOLD = 17.72% ($p=5.0e-7$), ASL-CBF = 23.45% ($p=4.5e-7$), Combined = 23.45% ($p=0.0014$)

Weekly. ASL-BOLD = 18.79% ($p=8.1e-6$), ASL-CBF = 17.97% ($p=1.0e-7$), Combined = 7.23% ($p=1.75e-5$)

The largest improvements in daily reproducibility occur in the CBF based functional connectivity, suggesting that in this particular case, little useful information is gained by including the ASL-BOLD signal.

3.2 Cross-Sectional Examination of Brain Connectivity in TBI

Neuroimaging Data. Our cohort consists of 41 participants (mean age 30.4 ± 10.2), including 22 patients with TBI (9 females), and 18 controls (9 females). No significant difference exist between age or education in the patient or control groups. The same image T1 acquisition as described above was used for these subjects.

Image Processing. Processing for this data is identical to that for the test-retest data described above. The pulsed ASL (PASL) images were acquired with 160 alternating tag/control images and 2 M0 images all with 14 contiguous slice of 7.5mm thickness, $FOV = 220 \times 220\text{mm}^2$, $matrix = 64 \times 64$. Additional acquisition parameters: $TI1 = 700\text{ms}$, $TI2 = 1900\text{ms}$.

Network Differences. To identify potential effects of TBI, connectivity matrices were calculated for all subjects using all metrics. In particular, the default mode network (DMN) is of interest as this network has been shown to be affected by TBI [12] and is relevant here as the data was acquired during the resting state. Here we used the following labeled regions for the DMN: posterior cingulate gyrus, hippocampus, frontal medial orbital cortex, and the angular gyrus. This network is illustrated in figure 2. Each subject’s time from injury was categorized as either: no injury, less than one year since injury, or more than one year since injury. For each metric of interest, connectivity values for intra-hemispheric connections in each hemisphere were extracted and R was used to examine the influence of diagnosis (TBI or control) on connectivity values using (in R language syntax):

$$Conn \sim 1 + Diagnosis + Age + Gender + YearsEducation + InjuryTime \quad (2)$$

All p-values for diagnosis were FDR corrected and connections with $q < 0.1$ were reported as potentially compromised connections. The region-averaging approach did not result in any reported connectivity differences between control and TBI, nor were any results reported for ASL-BOLD alone. Connectivity measured using SCCA on the ASL-CBF and the combined data both reported connectivity differences in the right hemisphere between the posterior cingulate gyrus and both the hippocampus ($q = 0.084$) and angular gyrus ($q = 0.091$).

4 Discussion

We detailed how SCCA may be used to fuse the ASL-CBF and ASL-BOLD signals to exploit both the multi-variate signal provided by ASL as well as the full information provided within each anatomical region. We demonstrated that the SCCA method provides a more repeatable measure of network connectivity than the classic region-averaged approach. However, the reproducibility gains in combined ASL-CBF and ASL-BOLD were less than either metric examined alone. This may be a result of equally weighting the signals. As these signals have different physiological origins, an adaptive weighting method may be more appropriate and will be examined in future work. An examination of TBI suggested that the SCCA method provides a measure of connectivity that is more sensitive to disruptions in the DMN. Future work will include exploring how additional modalities, such as standard BOLD fMRI, may be incorporated into the framework described here.

There are several caveats that must be kept in mind when interpreting these findings. One important issue in connectivity studies is the possible artifacts induced by motion. While we did not find significant differences in motion parameters between groups, this confound may not be entirely ruled out, although we note that its effect should be similar across all comparisons. Regarding our BOLD findings, we note that ASL sequences are not optimized for BOLD sensitivity; in general, our findings may differ for different types of ASL or other functional MRI sequences. Finally, we did not study every frequency range and

these may impact reliability in all of the studied signals. In future work, we will more carefully characterize the signal that is extracted by the SCCA approach in comparison to the ROI analysis. However, we believe that the novel findings reported in this work encourage further exploration of using SCCA to drive network analyses of the brain.

References

1. Wong, E.C., Buxton, R.B., Frank, L.R.: Implementation of quantitative perfusion imaging techniques for functional brain mapping using pulsed arterial spin labeling. *NMR Biomed.* 10(4-5), 237–249 (1997)
2. Aguirre, G.K., Detre, J.A., Zarahn, E., Alsop, D.C.: Experimental design and the relative sensitivity of bold and perfusion fMRI. *Neuroimage* 15(3), 488–500 (2002)
3. Chuang, K.H., van Gelderen, P., Merkle, H., Bodurka, J., Ikonomidou, V.N., Koretsky, A.P., Duyn, J.H., Talagala, S.L.: Mapping resting-state functional connectivity using perfusion MRI. *Neuroimage* 40(4), 1595–1605 (2008)
4. Zou, Q., Wu, C.W., Stein, E.A., Zang, Y., Yang, Y.: Static and dynamic characteristics of cerebral blood flow during the resting state. *Neuroimage* 48(3), 515–524 (2009)
5. Li, Z., Zhu, Y., Childress, A.R., Detre, J.A., Wang, Z.: Relations between BOLD fMRI-derived resting brain activity and cerebral blood flow. *PLoS One* 7(9), e44556 (2012)
6. Viviani, R., Messina, I., Walter, M.: Resting state functional connectivity in perfusion imaging: correlation maps with BOLD connectivity and resting state perfusion. *PLoS One* 6(11), e27050 (2011)
7. Avants, B.B., Tustison, N., Song, G.: Advanced normalization tools (ANTS). *Insight J.* (2009)
8. Hotelling, H.: Relations between two sets of variates. *Biometrika* 28(3/4), 321–377 (1936)
9. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54(3), 2033–2044 (2011)
10. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., et al.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* 15(1), 273–289 (2002)
11. van Wijk, B.C.M., Stam, C.J., Daffertshofer, A.: Comparing brain networks of different size and connectivity density using graph theory. *PLoS One* 5(10), e13701 (2010)
12. Johnson, B., Zhang, K., Gay, M., Horovitz, S., Hallett, M., Sebastianelli, W., Slobounov, S.: Alteration of brain default network in subacute phase of injury in concussed individuals: resting-state fMRI study. *Neuroimage* 59(1), 511–518 (2012)

Multi-resolutional Brain Network Filtering and Analysis via Wavelets on Non-Euclidean Space*

Won Hwa Kim¹, Nagesh Adluru¹, Moo K. Chung¹, Sylvia Charchut⁴,
Johnson J. GadElkarim², Lori Altshuler³, Teena Moody³,
Anand Kumar², Vikas Singh¹, and Alex D. Leow²

¹ University of Wisconsin, Madison

² University of Illinois, Chicago

³ University of California, Los-Angeles

⁴ Southeastern Louisiana University, Hammond

Abstract. Advances in resting state fMRI and diffusion weighted imaging (DWI) have led to much interest in studies that evaluate hypotheses focused on how brain connectivity networks show variations across clinically disparate groups. However, various sources of error (e.g., tractography errors, magnetic field distortion, and motion artifacts) leak into the data, and make downstream statistical analysis problematic. In small sample size studies, such noise have an unfortunate effect that the differential signal may not be identifiable and so the null hypothesis cannot be rejected. Traditionally, smoothing is often used to filter out noise. But the construction of convolving with a Gaussian kernel is not well understood on arbitrarily connected graphs. Furthermore, there are no direct analogues of scale-space theory for graphs — ones which allow to view the signal at multiple resolutions. We provide rigorous frameworks for performing 'multi-resolutional' analysis on brain connectivity graphs. These are based on the recent theory of non-Euclidean wavelets. We provide strong evidence, on brain connectivity data from a network analysis study (structural connectivity differences in adult euthymic bipolar subjects), that the proposed algorithm allows identifying statistically significant network variations, which are clinically meaningful, where classical statistical tests, if applied directly, fail.

1 Introduction

The development of diffusion weighted imaging (DWI) and functional magnetic resonance imaging (fMRI) have laid the groundwork for ambitious initiatives towards a full characterization of the human connectome (the brain's wiring diagram) to better understand the structural and functional aspects of brain connectivity. While such large scale projects will clearly push the frontiers of

* Correspondence: wonhwa@cs.wisc.edu; Research supported in part by NIH R01AG040396, NIH R01AG021155, NSF RI 1116584, NSF CAREER 1252725, the Wisconsin Partnership Proposal, UW ADRC, UW ICTR (1UL1RR025011), NC-CAM P01 AT004952-04, Waisman Core grant P30 HD003352-45, National Alliance for Research in Schizophrenia and Affective Disorders Young Investigator grant (AL) and by National Institute of Mental Health Grant R21 MH086104 (LA).

neuroscience forward, these efforts must necessarily go hand in hand with studies seeking to answer *more focused* questions pertaining to the variations in the connectivity structure in the context of *specific* neurodegenerative diseases, and how its manifestation is modulated by genetic and demographic factors. As such these studies operate in the *small sample size* regime, and the first order requirement on analysis methods appropriate for such applications is to maximize statistical power — in other words, the likelihood of observing a differential signal in the connectivity data given the limited size of the cohort.

Consider a brain connectivity network modeled as an undirected weighted graph denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \omega)$. The vertices, \mathcal{V} may denote anatomically meaningful parcellations [6] or regions exhibiting spatially contiguous BOLD activations where as the weighted edges may correspond to temporal correlations or strength of tract connectivity[1]. Now, our interest is to perform statistical analysis on a population of such brain connectivity networks in clinically disparate groups, to understand which connections are severely affected by the disease. The overwhelming majority of current literature suggests applying standard hypothesis testing at the level of individual network edges. This approach generally works well, but when the group-wise differences are weak to begin with, one finds that after correcting for multiple comparisons, a statistically significant signal may be unidentifiable. One may smooth out the noise variance in \mathcal{E} and \mathcal{V} , but it is still an issue in brain connectivity analysis. For example, even a small head motion in the scanner can influence the DTI connectivity information. [16]

The standard procedure to improve the signal to noise ratio is to smooth the input signal. In image processing, the measurements are defined on a uniformly sampled lattice (Euclidean space) where the standard notions of a convolution filter apply directly. The few instances in the literature which implement a smoothing process on brain network data essentially average the measurements within a the node or edge ROIs with isotropic Gaussian kernels [18,14], discarding the network structure. In computer vision, anisotropic diffusion kernels [15], heat kernels [20] and pyramids [11] have been used extensively, but mostly in the context of a regular lattice over pixels. Developments of these ideas have led to much work in scale space theory, towards deriving multi-resolution representations of the image by incremental smoothing. The question we investigate is whether such connections can be exploited to analyze brain connectivity network data with enhanced statistical sensitivity. The literature offers few strategies for filtering of signals on the edges of a brain connectivity network.

Key Contributions. The most natural mathematical tool which offers multi-resolution behavior, i.e., *wavelets*, was until recently, restricted to the Euclidean space. But the objects of interest here are networks with arbitrary topology (non-Euclidean). **a)** We make use of a recent harmonic analysis results to show how non-Euclidean wavelets provide tools for defining multi-scale representations of brain networks. **b)** We demonstrate an application to analyzing *structural connectivity* differences between euthymic bipolar disorder and healthy subjects. Our framework applies multi-resolutional analysis on the information defined on the edges, not on the vertices. The noise in raw connectivity data has the effect

that few edges show up as statistically significant after accounting for multiple comparisons correction. But applying the proposed method, clinically meaningful group differences can be detected at the Bonferroni threshold of $\alpha < 0.01$.

2 Non-euclidean Wavelets

Wavelet transform is conceptually similar to the well-known Fourier transform, however, it uses a certain shape of oscillating function as a basis with finite duration instead of the sine and cosine basis with infinite duration. The traditional construction of wavelets is defined by a mother wavelet function ψ and a scaling function ϕ , which are band and low-pass filters in the frequency domain.

The wavelet function ψ on x is a function defined by two parameters, the scale parameter s and translation parameters a

$$\psi_{s,a}(x) = \frac{1}{a} \psi\left(\frac{x-a}{s}\right). \tag{1}$$

Change in s varies the *dilation* of the wavelet, and together with a *translation* parameter a , approximates a signal in harmonics using wavelet expansion. The function $\psi_{s,a}(x)$ forms bases for the signal and can be used with other basis at different scales to decompose a signal. The wavelet transform of a signal $f(x)$ is defined as the inner product of the wavelet basis $\psi_{s,a}$ and $f(x)$,

$$W_f(s, a) = \langle f, \psi \rangle = \frac{1}{a} \int f(x) \psi^*\left(\frac{x-a}{s}\right) dx, \tag{2}$$

where $W_f(s, a)$ is the wavelet coefficient at scale s and at location a . The original signal $f(x)$ can be reconstructed from $W_f(s, a)$ and basis function without loss of information; the inverse transformation is

$$f(x) = \frac{1}{C_\psi} \iint W_f(s, a) \psi_{s,a}(x) da ds \tag{3}$$

where $C_\psi = \int \frac{|\Psi(j\omega)|^2}{|\omega|} d\omega$ is known as the *admissibility condition constant*, Ψ is the Fourier transform of the wavelet [9], and ω denotes the frequency domain.

Recent work in harmonic analysis [7] provides wavelet basis on structured data which expresses in a wide spectrum of frequencies. The solution in [7] relies on a graph Fourier transform to derive a spectral graph wavelet transform (SGWT). It is shown that SGWT formalization preserves the localization properties at fine scales as well as other wavelet specific properties, while addressing the bottleneck of defining *scales* on a domain where the space is non-Euclidean.

Let a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \omega\}$ be a undirected graph with a vertex set \mathcal{V} with N vertices, an edge set \mathcal{E} and corresponding edge weight $\omega \geq 0$. The adjacency matrix \mathbf{A} of \mathcal{G} is given as a $N \times N$ matrix whose elements a_{ij} are the edge weight ω_{ij} if i^{th} and j^{th} nodes are connected. The degree matrix \mathbf{D} is computed as a $N \times N$ diagonal matrix whose i^{th} diagonal is $\sum_j \omega_{ij}$. The graph Laplacian from these graph matrices is defined in the usual way as $\mathbf{L} = \mathbf{D} - \mathbf{A}$. Then, the complete orthonormal basis χ_l and eigenvalues $\lambda_l, l \in \{0, 1, \dots, N - 1\}$

obtained from the graph Laplacian, a self-adjoint operator, forms the basis for the graph Fourier transformation. Note that the λ_l are increasingly ordered, and are irrelevant to the order of vertex index in the graph domain. Using these basis, the forward and inverse graph Fourier transformation are defined using the eigenvalues and eigenvectors of \mathbf{L} as,

$$\hat{f}(l) = \langle \chi_l, f \rangle = \sum_{n=1}^N \chi_l^*(n) f(n) \quad \text{and} \quad f(n) = \sum_{l=0}^{N-1} \hat{f}(l) \chi_l(n) \quad (4)$$

Using these transforms, we construct spectral graph wavelets by applying band-pass filters at multiple scales and *localizing* it with an impulse function and low-pass filter for the scaling function.

Here, λ_l , the spectrum of the Laplacian, serves as an analog of the 1-D frequency domain, where scales can be easily defined. This directly provides the key component in obtaining a multi-resolutional view of the signal localized at n . Constructing a kernel function g which acts as band-pass filter in the frequency domain, when g is transformed back to the original graph domain, we directly obtain a representation of the signal for that scale. Repeating this procedure for multiple scales, the set of coefficients obtained for each $s \in S$ gives a multi-resolution representation for that particular vertex.

Since the transformed impulse function in the frequency domain is equivalent to a unit function, the wavelet ψ localized at vertex n can now be defined as,

$$\psi_{s,n}(m) = \sum_{l=0}^{N-1} g(s\lambda_l) \chi_l^*(n) \chi_l(m) \quad (5)$$

where m is a graph vertex. With this in hand, the wavelet coefficients of a given function $f(n)$ is given by the inner product of wavelets and the function,

$$W_f(s, n) = \langle \psi_{s,n}, f \rangle = \sum_{l=0}^{N-1} g(s\lambda_l) \hat{f}(l) \chi_l(n) \quad (6)$$

SGWT follows the same procedure of constructing wavelets as in the continuous wavelet transform. In the fine scale limit, SGWT maintains many of the properties of the traditional wavelet transform, including localization.

Remark. Wavelets in Euclidean space have a rich history in Signal processing. However, defining wavelets in non-Euclidean space is a recent development [7,5], and is especially interesting for network analysis in Neuroimaging.

3 Deriving a Multi-Resolution Perspective of a Network

Line Graphs. In graph theory, one defines the line graph $L(\mathcal{G})$ as a dual form of graph \mathcal{G} . The $L(\mathcal{G})$ is formed by interchanging the roles of \mathcal{V} and \mathcal{E} in \mathcal{G} . Two vertices in $L(\mathcal{G})$ are connected when the corresponding edges in \mathcal{G} share a common vertex. The line graph $L(\mathcal{G}) = \{\mathcal{V}_L, \mathcal{E}_L, \omega_L\}$ has a vertex set for the edges $\{\mathcal{E}, \omega\}$ and a edge set that corresponds to the vertices \mathcal{V} in \mathcal{G} [8].

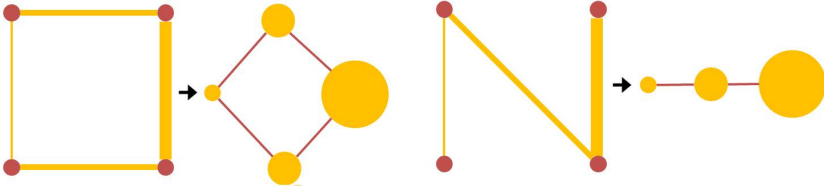


Fig. 1. Examples of graphs and the corresponding line graphs. Original graphs with vertices (red) and edges (yellow) with edge weights (thickness), and corresponding line graphs with vertices (yellow) with function (vertex size) and edges (red).

The transformation of $L(\mathcal{G})$ from a graph \mathcal{G} is defined as follows. Let g_{ij} be the elements in the adjacency matrix \mathbf{A}_L of $L(\mathcal{G})$, then

$$g_{ij} = \begin{cases} 1 & \text{if } v \in \mathcal{V}, v \sim e_i, e_j \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where v is a vertex in \mathcal{V} and e is an edge in \mathcal{E} . This means that when two edges share a common vertex in \mathcal{G} , these edges are connected to each other by the common vertex. After this transformation, the isolated vertices in \mathcal{G} are completely neglected in $L(\mathcal{G})$. If there are no isolated vertices in \mathcal{G} , then \mathcal{G} and $L(\mathcal{G})$ have equal number of components. After constructing a line graph $L(\mathcal{G})$ of a graph \mathcal{G} , the edges in \mathcal{G} form a completely new domain of analysis and the edge weight ω can be defined as a function defined on each vertex in \mathcal{V}_L , where the connection between each vertex in \mathcal{E}_L is given from \mathcal{V} . Toy examples of this transformation are shown in Fig. 1.

In a measured signal, the true signal tends to change smoothly while noise varies very rapidly in high frequencies. Using wavelets, smoothing can be efficiently performed by removing high frequency components tied to the finer scales, moreover, due to the bandpass property of wavelet, we can get a multi-resolutional view of the given signal. The multi-scale view comes from the inverse wavelet transformation of the resultant function that provides the estimate of the signal at various scales. Rewriting (3) in terms of the graph Fourier basis,

$$f(m) = \frac{1}{C_g} \sum_l \left(\int_0^\infty \frac{g^2(s\lambda_l)}{s} ds \right) \hat{f}(l)\chi_l(m) \tag{8}$$

which sums over the entire scale s . Limiting the scales to the coarse scales will reconstruct the smoothed approximation of the original signal, and the original signal can be reconstructed by adding finer scales.

In order to filter the network structure, it is necessary to bring the network connectivity information as a signal into another domain. As described above, the transformation of a graph domain \mathcal{G} to a line graph $L(\mathcal{G})$ enables us to view the edge weights as a signal defined in the domain of $L(\mathcal{G})$. We can therefore define the connectivity as a signal on each vertex of $L(\mathcal{G})$, and continue with the smoothing technique using wavelet. An illustrative example of the framework for the network smoothing is given in Fig. 2, where the edge weights are filtered along

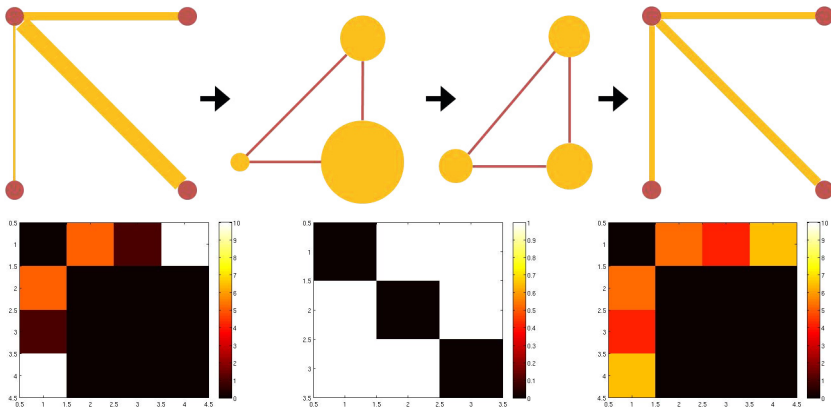


Fig. 2. A toy example of graph structure filtering. The top panel shows the graph filtering steps: (1) Construction of the line graph, (2) filtering the signal on the line graph vertices, (3) reconstructing the filtered graph. The bottom panel shows the corresponding adjacency matrices.

their connection and not losing the original topology of \mathcal{G} . The corresponding adjacency matrices are displayed at the bottom.

In addition to the filtering, we define *Wavelet Multi-scale Descriptor* (WMD) using the wavelet wavelet coefficients over the signal defined on each vertex as

$$\text{WMD}_f(n) = \{W_f(s, n) | s \in S\} \quad (9)$$

which characterizes the signal at multi-resolutions on the vertex according to the geometry of the graph[10]. While [19] uses a sphere to obtain the descriptor, which causes data distortion by mapping process, WMD is derived based on the eigenfunction of the original graph itself, and thus avoids 'ballooning'.

4 Connectivity Differences in Bipolar Disorder

Dataset. We scanned 25 healthy subjects (13 male and 12 female; age: 42.2 ± 10.8) and 25 gender and age matched bipolar subjects (14 male, 11 female; age: 41.7 ± 12.6). All bipolar subjects received comprehensive psychiatric evaluations using the structured clinical interview for DSM disorders (SCID) and met the DSM IV criteria for bipolar I disorder (at the time of image acquisition all subjects have been in an euthymic state for at least 30 days). A Siemens 3T Trio scanner was used to acquire the brain MRI data. High resolution T1-weighted images were acquired with MPRAGE sequence (FOV = $250 \times 250 \text{ mm}^2$; TR/TE = 1900/2.26 ms; flip angle = 9° ; voxel size = $1 \times 1 \times 1 \text{ mm}^3$). Diffusion weighted (DW) images were acquired using SS-SE-EPI sequences (FOV = $190 \times 190 \text{ mm}^2$; resolution $2 \times 2 \times 2 \text{ mm}^3$; TR/TE = 8400/93 ms; 64 gradient directions, $b = 1000 \text{ s/mm}^2$ and one minimally DW scan: $b=0$ image).

Structural brain networks were generated using a pipeline which integrates multiple image processing steps. First, DW images were eddy current corrected using FSL by registering all DW images to their corresponding $b=0$ images with

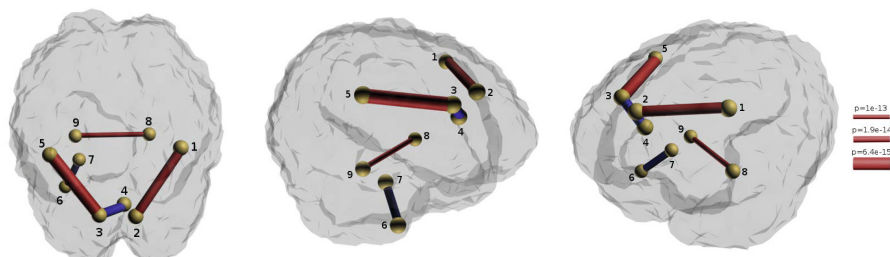


Fig. 3. Anatomical connectivity showing group differences between bipolar and controls after Bonferroni threshold at $\alpha = 10^{-7}$. Connection thickness represents the p -value in negative log scale; color gives sign of strength: red (and blue) are stronger in controls (and bipolar group). Region labels are: 1. ctx-rh-precentral, 2. ctx-rh-superiorfrontal, 3. ctx-lh-superiorfrontal, 4. ctx-rh-caudal anterior cingulate, 5. ctx-lh-precentral, 6. ctx-lh-temporalpole, 7. left amygdala, 8. right hippocampus, 9. left hippocampus.

12-parameter affine transformations. This was followed by the computation of diffusion tensors and then deterministic tractography using the FACT algorithm [13] built into the DTIStudio program (maximum bending angle 60 degrees; FA cut-off 0.25). T1-weighted images were used to generate label maps using Freesurfer. The number of tracts connecting 87 cortical/subcortical regions were used in constructing 87×87 connectivity matrix for each subject.

Group Analysis. There are total 3741 edges in the network. In order to detect connectivity differences between the two groups, we performed a Hotellings T^2 -test using WMD. WMD was realized by a Mexican-hat wavelet, which was defined at 5 scales over the spectrum of λ . Since typically noise lies in high frequency, we dropped 2 scales that correspond to larger λ , and used the rest for the statistical analysis. When using the raw edge weights, we could not detect any significant difference between the two groups after accounting for Bonferroni correction at $\alpha = 0.05$ significance level. However, after applying WMD on smoothed edge weights with the proposed method, we identified 5 connections over 9 different brain regions as having significant connectivity differences at a very conservative Bonferroni correction level, 10^{-7} .

Interpretation. Results showed that relative to control subjects, bipolar patients on average exhibited weaker strength for the connections within the frontal lobe (bilateral precentral to superior frontal) as well as in fiber tracts linking the bilateral hippocampus. These findings are consistent with past studies where abnormalities in the frontal, limbic, and callosal systems have been reported (for a review, see for example [12]). Additionally, previous fMRI studies of euthymic bipolar patients have also consistently revealed frontal hypoactivation [4,3,17]. In contrast, bipolar subjects exhibit a stronger connection, relative to controls, between the left amygdala and the left temporal pole and between the left superior frontal gyrus and the right caudal anterior cingulate. Although our subjects were in euthymia at the time of the scan, these stronger connections in bipolar may be related to amygdala activation during mania as reported in [2].

5 Conclusion

In this paper, we introduced a novel signal filtering approach for brain network data that takes into account the non-Euclidean nature of the structured data. Using a line graph construction from the original network, we perform band-pass filtering of signals defined on network edges to obtain a multi-resolutional view. The algorithm significantly improves the statistical sensitivity of connectivity differences using Hotelling's T^2 -tests and Bonferroni correction. We believe that adapting non-Euclidean wavelets for improving the statistical properties of brain connectivity networks may improve analysis of a much wider variety of studies.

References

1. Adluru, N., Chung, M.K., Lange, N.T., Lainhart, J.E., Alexander, A.L.: Applications of epsilon radial networks in neuroimage analyses. In: Ho, Y.-S. (ed.) PSIVT 2011, Part I. LNCS, vol. 7087, pp. 236–247. Springer, Heidelberg (2011)
2. Altshuler, L., Bookheimer, S., Proenza, M.A., et al.: Increased amygdala activation during mania: a functional magnetic resonance imaging study. *American Journal of Psychiatry* 162(6), 1211–1213 (2005)
3. Cerullo, M.A., Adler, C.M., Delbello, M.P., et al.: The functional neuroanatomy of bipolar disorder. *International Review of Psychiatry* 21(4), 314–322 (2009)
4. Chen, C., Suckling, J., Lennox, B.R., Ooi, C., et al.: A quantitative meta-analysis of fMRI studies in bipolar disorder. *Bipolar Disorders* 13(1), 1–15 (2011)
5. Coifman, R.R., Maggioni, M.: Diffusion wavelets. *Applied and Computational Harmonic Analysis* 21(1), 53–94 (2006)
6. Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C., et al.: Mapping the structural core of human cerebral cortex. *PLoS Biol* 6(7), e159 (2008)
7. Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. *App. and Comp. Harmonic Analysis* 30(2), 129–150 (2011)
8. Harary, F.: *Graph Theory*. Addison-Wesley, Reading (1969)
9. Haykin, S., Van Veen, B.: *Signals and Systems*, 2nd edn. Wiley (2005)
10. Kim, W.H., Pachauri, D., Hatt, C., et al.: Wavelet based multi-scale shape features on arbitrary surfaces for cortical thickness discrimination. In: NIPS (2012)
11. Lowe, D.G.: Object recognition from local scale-invariant features. In: ICCV, vol. 2, pp. 1150–1157. IEEE (1999)
12. Mahon, K., Burdick, K.E., Szeszko, P.R.: A role for white matter abnormalities in the pathophysiology of bipolar disorder. *Neuroscience and Biobehavioral Reviews* 34(4), 533–554 (2010)
13. Mori, S., Crain, B.J., Chacko, V.P., et al.: Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology* 45, 256–269 (1999)
14. Pachauri, D., Hinrichs, C., Chung, M.K., et al.: Topology-based kernels with application to inference problems in AD. *TMI* 30(10), 1760–1770 (2011)
15. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *TPAMI* 12(7), 629–639 (1990)
16. Power, J.D., Barnes, K.A., Snyder, A.Z., et al.: Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage* 59(3), 2142–2154 (2012)

17. Van der Schot, A., Kahn, R., Ramsey, N., et al.: Trait and state dependent functional impairments in bipolar disorder. *Psychiatry Research* 184(3), 135 (2010)
18. Van Hecke, W., Leemans, A., De Backer, S., et al.: Comparing isotropic and anisotropic smoothing for voxel-based DTI analyses: A simulation study. *Human Brain Mapping* 31, 98–114 (2010)
19. Yu, P., Grant, P.E., Qi, Y., et al.: Cortical surface shape analysis based on spherical wavelets. *TMI* 26(4), 582–597 (2007)
20. Zhang, F., Hancock, E.: Graph spectral image smoothing using the heat kernel. *Pattern Recognition* 41(11), 3328–3342 (2008)

Implications of Inconsistencies between fMRI and dMRI on Multimodal Connectivity Estimation

Bernard Ng^{1,2}, Gael Varoquaux¹, Jean Baptiste Poline¹, and Bertrand Thirion¹

¹Parietal team, Neurospin, INRIA Saclay, France

²FIND Lab, Stanford University, United States

bernardying@gmail.com

Abstract. There is a recent trend towards integrating resting state functional magnetic resonance imaging (RS-fMRI) and diffusion MRI (dMRI) for brain connectivity estimation, as motivated by how estimates from these modalities are presumably two views reflecting the same underlying brain circuitry. In this paper, we show on a cohort of 60 subjects that conventional functional connectivity (FC) estimates based on Pearson's correlation and anatomical connectivity (AC) estimates based on fiber counts are actually not that highly correlated for typical RS-fMRI (~7 min) and dMRI (~32 gradient directions) data. The FC-AC correlation can be significantly increased by considering sparse partial correlation and modeling fiber endpoint uncertainty, but the resulting FC-AC correlation is still rather low in absolute terms. We further exemplify the inconsistencies between FC and AC estimates by integrating them as priors into activation detection and demonstrating significant differences in their detection sensitivity. Importantly, we illustrate that these inconsistencies can be useful in fMRI-dMRI integration for improving brain connectivity estimation.

Keywords: brain activation, connectivity, dMRI, fMRI, multimodal integration.

1 Introduction

Recently, there is a growing interest in integrating resting state functional magnetic resonance imaging (RS-fMRI) and diffusion MRI (dMRI) for brain connectivity estimation [1, 2]. This is motivated by how connectivity inferred from these modalities are presumably two views resembling the same underlying wiring structure of the brain. To enable meaningful integration, it is important to understand the relationships between brain function and structure and the degree to which these relationships are reflected by the connectivity estimates derived from RS-fMRI and dMRI data. A number of studies suggest a strong positive correlation between RS-fMRI and dMRI connectivity estimates [3-5], but it is unclear how much this correlation depends on the data acquisition and analysis methods.

Pearson's correlation is, by far, the most widely-used estimate of functional connectivity (FC) for RS-fMRI studies. Due to confounds, such as head motions, scanner noise, and physiological artefacts, as well as the typically small sample size as limited by data acquisition rate of standard pulse sequences, Pearson's correlation is subject

to a high degree of estimation error [6]. Also, Pearson’s correlation cannot distinguish direct from indirect connections, which complicates connection structure identification [6]. As for inferring anatomical connectivity (AC) from dMRI data, fiber count and the average fractional anisotropy (FA) along fiber tracts between pairs of brain areas are typically used as estimates [7]. Approaches for estimating the probability of diffusion between brain area pairs have been put forth [7], but do not scale well computationally to whole-brain analysis. In general, the fundamental limitations of fiber tractography pose great challenges to accurate AC estimation [7], especially when applied to typical dMRI data with limited number of gradient directions due to acquisition time restrictions. Besides the problem of crossing fibers, which harms the accuracy of the fiber tracts’ trajectories, the location of the fiber endpoints holds high uncertainty, since the diffusion orientation is ambiguous near white-gray matter interface and tractography often terminates before reaching gray matter tissues [7].

In this paper, we assess the impact of the aforementioned limitations in data acquisition and analysis techniques on the consistency of FC and AC estimates. On typical RS-fMRI (~7 min) and dMRI (32 gradient directions) data collected from a cohort of 60 subjects, we take the correlation between conventional FC and AC estimates based on Pearson’s correlation and fiber count as baseline, and examine the effects of using sparse partial correlation [8] as the FC estimate, which controls for estimation errors and reduces indirect influences. We also propose a simple post-processing scheme for modeling fiber endpoint uncertainty and examine its influence on the FC-AC correlation. These comparisons are performed to determine whether and to what extent methodological improvements can compensate for limitations in data acquisition. In addition, we describe how these FC and AC estimates can be meaningfully incorporated as priors for task activation detection to further evaluate their consistency. The implications of combining the FC and AC estimates are also explored.

2 Methods

In this work, we are interested in contrasting Pearson’s correlation against sparse partial correlation as FC estimates (Section 2.1) and examining the effects of modeling the uncertainty in fiber endpoint location when estimating AC (Section 2.2). Consistency between various FC and AC estimates is evaluated based on their correlation and their impact as priors on task activation detection (Section 2.3).

2.1 Functional Connectivity Estimation

Pearson’s Correlation. Let \mathbf{Z} be a $t \times d$ matrix with normalized RS-fMRI time courses (i.e. demeaned with unit standard deviation) of d brain areas along the columns. The Pearson’s correlations between all brain area pairs are given by: $\mathbf{C} = \mathbf{Z}^T \mathbf{Z} / (t - 1)$. For typical RS-fMRI data, $t < d$, thus \mathbf{C} will contain high estimation errors [6]. Also, elements in \mathbf{C} are assumed independent from each other, which renders separation of direct connections from indirect connections impossible [6]. These limitations can be alleviated by adopting sparse partial correlation as an FC estimate, as discussed next.

Sparse Partial Correlation. Given a $d \times d$ empirical covariance matrix, \mathbf{S} , computed from samples drawn from a centered multivariate Gaussian distribution, we can estimate a well-conditioned sparse invariance covariance matrix, $\hat{\mathbf{\Lambda}}$, by minimizing the negative log-likelihood over the space of positive definite matrices, $\mathbf{\Lambda} > 0$, and imposing an l_1 penalty on the off diagonal elements of $\mathbf{\Lambda}$ [8]:

$$\min_{\mathbf{\Lambda} > 0} tr(\mathbf{S}\mathbf{\Lambda}) - \log \det(\mathbf{\Lambda}) + \lambda \|\mathbf{\Lambda}\|_1. \quad (1)$$

Sparse partial correlation, $\hat{\mathbf{\Gamma}}$, can then be computed by applying a simple normalization on $\hat{\mathbf{\Lambda}}$: $\hat{\Gamma}_{ij} = -\hat{\Lambda}_{ij} / \sqrt{\hat{\Lambda}_{ii}\hat{\Lambda}_{jj}}$ for $i \neq j$ and $\hat{\Gamma}_{ij} = 1$ for $i = j$. The level of sparsity on $\hat{\mathbf{\Lambda}}$ is governed by λ , which we select using a refined grid search strategy combined with cross-validation as in [1]. (1) can be efficiently solved using e.g. the QUadratic Inverse Covariance (QUIC) algorithm [8]. In the context of FC estimation, \mathbf{S} corresponds to \mathbf{C} of a given subject. We highlight that enforcing sparsity serves the dual purpose of controlling for estimation errors and imposing our prior knowledge that the connection structure of the brain is sparse [1]. Also, partial correlation provides a measure of connectivity between two brain areas with the effects from all other brain areas partialled out, thus discriminates direct connections from indirect connections.

2.2 Anatomical Connectivity Estimation

The fiber count between a pair of brain areas is widely-used as an estimate of AC, since this measure presumably reflects the anatomical capacity for functional interactions. This AC estimate requires the fiber tracts to terminate in gray matter brain areas. For terminating a fiber tract, low FA (e.g. 0.15) and large curvature (e.g. 45°) are typically used as the criteria [7]. Although gray matter tissues have relatively lower FA than white matter tissues, voxels containing crossing fibers or near white-gray matter boundaries also have low FA. Thus, fiber tracts might not necessarily terminate in gray matter voxels. One simple way to deal with this limitation is to extrapolate along the tangent direction of the fiber endpoints (e.g. < 10 mm from the endpoints) until the fiber tracts reach a gray matter voxel (Fig. 1(a), option 2). However, the location of the resulting endpoints could very well be incorrect, especially if the original fiber endpoints are near gyri (Fig. 1(a)) [7]. In fact, the restriction on curvature during tractography would bias fiber tracts towards option 2 in Fig. 1(a) if tracking is not prematurely terminated by low FA [7].

To model the uncertainty in fiber endpoint location, we propose the following schem: Let ω^p be a 27×1 weight vector associated with an endpoint p of a fiber tract. We assign a weight of 1 to the endpoint and a weight of $\exp(-\Delta_{ps})$ to its 26-connected neighbors but only if the neighbors are gray matter voxels. If no neighboring gray matter voxels are present, the given fiber is discarded. Δ_{ps} is the distance between the endpoint p and the neighbor s in voxel units. We then normalize ω^p such that its

elements sum to 1. The same procedure is performed for the other endpoint q of the given tract in generating ω_q . For each pair of elements $\omega^p(k)$ and $\omega^q(l)$ in which the two corresponding voxels lie in brain area i and j , respectively, we add $\omega^p(k) \cdot \omega^q(l)$ to \mathbf{A}_{ij} , where \mathbf{A}_{ij} is the $(i,j)^{\text{th}}$ element of the fiber count matrix, \mathbf{A} . In effect, we are placing a Gaussian kernel at each endpoint and partitioning a tract across spatially proximal brain areas to model endpoint uncertainty. If all neighbors of endpoint p lie within the same brain area i and all neighbors of endpoint q lie within the same brain area j , the scheme above reduces to simple fiber count. Note that the resulting AC estimates would depend on the choice of neighborhood size and the drop off rate of the Gaussian kernel. We defer sensitivity analysis on these parameters to future work.

2.3 Connectivity Consistency Assessment

We evaluate the consistency between the various FC and AC estimates described in Section 2.1 and 2.2 using two criteria. The first is simply the correlation between the FC and AC estimates averaged over subjects. The second is based on the incorporation of the FC and AC estimates as priors for task activation detection using the model that we proposed in [9]. The assumption is that if the FC and AC estimates are consistent, then using them as priors should result in the same brain areas detected as activated. To safely base our assessment on activation detection, we employed the maximum-t permutation test [10] to enforce strict control on false positive rate. Under the model in [9], the maximum a posteriori mean of the activation effects is given by:

$$\mathbf{M} = \left(\mathbf{V}_1^{-1} + \alpha \mathbf{V}_2^{-1}\right)^{-1} \mathbf{V}_1^{-1} \mathbf{Y} \mathbf{X}^T \left(\mathbf{X} \mathbf{X}^T\right)^{-1}, \tag{2}$$

where \mathbf{Y} is a $d \times n$ matrix containing task fMRI time courses of d brain areas of a given subject. \mathbf{X} is a $m \times n$ regressor matrix, where m is the number of experimental conditions. \mathbf{V}_1 and \mathbf{V}_2 are $d \times d$ covariance matrices of \mathbf{Y} and the activation effects, respectively. We assume $\mathbf{V}_1 = I_{d \times d}$ as conventionally done and \mathbf{V}_2 is where we inject different connectivity priors. The influence of the prior on \mathbf{M} is controlled by α , which we optimize based on model evidence [9]. We note that the model in [9] assumes \mathbf{V}_2^{-1} is positive definite, hence theoretically, only FC estimates, $\hat{\mathbf{A}}$, generated by (1) can be directly employed for this model. In practice, even if \mathbf{V}_2^{-1} is only positive semi-definite, with $\mathbf{V}_1 = I_{d \times d}$, the optimal α derived based on model evidence would ensure $\mathbf{V}_1^{-1} + \alpha \mathbf{V}_2^{-1}$ is invertible. This observation is particularly important, since as we will now show, it enables fiber count, ill-conditioned Pearson’s correlation matrix, and combinations of these connectivity estimates to be integrated into activation effect estimation. Specifically, let \mathbf{W} be a $d \times d$ matrix with non-negative elements, \mathbf{W}_{ij} , reflecting the connectivity between brain areas i and j , and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ be the corresponding weighted graph Laplacian, which is always positive semi-definite. \mathbf{D} is a $d \times d$ diagonal matrix with $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$. We claim here that \mathbf{L} is a useful prior that permits connectivity estimates to be meaningfully integrated into activation effect estimation when substituted into \mathbf{V}_2^{-1} . For intuition, consider the simplified scenario in which $m = 1$ and the columns of \mathbf{X} are orthonormal. One can easily show

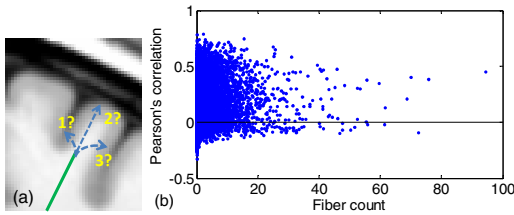
that the negative log posterior distribution of the activation effects comprises a term $\mathbf{a}^T \mathbf{L} \mathbf{a}$, where \mathbf{a} corresponds to activation effects. Since $\mathbf{a}^T \mathbf{L} \mathbf{a} = (\sum_{ij} (\mathbf{a}_i - \mathbf{a}_j)^2 \mathbf{W}_{ij})/2$, minimizing the negative log posterior probability penalizes discrepancies between the activation effect estimates of connected brain areas. Given this setup, AC estimates can be easily integrated into activation effect estimation by setting \mathbf{W} to fiber count. For Pearson's correlation, \mathbf{C} , one would normally take the matrix inverse and directly assign it to \mathbf{V}_2^{-1} . However, since $t < d$ for typical RS-fMRI data, matrix inversion would be unstable. We thus propose zeroing the negative elements of \mathbf{C} and using the resulting matrix as \mathbf{W} . The rationale is that negative FC estimates are usually associated with less anatomical support, i.e. lower fiber count, hence we hypothesize that they are more likely to be false correlations [5]. Building on this intuition, further zeroing elements of \mathbf{C} with no anatomical support might help remove some of the noise-induced correlations. Conversely, spurious fiber tracts induced by tractography errors have no functional relevance. Thus, retaining only elements of the fiber count matrix with functional support based on the sparsity pattern of $\hat{\mathbf{A}}$ might improve activation detection. We assess the effectiveness of these simple FC-AC integration schemes by setting \mathbf{W} to the corresponding connectivity estimates.

3 Materials

60 healthy subjects were recruited and scanned at multiple centers. Each subject performed 10 experimental tasks similar to those in [11], as fMRI data were acquired over a duration of ~5 min. ~7 min of RS-fMRI data were also collected. Scanning was performed using 3T scanners from multiple manufacturers with TR = 2200 ms, TE = 30 ms, and flip angle = 75°. The task fMRI data were corrected for slice timing and head motions, temporally detrended, and spatially normalized using the SPM8 software. The RS-fMRI data were similarly preprocessed except a band-pass filter with cutoff frequencies at 0.01 to 0.1 Hz was applied. White matter and cerebrospinal fluid confounds were regressed out from the gray matter voxel time courses.

We divided the brain into $N = 500$ parcels by concatenating RS-fMRI time courses across subjects and applying hierarchical Ward clustering. The gray matter voxel time courses within each parcel were averaged to generate parcel time courses. The time courses were then normalized by subtracting the mean and dividing by the standard deviation to account for scanner variability across imaging centers. The choice on N was based on a recent brain surface analysis and macaque monkey studies [12]. From the perspective of stable FC estimation with limited time samples, setting N to a lower value would be more suitable. We defer investigation on how to draw a balance between functional localization and stable FC estimation for future work.

dMRI data were collected from the same 60 subjects with TR = 15000 ms, TE = 104 ms, flip angle = 90°, 32 gradient directions, and b-value = 1300 s/mm². After correcting for eddy currents and head motions using FSL, tensor estimation and fiber tractography based on a single-tensor model were performed using MedINRIA. To better deal with crossing fibers, we also employed the unscented Kalman filter (UKF) tractography algorithm with a two-tensor model [13] for comparison. We warped our parcel template onto each subject's B₀ volume for fiber count computation.

**Table 1.** Correlation between FC and AC estimates.

		Endpoint Processing	Pearson's correlation	Sparse partial correlation
MedINRIA	Extrapolate along tangent		0.1223	0.1347
	Gaussian kernel		0.1455	0.1538
UKF	Extrapolate along tangent		0.0530	0.0591
	Gaussian kernel		0.0731	0.0779

Fig. 1. Connectivity estimate comparisons. (a) Fiber endpoint near gyri has high uncertainty. (b) Pearson's correlation vs. fiber count. (c) Correlation between FC and AC estimates.

4 Results and Discussion

The correlation between various FC and AC estimates are summarized in Fig. 1(c). Fiber count based on tracts generated by MedINRIA with endpoint extrapolation is taken as the baseline AC estimate. Conventional Pearson's correlation and fiber count are not very consistent (Fig. 1(b)) with an average correlation of only 0.1223 across subjects, which is much lower than that reported in [4]. In fact, for the same estimated AC, the estimated FC vary substantially. This finding suggests a high variability in the degree of functional interactions for the same anatomical capacity, but could also be attributed to limitations of the conventional connectivity estimates and data acquisition. Using sparse partial correlation as an FC estimate to control for estimation errors and indirect effects significantly increased the FC-AC correlation based on a Wilcoxon signed rank test (p -value < 0.01). Modeling fiber endpoint uncertainty further increased the FC-AC correlation (p -value < 0.01), but the overall correlation is still low compared to [4], in which RS-fMRI data of 35 min and dMRI data with 128 gradient directions were acquired from 5 subjects. Our results thus indicate that methodological improvements might not be adequate to compensate for the fundamental limitations in data acquisition. Further examining this point, we employed UKF tractography, which exploits tracking history to improve fiber estimation [13], but the resulting FC-AC correlation was lower than MedINRIA. We suspect that the number of gradient directions in our data was insufficient to benefit from UKF tractography.

FC-AC consistency results with activation detection as the assessment criterion are shown in Fig. 2. We used maximum-t permutation [10] to strictly control false positive rate, so that more detections would imply higher sensitivity. The average percentage of parcels detected over 10 experimental conditions and 21 contrasts between these conditions using ordinary least square (OLS) is taken as the baseline. Integrating FC estimated using Pearson's correlation significantly increased sensitivity (Fig. 2(a)) based on a permutation test described in [1]. Minor improvement over Pearson's correlation was observed at lower p -value thresholds by incorporating AC priors with tracts extracted using UKF tractography. A greater increase in sensitivity was found using AC priors with tracts generated by MedINRIA, which was slightly improved by modeling endpoint uncertainty. Overall, highest sensitivity was achieved with sparse inverse covariance as an FC prior. These differences in detection sensitivity again

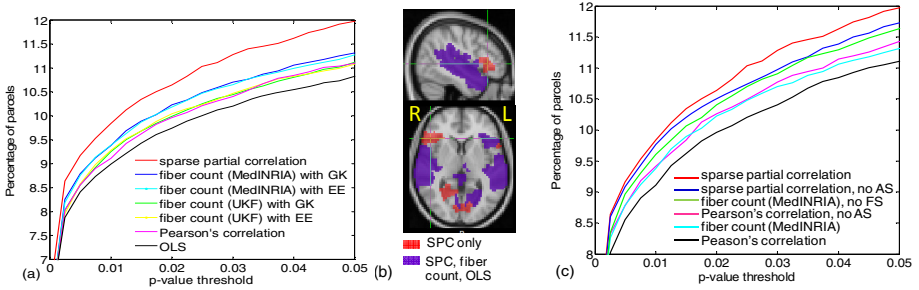


Fig. 2. Activation detection results. (a) % of parcels found activated vs. p-values with FC and AC estimates separately incorporated as priors. GK = Gaussian kernel, EE = endpoint extrapolation. (b) Activation map for an auditory task, p-value < 0.01. SPC = sparse partial correlation. (c) FC and AC estimates combined. AS = anatomical support, FS = functional support.

indicate that the FC and AC estimates are not very consistent. In fact, using AC priors generally detected less distant bilateral activation (Fig. 2(b)), which is likely due to difficulties in tracking fibers across the two hemispheres [7]. Removing Pearson's correlation with no anatomical support significantly increased sensitivity compared to directly using Pearson's correlation as an FC prior (Fig. 2(c)). We speculate that this increase arises from more false correlations and indirect effects being eliminated compared to true correlations on average. Retaining only fiber count with functional support also significantly enhanced sensitivity, which is likely due to spurious tracts being removed. However, when we zeroed out $\hat{\Lambda}_{ij}$ that has no anatomical support, detection sensitivity actually reduced. This suggests that the sparse inverse covariance estimates have false correlations and indirect effects reasonably controlled, thus nulling $\hat{\Lambda}_{ij}$ with no anatomical support actually removed important connections missed by the AC estimates. Hence, determining the presence of a functional connection completely based on AC estimates could be detrimental. Instead, penalizing FC estimates with no anatomical support [1] or probabilistically modeling the absence of anatomical connections [2] are likely better strategies for FC-AC integration.

5 Conclusions

We showed on a cohort of 60 subjects that FC and AC estimates are not very consistent for typical RS-fMRI and dMRI data. Integrating them as priors for task activation detection further demonstrated their inconsistencies. Importantly, we illustrated that isolating fibers with functional support as well as removing Pearson's correlation with no anatomical support enhanced activation detection compared to directly using fiber count and Pearson's correlation as priors. This suggests that the inconsistencies actually make integration of FC and AC estimates all the more valuable for connectivity estimation. However, when $\hat{\Lambda}_{ij}$ with no anatomical support was nulled, activation detection worsened. Thus, FC and AC estimates should be combined with caution.

Acknowledgements. This work was supported by the Genim ANR-10-BLAN-0128 grant, NSERC, and the Berkeley-INRIA-Stanford grant. The data were acquired within the IMAGEN project. Jean Baptiste Poline was partly funded by the IMAGEN project, which receives funding from the E.U. Community's FP6, LSHM-CT-2007-037286. This manuscript reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein.

References

1. Ng, B., Varoquaux, G., Poline, J.-B., Thirion, B.: A Novel Sparse Graphical Approach for Multimodal Brain Connectivity Inference. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 707–714. Springer, Heidelberg (2012)
2. Venkataraman, A., Rathi, Y., Kubicki, M., Westin, C.F., Golland, P.: Joint Modeling of Anatomical and Functional Connectivity for Population Studies. *IEEE Trans. Med. Imaging* 31, 164–182 (2012)
3. Damoiseaux, J.S., Greicius, M.D.: Greater than the Sum of its Parts: A Review of Studies Combining Structural Connectivity and Resting-state Functional Connectivity. *Brain Struct. Funct.* 213, 525–533 (2009)
4. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P.: Predicting Human Resting-state Functional Connectivity from Structural Connectivity. *Proc. Nat. Acad. Sci.* 106, 2035–2040 (2009)
5. Skudlarski, P., Jagannathan, K., Calhoun, V.D., Hampson, M., Skudlarska, B.A., Pearlson, G.: Measuring Brain Connectivity: Diffusion Tensor Imaging Validates Resting State Temporal Correlations. *NeuroImage* 43, 554–561 (2008)
6. Marrelec, G., Krainik, A., Duffau, H., Pelegrini-Issac, M., Lehericy, S., Doyon, J., Benali, H.: Partial Correlation for Functional Brain Interactivity Investigation in Functional MRI. *NeuroImage* 32, 228–237 (2006)
7. Jbabdi, S., Johansen-Berg, H.: Tractography: Where Do We Go from Here? *Brain Connectivity* 1, 169–183 (2011)
8. Hsieh, C.J., Sustik, M.A., Dhillon, I.S., Ravikumar, P.: Sparse Inverse Covariance Matrix Estimation Using Quadratic Approximation. In: *Advances in Neural Information Processing Systems*, vol. 24, pp. 2330–2338 (2011)
9. Ng, B., Abugharbieh, R., Varoquaux, G., Poline, J.B., Thirion, B.: Connectivity-informed fMRI Activation Detection. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 285–292. Springer, Heidelberg (2011)
10. Nichols, T., Hayasaka, S.: Controlling the Familywise Error Rate in Functional Neuroimaging: a Comparative Review. *Stat. Methods Med. Research* 12, 419–446 (2003)
11. Pinel, P., Thirion, B., Meriaux, S., Jobet, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S.: Fast Reproducible Identification and Large-scale Databasing of Individual Functional Cognitive Networks. *BioMed. Central Neurosci.* 8, 91 (2007)
12. Van Essen, D.C., Ugurbil, K.: The Future of the Human Connectome. *NeuroImage* 62, 1299–1310 (2012)
13. Malcolm, J.G., Shenton, M.E., Rathi, Y.: Filtered Multitensor Tractography. *IEEE Trans. Med. Imaging* 29, 1664–1675 (2010)

Erratum: Dynamic CT Reconstruction by Smoothed Rank Minimization

Angshul Majumdar^{1,2} and Rabab K. Ward²

¹ Indraprastha Institute of Information Technology, Delhi

² Department of Electrical Engineering, University of British Columbia
angshul@iiitd.ac.in, rababw@ece.ubc.ca

K. Mori et al. (Eds.): MICCAI 2013, Part III, LNCS 8151, pp. 131–138, 2013.
© Springer-Verlag Berlin Heidelberg 2013

DOI 10.1007/978-3-642-40760-4_82

- In the Experimental Evaluation section, paragraph one, it is mentioned that our technique has been compared with NCPICCS. This is true, but the corresponding reference is incorrect. The reference should be [5] and not [7].

Similarly in the last paragraph of the experimental results and the conclusion, it was mentioned that our work was compared with the PICCS technique. This was incorrect, as mentioned above our work was compared with the NCPICCS technique

- The data description for the in-vivo scan (p. 137, paragraph 2) is incorrect. The described data pertains to the dynamic perfusion scan used in [5]. However, this dynamic CT dataset is not publicly available and has not been used in our work. We synthesized a rotational motion from a single in-vivo image-frame published in paper [5]. This image was used for synthesizing the data used in our experiments. The frame-by-frame reconstruction error in Fig. 1 and the reconstructed image in Fig. 2. correspond to the synthesized data (rotational motion) and not the actual perfusion data

The original online version for this chapter can be found at
http://dx.doi.org/10.1007/978-3-642-40760-4_17

Author Index

- Aandal, Gunhild I-558
Abi-Jaoudeh, Nadine I-719
Abolmaesumi, Purang II-173, II-279,
II-361, III-461
Abraham, Alexandre II-607
Abràmoff, Michael D. II-436
Abugharbieh, Rafeef II-682
Acosta, Oscar II-387
Adluru, Nagesh I-300, III-643
Afacan, Onur III-1
Agartz, Ingrid A. II-690
Ahmadi, Seyed-Ahmad II-230
Ahmidi, Narges I-26
Akhondi-Asl, Alireza III-518
Alexander, Andrew L. I-300, I-598,
III-501
Alexander, Daniel C. I-74
Alhrishy, Mazen III-98
Ali, Sahirzeeshan I-412, II-238, III-396
Alipoor, Mohammad I-687
Aljabar, Paul I-574
Allard, Jeremie III-299
Allasonnière, Stéphanie III-592
Altshuler, Lori III-643
Amr, Ali I-1
An, Hongyu III-291, III-527
Anderson, Mark I-316
Ando, Takehiro III-388
Andreassen, Ole A. II-690
Arbel, Tal I-751, III-543
Ardon, Roberto II-99, III-82
Arevalo Ovalle, John Edison II-403
Arnold, Douglas L. III-543
Aro, Hannu T. II-428
Arridge, Simon R. I-147
Ashraf, Ahmed B. II-295
Asman, Andrew J. I-759
Atkins, M. Stella III-453
Atkinson, David I-147, II-132
Audigier, Chloé III-323
Avants, Brian B. III-635
Ayache, Nicholas I-267, I-372, II-328,
III-323, III-477
Aydogan, Dogu Baran II-428
Aylward, Stephen II-534
Azuma, Takashi III-388
Azzabou, Noura III-219
Bagci, Ulas II-559, III-115
Bai, Wenjia III-9
Bailey, James III-315
Bainbridge, Alan II-336
Ball, Gareth I-574
Balocco, Simone II-345
Bansal, Mayank II-567
Barber, Lauren III-380
Barnes, Anna I-147
Basavanhally, Ajay II-238
Bastin, Mark E. I-82
Basu, Sreetama I-396
Baudin, Pierre-Yves III-219
Bauer, Sebastian I-139, I-356, II-116
Bazin, Pierre-Louis II-51
Becker, Carlos I-526
Béjar, Benjamín I-26
Ben Ayed, Ismail I-509
Bendlin, Barbara B. III-501
Benseghir, Thomas I-179
Bhatia, Kanwal III-9
Bhavsar, Arnav III-139
Bhushan, Manav I-316
Bicknell, Colin II-369
Bieth, Marie II-35
Bijnens, Bart H. II-484
Birkbeck, Neil III-235
Block, Tobias III-17
Bloem, Johan L. I-106
Boag, Alexander III-461
Boctor, Emad III-323
Booth, Brian G. III-469
Börnert, Peter I-106
Bötzel, Kai II-230
Bouix, Sylvain I-50
Boumal, Nicolas I-695
Bradley, Andrew P. I-452
Brady, Sir Michael I-187, I-316
Bregasi, Alda III-57
Brieu, Matthias III-307

- Brosch, Tom II-633
 Brown, Michael S. III-437
 Buhmann, Joachim M. II-214
 Burgos, Ninon I-147
 Butakoff, Constantine I-251
- Caballero, Jose III-9
 Cabeen, Ryan P. I-82
 Cabeza, Rafael II-641
 Cai, Weidong I-284, II-303, II-452, II-625
 Camp, Jon J. I-9
 Cantoni, Marco III-413
 Cao, Tian I-259
 Cardoso, Manuel Jorge I-147, II-10, II-287, II-336
 Carey, Frank A. III-429
 Carlier, Pierre G. III-219
 Carnegie, Daniel III-323
 Carneiro, Gustavo I-452
 Carrell, Tom III-98, III-179
 Carvalho, Diego D.B. II-542
 Cash, David M. II-10
 Cates, Joshua II-19
 Cattin, Philippe C. I-18
 Caus, Carles II-345
 Cepek, Jeremy I-348
 Cerrolaza, Juan J. II-641
 Chang, Hang II-91
 Chang, Ping-Lin I-42
 Chang, Silvia II-279
 Chappell, Michael I-316
 Charchut, Sylvia III-643
 Chatelain, Pierre II-230
 Chefd'hotel, Christophe II-1
 Chen, Chao II-157
 Chen, Chen III-106
 Chen, Cheng III-227
 Chen, Elvis C.S. III-363
 Chen, Hanbo I-66, III-626
 Chen, Huaifu I-308, II-311
 Chen, Ken Chung III-251
 Chen, Mei I-420, II-444, II-452
 Chen, Mingqing III-90
 Chen, Tsuhan I-114
 Chen, Yasheng III-291, III-527
 Chen, Yen-Wei I-235
 Cheng, Bo I-275
 Cheng, Ching-Yu II-468
 Cheng, Jian I-155, I-639
- Cheng, Jun III-421
 Cheng, Xiaoyin III-155
 Cheung, Carol Y. II-468
 Chiu, Bernard I-534
 Cho, Baek Hwan I-227
 Choti, Michael III-323
 Chow, Ben III-251
 Christudass, Christhunesa III-396
 Chronik, Blaine A. I-348
 Chu, Chengwen II-165
 Chung, Albert C.S. II-550
 Chung, Moo K. I-300, I-598, III-501, III-643
 Ciompi, Francesco II-345
 Cireşan, Dan C. II-411
 Ciuciu, Philippe II-616
 Coats, Maria III-429
 Cobzas, Dana III-187
 Cohen, Laurent D. II-99, III-82
 Collins, D. Louis I-751, III-543
 Colliot, Olivier I-267
 Comaniciu, Dorin I-1, II-395, III-323
 Commandeur, Frédéric II-387
 Cool, Derek W. II-124
 Cootes, Tim F. II-181
 Cormack, Robert A. III-380
 Correa, Juan Carlos II-387
 Corso, Jason J. III-567
 Cosson, Michel III-307
 Cotin, Stéphane III-299, III-307, III-331
 Counsell, Serena J. I-574
 Criminisi, Antonio I-606, II-262, III-66
 Cruz-Roa, Angel Alfonso II-403
 Cuingnet, Remi II-99, III-82
 Czymmek, Kirk I-259
- d'Alché-Buc, Florence II-271
 Dam, Erik II-246
 Damato, Antonio III-380
 Da Mota, Benoit II-591
 Datar, Manasi II-19
 Dauguet, Julien I-404
 Davatzikos, Christos I-477
 Davidson, Richard J. I-300, I-598
 Davis, Brad III-584
 Davison, Andrew J. I-42
 Dawes, Tim III-9
 Daye, Dania II-295
 de Bruijne, Marleen II-287, II-542

- De Craene, Mathieu II-484
 de Crevoisier, Renaud II-387
 Dehaghani, Azar Rahimi I-501
 Delibaltov, Diana L. I-444
 Delingette, Hervé III-323
 De Luca, Valeria I-518
 Denk, Winfried II-419
 Depeursinge, Adrien II-353
 Dequidt, Jeremie III-307
 Deriche, Rachid I-639
 De Silva, Tharindu II-124
 Desjardins, Benoit III-195
 Detre, John A. III-635
 Deux, Jean-Francois III-267
 de Vecchi, Adelaide II-476
 de Zubicaray, Greig I. III-600
 Diallo, Mouhamadou Nassirou III-307
 Díaz Díaz, Jesús II-347
 Dickson, John I-147
 Dinov, Ivo D. I-655
 Dinse, Juliane II-51
 Dione, Donald P. III-57
 Dirksen, Asger II-287
 Djurovic, Srdjan II-690
 Dohmatob, Elvis II-607
 Donner, René I-219
 Dorfer, Matthias I-219
 Dréan, Gaël II-387
 Dregely, Isabel III-17
 Drew, Mark S. III-453
 Dubois, Jessica I-590
 Duchateau, Nicolas II-484
 Duda, Jeffrey T. III-635
 Duncan, James S. III-57
 Duncan, John S. I-147
 Dürichen, Robert II-108
 Duriez, Christian III-299, III-307,
 III-331
 Durrleman, Stanley I-267
 Dzyubachyk, Oleh I-106, III-147
- Eaton-Rosen, Zach II-336
 Eberl, Stefan I-284
 Edwards, A. David I-574
 Edwards, Philip "Eddie" I-42
 Eledath, Jayan II-567
 El-Zehiry, Noha I-122
 Epstein, Jonathan I. III-396
 Ernst, Floris II-108
- Falcon, Carles I-671
 Fallavollita, Pascal III-371
 Fan, Zheng III-291
 Fang, Ruogu I-114
 Fang, Tong I-122
 Fazli, Ladan III-461
 Feldman, Michael II-295
 Feng, Chaolu I-477
 Feng, Dagan I-284, II-303, II-452
 Fenster, Aaron I-195, I-348, I-534,
 II-124, II-198, II-279, III-559
 Ferizi, Uran I-74
 Ferrante, Enzo III-163
 Feußner, Hubertus I-139, I-356
 Fischl, Bruce I-131, I-631
 Foncubierta-Rodríguez, Antonio II-353
 Forbes, Florence II-616
 Forman, Christoph I-558, II-575, III-17
 Foster, Brent II-559
 Frangi, Alejandro I-251
 Frangi, Alejandro F. II-484
 Franke, Jochen III-227
 Franklin, Jamie I-316
 Freiman, Moti III-1
 Fritsch, Virgile II-591
 Frouin, Vincent II-591
 Fua, Pascal I-526, III-413
 Fuerst, Bernhard I-34
 Fujiwara, Keisuke III-388
 Fujiwara, Michitaka II-165
 Fulham, Michael I-284
 Funka-Lea, Gareth III-74
 Fürst, Sebastian III-17
 Furukawa, Kazuhiro I-332
- GadElkarim, Johnson J. III-643
 Gaed, Mena II-279
 Gagoski, Borjan III-510
 Galesh-Khale, Amir Khojaste II-279
 Gallardo-Pujol, David I-671
 Gambardella, Luca M. II-411
 Gao, Fei I-501
 Gao, Qinquan II-599
 Gao, Xinting II-468
 Gao, Yaozong II-254, II-378, II-649,
 III-251
 Gao, Yixin I-26
 Gardiazabal, José I-163
 Garvin, Mona K. II-436
 Gauthier, Susan A. I-622

- Gee, James C. I-428, II-75, III-635
 Georgescu, Bogdan I-1
 Germanaud, David I-590
 Geyer, Stefan II-51
 Ghanbari, Yasser I-90
 Ghosh, Pratim I-444
 Giannarou, Stamatia II-67
 Gibson, Eli II-279
 Giese, Daniel II-476
 Gilmore, John H. I-703
 Ginsburg, Shoshana II-238
 Giusti, Alessandro II-411
 Gleeson, Fergus I-316
 Glocker, Ben I-606, I-631, II-262, III-66
 Goh, Kim-Fat III-347
 Golland, Polina I-98, III-211
 Gómez, Alberto II-476
 Gomez, Jose A. II-279
 González Ballester, Miguel Angel II-641
 González Osorio, Fabio Augusto II-403
 Good, Sara I-122
 Goodman, Danny III-219
 Gori, Pietro I-267
 Gorman, Robert C. I-485
 Gorman III, Joseph H. I-485
 Gorospe, Giann I-469
 Goto, Hidemi I-332
 Grady, Leo I-122
 Grant, P. Ellen III-510
 Grayeli, Alexis Bozorg III-331
 Grbic, Sasa I-171, II-395
 Greiser, Andreas I-558
 Grimm, Robert I-558, II-575, III-17
 Grisan, Enrico III-259
 Groeschel, Samuel I-663
 Grützner, Paul A. III-227
 Gu, Irene Yu Hua I-687
 Gunn, Roger N. I-211
 Guo, Lei II-665, II-674, II-698, III-608,
 III-617, III-626
 Guo, Ping I-227
 Gupta, Vikash III-477
 Guyot, Alexis III-179
- Haas, Jan I-1
 Haase, Sven I-139, I-356
 Hacihaliloglu, Ilker II-361
 Hager, Gregory D. I-26, III-339
 Haignon, Pascal II-387
 Hajnal, Joseph V. I-574, I-582, II-599
- Hamarnah, Ghassan I-461, II-43,
 III-469
 Hamm, Jihun III-195
 Hamprecht, Fred A. II-419
 Hamy, Valentin II-132
 Han, Junwei II-665, II-698
 Hanson, Jamie L. I-300
 Hao, Zhihui I-227
 Harrison, Adam P. III-235
 Hartmann, Andreas I-267
 Hata, Nobuhiko I-364, III-355
 Hayashi, Yuichiro I-332, II-165
 Haynor, David R. II-262
 He, Tiancheng III-33
 Heimann, Tobias III-49
 Heinrich, Mattias Paul I-187, III-25
 Helbren, Emma II-132
 Herrezuelo, Noemí Carranza II-641
 Hertz-Pannier, Lucie I-590
 Hibar, Derrek P. II-690, III-600
 Higashiura, Keisuke I-235
 Hirooka, Yoshiki I-332
 Holmes III, David R. I-9
 Hong, Yi III-584
 Hoogendoorn, Corné I-251
 Hori, Masatoshi I-235, III-275
 Hornegger, Joachim I-139, I-171, I-356,
 I-558, II-116, II-575, III-17
 Hosseinbor, Ameer Pasha I-598, III-501
 Houle, Helene II-395
 Hu, Mingxing I-542
 Hu, Qiao II-436
 Hu, Xintao II-665, II-698, III-608
 Huang, Heng I-284, II-625
 Huang, Jin II-625
 Huang, Junzhou III-106
 Huang, Xiaojie III-57
 Huh, Seungil I-420, II-59, II-452
 Hüppi, Petra S. I-590
 Hutter, Jana I-558
 Hutter, Jana Maria II-575, III-17
 Hutton, Brian F. I-147
 Hwang, Youngkyoo I-227
 Hyttinen, Jari II-428
- Igel, Christian II-246
 Iglesias, Juan Eugenio I-631, I-727,
 III-576
 Inuma, Gen I-332
 Imani, Farhad II-279

- Intwali, Victor I-590
 Ioannou, Ioanna III-315
 Ionasec, Razvan I-171, II-395, III-49
 Ishikawa, Hiroshi II-444
 Islam, Ali I-509
 Itani, Kazunori III-388
- Jackson, Benjamin M. I-485
 Jägersand, Martin III-187
 Jahanshad, Neda III-600
 Jayarathne, Uditha L. III-363
 Jenkinson, Mark I-187, I-316
 Ji, Songbai III-291
 Jiang, Tianzi I-639
 Jiang, Xi II-674, III-608, III-617, III-626
 Jiao, Jieqing I-211
 Jie, Biao I-275, II-319
 Jiménez del Toro, Oscar Alfonso II-353
 Jin, Changfeng II-698
 Jin, Yan I-655
 John, Matthias I-171, II-395, III-49
 Jojic, Vladimir I-259
 Joskowicz, Leo II-206
 Jürgens, Philipp I-18
- Kadoury, Samuel I-719
 Kallem, Vinutha II-567
 Kamen, Ali I-1, III-323
 Kanade, Takeo I-420, II-59, II-452
 Kang, Xin III-259
 Kapur, Tina III-380
 Karimaghloo, Zahra III-543
 Karp, Karen II-567
 Kato, Takahisa I-364
 Katus, Hugo I-1
 Kayvanpour, Elham I-1
 Kazmitcheff, Guillaume III-331
 Kebebew, Electron III-283
 Kendall, Giles S. II-336
 Kennedy, Gregor III-315
 Keraudren, Kevin I-582
 Khudanpur, Sanjeev I-26
 Kiefer, Berthold III-17
 Kikinis, Ron I-647, II-303
 Kilgus, Thomas I-139, I-356
 Kim, DaeYoung III-243
 Kim, Jung Bae I-227
 Kim, Junghoon III-635
 Kim, Minjeong II-649
- Kim, Sungeun II-690
 Kim, SunHyung II-19
 Kim, Won Hwa III-643
 King, Andrew III-98
 Kishimoto, Jessica III-559
 Kitamoto, Asanobu III-437
 Kitasaka, Takayuki I-332, II-165
 Klein, Stefan II-542
 Knott, Graham III-413
 Knutsson, Hans I-679
 Kobayashi, Etsuko III-388
 Koethe, Ullrich II-419
 Kohannim, Omid III-600
 Köhler, Thomas I-139
 Kontos, Despina II-295
 Konukoglu, Ender I-131, I-606, I-631, II-262
 Kourtzi, Zoe II-657
 Krainik, Alexandre II-616
 Kranzfelder, Michael I-356
 Kriston, Andras I-324
 Kroeger, Thorben II-419
 Kronman, Achia II-206
 Kubicki, Marek I-647
 Kudo, Masayuki I-324
 Kulikova, Maria I-396
 Kumar, Anand III-643
 Kumar, M. Pawan III-219
 Kumar, Puneet III-219
 Kundrat, Dennis III-347
 Kwitt, Roland II-534, III-584
 Kyriakopoulou, Vanessa I-582
- Ladikos, Alexander I-34
 Laidlaw, David H. I-82
 Lamb, Hildo III-147
 Lamb, Peter I-324
 Landman, Bennett A. I-759
 Langs, Georg I-219
 Lasser, Tobias I-163
 Lauritsch, Guenter III-90
 Lauze, François II-246
 Lecomte, Sophie I-267
 Lecomte-Grosbras, Pauline III-307
 Ledig, Christian III-9
 Lee, George II-238, III-396
 Lee, Hyekyoung I-300
 Lee, Joohwi III-203
 Lee, Junghoon III-41

- Lee, Philip K.M. III-251
 Lee, Su-Lin II-369
 Lee, SunMi III-243
 Lee, Won Ki I-227
 Lefèvre, Julien I-590
 Lekadir, Karim I-251
 Lelieveldt, Boudewijn P.F. I-106,
 III-147
 Leow, Alex D. III-643
 Lepetit, Vincent I-526
 Letort, Véronique II-271
 Leviridge, Michael II-279
 Lewis, James I-412
 Li, Chunming I-477
 Li, Gang I-58, I-703, III-251
 Li, Junning I-655
 Li, Kaiming II-674, III-617
 Li, Kang II-157
 Li, Lingjiang II-698
 Li, Wenqi III-429
 Li, Xiang II-698, III-608, III-626
 Li, Xin II-149
 Li, Yang II-319
 Li, Yeqing III-106
 Liang, Xi III-437
 Liao, Hongen III-388
 Liao, Rui I-380
 Liao, Shu II-254, II-649, III-251, III-551
 Liao, Wei I-550
 Lilja, Ylva I-687
 Lin, Ben A. III-57
 Lin, Stephen II-468, III-445
 Lin, Weili I-58, I-703, III-291
 Lindner, Claudia II-181
 Linguraru, Marius George I-340, II-83,
 II-222, II-641, III-259, III-275
 Linte, Cristian A. I-9
 Litt, Harold I-477
 Liu, Feng I-308, II-311
 Liu, Hesheng I-98
 Liu, Jiamin I-243
 Liu, Jianfei I-340, II-83, II-518
 Liu, Jiang II-468, III-421, III-445
 Liu, Jindong II-369
 Liu, Liu I-542
 Liu, Nancy X. III-251
 Liu, Peter I-243
 Liu, Shubao II-526
 Liu, Sidong II-303
 Liu, Tianming I-66, II-665, II-674,
 II-698, III-608, III-617, III-626
 Liu, Xiaofeng I-743
 Liu, Yixun III-283
 Liu, Yu-Ying II-444
 Lombaert, Hervé II-35, II-492, II-509
 Lorenzen, Peter II-460
 Lorenzi, Marco II-328
 Lorza, Andrés M. Arias II-542
 Lu, Le II-518
 Lu, Zhi I-452
 Lucchi, Aurelien III-413
 Lucidarme, Olivier II-99, II-271
 Luo, Xiongbiao I-388
 Lv, Jinglei II-665, III-608, III-617,
 III-626
 Lv, Peili II-698
 Lyu, Ilwoo II-19, III-203
 Macq, Benoît I-695
 Madabhushi, Anant I-412, II-238,
 II-403, III-396
 Madooei, Ali III-453
 Madsen, Joseph I-98
 Maes, Frederik I-566
 Mahapatra, Dwarikanath II-214
 Mahdavi, S. Sara II-173
 Mahé, Jessie I-404
 Mahrooghy, Majid II-295
 Maier, Andreas I-558, II-575
 Maier-Hein, Lena I-139, I-356
 Majdani, Omid III-347
 Majorczyk, Vincent III-299
 Majumdar, Angshul III-131
 Makris, Nikos I-647
 Malandain, Grégoire I-179
 Mangin, Jean-François I-590
 Manjunath, B.S. I-444
 Mansi, Tommaso I-1, II-1, II-395,
 III-323
 Mansoor, Awais II-559
 Marchal, Damien III-307
 Marescaux, Jacques I-372
 Marlow, Neil II-336
 Marrakchi-Kacem, Linda I-267
 Marreiros, Filipe M.M. II-141
 Marron, J.S. III-584
 Martin, Aaron D. III-259
 Martin, Nicholas G. II-690, III-600
 Marvao, Antonio III-9

- Matsumoto, Yoichiro III-388
 Matthies, Philipp I-163
 Mauri, Josepa II-345
 Mayhew, Stephen II-657
 McKenna, Matthew I-243
 McKenna, Stephen J. III-429
 McLeod, A. Jonathan III-363
 McLeod, Kristin II-501
 McMahan, Katie L. II-690, III-600
 Meder, Benjamin I-1
 Medland, Sarah E. II-690
 Mehnert, Andrew J.H. I-687
 Mehrtash, Alireza III-380
 Meijering, Erik III-123
 Melbourne, Andrew II-336
 Mendonca, Paulo I-743, II-526
 Mendonça, Paulo R.S. I-324
 Mendoza, Carlos S. III-259
 Menys, Alex II-132
 Menze, Bjoern H. II-328
 Metaxas, Dimitris II-157
 Mewes, Philip II-1
 Miao, Shun I-380
 Michailovich, Oleg III-510
 Mies, Carolyn II-295
 Mihalef, Viorel III-323
 Mikula, Shawn II-419
 Milletari, Fausto III-371
 Mills, Monte D. II-567
 Miroir, Mathieu III-331
 Misawa, Kazunari I-332, II-165
 Miyahara, Ryoji I-332
 Modat, Marc I-147, II-10, II-287
 Modla, Shannon I-259
 Mollura, Daniel J. II-559, III-115
 Montgomery, Grant W. II-690, III-600
 Montillo, Albert I-743
 Moody, Teena III-643
 Mori, Kensaku I-332, I-388, II-165
 Moritz, Niko II-428
 Morris, William J. II-173
 Mory, Benoit II-99, III-82
 Mountney, Peter II-1, III-49
 Mousavi, Parvin II-279
 Moussa, Madeleine II-279
 Moussavi, Farshid II-460
 Mueller, Kerstin III-90
 Muenzer, Joseph III-291
 Mukherjee, Dipti Prasad I-235
 Mulhern, Robert V. III-1
 Müller, Henning II-353
 Murano, Emi Z. III-41
 Myers, Emmarie III-259
 Nambakhsh, Cyrus M.S. I-509
 Navab, Nassir I-1, I-34, I-163, I-171,
 II-230, II-395, III-155, III-371
 Nawano, Shigeru I-332
 Nayak, Nandita II-91
 Nekolla, Stephan III-17
 Neumann, Dominik I-171
 Ng, Bernard II-682, III-652
 Nguyen, Thanh D. I-622
 Nguyen, Yann III-331
 Nicolau, Stéphane II-1
 Nicolau, Stephane A. I-372
 Nie, Feiping II-625, III-551
 Nielsen, Mads II-246
 Niessen, Wiro J. II-542, III-123
 Niethammer, Marc I-259, I-292, II-27,
 II-328, II-534, III-584
 Nilsson, Daniel I-687
 Nimura, Yukitaka II-165
 Nitsch, Paige L. III-33
 Nolte, Lutz-P. I-18, III-227
 Nosrati, Masoud S. I-461
 Nouranian, Saman II-173
 O'Brien, Joan II-75
 Oda, Masahiro I-332, II-165
 Odille, Freddy II-132
 Oğuz, İpek III-203
 Okada, Kazunori II-222
 Okada, Toshiyuki I-235, II-190, III-275
 Oktay, Ozan II-1
 Okumura, Ichiro I-364
 Okur, Ash I-163
 O'Leary, Stephen III-315
 Onofrey, John A. III-171
 Ooi, Wei Tsang I-396
 Ordy, Vincent I-380
 O'Regan, Declan III-9
 Ortmaier, Tobias III-347
 Ospina, Juan David II-387
 Oto, Aytekin II-254
 Ourselin, Sebastien I-147, I-542, II-10,
 II-287, II-336
 Pace, Danielle II-534
 Padfield, Dirk II-526
 Papademetris, Xenophon III-171

- Papież, Bartłomiej W. I-187, III-25
 Paragios, Nikos III-163, III-219, III-267
 Park, JinHyeong III-243
 Parvin, Bahram II-91
 Patel, Nisha II-67
 Pauly, Olivier II-230
 Pearson, Denise J. II-567
 Pedemonte, Stefano I-147
 Peng, Zi-Wen II-319
 Pennec, Xavier II-328, II-501, III-477
 Penney, Graeme II-476, III-98, III-179
 Peplinski, Brandon II-518
 Pereañez, Marco I-251
 Pernelle, Guillaume III-380
 Peschke-Schmitz, Lara I-598
 Peter, Loïc II-230
 Peters, Craig A. III-259
 Peters, Terry M. I-509, III-363
 Petersen, Jens II-287, II-542
 Petersen, Kersten II-246
 Peyrat, Jean-Marc II-492
 Pfister, Marcus I-380
 Piella, Gemma II-484
 Piuze, Emmanuel II-509
 Plate, Annika II-230
 Plenge, Esben III-123
 Pohl, Kilian M. II-157, III-195
 Poline, Jean-Baptiste II-591
 Poline, Jean Baptiste III-652
 Pollak, Seth D. I-300
 Poot, Dirk H.J. III-123, III-147
 Popuri, Karteek III-187
 Porras, Antonio R. II-484
 Pouch, Alison M. I-485
 Poupon, Cyril I-267
 Powell, Debbie I-259
 Prabhu, Sanjay P. III-518
 Prasoon, Adhish II-246
 Prats-Galino, Alberto I-671
 Prčkovska, Vesna I-671
 Precup, Doina I-751
 Prevost, Raphael II-99, III-82
 Prince, Jerry L. III-41
 Pujol, Sonia II-303
 Punwani, Shonit II-132
 Puonti, Oula I-727
 Pushparajah, Kuberan II-476

 Qiu, Wu I-195, I-534, II-198, III-559
 Quinn, Graham E. II-567

 Racoceanu, Daniel I-396
 Radeva, Petia II-345
 Rafii-Tari, Hedyeh II-369
 Rahmouni, Alain III-267
 Raj, Ashish I-622
 Rajchl, Martin I-195, I-534, II-198
 Rao, Anil I-735
 Rapaka, Saikiran III-323
 Rashid, Sabrina III-461
 Rasoulian, Abtin II-361
 Rathi, Yogesh I-50, I-647, III-510
 Reader, Andrew J. II-35
 Rehg, James M. II-444
 Reich, Daniel S. I-759
 Reijnierse, Monique I-106
 Reisert, Marco III-493
 Rettmann, Maryam E. I-9
 Reyes, Mauricio I-18
 Rigamonti, Roberto I-526
 Ríos, Richard II-387
 Risser, Laurent I-203, III-25
 Rivaz, Hassan III-543
 Robb, Richard A. I-9
 Robben, David I-566
 Robertson, Nicola J. II-336
 Rodoplu, Volkan I-444
 Rodrigues, Paulo I-671
 Rohling, Robert N. II-361
 Rohr, Karl I-550
 Romagnoli, Cesare I-195, II-124, II-279
 Romain, Blandine II-99, II-271
 Rosa, Benoît I-404
 Rosen, Mark II-295
 Rosenbaum, Kenneth II-222
 Rossitti, Sandro II-141
 Rouet, Laurence II-99, II-271
 Roy, Sharmili III-437
 Rueckert, Daniel I-542, I-574, I-582,
 I-735, II-165, II-599, III-9
 Rutherford, Mary I-582

 Sabuncu, Mert Rory I-131, III-576
 Sadeghi, Maryam III-453
 Sadowski, Samira M. III-283
 Safdar, Nabile III-259
 Sahani, Dushyant V. I-324
 Sahin, Mustafa III-518
 Sahn, David J. I-493
 Sakuma, Ichiro III-388
 Salcudean, Septimiu E. II-173, III-461

- Samaras, Dimitris II-607
 Sanelli, Pina C. I-114
 Sasaki, Akira III-388
 Sasaki, Kosuke I-324
 Sato, Yoshinobu I-235, II-190, III-275
 Savadjiev, Peter I-50
 Sawhney, Harpreet II-567
 Saykin, Andrew J. II-625, II-690
 Schaefer, Stacey M. I-598
 Schaeffter, Tobias II-476
 Schäfer, Andreas II-51
 Schenck, John F. I-743
 Scherrer, Benoît I-695
 Scherrer, Benoit III-518
 Schirmer, Markus I-574
 Schmah, Tanya I-203
 Schmidhuber, Jürgen II-411
 Schmidt, Ehud III-380
 Schmitt, Peter I-558
 Schnabel, Julia A. I-187, I-211, I-316,
 II-10, III-25
 Schneider, Armin I-356
 Schneider, Torben I-74
 Schoebinger, Max II-395
 Schüffler, Peter J. II-214
 Schultz, Robert T. I-50, I-90
 Schultz, Thomas I-663
 Schuman, Joel S. II-444
 Schwaiger, Markus III-17
 Schwartzman, Armin III-518
 Schweikard, Achim II-108
 Sedaghat-Hamedani, Farbod I-1
 Seethamraju, Ravi Teja III-380
 Sehgal, Chandra M. I-485
 Seiler, Christof II-501
 Sermesant, Maxime II-501
 Setsompop, Kawin III-510
 Shah, Amit I-34
 Shahim, Kamal I-18
 Sharma, Kanishka I-34, I-163
 Sharma, Puneet III-323
 Sharp, Gregory C. III-211
 Shavakh, Shadi I-534
 Shen, Dinggang I-58, I-155, I-275, I-292,
 I-308, I-639, I-703, II-27, II-254, II-311,
 II-319, II-378, II-583, II-649, II-674,
 III-139, III-251, III-527, III-551
 Shen, Li II-625, II-690
 Shen, Steve GF III-251
 Shen, Xiaobo I-622
 Shen, Yuan II-657
 Shenton, Martha I-647
 Shi, Feng I-58, I-155, I-703, III-251
 Shi, Kuangyu III-155
 Shi, Wenzhe I-542, III-9
 Shi, Yonggang I-655, III-485
 Shiroishi, Toshihiko III-437
 Siddiqi, Kaleem II-35, II-509
 Siemens, D. Robert II-279
 Siemens, Robert III-461
 Simon, Antoine II-387
 Simpson, Ivor J.A. II-10
 Simpson, John II-476
 Singh, Vikas III-643
 Sinusas, Albert J. III-57
 Sitges, Marta II-484
 Skibbe, Henrik III-493
 Smedby, Örjan II-141
 Smith, Alex R. I-50, I-90
 Smith, Seth A. I-759
 Smith, William I-444
 Sofka, Michal III-235, III-243
 Soler, Luc I-372, II-1
 Song, Sang-Eun I-364, III-355
 Song, Xubo I-493
 Song, Yang I-284, II-303, II-452
 Spadinger, Ingrid II-173
 Spellman, Paul T. II-91
 Sporryng, Jon II-509
 Staib, Lawrence H. III-171
 Stambolian, Dwight II-75
 Staring, Marius I-106
 Steen, Henning I-1
 Stein, Jason L. II-690, III-600
 Sterkers, Olivier III-331
 Stone, Maureen III-41
 Stone, Richard A. II-567
 Stoyanov, Danail I-42
 Stufflebeam, Steven M. I-98
 Styner, Martin A. II-19, III-203
 Su, Hai I-436
 Su, Hang I-420
 Subbanna, Nagesh K. I-751
 Suetens, Paul I-566
 Sugano, Nobuhiko II-190
 Suito, Tomoaki I-332
 Suk, Heung-II II-311, II-583
 Summar, Marshall II-222
 Summers, Ronald M. I-243, I-340, II-83,
 II-518, III-275, III-283

- Sun, Yue I-195, II-198
 Sunaert, Stefan I-566
 Sutterer, Matt I-598
 Sweet, Andrew I-98
 Sze, Raymond II-222
 Székely, Gábor I-518
 Sznitman, Raphael III-413
- Tada, Yukio II-190
 Takabe, Manabu I-485
 Takagi, Shu III-388
 Takao, Masaki II-190
 Tam, Roger I-614, II-633
 Tamura, Masaru III-437
 Tan, Ek.T. I-743
 Tanaka, Naoro I-98
 Tang, Lisa Y.W. II-43
 Tanner, Christine I-518
 Tao, Dacheng III-421
 Tao, Lingling III-339
 Tao, Qian III-147
 Taquet, Maxime I-695, III-518
 Tardif, Christine Lucas II-51
 Tariq, Maira I-74
 Taylor, Stuart II-132
 Teare, Julian II-67
 Teh, Bin S. III-33
 Tek, Huseyin III-74
 Tempany, Clare III-355
 Thiagarajah, Shankar II-181
 Thijs, Vincent I-566
 Thirion, Bertrand II-591, II-607,
 III-592, III-652
 Thompson, Paul M. I-655, II-690,
 III-600
 Tielbeek, Jeroen A.W. II-214
 Tiño, Peter II-657
 Toga, Arthur W. I-655, III-600
 Tokuda, Junichi III-355
 Tomiyama, Noriyuki I-235, II-190,
 III-275
 Tong, Tong II-599
 Tran, Giang III-485
 Tschannen, Michael I-518
 Tuncali, Kemal III-355
 Tung, Leslie I-469
 Turner, Robert II-51
- Ukwatta, Eranga I-534, II-198, III-559
 Uzunbas, Mustafa Gökhan II-157
- Vaillant, Régis I-179
 Valdes, Pablo A. I-719
 van der Geest, Rob J. I-106, III-147
 van der Kouwe, Andre I-131
 van der Lugt, Aad II-542
 van Dijk, Anouk C. II-542
 Van Leemput, Koen I-631, I-727,
 III-576
 van Reekum, Carien M. I-598
 Vargas Gómez, María Isabel II-353
 Varnavas, Andreas III-98, III-179
 Varoquaux, Gaël II-591
 Varoquaux, Gael II-607, III-652
 Vedula, S. Swaroop I-26
 Veeman, Michael I-444
 Veltri, Robert III-396
 Vemuri, Anant Suraj I-372
 Venkataraman, Archana I-98
 Vercauteren, Tom I-404
 Verma, Ragini I-50, I-90
 Vialard, François-Xavier I-203
 Vidal, René I-26, I-469, III-339
 Villanueva, Arantxa II-641
 Villien, Marjorie II-616
 Villoslada, Pablo I-671
 Vincent, Thomas II-616
 Viswanathan, Akila III-380
 Vogel, Jakob I-163
 Voigt, Ingmar II-395
 Vos, Franciscus M. II-214
- Wachinger, Christian III-211
 Waehnert, Miriam II-51
 Wallis, Gillian A. II-181
 Wang, Chunliang II-141
 Wang, Danny J. I-655
 Wang, Hongzhi I-485, I-711, III-535
 Wang, Junchen III-388
 Wang, Li I-58, I-155, I-703, III-251
 Wang, Linwei I-501
 Wang, Lu II-567
 Wang, Qian II-649, III-551
 Wang, Qiang I-227
 Wang, Shijun I-243, I-340, II-83, II-518
 Wang, Wei III-380
 Wang, Xiaogang I-284, II-303
 Wang, Xiaotao I-227
 Wang, Yan I-428
 Wang, Yang III-90
 Wang, Yu II-460

- Ward, Aaron D. II-124, II-279
 Ward, Rabab K. III-131
 Warfield, Simon K. I-695, III-1, III-518
 Warnking, Jan II-616
 Wassermann, Demian I-647
 Wasza, Jakob I-139, I-356, II-116
 Wee, Chong-Yaw I-292, I-308, II-311, II-319
 Wei, Zhuoshi I-243, II-518
 Wein, Wolfgang I-34
 Weiner, Michael W. III-600
 Weisbrod, Allison B. III-283
 Weiss, Nick I-735
 Wells, Williams III-380
 Westin, Carl-Fredrik I-50, I-647, I-679, III-510
 Wheeler-Kingshott, Claudia A.M. I-74
 Whitaker, Ross II-19
 Wijewickrema, Sudanthi III-315
 Wilkinson, J. Mark II-181
 Wilms, Guy I-566
 Wissel, Tobias II-108
 Witz, Jean-Francois III-307
 Wollstein, Gadi II-444
 Wolz, Robin II-599
 Wong, Damon Wing Kee II-468, III-421, III-445
 Wong, Stephen T. III-33
 Wong, Tien Yin II-468, III-421
 Woo, Jonghye III-41
 Woolrich, Mark W. II-10
 Worbe, Yulia I-267
 Wörz, Stefan I-550
 Wright, Margaret J. II-690, III-600
 Wu, Guorong II-649, III-139, III-551
 Wu, Wei III-567
 Wu, Yu-Chien III-501

 Xia, James J. III-251
 Xiang, Bo III-267
 Xiao, Rui I-428
 Xie, Weiguo III-227
 Xing, Fangxu III-41
 Xing, Fuyong I-436, III-404
 Xu, Dong II-468, III-445
 Xu, Hao III-592
 Xu, Huanhuan II-149
 Xu, Jianrong I-534
 Xu, Jingjia I-501
 Xu, Yanwu II-468, III-421, III-445

 Xu, Ziyue II-559
 Xue, Zhong III-33

 Yan, Jin III-251
 Yan, Jingwen II-625
 Yan, Michelle I-122
 Yang, Guang-Zhong II-67, II-369
 Yang, Lin I-436, III-404
 Yao, Jianhua I-340, II-83, III-283
 Yap, Pew-Thian I-155, I-292, I-639, II-27, III-527
 Ye, Dong Hye I-606, III-195
 Ye, Menglong II-67
 Yin, Fengshou III-421
 Yokota, Futoshi II-190
 Yoldemir, Burak II-682
 Yonetsuji, Taizan III-388
 Yoo, Youngjin I-614
 Yoshinaka, Kiyoshi III-388
 Younes, Laurent I-469
 Yuan, Jing I-195, I-534, II-124, II-198, III-559
 Yushkevich, Paul A. I-485, I-711, III-535

 Zand, Dina J. II-222
 Zappella, Luca III-339
 Zenge, Michael O. II-575
 Zeppenfeld, Katja III-147
 Zetting, Oliver I-1
 Zhan, Yiqiang II-378
 Zhang, Daoqiang I-275, III-551
 Zhang, Hui I-74
 Zhang, Jianguo III-429
 Zhang, Li I-380, II-1
 Zhang, Pei I-292, II-27
 Zhang, Shaoting II-157
 Zhang, Shu III-626
 Zhang, Tuo I-66, II-665, II-674, III-608, III-617, III-626
 Zhang, Weidong II-518
 Zhang, Xin II-665, III-608
 Zhang, Zhijun I-493
 Zhao, Dazhe I-477
 Zhao, Liang III-567
 Zhao, Qian II-222
 Zheng, Guoyan III-227
 Zheng, Wei-Shi III-429
 Zheng, Yefeng III-74, III-90
 Zheng, Yuanjie I-428, II-75

Zhizhina, Elena I-396

Zhou, S. Kevin I-122, III-243

Zhou, Yun II-452, III-315

Zhu, Dajiang II-665, II-674, III-608,
III-617, III-626

Zhu, Ning II-550

Zhuang, Xiahai I-542, III-9

Ziegler, Sibylle I. III-17, III-155

Zikic, Darko I-606, I-631, II-262, III-66

Zubicaray, Greig I. de II-690