

Gábor Székely
Horst K. Hahn (Eds.)

LNCS 6801

Information Processing in Medical Imaging

22nd International Conference, IPMI 2011
Kloster Irsee, Germany, July 2011
Proceedings



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

Gábor Székely Horst K. Hahn (Eds.)

Information Processing in Medical Imaging

22nd International Conference, IPMI 2011
Kloster Irsee, Germany, July 3-8, 2011
Proceedings



Springer

Volume Editors

Gábor Székely
Swiss Federal Institute of Technology
Computer Vision Laboratory
Medical Image Analysis and Visualization Group
ETH-Zentrum, Sternwartstr. 7, 8092 Zurich, Switzerland
E-mail: szekely@vision.ee.ethz.ch

Horst K. Hahn
Fraunhofer MEVIS
Universitätsallee 29, 28359 Bremen, Germany
E-mail: horst.hahn@mevis.fraunhofer.de

ISSN 0302-9743
ISBN 978-3-642-22091-3
DOI 10.1007/978-3-642-22092-0
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-22092-0

Library of Congress Control Number: 2011929993

CR Subject Classification (1998): I.4, I.5, I.2.5-6, H.3, J.1, J.3, I.3

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

© Springer-Verlag Berlin Heidelberg 2011

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

The use of general descriptive names, registered names, trademarks, 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.

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

It is a great pleasure and an honor for us to present the proceedings of the 2011 International Conference on Information Processing in Medical Imaging (IPMI), the 22nd in the series after the successful meetings held in Kerkrade, The Netherlands, in 2007 and in Williamsburg, VA, USA, in 2009. Biannually, IPMI brings together excellent young investigators, experienced researchers and some of the old stagers in medical image formation, analysis and interpretation. After the first meeting in 1969, IPMI has developed into a highly prestigious event and is said to be the longest-running international scientific workshop on medical image analysis. Many important developments in the field were first presented at IPMI and, with its unique format that allows for intensive and comprehensive discussion of new ideas and a variety of clinical applications, the series was always committed to a rigorous scientific approach to information processing in medical imaging.

IPMI 2011 was held during July 3–8, 2011 at the Irsee Monastery in Bavaria, Germany. Topics of the conference include image and signal processing, shape representation and analysis, image registration and fusion, functional and molecular imaging, computational physiology, statistical and mathematical models, computer-aided detection and image interpretation, image reconstruction, objective assessment of image quality, data visualization, and novel image acquisition methods. Most of these topics were covered by a relatively small number of talks within single-track sessions plus a number of poster presentations. We received 224 full-length submissions before the deadline in December 2010, from which we selected 24 for oral presentation and 39 as posters during the all-plenary five-day conference. This corresponds to an overall acceptance rate of 28%. All papers were carefully judged by at least three reviewers, each of whom performed at least nine reviews and also provided a relative ranking of the reviewed papers. On this basis, the paper selection committee assessed all papers in a two-stage process that focused on clarity of presentation, justification of the methodological approach, scientific rigor, quality and depth of evaluation, and novelty. The selection of the best papers was a difficult task, but based on the high-quality reviews, a unanimous decision could finally be made. The quality of the submissions was very high, such that due to size limitations of the conference, a number of valuable submissions could unfortunately not be accepted.

The number of active researchers permitted to attend was limited to just above 100, and like past meetings, the unique study group concept was implemented to foster deep scientific exchange. Each participant was member of one study group, which focused on reading and discussing two papers in advance. Within its respective session, the study group led the discussion after each author's presentation before it was opened to the plenum. As an important IPMI rule, in-depth discussions were allowed to stretch far beyond the allocated session

time to permit a detailed understanding of the presented paper, including its limitations and comparison to existing methodology. In addition, this year we introduced the special focus session as a new concept. Selected poster contributions on topics of great common interest were presented in a summary fashion by the members of the study group, while offering the authors the opportunity to reply, clarify raised issues, and comment on the assessment before the discussion was opened to the audience. This highly interactive mode greatly contributed to intensifying and broadening the discussion of high-interest papers. As an additional stimulus, the François Erbsmann Prize was awarded for the best contribution by a young scientist giving an IPMI talk for the first time. This year, over half the participants attended their first IPMI and among the 24 first authors of oral presentations, 20 were eligible for the Erbsmann Prize.

As a tradition of IPMI, a remote location was chosen in order to intensify the participants' interchange of ideas. In the middle of Bavaria and at the foot of the mighty Alps lies the former Benedictine monastery Irsee. The Irsee Monastery, whose current buildings were constructed in the Baroque era, was founded by hermits on the Irsee Castle Mountain in 1182 and remained active until the nineteenth century. In addition to prayer and spirituality, the main focuses of monastery life at Irsee were scientific knowledge and education. In the eighteenth century, members of Irsee included renowned thinkers in the natural sciences, philosophy, and music. The monastery hosted an acclaimed mathematical museum, and this old wisdom lingered and inspired the IPMI attendees when walking the same corridors as the old monks. In addition to visiting the monastery's brewery, one afternoon was devoted to refreshing our minds and interacting informally during a hike through the beautiful surroundings of the monastery, while some attendees visited the fairy-tale castle Neuschwanstein 50 kilometers south of Irsee. We also held the traditional "US versus the Rest of the World" soccer match on Wednesday evening.

In these proceedings, IPMI 2011 papers are published in the order of their presentation at the meeting and we hope that they will remain an invaluable source of information and reminder for the participants. For those who could not attend, they provide an excellent overview of some of the best current research available in information processing in medical imaging and an encouragement to participate in the next IPMI, which will be held in the USA in 2013. Please visit www.ipmi-conference.org for up-to-date information.

July 2011

Gábor Székely
Horst K. Hahn

Acknowledgements

The organization of a meeting is always a large team effort and the Co-chairs of IPMI 2011 are grateful for the many individuals who enthusiastically supported the preparation of the meeting. First of all, we would like to thank the members of the Scientific Review Committee for the careful reviewing of a large number of manuscripts and their invaluable expert advice for the rather difficult task of identifying the best papers out of many excellent submissions. We are very grateful to the members of the Paper Selection Committee, Jerry Prince, Chris Taylor and Baba Vemuri, who spent enormous effort leading to an even-handed selection and the compilation of an outstanding final program. We also would like to thank many organizers of previous IPMI meetings, especially Jerry Prince, Chris Taylor, Nico Karssemeijer, Dzung Pham and Kyle Myers for their advice and support.

We greatly acknowledge Derek Jones taking all his time and effort to provide us with an outstanding keynote lecture on multi-spectral imaging of white matter and the challenges of quantifying connectivity.

The organization of this meeting would not have been possible without the support of many colleagues at the Computer Vision Laboratory of the ETH Zürich and Fraunhofer MEVIS. We would like to wholeheartedly thank Christoph Brachmann, Julien Egger, Ola Friman, Nils Papenberg, Bram Platel, Matthias Schneider and Christine Tanner for their enormous work and enthusiasm, spending countless hours on the numerous tasks related to all aspects of preparing IPMI 2011. We also thank our colleagues forming the IPMI Staff for spending a week at the conference venue helping to run the meeting in a smooth and effective fashion, as well as Iris Auer from the Irsee Monastery for hosting us so well.

Finally, we would like to acknowledge the following organizations for their generous financial support:

NCCR Co-Me, the Swiss National Center of Competence in Research on
Computer-Aided and Image-Guided Medical Interventions

Bayer HealthCare, Diagnostic Imaging, Berlin, Germany

MeVis Medical Solutions AG, Bremen, Germany

Siemens AG, Healthcare Sector, Erlangen, Germany

CIMST, Zurich Center of Imaging Science and Technology, Switzerland

François Erbsmann Prizewinners

1987 (Utrecht, The Netherlands): **John M. Gauch**, University of North Carolina, Chapel Hill, NC, USA.

J.M. Gauch, W.R. Oliver and S.M. Pizer: Multiresolution shape descriptions and their applications in medical imaging.

1989 (Berkeley, CA, USA): **Arthur F. Gmitro**, University of Arizona, Tucson, AZ, USA.

A.F. Gmitro, V. Tresp, V. Chen, Y. Snell and G.R. Gindi: Video-rate reconstruction of CT and MR images.

1991 (Wye, Kent, UK): **H. Isil Bozma**, Yale University, New Haven, CT, USA.

H.I. Bozma and J.S. Duncan: Model-based recognition of multiple deformable objects using a game-theoretic framework.

1993 (Flagstaff, AZ, USA): **Jeffrey A. Fessler**, University of Michigan, Ann Arbor, MI, USA.

J.A. Fessler: Tomographic reconstruction using information-weighted spline smoothing.

1995 (Brest, France): **Maurits K. Konings**, University Hospital, Utrecht, The Netherlands.

M.K. Konings, W.P.T.M. Mali and M.A. Viergever: Design of a robust strategy to measure intravascular electrical impedance.

1997 (Poultney, VT, USA): **David Atkinson**, Guys Hospital, London, UK.

D. Atkinson, D.L.G. Hill, P.N.R. Stoyke, P.E. Summers and S.F. Keevil: An autofocus algorithm for the automatic correction of motion artifacts in MR images.

1999 (Visegrad, Hungary): **Liana M. Lorigo**, Massachusetts Institute of Technology, Cambridge, MA, USA.

L.M. Lorigo, O. Faugeras, W.E.L. Grimson, R. Keriven, R. Kikinis and C.-F. Westin: Codimension 2 geodesic active contours for MRA segmentation.

2001 (Davis, CA, USA): **Viktor K. Jirsa**, Florida Atlantic University, FL, USA.

V.K. Jirsa, K.J. Jantzen, A. Fuchs and J.A. Scott Kelso: Neural field dynamics on the folded three-dimensional cortical sheet and its forward EEG and MEG.

2003 (Ambleside, UK): **Guillaume Marrelec**, INSERM, France.

G. Marrelec, P. Ciuciu, M. Pelegrini-Issac and H. Benali: Estimation of the hemodynamic response function in event-related functional MRI: directed acyclic graphs for a general Bayesian inference framework.

2005 (Glenwood Springs, Colorado, USA): **Duygu Tosun**, Johns Hopkins University, Baltimore, USA.

D. Tosun and J.L. Prince: Cortical surface alignment using geometry-driven multispectral optical flow.

2007 (Kerkrade, The Netherlands): **Ben Glocker**, Technische Universität München, Garching, Germany.

B. Glocker, N. Komodakis, N. Paragios, G. Tziritas and N. Navab: Inter- and intramodal deformable registration: continuous deformations meet efficient optimal linear programming.

2009 (Williamsburg, Virginia, USA): **Maxime Descoteaux**, NeuroSpin, IFR 49 CEA Saclay, France.

M. Descoteaux, R. Deriche, D. Le Bihan, J.-F. Mangin and C. Poupon: Diffusion propagator imaging: using Laplace's equation and multiple shell acquisitions to reconstruct the diffusion propagator.

Organization

Chairs

Gábor Székely	Swiss Federal Institute of Technology Zürich, Switzerland
Horst K. Hahn	Fraunhofer MEVIS, Institute for Medical Image Computing, Bremen, Germany

Paper Selection Committee

Jerry L. Prince	Johns Hopkins University, USA
Chris Taylor	The University of Manchester, UK
Baba C. Vemuri	University of Florida, Gainesville, USA

Scientific Review Committee

Daniel C. Alexander	University College London, UK
Amir A. Amiri	University of Louisville, USA
Suyash P. Awate	University of Utah, USA
Christian Barillot	IRISA/CNRS, France
Pierre-Louis Bazin	Johns Hopkins University, USA
Sylvain Bouix	Brigham and Women's Hospital, Harvard Medical School, USA
Djamal Boukerroui	Université de Technologie de Compiègne, France
Michael Brady	University of Oxford, UK
Elizabeth Bullitt	University of North Carolina, USA
Owen T. Carmichael	University of California, Davis, USA
Philippe Cattin	University of Basel, Switzerland
Gary E. Christensen	University of Iowa, USA
Ela Claridge	The University of Birmingham, UK
Christos Davatzikos	University of Pennsylvania, USA
Benoit M. Dawant	Vanderbilt University, USA
Marleen de Bruijne	Erasmus MC - University Medical Center Rotterdam, The Netherlands
James S. Duncan	Yale University, USA
Bernd Fischer	Institute of Mathematics and Image Computing, University of Lübeck, Germany
P. Thomas Fletcher	University of Utah, USA
Luc Florack	Eindhoven University of Technology, The Netherlands

Alejandro F. Frangi	Universitat Pompeu Fabra, Spain
Ola Friman	Fraunhofer MEVIS, Bremen, Germany
James C. Gee	University of Pennsylvania, USA
Guido Gerig	University of Utah, USA
Polina Golland	Massachusetts Institute of Technology, USA
Michael L. Goris	Stanford University School of Medicine, USA
Matthias Guenther	Fraunhofer MEVIS, Bremen, Germany
Jo Hajnal	Imperial College London, UK
Joachim Hornegger	University of Erlangen, Germany
Nico Karssemeijer	Radboud University Nijmegen Medical Centre, The Netherlands
Frithjof Kruggel	University of California, Irvine, USA
Jan Kybic	Czech Technical University, Czech Republic
Rasmus Larsen	Technical University of Denmark, DTU, Denmark
Richard M. Leahy	University of Southern California, USA
Boudewijn P.F. Lelieveldt	Leiden University Medical Center, The Netherlands
Gabriele Lohmann	Max Planck Institute for Human Cognitive and Brain Sciences, Germany
Frederik Maes	Katholieke Universiteit Leuven, Belgium
Jean-Francois Mangin	CEA, France
Charles R. Meyer	University of Michigan, USA
Mads Nielsen	University of Copenhagen, Denmark
Wiro Niessen	Erasmus Medical Center Rotterdam, The Netherlands
Alison Noble	University of Oxford, UK
Sebastien Ourselin	Centre for Medical Image Computing, University College London, UK
Xenophon Papademetris	Yale University, USA
Nikos Paragios	Ecole Centrale de Paris, France
Xavier Pennec	INRIA, France
Franjo Pernus	University of Ljubljana, Slovenia
Dzung L. Pham	Johns Hopkins University, USA
Stephen M. Pizer	The University of North Carolina at Chapel Hill, USA
Hill Kilian Pohl	University of Pennsylvania, USA
Jinyi Qi	University of California, Davis, USA
Anand Rangarajan	University of Florida, USA
Joseph M. Reinhardt	University of Iowa, USA
Torsten Rohlfing	SRI International, USA
Karl Rohr	DKFZ Heidelberg, Germany
Daniel Rueckert	Imperial College London, UK
Julia A. Schnabel	University of Oxford, UK

Dinggang Shen	The University of North Carolina at Chapel Hill, USA
Pengcheng Shi	Rochester Institute of Technology, USA
Lawrence H. Staib	Yale University, USA
Martin Andreas Styner	The University of North Carolina at Chapel Hill, USA
Jean-Philippe Thiran	Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland
Klaus D. Toennies	Universität Magdeburg, Germany
Carole J. Twining	The University of Manchester, UK
Bram van Ginneken	Radboud University Nijmegen Medical Centre, The Netherlands
Koen Van Leemput	Harvard Medical School, Massachusetts General Hospital, USA
Marijn van Stralen	Image Sciences Institute, The Netherlands
Dirk Vandermeulen	Katholieke Universiteit Leuven, Belgium
Stefan Wörz	DKFZ Heidelberg, Germany
Simon K. Warfield	Harvard Medical School and Children's Hospital, USA
William M. Wells	Harvard Medical School and Brigham and Women's Hospital, USA
Carl-Fredrik Westin	Harvard Medical School and Brigham and Women's Hospital, USA
Guoyan Zheng	University of Bern, Switzerland
Lilla Zollei	Massachusetts General Hospital, USA

IPMI 2011 Board

Harrison H. Barrett
 Christian Barillot
 Aaron B. Brill
 Gary E. Christensen
 Alan C.F. Colchester
 James S. Duncan
 Michael L. Goris
 Nico Karssemeijer
 Richard M. Leahy
 Stephen M. Pizer
 Jerry L. Prince
 Gábor Székely
 Chris Taylor
 Andrew Todd-Pokropek

Table of Contents

Segmentation

Segmentation of Brain Images Using Adaptive Atlases with Application to Ventriculomegaly	1
<i>Navid Shiee, Pierre-Louis Bazin, Jennifer L. Cuzzocreo, Ari Blitz, and Dzung L. Pham</i>	
Adaptive Riemannian Metrics for Improved Geodesic Tracking of White Matter	13
<i>Xiang Hao, Ross T. Whitaker, and P. Thomas Fletcher</i>	
Combining Generative and Discriminative Models for Semantic Segmentation of CT Scans via Active Learning	25
<i>Juan Eugenio Iglesias, Ender Konukoglu, Albert Montillo, Zhuowen Tu, and Antonio Criminisi</i>	
Segmentation of 3D RF Echocardiography Using a Multiframe Spatio-temporal Predictor	37
<i>Paul C. Pearlman, Hemant D. Tagare, Ben A. Lin, Albert J. Sinusas, and James S. Duncan</i>	
Optimal Graph Based Segmentation Using Flow Lines with Application to Airway Wall Segmentation	49
<i>Jens Petersen, Mads Nielsen, Pechin Lo, Zaigham Saghir, Asger Dirksen, and Marleen de Bruijne</i>	
Surface-Region Context in Optimal Multi-object Graph-Based Segmentation: Robust Delineation of Pulmonary Tumors	61
<i>Qi Song, Mingqing Chen, Junjie Bai, Milan Sonka, and Xiaodong Wu</i>	

Statistical Methods

Optimal Weights for Multi-atlas Label Fusion	73
<i>Hongzhi Wang, Jung Wook Suh, John Pluta, Murat Altinay, and Paul Yushkevich</i>	
Characterizing Spatially Varying Performance to Improve Multi-atlas Multi-label Segmentation	85
<i>Andrew J. Asman and Bennett A. Landman</i>	
Discovering Dense and Consistent Landmarks in the Brain	97
<i>Dajiang Zhu, Degang Zhang, Carlos Faraco, Kaiming Li, Fan Deng, Hanbo Chen, Xi Jiang, Lei Guo, L. Stephen Miller, and Tianming Liu</i>	

Auto-alignment of Knee MR Scout Scans through Redundant, Adaptive and Hierarchical Anatomy Detection 111
Yiqiang Zhan, Maneesh Dewan, and Xiang Sean Zhou

Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis 123
Sandy Durrleman, Marcel Prastawa, Guido Gerig, and Sarang Joshi

Learning an Atlas of a Cognitive Process in Its Functional Geometry . . . 135
Georg Langs, Danial Lashkari, Andrew Sweet, Yanmei Tie, Laura Rigolo, Alexandra J. Golby, and Polina Golland

Shape Analysis

Parameterization-Invariant Shape Statistics and Probabilistic Classification of Anatomical Surfaces 147
Sebastian Kurtek, Eric Klassen, Zhaohua Ding, Malcolm J. Avison, and Anuj Srivastava

On the Extraction of Topologically Correct Thickness Measurements Using Khalimsky’s Cubic Complex 159
M. Jorge Cardoso, Matthew J. Clarkson, Marc Modat, and Sebastien Ourselin

Poster Session I (Segmentation, Shape Analysis, Statistical Methods, Image Reconstruction, Microscopic Image Analysis)

A Convex Max-Flow Segmentation of LV Using Subject-Specific Distributions on Cardiac MRI 171
Mohammad Saleh Nambakhsh, Jing Yuan, Ismail Ben Ayed, Kumaradevan Punithakumar, Aashish Goela, Ali Islam, Terry Peters, and Shuo Li

Entangled Decision Forests and Their Application for Semantic Segmentation of CT Images 184
Albert Montillo, Jamie Shotton, John Winn, Juan Eugenio Iglesias, Dimitri Metaxas, and Antonio Criminisi

Venous Tree Separation in the Liver: Graph Partitioning Using a Non-Ising Model 197
Thomas O’Donnell, Jens N. Kaftan, Andreas Schuh, Christian Tietjen, Grzegorz Soza, and Til Aach

Efficient Algorithms for Segmenting Globally Optimal and Smooth Multi-surfaces 208
Lei Xu, Branislav Stojkovic, Yongding Zhu, Qi Song, Xiaodong Wu, Milan Sonka, and Jinhui Xu

Graph Cuts with Invariant Object-Interaction Priors: Application to Intervertebral Disc Segmentation	221
<i>Ismail Ben Ayed, Kumaradevan Punithakumar, Gregory Garvin, Walter Romano, and Shuo Li</i>	
CoRPORATE: Cortical Reconstruction by Pruning Outliers with Reeb Analysis and Topology-Preserving Evolution	233
<i>Yonggang Shi, Rongjie Lai, and Arthur W. Toga</i>	
Globally Optimal Tumor Segmentation in PET-CT Images: A Graph-Based Co-segmentation Method	245
<i>Dongfeng Han, John Bayouth, Qi Song, Aakant Taurani, Milan Sonka, John Buatti, and Xiaodong Wu</i>	
Approximations of the Diffeomorphic Metric and Their Applications in Shape Learning	257
<i>Xianfeng Yang, Alvina Goh, and Anqi Qiu</i>	
Anisotropic Diffusion of Tensor Fields for Fold Shape Analysis on Surfaces	271
<i>Maxime Boucher, Alan Evans, and Kaleem Siddiqi</i>	
A Novel Longitudinal Atlas Construction Framework by Groupwise Registration of Subject Image Sequences	283
<i>Shu Liao, Hongjun Jia, Guorong Wu, and Dinggang Shen</i>	
A Probabilistic Framework to Infer Brain Functional Connectivity from Anatomical Connections	296
<i>Fani Deligianni, Gael Varoquaux, Bertrand Thirion, Emma Robinson, David J. Sharp, A. David Edwards, and Daniel Rueckert</i>	
Optimal Initialization for 3D Correspondence Optimization: An Evaluation Study	308
<i>Matthias Kirschner, Sebastian T. Gollmer, Stefan Wesarg, and Thorsten M. Buzug</i>	
White Matter Bundle Registration and Population Analysis Based on Gaussian Processes	320
<i>Demian Wassermann, Yogesh Rathi, Sylvain Bouix, Marek Kubicki, Ron Kikinis, Martha Shenton, and Carl-Fredrik Westin</i>	
Personalization of Pictorial Structures for Anatomical Landmark Localization	333
<i>Vaclav Potesil, Timor Kadir, Günther Platsch, and Michael Brady</i>	

Joint Restoration of Bi-contrast MRI Data for Spatial Intensity Non-uniformities	346
<i>Stathis Hadjidemetriou, Martin Buechert, Ute Ludwig, and Juergen Hennig</i>	
The 2D Analytic Signal on RF and B-Mode Ultrasound Images	359
<i>Christian Wachinger, Tassilo Klein, and Nassir Navab</i>	
A Compressed Sensing Approach for MR Tissue Contrast Synthesis	371
<i>Snehashis Roy, Aaron Carass, and Jerry Prince</i>	
Restoring DIC Microscopy Images from Multiple Shear Directions	384
<i>Zhaozheng Yin, Dai Fei Elmer Ker, and Takeo Kanade</i>	
Identifying Nuclear Phenotypes Using Semi-supervised Metric Learning	398
<i>Shantanu Singh, Firdaus Janoos, Thierry Pécot, Enrico Caserta, Gustavo Leone, Jens Rittscher, and Raghu Machiraju</i>	
Actin Filament Segmentation Using Dynamic Programming	411
<i>Hongsheng Li, Tian Shen, and Xiaolei Huang</i>	

Registration

Multimodal Registration via Spatial-Context Mutual Information	424
<i>Zhao Yi and Stefano Soatto</i>	
Generalized Partial Volume: An Inferior Density Estimator to Parzen Windows for Normalized Mutual Information	436
<i>Sune Darkner and Jon Sporring</i>	
Large Deformation Diffeomorphic Metric Mapping of Orientation Distribution Functions	448
<i>Jia Du, Alvina Goh, and Anqi Qiu</i>	
Schild's Ladder for the Parallel Transport of Deformations in Time Series of Images	463
<i>Marco Lorenzi, Nicholas Ayache, Xavier Pennec, and the Alzheimer's Disease Neuroimaging Initiative</i>	

Poster Session II (Computer Aided Diagnosis, Diffusion Imaging, Functional Brain Analysis, Registration, Other)

Dissimilarity-Based Classification of Anatomical Tree Structures	475
<i>Lauge Sørensen, Pechin Lo, Asger Dirksen, Jens Petersen, and Marleen de Bruijne</i>	

Automated Detection of Junctions Structures and Tracking of Their Trajectories in 4D Images	486
<i>Guanglei Xiong and Lei Xing</i>	
Systematic Assessment of Performance Prediction Techniques in Medical Image Classification: A Case Study on Caliac Disease	498
<i>Sebastian Hegenbart, Andreas Uhl, and Andreas Vécsei</i>	
Detecting and Classifying Linear Structures in Mammograms Using Random Forests	510
<i>Michael Berks, Zezhi Chen, Sue Astley, and Chris Taylor</i>	
A Unified Framework for Joint Segmentation, Nonrigid Registration and Tumor Detection: Application to MR-Guided Radiotherapy	525
<i>Chao Lu, Sudhakar Chelikani, and James S. Duncan</i>	
Detection of Crossing White Matter Fibers with High-Order Tensors and Rank- k Decompositions	538
<i>Fangxiang Jiao, Yaniv Gur, Chris R. Johnson, and Sarang Joshi</i>	
Nonnegative Factorization of Diffusion Tensor Images and Its Applications	550
<i>Yuchen Xie, Jeffrey Ho, and Baba C. Vemuri</i>	
Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity	562
<i>Gael Varoquaux, Alexandre Gramfort, Fabian Pedregosa, Vincent Michel, and Bertrand Thirion</i>	
Activated Fibers: Fiber-Centered Activation Detection in Task-Based fMRI	574
<i>Jinglei Lv, Lei Guo, Kaiming Li, Xintao Hu, Dajiang Zhu, Junwei Han, and Tianming Liu</i>	
State-Space Models of Mental Processes from fMRI	588
<i>Firdaus Janoos, Shantanu Singh, Raghu Machiraju, William M. Wells, and Istvan Á. Mórocz</i>	
Functional Brain Imaging with M/EEG Using Structured Sparsity in Time-Frequency Dictionaries	600
<i>Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti Hamalainen, and Matthieu Kowalski</i>	
Generalized Sparse Regularization with Application to fMRI Brain Decoding	612
<i>Bernard Ng and Rafeef Abugharbieh</i>	
A Multi-scale Kernel Bundle for LDDMM: Towards Sparse Deformation Description across Space and Scales	624
<i>Stefan Sommer, Mads Nielsen, François Lauze, and Xavier Pennec</i>	

Automatic Part Selection for Groupwise Registration 636
Pei Zhang and Timothy F. Cootes

Temporal Groupwise Registration for Motion Modeling 648
Mehmet Yigitsoy, Christian Wachinger, and Nassir Navab

Fast Brain Matching with Spectral Correspondence 660
Herve Lombaert, Leo Grady, Jonathan R. Polimeni, and Farida Cheriet

Landmark Matching Based Automatic Retinal Image Registration with Linear Programming and Self-similarities 674
Yuanjie Zheng, Allan A. Hunter III, Jue Wu, Hongzhi Wang, Jianbin Gao, Maureen G. Maguire, and James C. Gee

Reconstruction of 4D-CT from a Single Free-Breathing 3D-CT by Spatial-Temporal Image Registration 686
Guorong Wu, Qian Wang, Jun Lian, and Dinggang Shen

Probabilistic Elastography: Estimating Lung Elasticity 699
Petter Risholm, James Ross, George R. Washko, and William M. Wells

Diffusion Imaging

Rotation Invariant Completion Fields for Mapping Diffusion MRI Connectivity 711
Parya MomayyezSiahkal and Kaleem Siddiqi

A Polynomial Approach for Maxima Extraction and Its Application to Tractography in HARDI 723
Aurobrata Ghosh, Demian Wassermann, and Rachid Deriche

Disease Progression Modeling

A Generative Approach for Image-Based Modeling of Tumor Growth ... 735
Bjoern H. Menze, Koen Van Leemput, Antti Honkela, Ender Konukoglu, Marc-André Weber, Nicholas Ayache, and Polina Golland

An Event-Based Disease Progression Model and Its Application to Familial Alzheimer’s Disease..... 748
Hubert M. Fonteijn, Matthew J. Clarkson, Marc Modat, Josephine Barnes, Manja Lehmann, Sebastien Ourselin, Nick C. Fox, and Daniel C. Alexander

Computer Aided Diagnosis

The Ideal Observer Objective Assessment Metric for Magnetic Resonance Imaging: Application to Signal Detection Tasks	760
<i>Christian G. Graff and Kyle J. Myers</i>	
3D Shape Analysis for Early Diagnosis of Malignant Lung Nodules	772
<i>Ayman El-Baz, Matthew Nitzken, Fahmi Khalifa, Ahmed Elnakib, Georgy Gimel'farb, Robert Falk, and Mohammed Abo El-Ghar</i>	

Erratum

Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis	E1
<i>Sandy Durrleman, Marcel Prastawa, Guido Gerig, and Sarang Joshi</i>	

Author Index	785
-------------------------------	-----

Segmentation of Brain Images Using Adaptive Atlases with Application to Ventriculomegaly

Navid Shiee¹, Pierre-Louis Bazin^{1,2}, Jennifer L. Cuzzocreo¹,
Ari Blitz², and Dzung L. Pham^{1,2,3}

¹ The Laboratory of Medical Image Computing, Johns Hopkins University, USA

² Department of Radiology and Radiological Science, Johns Hopkins University, USA

³ Center for Neuroscience and Regenerative Medicine,
Henry M. Jackson Foundation, USA

Abstract. Segmentation of brain images often requires a statistical atlas for providing prior information about the spatial position of different structures. A major limitation of atlas-based segmentation algorithms is their deficiency in analyzing brains that have a large deviation from the population used in the construction of the atlas. We present an expectation-maximization framework based on a Dirichlet distribution to adapt a statistical atlas to the underlying subject. Our model combines anatomical priors with the subject’s own anatomy, resulting in a subject specific atlas which we call an “adaptive atlas”. The generation of this adaptive atlas does not require the subject to have an anatomy similar to that of the atlas population, nor does it rely on the availability of an ensemble of similar images. The proposed method shows a significant improvement over current segmentation approaches when applied to subjects with severe ventriculomegaly, where the anatomy deviates significantly from the atlas population. Furthermore, high levels of accuracy are maintained when the method is applied to subjects with healthy anatomy.

1 Introduction

Automated algorithms for the segmentation of magnetic resonance (MR) brain images provide valuable tools for analyzing human brain structure. The incorporation of prior information is often required in many of these algorithms. One of the most commonly used types of prior information are statistical atlases that use a selection of training examples (i.e. manual delineations) from multiple subjects to model the spatial variability of the structures of interest [1,2,8,17]. These atlases can be utilized within a Bayesian framework to guide the algorithm as to where a structure is likely to appear within the image. Such approaches offer several advantages, including enhanced stability and convergence, as well as providing the ability to distinguish between structures with similar intensities. However, one of the major drawbacks of algorithms that use such statistical atlases is their inability to accurately model subjects whose brain anatomy deviates from the atlas population to a great extent. In some diseases and neurodegenerative conditions such as hydrocephalus, large changes take place in the geometry

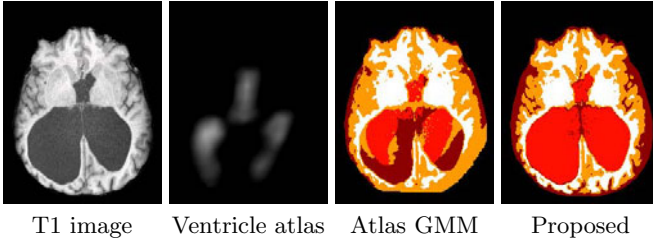


Fig. 1. Results of an atlas-based GMM segmentation algorithm on a subject with large ventricles (sulcal-CSF, ventricles, GM, and WM are represented by dark red, light red, orange, and white, respectively)

of the brain, potentially resulting in wildly inaccurate segmentations when using atlases derived from healthy populations.

Few methods have been proposed to effectively employ spatial atlases when population-specific training data is unavailable. Bhatia et al. [3] proposed a combined segmentation-registration method for atlas based brain segmentation in a group of subjects that are anatomically different from the atlas subjects. Riklin-Raviv et al. [19] avoided the use of a statistical atlas by introducing latent atlases generated from an image ensemble. The main limitation of these approaches is their dependency on the existence of a group of images that implicitly are needed to represent similar anatomy. Liu et al. utilize a generative segmentation model combined with a discriminative classifier to reduce dependency on an atlas [10]. However, this method focused on diseases that possess relatively modest geometric changes in brain structure.

Our approach for the segmentation of MR brain images is based on Gaussian Mixture Models (GMMs), similar to many other approaches (cf. [1,22]). In GMMs, the intensity values of each cluster are assumed to have a Gaussian distribution whose parameters can be estimated by maximum likelihood (ML) using the expectation maximization (EM) algorithm [6]. The mixing coefficients of the GMM in brain segmentation algorithms can be given by a statistical atlas registered to the image space [22]. Consequently, if the difference between the subject and the training data used in the construction of the atlas can not be captured by the registration algorithm, these algorithms are unable to generate an accurate segmentation (Fig. 1). In other image processing applications, several approaches have been introduced to derive the mixing coefficients from the image itself. The concept of spatially varying mixing coefficients was first introduced by Sanjay-Gopal and Hebert [20]. In their work, a Gibbs MRF-based prior was assumed on the mixing coefficients whose maximum *a posteriori* (MAP) estimates were computed by means of a generalized EM algorithm. Several other approaches have been proposed based on this model (see [12,13]). Although these approaches are not biased to a statistical atlas, they can not effectively model separate structures with similar intensity values, such as sulcal and ventricular cerebrospinal fluid (CSF). This is a major advantage of methods employing a statistical atlas.

In this work, we propose a new framework that combines the desirable properties of both of these models. We model the mixing coefficients by a Dirichlet distribution whose parameters are derived from a statistical atlas of healthy subjects. Employing this model within an EM framework, we estimate an “*adaptive*” atlas which is updated iteratively by combining the original statistical atlas with information from the image under study. Since the adaptive atlas is informed by a statistical atlas, clusters with similar intensities can be embedded in our adaptive atlas. At the same time, the influence of the data on the adaptive atlas removes the bias toward the statistical atlas to a great extent. The resulting adaptive atlas is specific to the subject and therefore, does not limit the atlas based segmentation of the image. Unlike the approaches of [3,19], our model does not require an ensemble of images and automatically adapts the statistical atlas to a single subject. It is worth mentioning that two recent works have incorporated a Dirichlet model in image segmentation algorithms in a different context [13,11].

We test and validate our method both on subjects with normal brain anatomy and subjects suffering from hydrocephalus. Hydrocephalus patients can have severe ventriculomegaly which are not well modeled by statistical atlases generated from normal subjects. Because of the specific interest in ventricular size and shape in hydrocephalus patients [18,4], traditional three-class brain segmentation approaches are not appropriate. For this application, we employ a four tissue class model, consisting of gray matter (GM), white matter (WM), sulcal CSF, and ventricular CSF. Our results demonstrate that our method applied to brains with severe ventriculomegaly has superior performance over state-of-the-art atlas-based methods, while it achieves similar accuracy in the segmentation of brains with normal anatomy.

2 Methods

In this section, we first review the Gaussian mixture model (GMM) with spatially varying mixing coefficients introduced in [20] before presenting our work on incorporating adaptive atlases in the segmentation of brain MR images. In a standard atlas approach, the mixing coefficients represent our atlas, providing the prior probability that a tissue class occurs at a voxel. The central idea of this paper is to employ a Dirichlet distribution as a prior to the mixing coefficients, where the mode of the distribution is given by a statistical atlas constructed from healthy subjects. In so doing, the estimated mixing coefficients can deviate significantly from the original statistical atlas and adapt to the subject data.

2.1 Gaussian Mixture Model with Spatially Varying Mixing Coefficients

Our goal is to segment the brain into K structures. We assume that the intensity distribution of each structure follows a Gaussian distribution. Hence the observed MR image(s) can be modeled by a GMM. Let \mathbf{x}_j be the $C \times 1$ vector of the observed intensities where $j \in \{1, 2, \dots, N\}$ represents the voxel location in the

image and C represents the number of input channels. Also let $\theta_k = \{\mu_k, \Sigma_k\}$ be the parameters of the Gaussian distribution associated with structure k ($k \in \{1, 2, \dots, K\}$). Here, μ_k is the $C \times 1$ mean vector and Σ_k is the $C \times C$ covariance matrix.

The label of each voxel j is denoted by a $K \times 1$ vector \mathbf{z}_j . If voxel j belongs to structure k , $\mathbf{z}_j = e_k$ where e_k is a $K \times 1$ vector whose k th component is 1 and all of its other components are 0. We model the prior distribution of \mathbf{z}_j 's as a multinomial distribution with the $K \times 1$ vector parameter π_j , i.e. $f(z_{jk} = 1) = \pi_{jk}$ where the second letter in the subscript denotes the vector component index. The π_{jk} 's are called the mixing coefficients of GMM and they represent the possibility that voxel j belongs to structure k *a priori* (by construction, $\sum_{k=1}^K \pi_{jk} = 1, \forall j$). With the assumption of the independence of \mathbf{z}_j 's, we can write the prior on $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$ as:

$$f(\mathbf{z}) = \prod_{j=1}^N \prod_{k=1}^K \pi_{jk}^{z_{jk}}. \quad (1)$$

In [20], it was assumed that the observations \mathbf{x}_j 's are independent, hence the conditional distribution of the *complete* data $\mathbf{y} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$ is given by:

$$f(\mathbf{y}|\Psi) = \prod_{j=1}^N \prod_{k=1}^K [\pi_{jk} G(\mathbf{x}_j; \theta_k)]^{z_{jk}}, \quad (2)$$

where $\Psi = (\pi_1, \pi_2, \dots, \pi_K, \theta_1, \theta_2, \dots, \theta_K)$ and $G(\cdot; \theta_k)$ is a Gaussian distribution with parameter θ_k . Sanjay-Gopal and Hebert derived an EM algorithm for ML estimation of the Gaussian parameters θ_k 's and π_{jk} 's in [2]:

$$w_{jk} = \frac{\pi_{jk} G(\mathbf{x}_j; \theta_k)}{\sum_{l=1}^K \pi_{jl} G(\mathbf{x}_j; \theta_l)} \quad (3)$$

$$\pi_{jk} = w_{jk}, \quad \mu_k = \frac{\sum_{j=1}^N w_{jk} \mathbf{x}_j}{\sum_{j=1}^N w_{jk}}, \quad \Sigma_k = \frac{\sum_{j=1}^N w_{jk} (\mathbf{x}_j - \mu_k)(\mathbf{x}_j - \mu_k)^T}{\sum_{j=1}^N w_{jk}} \quad (4)$$

The posterior probabilities given by (3) are used to associate each voxel to one of the structures by choosing the one that has the largest posterior probability w_{jk} .

It is worth mentioning here that the two other common assumptions on π_{jk} 's are:

- (i) $\pi_{jk} = \pi_k, \forall j$, which is the classical mixture model approach.
- (ii) $\pi_{jk} = p_{jk}$, where p_{jk} is given by a statistical atlas and $\sum_{k=1}^K p_{jk} = 1, \forall j$.

2.2 GMM with Adaptive Atlases

A major drawback of using (2) in the segmentation of the brain structures is the insufficiency of this model in distinguishing structures with similar intensities from one another. Assumption (ii) in the previous section resolves this problem

to some extent. However, if the underlying image largely differs from the subjects used in the construction of the atlas, this assumption hinders the correct segmentation of the image. Here we describe a MAP framework to solve this limitation of atlas-based segmentation algorithms.

To incorporate a statistical atlas within a framework that estimates the mixing coefficients based on the subject data, we treat π_j as a random vector with a distribution whose parameters are derived from the statistical atlas. As the vector π_j lives on a K -dimensional simplex, we should choose a distribution with simplex support as a prior on π_j . The natural choice for such a distribution is a *Dirichlet* distribution :

$$Dir(\pi_j; \alpha_j) = \frac{\prod_{k=1}^K \pi_{jk}^{(\alpha_{jk}-1)}}{B(\alpha_j)}, \quad (5)$$

where $B(\cdot)$ is a Beta function and α_j is the $K \times 1$ vector parameter of the distribution. In our model, we have a statistical atlas generated from training data p_{jk} , that provides us the *a priori* probability of a structure j occurring at voxel k in a healthy population (as will be described later in Sec 2.4). We therefore employ a Dirichlet prior with parameter $\alpha_{jk} = 1 + \delta p_{jk}, \forall k$. The parameter δ is non-negative and we describe its selection later in this section. For the case of $\delta = 1$, the values $\mathbf{p}_j = (p_{j,1}, p_{j,2}, \dots, p_{j,K})$ represent the mode of the Dirichlet prior probability function.

As the brain structures often form a connected region, we incorporate a Markov Random Field (MRF) as a prior on \mathbf{z}_j 's. In this work, we utilize a model similar to the one employed in [22,16], which yields the following prior on \mathbf{z} :

$$f(\mathbf{z}|\pi) = \frac{1}{Z_{MRF}} \prod_{j=1}^N \prod_{k=1}^K \pi_{jk}^{z_{jk}} \exp(-\beta \sum_{i \in N_j} \sum_{l=1, l \neq k}^K z_{jk} z_{il}), \quad (6)$$

where $\pi = (\pi_1, \pi_2, \dots, \pi_N)$, N_j is the 6-connected neighborhood of voxel j , β is the parameter controlling the effect of the prior, and Z_{MRF} is the normalizing factor. It is worth mentioning that we assume uniform prior on θ_k 's.

With these assumptions, the conditional distribution of the complete data in our model is given by:

$$f(\mathbf{y}|\Psi) = \frac{1}{Z_{MRF}} \prod_{j=1}^N \prod_{k=1}^K (\pi_{jk} G(\mathbf{x}_j; \theta_k))^{z_{jk}} \exp(-\beta \sum_{i \in N_j} \sum_{l=1, l \neq k}^K z_{jk} z_{il}) \quad (7)$$

2.3 Estimation Algorithm

In this section we derive the estimation algorithm for the parameters in (7) using a generalized EM algorithm. The EM algorithm is an iterative approach for the estimation of model parameters that involves two steps: (i) an E-step in which, given the parameter estimates from the previous iteration, the conditional expectation of the complete data likelihood function is computed; and (ii) a M-step in which the new estimate of the model parameters is computed to increase

or maximize the conditional expectation computed in E-step [6]. We provide here the E-step and M-step for (7). If $\Psi^{(t)}$ is the set of estimated parameters from iteration t , then we have:

E-step:

$$Q(\Psi|\Psi^{(t)}) = E \left\{ \ln f(\mathbf{y}|\Psi)|\mathbf{x}, \Psi^{(t)} \right\} + \ln f(\pi, \theta_1, \theta_2, \dots, \theta_K) \quad (8)$$

The main component in (8) is computing $w_{jk}^{(t)} = E \{ z_{jk} | \mathbf{x}, \Psi^{(t)} \}$. Due to the MRF prior on \mathbf{z} , computing $w_{jk}^{(t)}$ analytically is computationally intractable. Using a mean field approximation yields (see [16]):

$$\begin{aligned} w_{jk}^{(t)} &= E \left\{ z_{jk} | \mathbf{x}, \Psi^{(t)} \right\} \approx E \left\{ z_{jk} | \mathbf{x}, \Psi^{(t)}, E\{\mathbf{z}_i\}, i \in N_j \right\} \\ &= f \left\{ z_{jk} = 1 | \mathbf{x}, \Psi^{(t)}, E\{\mathbf{z}_i\}, i \in N_j \right\} \\ &= \frac{f(\mathbf{x} | z_{jk} = 1, \Psi^{(t)}) f(z_{jk} = 1 | E\{\mathbf{z}_i\}, i \in N_j)}{\sum_{k=1}^K f(\mathbf{x} | z_{jk} = 1, \Psi^{(t)}) f(z_{jk} = 1 | E\{\mathbf{z}_i\}, i \in N_j)}. \end{aligned} \quad (9)$$

M-step: In this step, we want to find the set of parameters $\Psi^{(t+1)}$ that maximizes (8). This can be done independently for $\pi_j^{(t+1)}$ and $\theta_k^{(t+1)}$. For $\theta_k^{(t+1)}$, we first derive the equation for $\mu_k^{(t+1)}$ and then for $\Sigma_k^{(t+1)}$. As we assume uniform priors on θ_k 's, the solution for $\theta_k^{(t+1)}$ is similar to (4):

$$\mu_k^{(t+1)} = \frac{\sum_{j=1}^N w_{jk}^{(t)} \mathbf{x}_j}{\sum_{j=1}^N w_{jk}^{(t)}}, \quad \Sigma_k^{(t+1)} = \frac{\sum_{j=1}^N w_{jk}^{(t)} (\mathbf{x}_j - \mu_k^{(t+1)}) (\mathbf{x}_j - \mu_k^{(t+1)})^T}{\sum_{j=1}^N w_{jk}^{(t)}} \quad (10)$$

Because of the constraint on π_{jk} 's, we use Lagrange multiplier to maximize (8) with respect to π_{jk} :

$$\begin{aligned} &\frac{\partial}{\partial \pi_{jk}} \left\{ Q(\Psi|\Psi^{(t)}) + \lambda(1 - \sum_{k=1}^K \pi_{jk}) \right\} \Big|_{\pi_{jk}^{(t+1)}} \\ &= \frac{\partial}{\partial \pi_{jk}} \left\{ (\delta p_{jk} + w_{jk}^{(t)}) \ln \pi_{jk} - \lambda \pi_{jk} \right\} \Big|_{\pi_{jk}^{(t+1)}} = 0. \end{aligned} \quad (11)$$

Hence:

$$\pi_{jk}^{(t+1)} = \frac{\delta p_{jk} + w_{jk}^{(t)}}{\lambda}. \quad (12)$$

Using the constraint on π_{jk} 's and p_{jk} 's and (9) λ is given by:

$$\lambda = \sum_{k=1}^K \delta p_{jk} + \sum_{k=1}^K w_{jk}^{(t)} = \delta + 1. \quad (13)$$

Therefore the update for $\pi_{jk}^{(t+1)}$ is given by:

$$\pi_{jk}^{(t+1)} = \frac{\delta p_{jk} + w_{jk}^{(t)}}{\delta + 1}. \quad (14)$$

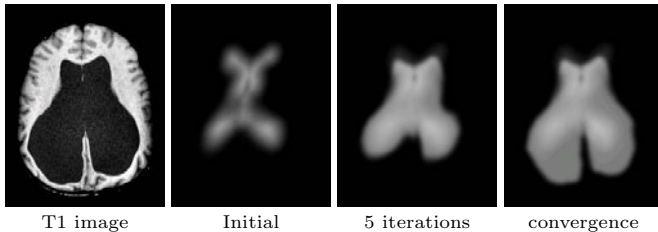


Fig. 2. Evolution of the adaptive atlas from the initial statistical atlas. It originally is biased by the training data but eventually converges to the subject’s geometry.

The statistical atlas, by construction, should be a smooth probability map that represents the population under study. Although p_{jk} ’s are given by a smooth statistical atlas, w_{jk} ’s do not have this desirable smooth property. To enforce the spatial smoothness on w_{jk} ’s and make the mixing coefficients reflective of a population that the subject is drawn from, we convolve w_{jk} ’s with a spatial Gaussian kernel in (14):

$$\pi_{jk}^{(t+1)} \approx (1 - \kappa)p_{jk} + \kappa(G * w_{jk}^{(t)}), \quad (15)$$

where G is a Gaussian kernel (we used a kernel size of 2.5 mm in all our experiments). This provides our model for generating the *adaptive atlas*. Although the Gaussian filtering deviates from our statistical framework, simple smoothing of parameter estimates has previously been shown to offer good convergence properties and computational advantages over explicit smoothing priors when used within the EM algorithm [14].

Eq. (15) has an intuitive interpretation of being a weighted average of the original atlas with the subject atlas. We call $\kappa = \frac{1}{\delta+1}$ in (15) the *adaptation factor*, which controls the influence of the image on the adaptive atlas. For instance, $\kappa = 0$ results in case (ii) of Sec 2.1 (a non-adapted statistical atlas), whereas $\kappa = 1$ results in a smooth version of (4) (no influence from the statistical atlas). In particular, $\kappa = 0.5$, which we used in all our experiments, enforces the mode of the Dirichlet prior on π_j to be equal to \mathbf{p}_j . Fig. 2 shows how the adaptive atlas changes during the segmentation algorithm.

2.4 Implementation Details

Statistical atlas: The statistical atlas we used in this work is built from a set of 18 manual delineations of the structures of interest, based on the IBSR data set [23]. For each image in the atlas, the delineation is rigidly aligned with the current atlas image, and a smooth approximation of the probabilities is accumulated. The smoothing replaces the step edge at the boundary of each structure in their binary delineation by a linear ramp over a band of size ϵ (which we set to 10 mm in our work) [2]. In this work, the statistical atlas contains prior information for sulcal-CSF, ventricles, GM, and WM, however it can be easily extended to more structures.

Atlas registration: The statistical atlas needs to be registered to the image space for the segmentation task. We used a joint segmentation and registration technique that alternates between estimating the segmentation given the current atlas position, and then updating the transformation given the current segmentation. The registration maximizes the correlation of the mixing coefficients with the posterior probabilities. The transformation is updated at each iteration of the algorithm. We used affine transformations in this work.

Initializing the EM algorithm: To initialize the intensity centroids of the different structures, we first estimate the robust minimum and maximum of the intensity (the intensity values at 5% and 95% of the histogram, respectively), and normalize the profiles such that 0 corresponds to the minimum and 1 to the maximum. We then initialized the centroids to empirically determined values stemming from the expected intensities of the tissue classes for the appropriate pulse sequence.

Convergence criteria: We used the maximum amount of change in posterior probabilities as the measure of convergence of the EM algorithm. We set a threshold of 0.01 on this maximum in our work.

3 Experiments

We evaluated the performance of the introduced method on both brains with healthy anatomy and hydrocephalus brains which suffer from ventriculomegaly. We compared the performance of our method to an in-house implementation of a standard atlas-based EM segmentation algorithm (which we refer to as the atlas EM segmentation algorithm in this section). The only difference between this implementation and our proposed method lies in the use of a conventional statistical atlas instead of the introduced adaptive atlas. Brain images were pre-processed to extract the brain and correct for inhomogeneities. Some manual refinement of the brain mask was necessary on some cases. We also compared the performance of our method in the segmentation of the ventricles of hydrocephalus brains with the Freesurfer software [8], which employs a nonstationary MRF model and a probabilistic atlas updated using nonlinear registration. Finally we compared our method to a registration based segmentation algorithm which uses Hammer [21], a state-of-the-art deformable registration algorithm, to label the ventricles. We used Dice overlap coefficient [7] and false negative ratio (for the ventricle segmentation experiment) between the automated segmentation and the ground truth as the accuracy measures.

3.1 Normal Brains

To study the performance of the algorithm on brains with normal anatomy, we used both simulated and real data sets. In our first experiment, we used the simulated T1 brain from the Brainweb phantom [5] with 3% noise and no inhomogeneity. As Brainweb truth model does not have a separate segmentation

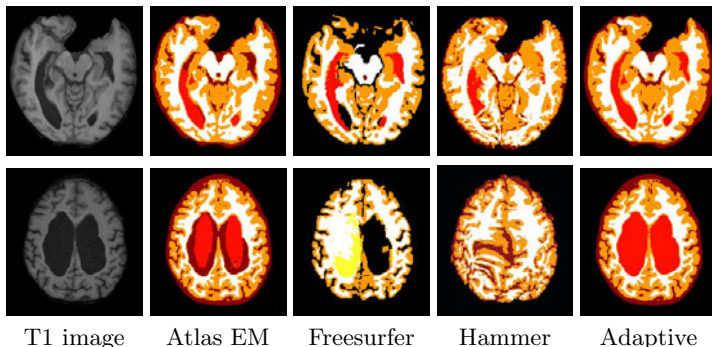


Fig. 3. Comparison of the adaptive atlas approach and three other atlas-based segmentation algorithms on hydrocephalus subjects with moderate (top row) and marked (bottom row) ventriculomegaly (sulcal-CSF, ventricles, GM, and WM are represented by dark red, light red, orange, and white, respectively. Yellow represents WM-hypointensity in Freesurfer segmentation).

for the ventricles, we manually separated the sulcal-CSF from the ventricles on the ground truth image. As Table I demonstrates, our method has slightly better overall accuracy in segmenting the simulated brain in comparison to the non-adaptive EM segmentation algorithm.

In the second experiment, we used the IBSR database [23] which contains MR brain images of 18 real subjects. As we also use this database to create our statistical atlas, we separated the subjects to two non-overlapping groups of 8 and 10 subjects. We used the first group to create the statistical atlas and then used that atlas to validate the algorithms on the 10 other subjects. As the manual segmentation of this data set does not contain sulcal-CSF and includes it inside the GM segmentation, we combined the sulcal-CSF and GM labels in the automated segmentation results before comparing to the manual segmentation. The accuracy of our method and the non-adaptive atlas EM segmentation algorithm in the segmentation of this data set are very similar (Table II).

3.2 Brains with Ventriculomegaly

The main advantage of our approach using the adaptive atlases over other atlas-based segmentation algorithms is its ability in the segmentation of images that largely deviate from the atlas population. To study this unique aspect of our

Table 1. Comparison of the non-adaptive EM segmentation algorithm and the proposed method on brains with normal anatomy (Brainweb and IBSR data sets), measured by Dice overlap coefficient. For IBSR data set, the mean dice value over 10 subjects is reported.

Algorithm	Brainweb				IBSR*		
	CSF	VENT	GM	WM	VENT	GM	WM
Atlas EM	0.938	0.953	0.942	0.952	0.851	0.931	0.877
Adaptive atlas	0.945	0.957	0.945	0.951	0.835	0.933	0.878

adaptive method, we evaluated the performance of the proposed method, the atlas EM segmentation algorithm, Freesurfer, and Hammer on the segmentation of ventricles of 14 patients with hydrocephalus. Our data set has 9 subjects with moderate and 5 subjects with marked ventricular dilatation, as diagnosed by a neuroradiologist. This allows us to be able to study the effect of large deviations from the atlas population on the performance of each algorithm more thoroughly. Also, in addition to Dice overlap coefficient, we computed the false negative ratio ($FNR = \frac{|Ref| - |Ref \cap Seg|}{|Ref|}$) between the automatically segmented ventricles (Seg) and the manual delineation by an expert (Ref) for each method. We found FNR very informative in our validation, as the errors in the segmentation of the ventricles in this data set are mostly due to a misclassification as other structures. The validation results (Table 2) show that our method has superior accuracy on all the subjects in comparison to other methods. As expected, the atlas EM segmentation algorithm, Freesurfer, and the Hammer based segmentation perform significantly worse on subjects with marked ventriculomegaly, whereas the performance of our approach is not altered significantly by the amount of the dilatation. Freesurfer failed (did not complete or resulted in a nearly empty image) in the processing of 3 subjects with severe structural anomalies due to problems in its brain extraction step. We also note that Freesurfer attempts to segment a much larger number of structures and is therefore at a disadvantage trying to solve a more difficult problem. Also it is worth mentioning that for the approach based on Hammer, we first used FANTASM [15] to classify the image voxels as CSF, GM, or WM. We then used the atlas provided with the Hammer software [9] to segment the ventricles. Although this atlas contains 101 regions of interest, we grouped all of these regions into sulcal-CSF, ventricles, GM, and WM in our experiments.

Table 2. Segmentation accuracy measures on hydrocephalus data set

Subject	Dice overlap coefficient				False negative ratio			
	Atlas EM	Freesurfer	Hammer	Adaptive	Atlas EM	Freesurfer	Hammer	Adaptive
Moderate 1	0.908	0.884	0.786	0.977	0.151	0.156	0.320	0.015
Moderate 2	0.964	0.929	0.889	0.972	0.050	0.053	0.154	0.038
Moderate 3	0.961	0.930	0.864	0.967	0.040	0.043	0.187	0.023
Moderate 4	0.959	0.911	0.827	0.970	0.056	0.061	0.251	0.026
Moderate 5	0.953	0.889	0.824	0.962	0.051	0.108	0.236	0.030
Moderate 6	0.941	0.890	0.847	0.949	0.063	0.083	0.203	0.035
Moderate 7	0.898	0.753	0.864	0.970	0.163	0.362	0.194	0.026
Moderate 8	0.937	0.923	0.860	0.958	0.106	0.065	0.191	0.073
Moderate 9	0.951	0.922	0.576	0.977	0.080	0.079	0.580	0.022
Mean	0.941	0.892	0.815	0.967	0.084	0.112	0.257	0.032
Marked 1	0.496	Failed*	0.119	0.977	0.670	Failed*	0.936	0.043
Marked 2	0.494	Failed*	0.154	0.962	0.671	Failed*	0.915	0.064
Marked 3	0.561	Failed*	0.111	0.967	0.608	Failed*	0.941	0.046
Marked 4	0.798	0.152	0.190	0.970	0.333	0.917	0.893	0.044
Marked 5	0.863	0.258	0.417	0.981	0.233	0.849	0.731	0.021
Mean	0.642	0.205	0.198	0.971	0.503	0.883	0.883	0.044
Mean (All)	0.835	0.767	0.595	0.968	0.234	0.252	0.481	0.036

* Algorithm failed due to problems in the skull-stripping step.

Fig. 3 shows that even in the cases of moderate ventriculomegaly, the atlas EM segmentation algorithm, Freesurfer and Hammer are not able to segment the ventricles as accurately as the adaptive approach.

4 Discussion

We have presented a new statistical framework for the atlas based segmentation of MR brain images based on adaptive atlases. Our approach addresses the segmentation of images that deviate from the atlas population to a great extent. The validation results confirm that our algorithm has a superior performance in the segmentation of such subjects, while maintaining high level of accuracy in the segmentation of brains with healthy anatomy. On the hydrocephalus data, our method was shown to have some advantages when compared to a deformable registration approach (Hammer), as well as an atlas-based approach that employs deformable registration to update its atlas (Freesurfer). Deformable registration approaches are prone to local optima when the target is highly dissimilar from the template. Furthermore, the proposed approach leads to computationally efficient atlas updating (a Gaussian filter and simple averaging) when compared to deformable registration. The introduced adaptive atlas provides a general model for the computation of the mixing coefficients of a GMM model; for instance the conventional approach of using a statistical atlas as the mixing coefficients is a special case of this model. Although we described the adaptive atlas concept as part of a GMM model, our approach can be easily incorporated in other atlas-based probabilistic segmentation methods.

In computing the adaptive atlas, we included a smoothing step within EM iterations to make the adaptive atlas reflective of a population that the subject is drawn from. We will investigate explicitly modeling this property as a hierarchical prior on the mixing coefficients.

References

1. Ashburner, J., Friston, K.: Multimodal image coregistration and partitioning—a unified framework. *NeuroImage* 6(3), 209–217 (1997)
2. Bazin, P.-L., Pham, D.L.: Homeomorphic brain image segmentation with topological and statistical atlases. *Med. Image Anal.* 12(5), 616–625 (2008)
3. Bhatia, K.K., Aljabar, P., Boardman, J.P., Srinivasan, L., Murgasova, M., Counsell, S.J., Rutherford, M.A., Hajnal, J.V., Edwards, A.D., Rueckert, D.: Groupwise combined segmentation and registration for atlas construction. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 532–540. Springer, Heidelberg (2007)
4. Clarke, M.J., Meyer, F.B.: The history of mathematical modeling in hydrocephalus. *Neurosurg Focus* 22(4), E3 (2007)
5. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging* 17(3), 463–468 (1998)

6. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc.* 39(1), 1–38 (1977)
7. Dice, L.: Measures of the amount of ecologic association between species. *Ecology* 25(3), 297–302 (1945)
8. Fischl, B., Salat, D.H., van der Kouwe, A.J.W., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M.: Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23(suppl. 1), S69–S84 (2004)
9. Kabani, N., McDonald, D., Holmes, C.J., Evans, A.C.: 3D anatomical atlas of the human brain. *Proc of HBM, NeuroImage* 7(4), S717 (1998)
10. Liu, C.Y., Iglesias, J.E., Toga, A., Tu, Z.: Fusing adaptive atlas and informative features for robust 3D brain image segmentation. In: *Proc of ISBI*. pp. 848–851 (2010)
11. Maddah, M., Zollei, L., Grimson, W., Wells, W.: Modeling of anatomical information in clustering of white matter fiber trajectories using Dirichlet distribution. In: *Proc of MMBIA*, pp. 1–7 (2008)
12. Nikou, C., Galatsanos, N.P., Likas, A.C.: A class-adaptive spatially variant mixture model for image segmentation. *IEEE Trans. Image Process.* 16(4), 1121–1130 (2007)
13. Nikou, C., Likas, A.C., Galatsanos, N.P.: A bayesian framework for image segmentation with spatially varying mixtures. *IEEE Trans. Image Process.* 19(9), 2278–2289 (2010)
14. Nychka, D.: Some properties of adding a smoothing step to the EM algorithm. *Stat. Probabil. Lett.* 9(2), 187–193 (1990)
15. Pham, D.L., Prince, J.L.: Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans. Med. Imaging* 18(9), 737–752 (1999)
16. Pham, D., Bazin, P.L.: Unsupervised tissue classification. In: Bankman, I. (ed.) *Handbook of Medical Image Processing and Analysis*, 2nd edn., pp. 209–221. Elsevier, Amsterdam (2008)
17. Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M.: A hierarchical algorithm for MR brain image parcellation. *IEEE Trans. Med. Imaging* 26(9), 1201–1212 (2007)
18. Preul, C., Hübsch, T., Lindner, D., Tittgemeyer, M.: Assessment of ventricular reconfiguration after third ventriculostomy: what does shape analysis provide in addition to volumetry? *AJNR Am. J. Neuroradiol.* 27(3), 689–693 (2006)
19. Riklin-Raviv, T., Van Leemput, K., Menze, B.H., Wells, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. *Med. Image Anal.* 14(5), 654–665 (2010)
20. Sanjay-Gopal, S., Hebert, T.J.: Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *IEEE Trans. Image Process.* 7(7), 1014–1028 (1998)
21. Shen, D., Davatzikos, C.: HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging* 21(11), 1421–1439 (2002)
22. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* 18(10), 897–908 (1999)
23. Worth, A.: Internet brain segmentation repository (1996), <http://www.cma.mgh.harvard.edu/ibsr/>

Adaptive Riemannian Metrics for Improved Geodesic Tracking of White Matter

Xiang Hao, Ross T. Whitaker, and P. Thomas Fletcher

Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT

Abstract. We present a new geodesic approach for studying white matter connectivity from diffusion tensor imaging (DTI). Previous approaches have used the inverse diffusion tensor field as a Riemannian metric and constructed white matter tracts as geodesics on the resulting manifold. These geodesics have the desirable property that they tend to follow the main eigenvectors of the tensors, yet still have the flexibility to deviate from these directions when it results in lower costs. While this makes such methods more robust to noise, it also has the serious drawback that geodesics tend to deviate from the major eigenvectors in high-curvature areas in order to achieve the shortest path. In this paper we formulate a modification of the Riemannian metric that results in geodesics adapted to follow the principal eigendirection of the tensor even in high-curvature regions. We show that this correction can be formulated as a simple scalar field modulation of the metric and that the appropriate variational problem results in a Poisson's equation on the Riemannian manifold. We demonstrate that the proposed method results in improved geodesics using both synthetic and real DTI data.

1 Introduction

Front-propagation approaches [1,10,4,8,13,3] in diffusion tensor imaging (DTI) infer the pathways of white matter by evolving a level set representing the time-of-arrival of paths emanating from some starting region. The direction and speed of this evolving front at each point is determined by some cost function derived from the diffusion tensor data. One such method, first proposed by O'Donnell et al. [10], is to treat the inverse of the diffusion tensor as a Riemannian metric, and the paths in the propagating front as geodesics, i.e., shortest paths, under this metric. This makes intuitive sense: traveling along the large axis of the diffusion tensor results in shorter distances, while traveling in the direction of the small axes results in longer distances. Therefore, shortest paths will tend to prefer to remain tangent to major principal eigenvector of the diffusion tensor. While this is a powerful framework for computing white matter pathways, these geodesics have the serious deficiency that in high-curvature tracts they tend to deviate from the eigenvector directions and take straighter trajectories than is desired. That is, in high-curvature regions, the incremental cost of following the tensor field is overcome by the cost associated with the longer (more curved) path. In this paper we develop a new Riemannian metric, that relies on diffusion tensor

data but resolves this problem by adapting to high-curvature tracts, resulting in geodesic paths that more faithfully follow the principal eigenvectors.

1.1 Background

Front-propagation methods offer several advantages over conventional tractography [91], in which streamlines (sometimes called tracts) are computed by forward integration of the principal eigenvector of the tensor. One major problem with tractography is that imaging noise causes errors in the principal eigenvector direction, and these errors accumulate in the integration of the streamlines. The front-propagation algorithms are more robust to noise than tractography because the wavefront is not constrained to exactly follow the principal eigenvector of the tensors. Although the principal eigenvector of the tensor is the preferred direction for paths to travel, the minimal-cost paths may deviate from these directions if it decreases the overall cost.

Another disadvantage to tractography is that it has difficulty in cases where the goal is to find pathways between two regions. In this scenario, streamlines begin in one of the regions and are accepted only if they eventually pass through the desired ending region. However, several factors conspire to often result in only a small fraction of fibers being accepted. These factors include accumulated errors in the streamlines throwing off the final destination and stopping criteria being triggered, either by low anisotropy tensors, due to noise or partial voluming, or sudden direction changes caused by noise. As shown by Fletcher et al. [3], front propagation methods can be used to segment white matter tracts by solving the geodesic flow from both regions and combining the resulting cost functions. This has the advantage that the solution will not get stuck in regions of noisy data or low anisotropy. This type of analysis is only appropriate if the endpoint regions are well known to be connected by a white matter tract because a white matter path will always be found. Although, if a “false positive” connection is found, this should be detectable as an unusually high cost function incurred by that pathway.

An alternative approach that deals with the problems arising from image noise is probabilistic tractography [6,21,27], in which large numbers of streamlines are initiated from each seed voxel and are integrated along directions determined stochastically at each point. However, this is a computationally-intensive procedure (typically requiring several to many hours), whereas efficient implementations of front-propagation solvers are much faster (typically requiring several seconds). The graphics processing unit (GPU) implementation by Jeong et al. [5] even runs at near real-time speeds. Also, probabilistic tractography suffers from the same problems with streamlines stopping in noisy or low-anisotropy regions, leading to artificially low (or even zero) probabilities of connection.

1.2 Properties of Front-Propagation

Despite the advantages that front-propagation methods have over tractography, there is one serious drawback. Figure 1 shows a diagram illustrating the problem. In a curved tensor field, one would typically prefer a path that follows,

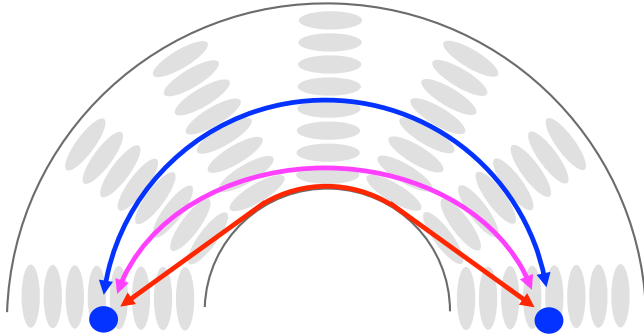


Fig. 1. Diagram of various pathways between two points in a curved tensor field: the desired path following the principal eigenvectors (blue), the shortest path under the Euclidean metric (red), and the compromise path taken when using the inverse tensor field as metric (magenta).

to whatever extent possible, the major eigenvectors of the tensors (shown in blue). The shortest path, using a Euclidean metric (i.e., ignoring the tensors) follows a straight line, except at constraints (shown in red). The typical geodesic with a local, anisotropic metric (e.g., using the inverse tensors as metric), will find a compromise between these two (shown in magenta). Although the magenta geodesic is taking infinitesimally higher-cost steps than the blue curve, its overall length under the inverse-tensor metric is shorter.

This issue has been addressed previously [3] by “sharpening” the tensor, i.e., increasing the anisotropy by taking the eigenvalues to some power and renormalizing them. This increases the cost of moving in directions other than the principal eigenvector. In fact, the first front-propagation algorithm proposed by Parker et al. [11] essentially takes this sharpening strategy to its limit, which results in a cost function that is the dot product of the level set velocity with the principal eigenvector. However, the amount of sharpening is an ad hoc parameter, and sharpening is applied equally across the image, rather than taking the curvature of the tract into account. Sharpening that increases with the curvature of the tract could be more effective. Another downside of sharpening is that it changes the shape of the tensor and reduces the ability to deviate from the principal direction, thus decreasing the desired robustness to noise. It is not clear how to set the amount of sharpening to find the best balance between robustness to noise versus faithful following of the eigenvectors.

Our proposed solution to this problem is to develop a new Riemannian metric that is a modulated version of the inverse diffusion tensor field. This metric is able to adaptively correct the geometry of geodesic curves in high-curvature regions so that they more closely follow the principal eigenvectors of the tensors. The resulting algorithm requires solving for an unknown scalar field, which requires solving a Poisson equation on the Riemannian manifold—however it does not require any arbitrary choice of parameters. We show that this solution

is sufficient to eliminate the problem with geodesics in high-curvature regions described above and illustrated in Figure [II](#), and we demonstrate the corrected behavior of geodesics on both synthetic and real DTI data.

2 Adaptive Riemannian Metrics

In this section we derive a procedure for computing geodesic flows in diffusion tensor data that resolves the major drawback of front-propagation approaches outlined above. Namely, the geodesics generated by our method more closely conform to the principal eigenvector field. Rather than directly using the inverse of the diffusion tensor as the Riemannian metric, as is typically done, we compute a spatially-varying scalar function that modulates the inverse tensor field at each point and use this as our metric. We show that this scalar field can be chosen in such a way that the resulting geodesic flows have the desired property of following the eigenvector directions. This entails solving the classical variational problem for geodesic curves, with the exception that the Riemannian metric is scaled by a positive function. In the resulting Euler-Lagrange equation, we then solve for the particular scaling function that causes geodesics to follow the desired directions. In the end, we see that the appropriate function is computed by solving a Poisson equation on the Riemannian manifold.

2.1 The Metric Modulating Function

On a Riemannian manifold, M , the geodesic between two points $p, q \in M$ is defined by the minimization of the energy functional

$$E(\gamma) = \int_0^1 \langle T(t), T(t) \rangle dt,$$

where $\gamma : [0, 1] \rightarrow M$ is a curve with fixed endpoints, $\gamma(0) = p$, $\gamma(1) = q$, $T = d\gamma/dt$, and the inner product is given by the Riemannian metric. In our case the manifold $M \subset \mathbb{R}^3$ is the image domain, and the Riemannian metric can be equated with a smoothly-varying, positive-definite matrix $g(x)$ defined at each point $x \in M$. Letting $T_x M$ denote the tangent space at a point $x \in M$, the inner product between two tangent vectors $u, v \in T_x M$ is given by $\langle u, v \rangle = u^t g(x) v$. As mentioned above, previous front-propagation approaches to DTI have used the inverse of the diffusion tensor field as a metric, i.e., $g(x) = D(x)^{-1}$ (or a sharpened or modified version), and this choice of metric leads to geodesics that bend inwards around curves. To rectify this problem, we will scale the Riemannian metric by a positive function $e^{\alpha(x)}$, which results in the new geodesic energy functional

$$E_\alpha(\gamma) = \int_0^1 e^{\alpha(\gamma(t))} \langle T(t), T(t) \rangle dt. \quad (1)$$

We call the function e^α the *metric modulating function* because it scales the Riemannian metric at each point. The exponentiation of α is to both ensure that

this scaling factor is positive and to make the solution to the variational problem comes out simpler in the end. While it is possible to envision more complicated modifications of the metric tensor, there are two reasons why we choose to modify the metric in this fashion. First, the shape of the diffusion tensor provides information about the relative preference in diffusion directions, and a scaling operation allows us to keep this information intact. Second, the modification in (II) is sufficient to correct for the effects of curvature. In other words, if the tensors are following a curved path, but not changing shape, the metric modulating function can be chosen in such a way that the resulting geodesics perfectly follow the principal eigenvector. We demonstrate this property empirically using a synthetic example in Section 3.

2.2 Computing the Geodesic Equation

To minimize the new geodesic energy functional given in (II), we use two tools of Riemannian geometry. The first is the affine connection $\nabla_X Y$, which is the derivative of a vector field Y in the direction of a vector field X . We'll write the vector fields X, Y in terms of a coordinate system (x^1, x^2, \dots, x^n) ; note that superscripts here are indices, not exponentiation. We write $X = \sum a^i E_i$ and $Y = \sum b^j E_j$, where $E_i = \frac{\partial}{\partial x^i}$ are the coordinate basis vectors, and a^i and b^j are smooth coefficients functions. Then the affine connection is given by

$$\nabla_X Y = \sum_k \left(\sum_i a^i \frac{\partial b^k}{\partial x^i} + \sum_{i,j} \Gamma_{ij}^k a^i b^j \right) E_k.$$

The terms Γ_{ij}^k are the Christoffel symbols, which are defined as

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^n g^{kl} \left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l} \right),$$

where g_{ij} denotes the entries of the Riemannian metric, g , and g^{ij} denotes the entries of the inverse metric, g^{-1} . Again, the intuition behind this affine connection is that it is like a directional derivative of vector fields. In the special case of $Y = X$, $\nabla_X X$ measures how the vector field X bends along its integral curves.

The second tool that we employ is the Riemannian gradient of a smooth function f , which we denote $\text{grad } f$. The gradient of a function on a Riemannian manifold looks like the standard Euclidean gradient, except with a multiplication by the inverse of the metric, i.e.,

$$\text{grad } f = g^{-1} \left(\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \dots, \frac{\partial f}{\partial x^n} \right).$$

The gradient is defined in this way so that the inner product with a unit vector u results in the usual directional derivative, $\nabla_u f = \langle \text{grad } f, u \rangle$.

Using the affine connection and Riemannian gradient, we take the variational of the energy (II). Let W be a vector field defined along the curve γ that represents

an arbitrary perturbation of γ , keeping the endpoints fixed, i.e., $W(0) = W(1) = 0$. To simplify notation, we will suppress the parameter t in most of the following. Then the variational of the energy functional is

$$\begin{aligned}
\nabla_W E_\alpha(\gamma) &= \nabla_W \int_0^1 e^\alpha \langle T, T \rangle dt \\
&= \int_0^1 \nabla_W e^\alpha \cdot \langle T, T \rangle + e^\alpha \nabla_W \langle T, T \rangle dt \\
&= \int_0^1 \langle W, \text{grad } e^\alpha \rangle \cdot \langle T, T \rangle + 2 \langle \nabla_W T, e^\alpha T \rangle dt \\
&= \int_0^1 \langle W, e^\alpha \|T\|^2 \text{grad } \alpha \rangle - 2 \langle W, \nabla_T (e^\alpha T) \rangle dt \\
&= \int_0^1 \langle W, e^\alpha \|T\|^2 \text{grad } \alpha - 2e^\alpha d\alpha(T) \cdot T - 2e^\alpha \nabla_T T \rangle dt.
\end{aligned}$$

Now, setting this last line to zero and dividing through by e^α , results in the geodesic equation

$$\text{grad } \alpha \cdot \|T\|^2 = 2\nabla_T T + 2d\alpha(T) \cdot T. \quad (2)$$

If we assume, without loss of generality, that geodesics have unit-speed parameterization, i.e., $\|T\| = 1$, then $\nabla_T T$ will be normal to T . Now, assuming this parameterization and taking the inner product with T on both sides of (2), we obtain

$$\langle \text{grad } \alpha, T \rangle = 2d\alpha(T) = 2\langle \text{grad } \alpha, T \rangle.$$

This can only hold if the tangential component, $\langle \text{grad } \alpha, T \rangle = 0$. Therefore, the last term in (2) must vanish, and we get the final, simplified geodesic equation

$$\text{grad } \alpha = 2\nabla_T T. \quad (3)$$

2.3 Computing the Metric Modulating Function

Now that we have the geodesic equation for the modulated Riemannian metric, we introduce the property that we would like to enforce: that the tangent vectors T follow the unit principal eigenvector directions, V . Satisfying this property directly would result in the equation $\text{grad } \alpha = 2\nabla_V V$, which we would need to solve for α . However, given an arbitrary unit vector field V , there may not exist such a function with the desired gradient field.

Instead we minimize the squared error between the two vector fields, i.e., we minimize the functional

$$F(\alpha) = \int_M \|\text{grad } \alpha - 2\nabla_V V\|^2 dx. \quad (4)$$

As before, the norm here is given by the Riemannian metric. The Euler-Lagrange solution to this problem is derived similarly to the classical Poisson equation,

with the exception that the div and grad operators used are the Riemannian versions. The divergence of X on M is defined in coordinates as

$$\text{div}(X) = \frac{1}{\sqrt{|g|}} \sum_i \frac{\partial}{\partial x^i} \left(\sqrt{|g|} a^i \right),$$

where $|g|$ is the determinant of the Riemannian metric, which represents the appropriate volume element. Finally, the equation of the metric modulating function that minimizes (4) is given by

$$\Delta\alpha = 2 \text{div}(\nabla_V V), \quad (5)$$

where $\Delta\alpha = \text{div}(\text{grad } \alpha)$ is the Laplace-Beltrami operator on M . The appropriate boundary conditions for this problem are the Neumann conditions,

$$\frac{\partial\alpha}{\partial \vec{n}} = \langle \text{grad } \alpha, \vec{n} \rangle = \langle 2\nabla_V V, \vec{n} \rangle.$$

A closer look at (5) reveals that it is nothing but an anisotropic Poisson equation on the image domain. The right-hand side is constant in α , and the Laplace-Beltrami operator on the left-hand side can be expressed as $\nabla \cdot (A \nabla \alpha)$, where A is a symmetric positive-definite matrix and $\nabla \cdot$ and ∇ are the usual Euclidean divergence and gradient operators in the image domain. We solve this equation using a finite-difference scheme with a Jacobi iteration. There are more efficient solvers, such as conjugate gradient or multigrid methods, but the application of these methods to the proposed anisotropic operator with irregular boundary conditions remains an area of future development.

3 Results

In this section we demonstrate the improvement of geodesic flows generated by our metric modulating method compared to those computed with the inverse-tensor metric using both synthetic and real DTI data (Figure 2). Our measure of quality is how well the geodesics from the two methods follow the principal eigenvectors of the tensors. However, front-propagation methods do not explicitly compute the geodesic curves, but instead compute a function $u(x)$, which is the time-of-arrival of the geodesic flow at the point x . The characteristic vectors of $u(x)$ give the tangent vectors along the geodesic. In the case of the inverse-tensor metric, the characteristic vectors are given by $T(x) = D(x)^{-1} \nabla u(x)$. In the case of our modulated metric, the characteristic vectors are given by $T(x) = e^{\alpha(x)} D(x)^{-1} \nabla u(x)$. Here $\nabla u(x)$ indicates the Euclidean gradient, which we approximate with finite differences, as described in 4.

We compute $u(x)$ by solving a Hamilton-Jacobi equation using the Fast Iterative Method, as described in 3. For visualization purposes, we compute the geodesics from both methods by integrating their characteristic vectors. Because these vectors always point away from the source region, we compute geodesic curves by integrating the characteristic vectors backward from any target point in the tensor field. These integral curves of the negative characteristic vectors are guaranteed to end up in the source region.

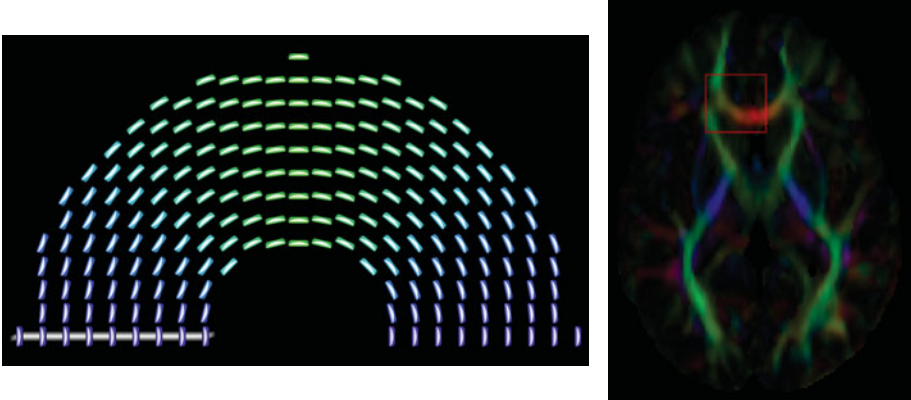


Fig. 2. A slice of the synthetic curved tensor field (left). We subsample the tensor field by a factor of 4 both horizontally and vertically in order to visualize it. A slice of the color-coded principal eigenvector image (right). The area inside the red box is the region of interest used in the experiments.

3.1 Synthetic Data

To test our method, we generate a synthetic curved tensor field which has similar properties to many white matter tracts in the brain. The synthetic data is the top half of a solid torus, where the tensors rotate along the large circle of the torus. The torus has smaller radius of 16 voxels and larger radius of 48 voxels. Each tensor in the tensor field has the same eigenvalues (3,1,1). A middle slice of the tensor field is shown in Figure 2. The source region for the geodesic front-propagation is shown in white.

In Figure 3, we compare the characteristic vector field (shown in blue) with the principal eigenvector field (shown in red). On the left, the characteristic

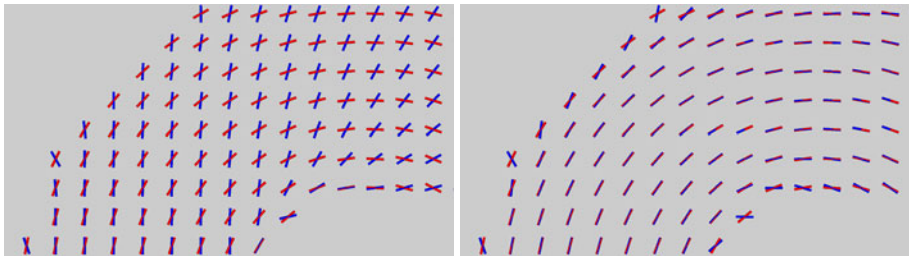


Fig. 3. Tangent vectors of the geodesics (blue) under the inverse-tensor metric without modulation (left) and with modulation (right). The red vectors are the principal eigenvectors of the diffusion tensors. We subsample the vector field by a factor of 4 both horizontally and vertically in order to visualize it.

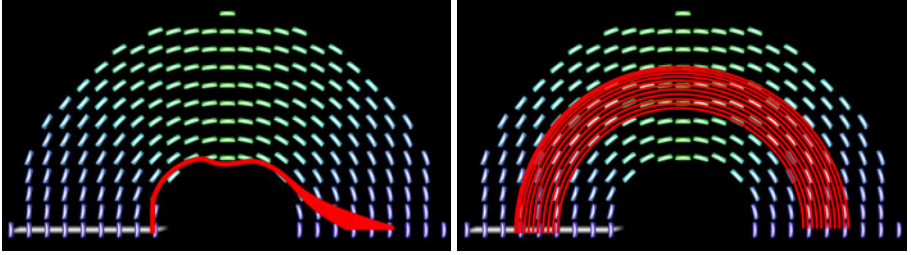


Fig. 4. The geodesics emanating from the targets points (right side of the torus) to the source region (white). We subsample the tensor field by a factor of 4 both horizontally and vertically in order to visualize it.

vector field is computed using the inverse-tensor metric without modulation. On the right, the characteristic vector field is computed using the metric scaled by our computed metric modulating function. Comparing the two pictures, we can clearly see the characteristic vectors T follow the principal eigenvectors V much better in the right picture. Around the boundary, some characteristic vectors are pointing outwards in both cases. This is caused by aliasing: geodesics at the boundary must take sharp turns around voxel corners.

In this synthetic example, we can compute the analytic solution of $\alpha(x)$, which is $\alpha(x) = -2 \ln r(x) + C$, where $r(x)$ is the distance from x to the center of the torus, and C is some constant. We computed the difference between our numerical solution and the analytic $\alpha(x)$, and the result is within numerical error. We also compute the root mean square error (RMSE) of the angle between the geodesic tangent vectors and principal eigenvectors. The RMSE with modulation is 10.6 degrees compared to 54.0 degrees without modulation. Most of the RMSE with modulation is caused by the aliasing artifacts at the boundary.

In Figure 4 we visualize the integrated geodesics between some target points (on the right side of the torus) and the source region (shown in white). The left picture shows the geodesics under the metric as the inverse diffusion tensor field. The right picture shows the geodesics under our modulated metric. Under the modulated metric, the geodesics follow the principal eigenvectors of the tensor field and arrive at a point inside the source region. In contrast, the geodesics under the inverse-tensor metric without modulation, starting from the same target points, take a shortcut and end up at the closest point inside the source region by closely following the boundary constraints.

3.2 Real Data

We now show the results of our method applied to a corpus callosum tract from a DTI of a healthy volunteer. DTI data were acquired on a Siemens Trio 3.0 Tesla Scanner with an 8-channel, receive-only head coil. DTI was performed using a single-shot, spin-echo, EPI pulse sequence and SENSE parallel imaging (undersampling factor of 2). Diffusion-weighted images were acquired in twelve

non-collinear diffusion encoding directions with diffusion weighting factor $b=1000$ s/mm² in addition to a single reference image ($b=0$). Data acquisition parameters included the following: contiguous (no-gap) fifty 2.5mm thick axial slices with an acquisition matrix of 128 x 128 over a FOV of 256 mm (2×2 mm² in-plane resolution), 4 averages, repetition time (TR) = 7000 ms, and echo time (TE) = 84 ms. Eddy current distortion and head motion of each data set were corrected using an automatic image registration program [14]. Distortion-corrected DW images were interpolated to $1 \times 1 \times 1.25$ mm³ voxels, and six tensor elements were calculated using least squares. The tensor upsampling is done only for the purposes of numerical computations on the voxel grid; a finer grid results in higher numerical accuracy.

In Figure 5, we compare the characteristic vector field T (shown in blue) with the principal eigenvector field V (shown in red) of the corpus callosum. In the left picture, the characteristic vector field is computed using the inverse-tensor

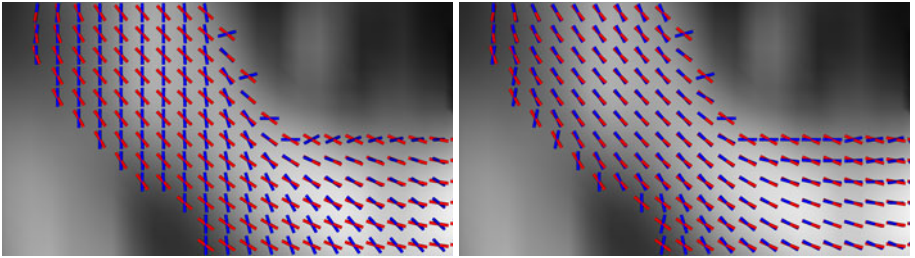


Fig. 5. Tangent vectors of the geodesics (blue) under the inverse-tensor metric without modulation (left) and with modulation (right) for a part of the corpus callosum. The red vectors are the principal eigenvectors of the diffusion tensors. The fractional anisotropy (FA) image is shown in the background.

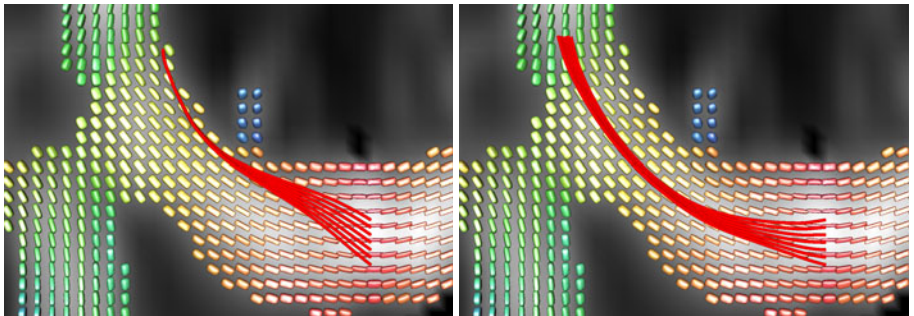


Fig. 6. The geodesic flow in the corpus callosum from the target points (genu of the corpus callosum) to the source region (in the right frontal forcep). The background images are the FA image and the diffusion tensor field.

metric. In the right picture, the characteristic vector field is computed using our modulated metric. After the modulation, the characteristic vectors tend to follow the main eigendirections. As in the synthetic example, some characteristic vectors are pointing outwards near the boundary, which is caused by aliasing. Again, we compute the root mean square error (RMSE) of the angle between the geodesic tangent vectors and principal eigenvectors. The RMSE with modulation is 23.0 degrees compared to 37.3 degrees without modulation. Much of the RMSE with modulation is caused by the aliasing artifacts at the boundary.

In Figure 6, as in the synthetic example, we track backward from the target points, which are in the middle of corpus callosum in this case to the source region on the upper-left of the pictures. Again, the geodesics under the inverse-tensor metric take a shortcut and merge into the closest point in the source region. In contrast, the geodesics under our modulated metric more faithfully follow the tensor directions. In the latter case, geodesics are drawn together slightly because the tensor field is thinner around the corner of the corpus callosum.

4 Conclusion and Future Work

We presented a new geodesic front-propagation method for computing white matter pathways in DTI and showed that it results in geodesics that more faithfully follow the principal eigenvectors of the diffusion tensor field, especially in tracts with high curvature. There are two areas we have identified as potential future work. First, the aliasing artifacts along the white matter boundary described in Section 3 have, to the best of our knowledge, not been addressed in the literature. One possible solution to this problem would be to use a fuzzy boundary where the cost function increases to infinity along the normal direction. Currently the cost function changes instantaneously to infinity at the boundary (i.e., moving outside the boundary is infinite cost). Another issue is that the geodesics in front-propagation techniques can, in some cases, cross an edge between two adjacent tracts. We can envision a modification to our metric modulating function, e^α , that would increase the penalty for passing across such edges. This could be achieved by scaling the metric by a larger amount at edges, i.e., increasing the distances in these directions.

Acknowledgement

We would like to thank Dr. Janet Lainhart for providing the image data, funded by NIH Grant R01 MH080826, and Fangxiang Jiao for help with visualization. This work was supported by NIH Grant R01 MH084795 and the National Alliance for Medical Image Computing (NAMIC): NIH Grant U54 EB005149.

References

1. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In-vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* 44, 625–632 (2000)

2. Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine* 50, 1077–1088 (2003)
3. Fletcher, P.T., Tao, R., Jeong, W.-K., Whitaker, R.T.: A volumetric approach to quantifying region-to-region white matter connectivity in diffusion tensor MRI. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007*. LNCS, vol. 4584, pp. 346–358. Springer, Heidelberg (2007)
4. Jackowski, M., Kao, C.Y., Qiu, M., Constable, R.T., Staib, L.H.: Estimation of anatomical connectivity by anisotropic front propagation and diffusion tensor imaging. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3217, pp. 663–670. Springer, Heidelberg (2004)
5. Jeong, W.K., Fletcher, P.T., Tao, R., Whitaker, R.T.: Interactive visualization of volumetric white matter connectivity in diffusion tensor MRI using a parallel-hardware Hamilton-Jacobi solver. *IEEE Transactions on Visualization and Computer Graphics* 13(6), 1480–1487 (2007)
6. Koch, M.A., Norris, D.G., Hund-Georgiadis, M.: An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* 16, 241–250 (2002)
7. Lazar, M., Alexander, A.L.: Bootstrap white matter tractography (BOOT-TRAC). *NeuroImage* 24, 524–532 (2005)
8. Melonakos, J., Mohan, V., Niethammer, M., Smith, K., Kubicki, M., Tannenbaum, A.: Finsler tractography for white matter connectivity analysis of the cingulum bundle. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I*. LNCS, vol. 4791, pp. 36–43. Springer, Heidelberg (2007)
9. Mori, S., Crain, B.J., Chacko, V.P., van Zijl, P.C.M.: Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology* 45, 265–269 (1999)
10. O’Donnell, L., Haker, S., Westin, C.-F.: New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic pDEs and geodesics in a tensor-warped space. In: Dohi, T., Kikinis, R. (eds.) *MICCAI 2002*. LNCS, vol. 2488, pp. 459–466. Springer, Heidelberg (2002)
11. Parker, G., Wheeler-Kingshott, C., Barker, G.: Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. *IEEE Transactions on Medical Imaging* 21, 505–512 (2002)
12. Parker, G.J.M., Haroon, H.A., Wheeler-Kingshott, C.A.M.: A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. *J. Magn. Reson. Im.* 18, 242–254 (2003)
13. Pichon, E., Westin, C.-F., Tannenbaum, A.R.: A hamilton-jacobi-bellman approach to high angular resolution diffusion tractography. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3749, pp. 180–187. Springer, Heidelberg (2005)
14. Rohde, G., Barnett, A., Basser, P., Marengo, S., Pierpaoli, C.: Comprehensive approach for correction of motion and distortion in diffusion-weighted mri. *Magnetic Resonance in Medicine* 51, 103–114 (2004)

Combining Generative and Discriminative Models for Semantic Segmentation of CT Scans via Active Learning

Juan Eugenio Iglesias^{1,2}, Ender Konukoglu², Albert Montillo^{3,2},
Zhuowen Tu¹, and Antonio Criminisi²

¹ University of California, Los Angeles, USA

jeiglesias@ucla.edu, zhuowen.tu@loni.ucla.edu

² Microsoft Research, Cambridge, UK

enderk@microsoft.com, antcrim@microsoft.com

³ GE Global Research Center, Niskayuna, NY USA
montillo@ge.com

Abstract. This paper presents a new supervised learning framework for the efficient recognition and segmentation of anatomical structures in 3D computed tomography (CT), with as little training data as possible. Training supervised classifiers to recognize organs within CT scans requires a large number of manually delineated exemplar 3D images, which are very expensive to obtain. In this study, we borrow ideas from the field of active learning to optimally select a minimum subset of such images that yields accurate anatomy segmentation. The main contribution of this work is in designing a combined generative-discriminative model which: i) drives optimal selection of training data; and ii) increases segmentation accuracy. The optimal training set is constructed by finding unlabeled scans which maximize the disagreement between our two complementary probabilistic models, as measured by a modified version of the Jensen-Shannon divergence. Our algorithm is assessed on a database of 196 labeled clinical CT scans with high variability in resolution, anatomy, pathologies, etc. Quantitative evaluation shows that, compared with randomly selecting the scans to annotate, our method decreases the number of training images by up to 45%. Moreover, our generative model of body shape substantially increases segmentation accuracy when compared to either using the discriminative model alone or a generic smoothness prior (e.g. via a Markov Random Field).

1 Introduction

Large field-of-view CT scans are widely used in the diagnosis of systemic diseases, which affect several organs. Automatic segmentation of body structures has application in anomaly detection, disease assessment, change tracking, registration, navigation, and further organ-specific analysis. In this study, we present an algorithm for simultaneous segmentation of nine anatomical structures in clinical CT scans: heart, liver, spleen, l/r pelvis, l/r kidney and l/r lung. The segmentation task is cast as voxel-wise classification. In clinical CT this is

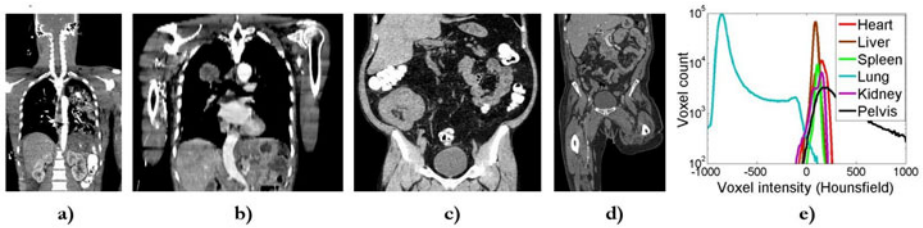


Fig. 1. Variability in our 196 clinical scans. a) Coronal view of a subject with pulmonary anomaly. b) Subject with lung tumor. c) Subject with oral contrast and abnormal kidney shape. d) Subject with amputated left leg. e) In a typical full-body scan, the tissue densities of different organs overlap considerably with one another. All these sources of variability make automatic anatomy segmentation challenging.

challenging due to similar density in different organs, presence of contrast agents, varying resolution and noise levels, variability in anatomy and small field of view (see Fig. 1). Thus, machine learning-based, supervised classifiers require a large amount of expensive, labeled data to generate high-quality segmentations. While that is the ultimate goal of our system, this study addresses the problem of finding the minimal sufficient training data within a pool of unlabeled CT scans. To do so we borrow ideas from the field of active learning.

Background. Active learning [14] studies the problem of training a robust supervised system with as little manual labeling as possible, i.e. which data, once labeled, yield the largest increase in accuracy? Most techniques are iterative. Starting from a pool of unlabeled data, a small set of samples is selected and labeled by a human expert. These enable training an initial classifier which in turn is used to select the most informative instances from the unlabeled pool, which often correspond to data that is under-represented in the current training set e.g. anatomical variations, pathologies, etc. Then, the expert provides ground truth for this new set and the classifier is updated, beginning a new iteration. Active learning suits the medical imaging domain well because unlabeled data are readily available whereas expert annotations are expensive.

There are two families of criteria for selecting samples to be annotated, based on: 1) maximizing a predefined informativeness measure (e.g. [8]); and 2) reducing the “version space” i.e. the set of all classification hypotheses consistent with the ground-truth. The most representative example of the latter is “query by committee” (QBC) [6], in which two or more classifiers sampled from the version space are used to classify an unlabeled sample. Manual labeling is then performed on the data for which the models outputs disagree. In co-testing [11], the members of the committee are trained on independent views of the data.

Contribution. Our framework combines two complementary models: a classifier that focuses on the appearance of the organs, and a generative model, which captures organ relative location and thus global, probabilistic shape information. As in QBC and co-testing, the models are jointly trained iteratively. At each

iteration, the classifier is tested on the unlabeled data and the output is evaluated under the generative model. The joint confidence of the probabilistic predictions is used to select the scans to be labeled next. As opposed to QBC and co-testing, our method: 1. focuses on increasing the prediction confidence of the two models; 2. does not require multiple views of the data; and 3. exploits long-range context via a generative model of anatomy. Additionally, we demonstrate how such shape prior increases the segmentation accuracy substantially.

Further related literature. Most approaches to joint segmentation of multiple structures rely on registration. Segmentation is achieved by deforming an atlas (or a set thereof) to the target scan. The atlas and deformation can be seen as a generative model of anatomy. Registration is computationally expensive, but works very well for organs such as the brain. However, it falters on structures whose relative locations can change substantially across subjects, which is the case in multi-organ segmentation. Post-processing is then required to refine the result [9,15,12]. On the other hand, Seifert et al. [13] make use of increasingly popular machine learning methods to approach the problem. They use a cascade of classifiers to detect key slices, then landmarks, and finally segment six organs.

2 Materials: The Labeled CT Database

Clinical CT scans from 196 different subjects acquired at different hospitals were used in this study. Most of the scans are challenging to segment due to variability in anatomy, acquisition protocols (resolution, filters, dose, etc), pathology, contrast agents, and even limb amputations (see Fig. 1). Most of the scans cover the thoracic or abdominal region, or both. Some also include the head or legs. Manual ground-truth segmentations were achieved via a semi-automatic approach [5]. All scans are resampled to $6mm$ isotropic resolution for faster calculations.

3 Methods

The proposed framework combines a discriminative and a generative model for obtaining high-quality segmentations and driving the selection of training images. Henceforth, we use the following notation: i indexes the scans in the training set, and $o \in \{1, 2, \dots, N_o\}$ indexes the $N_o = 9$ organs of interest. Appending the **background** class produces an extended set with $N_c = N_o + 1$ classes, indexed by $c \in \{1, 2, \dots, N_o, N_c\}$ (where $c = N_c$ corresponds to **background**). Subscripts d , g , and cog represent “discriminative”, “generative” and “center of gravity”.

3.1 Discriminative Voxel Classification

Following the work in [7], we have applied random forest classification [3] to the task of assigning organ class probabilities to all voxels of a previously unseen CT scan. The classifier is based on box-shaped visual features. Details are out of the scope of this paper, and can be found in the referenced work. Any other type of probabilistic classifier could also be used in our framework.

3.2 Generative Model of CT Scans

Although random forests do capture some level of context, they fail at modeling the long-range spatial relationships between organs (see Fig 3b). We address this issue by introducing a generative graphical model which captures relative organ positions and organ shapes probabilistically.

The model. The graphical model used here is shown in Fig. 2a. Its relatively few parameters allow it to be trained with very few scans, making it possible to start selecting scans actively early in learning. The model represents a CT scan by a collection of organs and the background. Each organ is represented by its centroid location \mathbf{k}_o and a probabilistic atlas of shape $A_o(\mathbf{r})$ such that the probability that the voxel at location \mathbf{r} is inside the organ is $A_o(\mathbf{r} - \mathbf{k}_o)$.

There are two coordinate systems in the model: a reference frame in which the sets of centroids from the training dataset are jointly aligned and a physical coordinate system in which all CT scans are defined. The coordinates \mathbf{k}'_o of the N_o centroids of a scan in the reference frame, stacked into a vector $\mathbf{x}'_{cog} = [\mathbf{k}'_1, \dots, \mathbf{k}'_{N_o}]^t$, follow a multivariate Gaussian distribution: $\mathbf{x}'_{cog} \sim \mathcal{N}(\bar{\mathbf{x}}'_{cog}, \Sigma_{cog})$. These coordinates are mapped to the physical frame by a simple rigid transform $\mathbf{x}_{cog} = [\mathbf{k}_1, \dots, \mathbf{k}_{N_o}]^t = s\mathbf{x}'_{cog} + \mathbf{t}$, where the scaling factor s is log-normal $\log s \sim \mathcal{N}(\overline{\log s}, \sigma_{\log s}^2)$ and the translation \mathbf{t} is improper uniform i.e. free. Finally, the organ probabilities for each of the V_{tot} voxels in the CT volume $\mathbf{p}_g = [p_{g,1}, \dots, p_{g,N_c}]^t$ are obtained by integrating the organ atlases and the centroid model. Since the atlases represent binary tests “organ o vs. rest”, \mathbf{p}_g is given by the product rule (assuming independence between the atlases):

$$p_{g,o}(\mathbf{r}) = \frac{1}{Z_g(\mathbf{r})} A_o(\mathbf{r} - \mathbf{k}_o) \prod_{o'=1, o' \neq o}^{N_o} [1 - A_{o'}(\mathbf{r} - \mathbf{k}_{o'})], \quad o \in [1, \dots, N_o]$$

$$p_{g,N_c}(\mathbf{r}) = \frac{1}{Z_g(\mathbf{r})} \prod_{o=1}^{N_o} [1 - A_o(\mathbf{r} - \mathbf{k}_o)] \quad (\text{for the background})$$

where $Z_g(\mathbf{r})$ ensures that $\sum_c p_{g,c}(\mathbf{r}) = 1, \forall \mathbf{r}$. The background requires special treatment because it has neither a centroid nor a probabilistic atlas in the model.

Learning the model. The first step to learn the centroid distributions is to create the reference coordinate system in which the sets of centroids are aligned. For each training scan i , we have $\mathbf{x}'_{i,cog} = (\mathbf{x}_{i,cog} - \mathbf{t}_i)/s_i$, where the scalings s_i and translations \mathbf{t}_i maximize the joint alignment of the centroid sets. This simple transform suffices to align the data because the scanned subject is always well aligned with the scanner bed. Moreover, the low number of parameters of the transform makes it very robust. The values of s_i and \mathbf{t}_i are obtained through an iterative, maximal agreement Procrustes alignment algorithm [16]. Then, the parameters $\overline{\log s}$ and $\sigma_{\log s}^2$ are just the sample mean and variance of $\{\log s_i\}$.

From the coordinates in the reference frame $\mathbf{x}_{i,cog}$, we can estimate the parameters of their multivariate Gaussian model. We use probabilistic principal

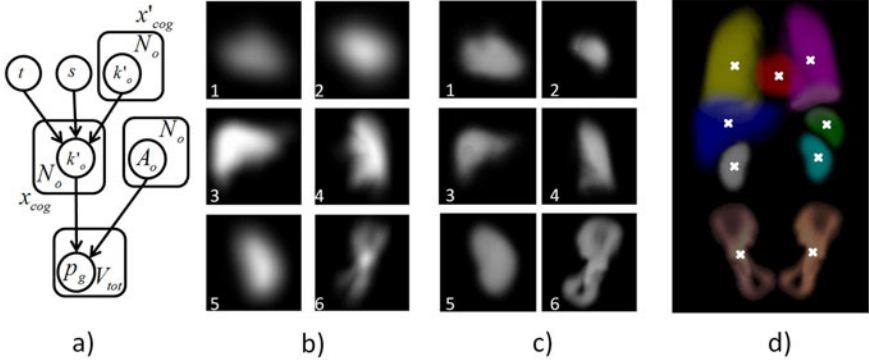


Fig. 2. Generative model: a) Graphical model. b) 3D rendering of probabilistic organ atlases, trained with 5 scans: 1.heart; 2.spleen; 3.liver; 4.left lung; 5.left kidney; and 6.left pelvis. c) Same atlases, trained on 20 scans. d) 3D rendering from a random sample of the complete model p_g (trained with 20 scans, centroids marked with crosses).

component analysis (PPCA [17]) to deal with missing data (organs out of the field of view): $\mathbf{x}'_{cog} = \bar{\mathbf{x}}'_{cog} + \Phi \mathbf{b} + \varepsilon$, where Φ is the matrix with the orthonormal principal components, $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\lambda}))$ is the shape vector, $\boldsymbol{\lambda}$ is a vector with the variances of the principal components, and $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$ is a noise vector with spherical covariance. The parameters $\bar{\mathbf{x}}'_{cog}, \Phi, \boldsymbol{\lambda}, \sigma_\varepsilon^2$ are estimated by PPCA.

Finally, the probabilistic atlases are constructed independently for each organ as follows: 1. scaling each manually labeled organ mask by s_i , to be consistent with the Procrustes analysis; 2. aligning the centroids of the masks; 3. taking the voxel-wise average; and 4. blurring the result with a Gaussian kernel. Blurring accounts for the fact that the generative model is in general not perfectly aligned to the anatomy when testing an unseen test scan (see Section 3.3). Fig. 2b-c displays the probabilistic atlases of the organs of interest trained on different numbers of scans, in order to illustrate their evolution with the size of the training dataset. A random draw from the full generative model is displayed in Fig. 2d.

3.3 Segmenting Previously Unseen Scans

This section describes how to achieve high-quality segmentations by combining the two models. This requires: 1. fitting the generative model to the classifier output; and 2. combining the models via Bayes' theorem.

Aligning the models. Given voxel-wise class probabilities from the forest classifier $\mathbf{p}_d(\mathbf{r}) = [p_{d,1}, \dots, p_{d,N_c}]^t$, the parameter space of the generative model ($s, \mathbf{b}, \mathbf{t}, \varepsilon$) is explored to maximize a similarity metric between the class distributions from the two models $\mathbf{p}_g(\mathbf{r})$ and $\mathbf{p}_d(\mathbf{r})$. Here, we minimize the Jensen-Shannon divergence $JS(P\|Q) = (1/2)[KL(P\|R) + (KL(Q\|R))]$, where $R = \frac{(P+Q)}{2}$ and $KL(P\|Q) = \sum_j P_j(\log P_j - \log Q_j)$ is the Kullback-Leibler divergence. Unlike KL, JS defines a bounded metric. The problem is then formulated as:

Table 1. Algorithm to align the generative model with the output of the classifier

1. The CT volumes are downsampled to 20mm isotropic resolution.
2. Assuming $\mathbf{b} = \mathbf{0}$ (mean centroid shape), do exhaustive search across s and t_z . We explore $t_z \in [t_{z,min}, t_{z,max}]$, with $t_{z,min}$ and $t_{z,max}$ such that translations with no overlap between the scan and the generative model are disregarded. For the scale s , we explore the interval $s \in [\exp(\overline{\log s} - 2.5\sigma_{\log s}), \exp(\overline{\log s} + 2.5\sigma_{\log s})]$. For each value of s , t_x and t_y are designed to match the x - y c.o.g. of the scan and the x - y c.o.g. of the centroids of the organs.
3. From the optimal point from the previous step, do coordinate descent on $\{s, \mathbf{b}, \mathbf{t}\}$, with $s \in [s_{min}, s_{max}]$ and \mathbf{b} constrained to stay in the ellipsoid that embraces 95% of the probability mass of the Gaussian density function.
4. From the solution from the previous step, optimize the noise ε independently for each organ, searching a sphere of radius $\sqrt{\sigma_\varepsilon^2 [\chi_3^2]_{0.95}}$ around each centroid, where $[\chi_3^2]_{0.95}$ is the 95% quantile of a Chi-square distribution with 3 degrees of freedom.

$$\theta^* = \{s^*, \mathbf{b}^*, \mathbf{t}^*, \varepsilon^*\} = \underset{\theta}{\operatorname{argmin}} \sum_{\mathbf{r}} JS(\mathbf{r}, \theta) = \underset{\theta}{\operatorname{argmin}} \frac{1}{2V_{tot}} \sum_{\mathbf{r}} \sum_{c=1}^{N_c} \left(p_{g,c}(\mathbf{r}) \log \frac{2p_{g,c}(\mathbf{r})}{p_{g,c}(\mathbf{r}) + p_{d,c}^\theta(\mathbf{r})} + p_{d,c}^\theta(\mathbf{r}) \log \frac{2p_{d,c}^\theta(\mathbf{r})}{p_{g,c}(\mathbf{r}) + p_{d,c}^\theta(\mathbf{r})} \right) \quad (1)$$

where $\theta^* = \{s^*, \mathbf{b}^*, \mathbf{t}^*, \varepsilon^*\}$ are the optimal parameters and V_{tot} is the number of voxels in the scan. This step uses well-behaved probability maps (particularly $p_{g,c}^\theta$, see Fig. 2bc) rather than pixel intensities. This makes the JS divergence smooth, which, next to the low number of degrees of freedom of the generative model, makes the optimization fast and robust. First, exhaustive search at 20mm resolution is used to initialize the two most sensitive parameters: axial translation (t_z) and scale (s). From that point, coordinate descent is used to optimize \mathbf{b} and refine \mathbf{t} and s . Finally, the noise ε is optimized independently for each organ. The algorithm is detailed in Table 1 and illustrated in Fig. 3a-c.

Bayesian semantic segmentation. The aligned generative model can be interpreted as a location prior in a Bayesian framework. The posterior probability of label L at location \mathbf{r} is therefore given by Bayes' theorem:

$$p[L(\mathbf{r}) = c] = \frac{p_{d,c}(\mathbf{r}) \cdot p_{g,c}^{\theta^*}(\mathbf{r})}{Z_L(\mathbf{r})} \quad (2)$$

where the partition function $Z_L(\mathbf{r})$ ensures that the probabilities add to one. The final hard segmentation can be obtained as the voxel-wise MAP estimate of the class probabilities i.e. $\operatorname{argmax}_c p[L(\mathbf{r}) = c]$. The effect of the generative model on the segmentation is illustrated in Fig. 3, where erroneous probability masses are removed from $p_{d,c}(\mathbf{r})$, improving the quality of the overall segmentation.

3.4 Training Set Construction through Active Learning

The second task at hand is that of building a minimal set of manually labeled CT images. Detailed algorithmic steps are presented in Table 2. In the spirit of

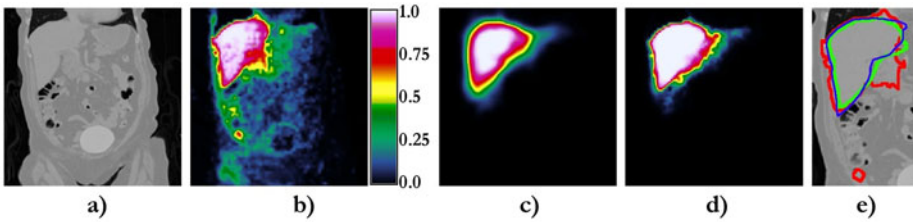


Fig. 3. Bayesian anatomy segmentation. a) Coronal slice of a sample test scan. b) Probability for the liver as output of the classifier (20 training scans). c) Aligned generative model. d) Posterior probability for liver as in Equation 2. The posterior map is much more accurate than in (b). e) Hard segmentations provided by the discriminative model alone (red) and the Bayesian model (green), as well as ground truth (blue). The generative spatial prior has a positive effect on the quality of the final segmentation (green). Please note that the segmentation (i.e. the MAP estimate) is not the 0.5-level of the probability map; that holds in binary classification, but not in our N_c -ary case.

Table 2. Active learning for optimal construction of manually segmented database

1. The generative and discriminative models are built starting with 2 labeled scans.
2. The remaining unlabeled scans are fed to the classifier, yielding multi-class probability maps for each voxel.
3. Align the generative model by minimizing JS in (1).
4. Compute disagreement via the weighted JS divergence as in (3).
5. Rejection of outlying scans via the local outlier factor (LOF) [4] on JS_w .
6. Select the unlabeled scan that maximizes JS_w and obtain its manual ground truth from a human expert.
7. Update the classifier and the generative model.
8. If the testing segmentation accuracy is satisfactory then stop. Otherwise, goto 2.

QBC and co-testing, at each iteration an expert labels the scan which maximizes the disagreement between the discriminative and generative models, which is the one with largest potential of improving the overall segmentation accuracy. To remove the bias towards larger organs, desirable when aligning the models but detrimental in active learning, we use a modified version of the JS divergence to measure the disagreement. We adapt the concept of weighted entropy of a distribution [1]: $H_w = \sum_c u_c p_c \log p_c$, where class c has an utility cost u_c and probability p_c . Making the utilities inversely proportional to the average volumes of the organs $u_c = 1/V_c$ weighs them uniformly. We define $V_c = \sum_{\mathbf{r}} A_c(\mathbf{r})$, $c \in [1, \dots, N_o]$, and V_{N_c} as the average volume of the background class in the training scans. Switching the order of $\sum_{\mathbf{r}}$ and \sum_o in (1) gives the weighted JS divergence:

$$JS_w := \frac{1}{2} \sum_{c=1}^{N_c} \frac{1}{V_c} \sum_{\mathbf{r}} \left(p_{d,c}(\mathbf{r}) \log \frac{2p_{d,c}(\mathbf{r})}{p_{d,c}(\mathbf{r}) + p_{g,c}^*(\mathbf{r})} + p_{g,c}^*(\mathbf{r}) \log \frac{2p_{g,c}^*(\mathbf{r})}{p_{d,c}(\mathbf{r}) + p_{g,c}^*(\mathbf{r})} \right) \quad (3)$$

Another important component of the algorithm is the outlier rejection strategy. Here we identify as outliers unlabeled scans for which the JS_w measure is far away from the rest of the population using the local outlier factor (LOF [4]).

LOF compares the density of the data around each point with the density at a number of nearest neighbors, computing an index which can be thresholded to detect outliers. Here, it is convenient to use an aggressive threshold (2.0) to be certain that all the outliers are detected, since the cost of leaving out informative inliers is lower than the toll of including outliers in the training set.

4 Experiments and Results

Experimental setup. This section assesses two aspects of our work: i) the accuracy of our Bayesian segmentation approach versus the discriminative classifier alone; and ii) the validity of our database construction strategy as compared to alternative techniques. On the second task, five algorithms are evaluated:

- A1. The proposed active learning approach.
- A2. Same as 1, but randomly selecting scans from the unlabeled pool.
- A3. Uncertainty sampling [8], in which the scan that maximizes the mean voxel entropy $H_{av} = (V_{tot}^{-1}) \sum_{\mathbf{r}} \sum_{c=1}^{N_c} p_{d,c}(\mathbf{r}) \log p_{d,c}(\mathbf{r})$ is selected. Our generative model is thus not used for data selection, but it is still used in segmentation.
- A4. Same as 2, but the generative model is replaced by a generic Markov Random Field (MRF) prior in segmentation. Graph cuts [2] are used to minimize:

$$\xi[A(\mathbf{r})] = - \sum_{\mathbf{r}} \log p_{d,\Lambda(\mathbf{r})} + \gamma \sum_{(\mathbf{r}_i, \mathbf{r}_j) \in \mathcal{N}} \delta[A(\mathbf{r}_i) = A(\mathbf{r}_j)]$$

where \mathcal{N} is a 6-neighborhood system and γ is the smoothness of the MRF.

- A.5 Same setup as in 2, but without any generative model at all. For organ o , the hard segmentation is computed as the largest connected component of the binary volume $\operatorname{argmax}_c p_{d,c}(\mathbf{r}) == o$ (i.e. the MAP estimate).

The following experiment was repeated 30 times and the results averaged: two scans (at least one with all the organs of interest) are randomly selected to form the initial training set. The remaining 194 are randomly split into unlabeled pool and test data. Then, unlabeled scans are iteratively added to the system according to the five algorithms, and the performance recorded using Dice's coefficient $D(A, B) = 2 \frac{|A \cap B|}{|A| + |B|}$ and Hausdorff distance (i.e. maximal surface-to-surface distance). These two metrics provide complementary perspectives: gross overlap and robustness, respectively. In case of total miss of an organ, the Hausdorff distance for an algorithm is assumed to be equal to the maximal Hausdorff distance of the other tested methods for that organ in that subject.

The system parameters were set to the following values: $T = 14$ trees, max. tree depth = 16, features tested per node = 200, training voxels = $4 \cdot 10^5$, min. info. gain = 0.01, width of kernel to blur prob. atlases = $20 \cdot n_{atl}^{-3/2}$, MRF smoothness $\gamma = 1/3$, LOF threshold = 2.0, LOF nearest neighbors = $\lceil \frac{n_{uni}}{20} \rceil$. The random forest parameters were partially motivated by computational limitations. The rest of the parameters were tuned by pilot experiments.

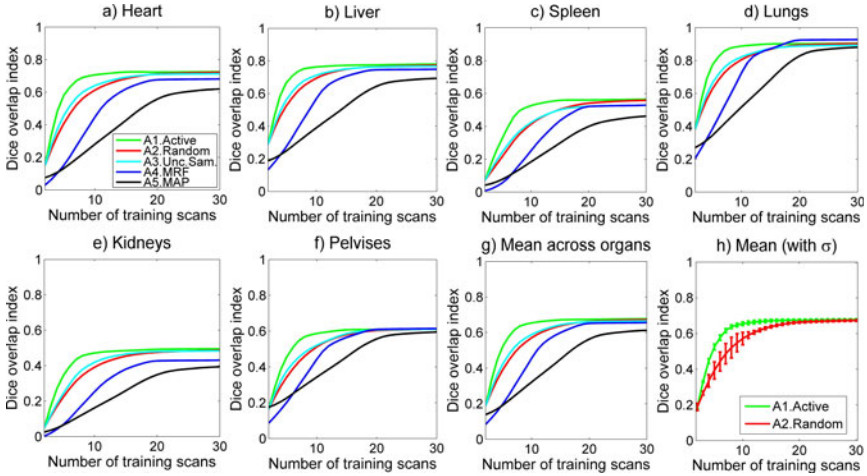


Fig. 4. Segmentation accuracy (Dice) vs. training set size for different database construction approaches. Plots a-f) are organ specific. Plot g) displays the mean for all organs. Plot h) is a zoom-in of f) that displays standard deviations (but only shows random selection and active learning for clarity).

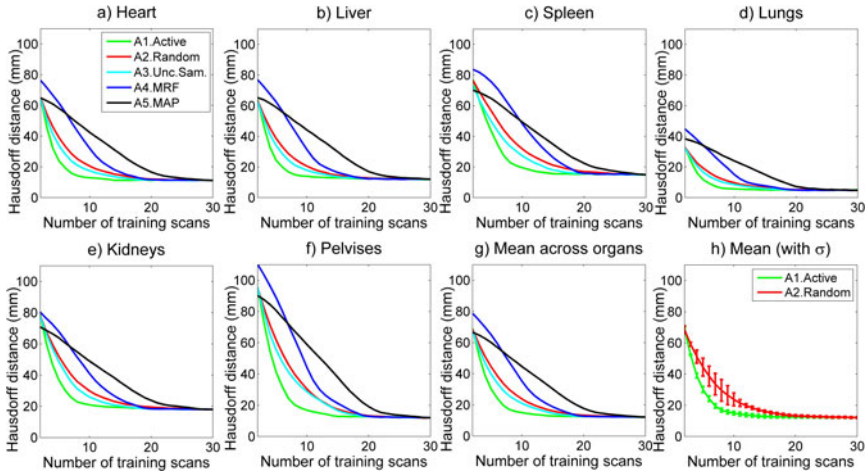


Fig. 5. Hausdorff distance vs. training set size for different database construction approaches. See caption of Figure 4 for the explanation of the plots.

4.1 Bayesian Segmentation Results

Figures 4 and 5 display the Dice's coefficient and the Hausdorff distance vs. the number of training examples for the five algorithms, whereas Fig. 6 shows the segmentations of three test scans. The accuracy is not high in absolute terms due to the difficulty of the dataset (pathologies, large variations). However, the

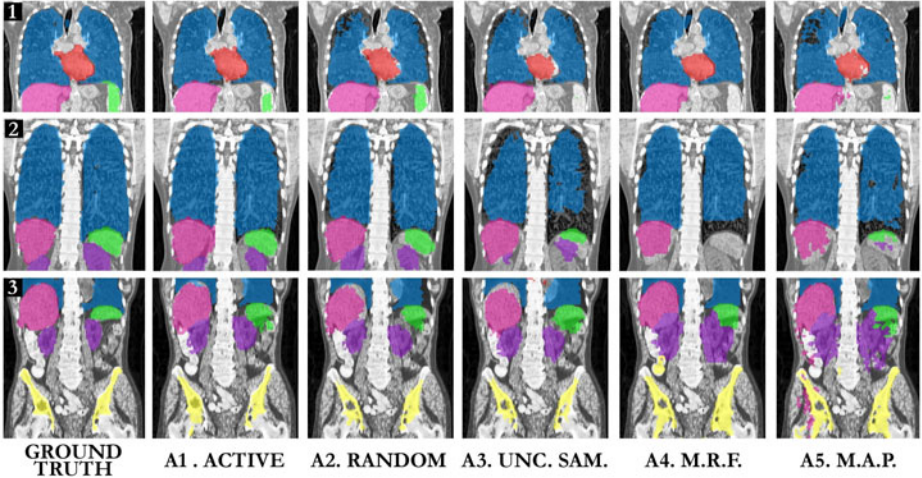


Fig. 6. Comparing segmentation results for different algorithms. Coronal slices of CT segmentations of three test scans. Different colors indicate different segmented organs. Only ten labeled scans were used for training. Our algorithm (A1. Active) produces better segmentation than other techniques, for the same training set size.

differences between active learning and random selection are still illustrated. Comparing curves A.2 and A.4, we see that our model outperforms the MRF for almost every organ, attributed to having been specifically designed for this task. For example, the MRF misses the spleen and kidneys in scans 1 and 2 in Fig. 6, whereas our method does not. Our model is also useful at the interface of neighboring organs, such as the liver and right kidney in scan 3 in the figure. The MRF only offers good results for the lung and pelvis, whose high image contrast makes discrimination possible with little contextual information. Finally, the MRF produces better segmentations than the direct MAP estimates, since the former is able to clean $p_d(\mathbf{r})$ to some extent (e.g. the liver in scan 3 in Fig. 6).

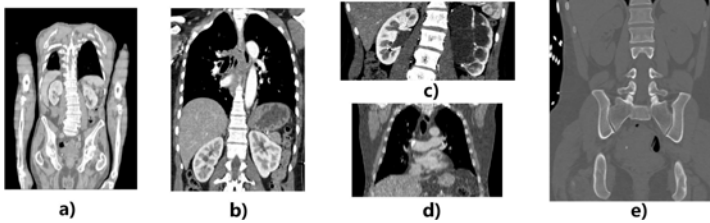
4.2 Training Database Construction Results

Figures 4 and 5 also show that our framework reduces the number of labeled scans that are necessary to achieve a desired accuracy (Table 3). Active selection is consistently better than random for every organ. At higher accuracies (Dice ≥ 0.6), the decrease in labeling effort is 40-45%. The improvement is approximately the same with respect to uncertainty sampling, which is marginally better than random selection when Dice ≤ 0.65 and marginally worse above 0.65.

Finally, Fig. 7 illustrates the scan selection process in our framework. The initial training set consists of the scans in Fig. 7a-b. Fig. 7c shows the scan with the highest disagreement (because of the cyst in the left kidney and the lack of context due to reduced field of view). Adding it to the training set could negatively affect the performance of the system, therefore the importance of the outlier

Table 3. Required number of scans to achieve different levels of accuracy using our active learning framework, random selection and uncertainty sampling

Target Dice's index	0.40	0.50	0.60	0.65	0.66	0.67
Number of training scans for A1. Active	3.6	4.7	8.2	9.4	11.1	14.3
Number of training scans for A2. Random	5.1	7.4	15.2	16.8	18.9	25.8
Number of training scans for A3. Unc.sam.	4.6	6.6	10.7	16.1	19.4	>30

**Fig. 7. Active selection of training scans.** a-b) initial training set. c) Scan with minimal weighted JS score at the first iteration, which displays a kidney with a large cyst and is rejected by LOF. d-e) Scans actually selected in the first two iterations.

rejection stage. Fig. 7d-e displays the two first scans from the unlabeled pool actually selected by our algorithm. Next to the initial two, they represent the main types of scan present in the dataset (full body, chest, abdominal, pelvic).

5 Discussion and Conclusion

A new framework for the automatic segmentation of anatomy within CT scans has been presented. A joint discriminative-generative model has been introduced to address two tasks: estimating voxel segmentation posteriors and constructing a minimal training dataset of manually segmented CT scans. Quantitative experiments demonstrated that the joint model considerably helps in: 1. attaining higher segmentation accuracy than generic smoothness priors (e.g. MRF); and 2. reducing the size of the training dataset by $\sim 45\%$ compared to alternatives.

In our Matlab / C# implementation, forest training takes 3 – 10 min. on a desktop. Segmenting a new CT scan takes ≈ 4 s. (≈ 1 s for forest testing and ≈ 3 s for model alignment). In active learning, ranking all unlabeled scans takes $\approx 4N_{unlabeled}$ s. Next we wish to exploit the algorithm's parallelism to reduce execution times further, combine the method with complementary work [10] to improve segmentation accuracy, and apply it to other image modalities, e.g. MR.

Acknowledgements

The authors would like to acknowledge Dr. S. Pathak and Dr. S. White for the data, and Dr. C. Rother and Dr. J. Wortman Vaughan for useful suggestion. J.E.

Iglesias and Dr. Tu were partially supported by NSF career award IIS-0844566 and ONR grant N000140910099.

References

1. Belis, M., Guiasu, S.: A quantitative-qualitative measure of information in cybernetic systems. *IEEE Trans. Inf. Theory* 14(4), 593–594 (1968)
2. Boykov, Y., Veksler, O., Zabih, R.: Efficient approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(12), 1222–1239 (2001)
3. Breiman, L.: Random forests. *Mach. Learn.* 45(1), 5–32 (2001)
4. Breunig, M., Kriegel, H., Ng, R., Sander, J.: LOF: identifying density-based local outliers. *Sigmod Rec.* 29(2), 93–104 (2000)
5. Criminisi, A., Sharp, T., Blake, A.: GeoS: Geodesic image segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part I. LNCS*, vol. 5302, pp. 99–112. Springer, Heidelberg (2008)
6. Freund, Y., Seung, H., Shamir, E., Tishby, N.: Selective sampling using the query by committee algorithm. *Mach. Learn.* 28(2), 133–168 (1997)
7. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for MS lesion segmentation in multi-channel MR images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010. LNCS*, vol. 6361, pp. 111–118. Springer, Heidelberg (2010)
8. Lewis, D., Gale, W.: A sequential algorithm for training text classifiers. In: *Proc. ACM SIGIR Conf. Res. and Dev. in Inf.*, pp. 3–12 (1994)
9. Linguraru, M.G., Sandberg, J.K., Li, Z., Pura, J.A., Summers, R.M.: Atlas-based automated segmentation of spleen and liver using adaptive enhancement estimation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 1001–1008. Springer, Heidelberg (2009)
10. 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), (this volume)
11. Muslea, I., Minton, S., Knoblock, C.: Active learning with multiple views. *J. Artif. Intell. Res.* 27(1), 203–233 (2006)
12. Park, H., Bland, P., Meyer, C.: Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Trans. Med. Im.* 22(4), 483–492 (2003)
13. Seifert, S., Barbu, A., Zhou, S., Liu, D., Feulner, J., Huber, M.S.M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: *Proc. of SPIE.*, vol. 7258, pp. 725902–725909 (2009)
14. Settles, B.: Active learning literature survey. *Computer Sciences Technical Report 1648*, University of Wisconsin–Madison (2009)
15. Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H., Nawano, S., Smutek, D.: Segmentation of multiple organs in non-contrast 3D abdominal CT images. *Int. J. Comput. Assisted Radiol. and Surg.* 2(3), 135–142 (2007)
16. Ten Berge, J.: Orthogonal Procrustes rotation for two or more matrices. *Psychometrika* 42(2), 267–276 (1977)
17. Tipping, M., Bishop, C.: Probabilistic principal component analysis. *J. R. Stat. Soc.: Series B* 61(3), 611–622 (1999)

Segmentation of 3D RF Echocardiography Using a Multiframe Spatio-temporal Predictor*

Paul C. Pearlman¹, Hemant D. Tagare^{1,2,3}, Ben A. Lin⁴,
Albert J. Sinusas^{3,4}, and James S. Duncan^{1,2,3}

Departments of ¹ Electrical Engineering, ² Biomedical Engineering,
³ Diagnostic Radiology, and ⁴ Internal Medicine
Yale University, New Haven, CT, USA
paul.pearlman@yale.edu

Abstract. We present an approach for segmenting left ventricular endocardial boundaries from RF ultrasound. Segmentation is achieved jointly using an independent identically distributed (i.i.d.) spatial model for RF intensity and a multiframe conditional model. The conditional model relates neighboring frames in the image sequence by means of a computationally efficient linear predictor that exploits spatio-temporal coherence in the data. Segmentation using the RF data overcomes problems due to image inhomogeneities often amplified in B-mode segmentation and provides geometric constraints for RF phase-based speckle tracking. The incorporation of multiple frames in the conditional model significantly increases the robustness and accuracy of the algorithm. Results are generated using between 2 and 5 frames of RF data for each segmentation and are validated by comparison with manual tracings and automated B-mode boundary detection using standard (Chan and Vese-based) level sets on echocardiographic images from 27 3D sequences acquired from 6 canine studies.

1 Introduction

Quantitative analysis of regional left ventricular deformation from noninvasive imaging sequences (e.g., echocardiography) can detect ischemia and ischemic injury in patients and offers important prognostic information. To this end, it has been well established that speckle tracking from high frame-rate radio-frequency (RF) ultrasound data can be used to generate dense estimates of displacement used to solve for strain and strain rate [4]. Speckle tracking is a computationally intensive process that works well only on spatially homogeneous regions of the myocardium. RF ultrasound segmentation prior to speckle tracking can provide a geometric constraint that simplifies these computations. The segmentation mask is used to remove the blood pool from image regions where speckle tracking is performed, thus significantly reducing the amount of data to be processed.

* This work is supported by grants 5R01HL082640-04 and 5R01HL077810-04.

In ultrasound, segmentation of the endocardium is particularly challenging due to characteristics artifacts such as attenuation, shadows, and dropout and is further complicated by spatially varying contrast [6]. Signal dropout is particularly challenging as it leads to gaps in boundaries that must be bridged by segmentation algorithms [8]. Previous work established that analysis of high frame rate (> 30 fps) RF images introduces important features that can be exploited for segmentation [7]. The primary benefit of using the RF signal, as opposed to B-mode datasets for boundary discrimination, is that due to the regular fiber structure of the myocardium, the phase of the RF exhibits a temporal coherence not present in the blood pool. This coherence is exploited by state-of-the-art RF speckle tracking through the analysis of complex correlations. Its advantages are documented in the ultrasound physics literature [4]. A similar advantage exists in using RF data for segmentation.

Few approaches have used RF data for segmentation. The maximum correlation coefficient (MCC) image, a derived parameter in speckle tracking, has been shown to be a useful feature for segmentation [11,5], but is computationally expensive. Also of interest is [2], where the authors introduce a spectral autoregressive model and a velocity based model for RF segmentation. This method also suffers from the expense of computing velocity estimates and is thus as computationally burdensome as the MCC based methods. The shared trait between these approaches is that they exploit differing patterns in the temporal coherence of the myocardium and blood pool in RF echocardiography. Based on similar reasoning, in [7] we introduced Segmentation using a Linear Predictor (SLiP), where a computationally efficient linear model was employed for each tissue over two frames, with the residue used as a segmentation feature. While our previous method achieved some success at preventing the contour from leaking through moderate dropout, there were the following limitations to the model. First, solving the linear predictor equations required significant memory allocation which was compensated for by decimation of predictor parameters, thus negatively influencing the segmentation results. Second, persistent dropout and particularly poor contrast still lead to poor segmentations. The two-frame model did not adequately compensate for these effects.

The model proposed in this work generalizes the approach in [7] and introduces a new and innovative model that leverages an arbitrary number of frames. The coefficients of the predictor are also modeled as a tensor product to greatly simplify computations. The image is modeled with a conditional probability relating the frame we wish to segment to an ensemble of future time points. We also include a spatial term for the frame being segmented. This term is designed to address the tendency of the linear prediction based terms to underestimate the blood pool. The primary contribution of our new multiframe predictor based algorithm (mSLiP) is that it extends the theory behind our previous model such that our segmentation feature makes use of significantly more data, greatly improving the segmentation accuracy. We adopt a *maximum a posteriori* (MAP) approach with a level set active contour. We compare our method to a traditional Chan-Vese intensity based level set segmenter on B-mode images [1]. We also

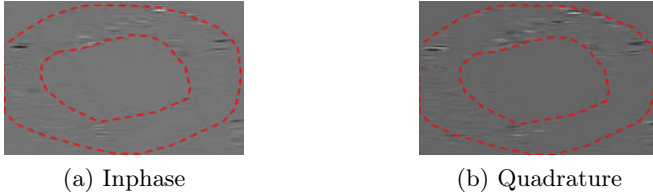


Fig. 1. Slice of 3D analytic image containing myocardium and blood pool (vertical is axial direction and horizontal is angular)

analyze the benefit of using between two and five frames in the mSLiP segmentation to determine how many frames contribute to more accurate segmentations of our data set. Additional validation is performed by comparison with manual segmentations.

2 Materials and Methods

2.1 Signal Model

Data Set. The interaction of the base-band RF signal with the blood pool and the myocardium is of interest for this work. The blood pool is made up of plasma and other elements that are much smaller than the wavelength of the ultrasound, while the myocardium is made up of primarily muscular fibers interspersed with blood vessels. The fibers make up 90% of the myocardial tissue and are thus responsible for most of the received signal [9]. The regular structure of the fibers is responsible for the temporal coherence in the data that we exploit with our algorithm. Likewise, the lack of regular structure in the blood pool causes the signal to rapidly decorrelate. It is of particular interest to us that our data is acquired at a high frame rate so that the motion of the speckle between frames is of the same order as the motion of the boundaries. As a result, a single linear predictor for each medium (blood pool and myocardium) suffices for segmentation.

Preprocessing. The input pulse produced by the ultrasound machine is a real, bandlimited signal. Since the effects of transmission and reflection in ultrasound are linear, the signal we record is also real and bandlimited and can thus be recovered by coherent demodulation [3]. The demodulation results in a vector image consisting of an in-phase and quadrature component, and is referred to as the analytic signal. It is this vector image we are interested in segmenting. An example slice of an analytic image containing myocardium and blood pool is shown in Figure 1.

2.2 MAP Estimation

We are interested in segmenting the current complex valued 3D frame, I^1 , given a sequence of subsequent frames, $\{I^2, \dots, I^l\}$. The domain of I^1 is Ω and C

is a boundary between two disjoint regions in the image, Ω_C and $\tilde{\Omega}_C$, where $\Omega = \Omega_C \cup \tilde{\Omega}_C$. The voxels of $\{I^1, \dots, I^l\}$ are indexed by k , where k runs over the voxels of $\{I^1, \dots, I^l\}$ in a scan line fashion. We represent C with an implicit level set function, ϕ , where $C = \{k \mid \phi(k) = 0\}$. The linear model we introduce in the following captures spatio-temporal correlation, but the probabilities that arise in the segmentation model are on the residues of the model. Since the residues are uncorrelated, the conditional model for I^1 is independent identically distributed (i.i.d.). We thus have the following log likelihood

$$\begin{aligned} \log p(I^1 \mid I^2, \dots, I^l, \phi, \Theta) &= \sum_{k \in \Omega} H(\phi_k) \log p_1(I_k^1 \mid I_k^2, \dots, I_k^l, \Theta_1) \\ &+ \sum_{k \in \tilde{\Omega}} (1 - H(\phi_k)) \log p_2(I_k^1 \mid I_k^2, \dots, I_k^l, \Theta_2) \end{aligned} \quad (1)$$

where Θ is a vector of parameters. We now write the probability for the voxels in each region, $n = \{1, 2\}$, as

$$p_n(I_k^1 \mid I_k^2, \dots, I_k^l, \Theta_n) \propto p_n(I_k^1, \dots, I_k^l \mid I_k^1, \alpha_n, \beta_n, \sigma_n) \cdot \Pi_n(I_k^1 \mid \mu_n, \tau_n). \quad (2)$$

where α_n and β_n are the coefficients of the linear predictor described in Section 2.5. σ_n , μ_n , and τ_n are the parameters of the probability distributions.

2.3 Probability for Single Frame Speckle

Separating I_k^1 into magnitude and phase components, the prior is given by

$$\begin{aligned} \Pi_n(I_k^1 \mid \mu_n, \tau_n) &= \Pi_n(|I_k^1| \mid \mu_n, \tau_n) \cdot \Pi(\angle I_k^1) \\ &\propto \Pi_n(|I_k^1| \mid \mu_n, \tau_n). \end{aligned} \quad (3)$$

The phase, $\angle I_k^1$, is i.i.d. with respect to k and uniformly distributed on $[0, 2\pi]$, so it does not contribute to the segmentation. Because the RF is assumed to be circular Gaussian, $\Pi_n(|I^1| \mid \mu_n, \tau_n)$ is Rician. Previous experimental work shows that many different unimodal models yield comparable segmentation accuracy [10], so for convenience we let $|I^1| \sim N(\mu_n, \tau_n)$.

2.4 Multiframe Conditional

We now seek to reduce the complexity of the multiframe conditional probability by assuming a Markov relation between all frames. By means of successive application of Bayes Rule and the Markov relation, we have the following:

$$\begin{aligned} &p_n(I_k^l, \dots, I_k^2 \mid I_k^1, \alpha_n, \beta_n, \sigma_n) \\ &\propto p_n(I_k^l \mid I_k^{l-1}, \alpha_n, \beta_n, \sigma_n) \cdot p_n(I_k^{l-1} \mid I_k^{l-2}, \alpha_n, \beta_n, \sigma_n) \cdot \dots \\ &\quad \cdot p_n(I_k^{r+1} \mid I_k^r, \alpha_n, \beta_n, \sigma_n) \cdot \dots \cdot p_n(I_k^2 \mid I_k^1, \alpha_n, \beta_n, \sigma_n) \end{aligned} \quad (4)$$

2.5 Linear Predictor

We model the complex value I_k^r linearly using a neighborhood of voxels around I_k^{r+1} with complex coefficients and residue. A diagram of the predictor is shown in Fig. 2. We assume that the process, $\{I^1, \dots, I^l\}$, can be treated as locally stationary in time because the sequence is acquired at a high frame rate (> 30 fps) relative to the motion of the structures being imaged. Thus, the same prediction coefficients are used for all predictors. The residues of the linear predictors are thus given by

$$e_k^r = \begin{cases} I_k^{r+1} - \sum_j \alpha_{1,j} \mathcal{R}e \left\{ I_{k,j}^r \right\} \\ \quad - i \sum_j \beta_{1,j} \mathcal{I}m \left\{ I_{k,j}^r \right\}, & k \in \Omega_C, \\ I_k^{r+1} - \sum_j \alpha_{2,j} \mathcal{R}e \left\{ I_{k,j}^r \right\} \\ \quad - i \sum_j \beta_{2,j} \mathcal{I}m \left\{ I_{k,j}^r \right\}, & k \in \tilde{\Omega}_C. \end{cases} \quad (5)$$

There are thus two predictors for each pair of frames where Ω_C represents the blood pool and $\tilde{\Omega}_C$ represents the myocardium. Predictor coefficients are indexed by j , $I_{k,j}^r$ is the j^{th} neighbor of voxel I_k^r , and $i = \sqrt{-1}$. Since the probability density function of the RF is assumed to be circular Gaussian, we have

$$p_n(I_k^1, \dots, I_k^2 | I_k^1, \alpha_n, \beta_n, \sigma_n) \propto \prod_{r=1}^{l-1} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{1}{2\sigma_n^2} \mathcal{R}e \left\{ e_k^{r\top} e_k^r \right\}\right). \quad (6)$$

2.6 Separable Coefficient Regression

Incorporation of multiple frames in the prediction leads to a large design matrix resulting in the regression algorithm becoming computationally burdensome and memory intensive. To reduce the size of the design matrix, we enforce a structure on our coefficients by assuming that each dimension of the neighborhood of voxels used for the predictor (i.e. one scan line dimension and two angular dimensions) can be treated independently. Thus, the coefficients of our predictor are the tensor product of the coefficients estimated in three simpler regressions.

Let the predictor coefficients be indexed over each dimension by x, y , and z . We then have $\alpha_{xyz} = a_x a_y a_z$ and $\beta_{xyz} = b_x b_y b_z$. A diagram of the model is shown in Fig. 3. To estimate regression parameters for each dimension, we first collapse the other two dimensions by means of weighted summation. The weights are the most current regression parameters for the other two dimensions. The model is then updated for the current dimension.

This tensor product simplifies the complexity of our regression by greatly reducing the space in which the regression is performed. Assuming that our search window has dimensions $X \times Y \times Z$, there are XYZ parameters in our predictor. By using the separable coefficient approach, we need only calculate three coefficient tensors of length X, Y , and Z , so we reduce the number our parameters calculated in our regression to $X + Y + Z$.

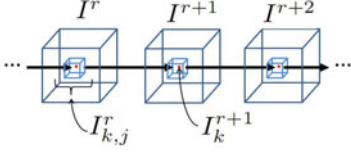


Fig. 2. Linear predictor for point k

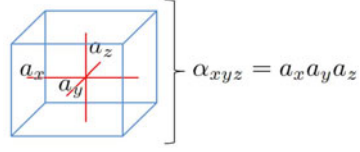


Fig. 3. Tensor product for inphase regression coefficients

2.7 Segmentation Model

Segmentation is performed by maximizing the following log posterior probability with respect to ϕ .

$$\begin{aligned}
& \max \log p(I^1 | I^2, \dots, I^l, \phi, \Theta) \\
&= \max \sum_{k \in \Omega} H(\phi_k) \left[\sum_{r=1}^{l-1} \log p_1(I_k^{r+1} | I_k^r, \alpha_1, \beta_1, \sigma_1) \right. \\
&\quad \left. + \log \Pi_1(|I_k^r| | \mu_1, \tau_1) \right] \\
&+ \sum_{k \in \Omega} (1 - H(\phi_k)) \left[\sum_{r=1}^{l-1} \log p_2(I_k^{r+1} | I_k^r, \alpha_2, \beta_2, \sigma_2) \right. \\
&\quad \left. + \log \Pi_2(|I_k^r| | \mu_2, \tau_2) \right] \\
&= \max \sum_{k \in \Omega} H(\phi_k) \left[\sum_{r=1}^{l-1} \left(\log \frac{\sigma_2}{\sigma_1} - \frac{(e_k^r)^2}{2\sigma_1^2} + \frac{(e_k^r)^2}{2\sigma_2^2} \right) \right. \\
&\quad \left. + \log \frac{\tau_2}{\tau_1} - \frac{(|I_k^1| - \mu_1)^2}{2\tau_1^2} + \frac{(|I_k^1| - \mu_2)^2}{2\tau_2^2} \right] \\
&- \left[\sum_{r=1}^{l-1} \left(\log \sqrt{2\pi}\sigma_2 - \frac{(e_k^r)^2}{2\sigma_2^2} \right) + \log \sqrt{2\pi}\tau_2 - \frac{(|I_k^1| - \mu_2)^2}{2\tau_2^2} \right] \quad (7)
\end{aligned}$$

As in [7], we introduce a prior that promotes the smoothness of the predictor coefficients. This prior is given by

$$\begin{aligned}
E_S(\alpha_1, \beta_1, \alpha_2, \beta_2) &= \frac{1}{2}\alpha_1 \mathbf{W}^T \mathbf{W} \alpha_1 + \frac{1}{2}\alpha_2 \mathbf{W}^T \mathbf{W} \alpha_2 \\
&\quad + \frac{1}{2}\beta_1 \mathbf{W}^T \mathbf{W} \beta_1 + \frac{1}{2}\beta_2 \mathbf{W}^T \mathbf{W} \beta_2 \quad (8)
\end{aligned}$$

where \mathbf{W} is a first order finite difference matrix. We also use a prior on the arc-length of the propagating front, $E_{AL}(\phi_k)$, as in [11]. Maximizing the likelihood is equivalent to minimizing its negative, so the overall energy is thus given by

$$E(I_k^1, \dots, I_k^l, \phi_k, \Theta) = -l(I_k^1, \dots, I_k^l, \phi_k, \Theta) + \lambda_{AL} E_{AL} + \lambda_S E_S \quad (9)$$

where λ_S and λ_{AL} are weights on the smoothness priors.

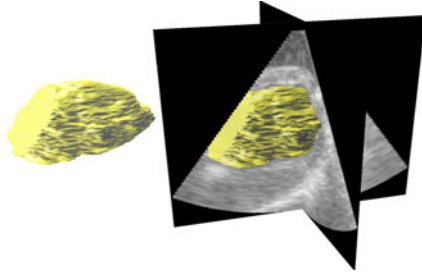


Fig. 4. A typical segmentation surface

2.8 Optimization

We iteratively minimize the energy functional as follows: 1) Initialize ϕ inside the blood pool. 2) Update the non-level set parameters of the model. 3) Update ϕ . We then iterate steps 2 and 3 until a local minima of the energy functional is reached. The level set is updated by gradient descent, α_n and β_n are updated by conjugate gradient descent on the coefficient tensors for each dimension, and closed form solutions are employed to estimate other model parameters. These solutions are given by

$$\hat{\mu}_n = \frac{1}{N} \sum_{k=1}^N |I_k^1| \quad \hat{\tau}_n^2 = \frac{1}{N} \sum_{k=1}^N (|I_k^1| - \hat{\mu}_n)^2 \quad \hat{\sigma}_n^2 = \frac{1}{N} \sum_{k=1}^N (e_k^r)^2 \quad (10)$$

where N is the number of voxels in the image. These are the maximum likelihood estimates of μ_n, τ_n , and σ_n for normally distributed random variables.

The introduction of separable coefficient regression by means of the previously mentioned tensor product greatly simplifies the computation of the predictor coefficients by significantly reducing the number of coefficients being computed. Nonetheless, we are presented with solving a very large system of equations which is computationally difficult to solve by inversion, so we use conjugate gradient descent. Initially we assume that all parameters are equal and alternately update the contour and prediction coefficients until convergence. The conjugate gradient method is particularly well suited for our linear model and typically converges in just a few iterations.

3 Experiments and Results

3.1 Surgical Preparation and Ultrasound Acquisition

We obtained 27 RF sequences from 6 canines (with an average weight of 20 kg) using a Philips iE33 4D ultrasound imaging system (Philips Health care, Andover, MA) at sampling rates of 51-89 volumes per second. All animal imaging studies were performed with approval of the Institutional Animal Care and Use Committee in anesthetized open-chested animals supported on mechanical ventilation. Images were acquired during brief suspensions of mechanical ventilation

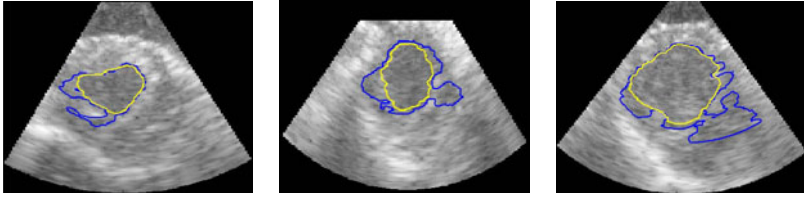


Fig. 5. Slices from 3D images where Chan-Vese contours leak through moderate dropout

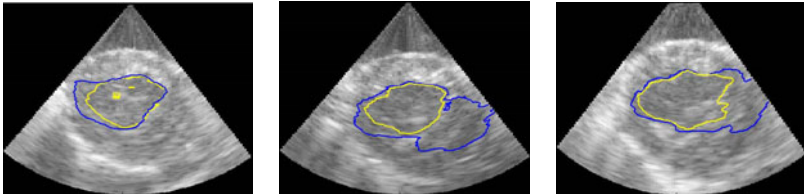


Fig. 6. Slices from 3D images where Chan-Vese contours leak through significant dropout

with an X7-2 transducer at 4.4 MHz fixed in a water bath over the heart. Time points included both baseline and one hour after surgical occlusion of the left anterior descending coronary artery. It was established in [7] that the status of regional and global function after coronary occlusion does not affect our segmentation; thus it is not taken into account in our results. It was also shown in [7] that SLiP worked comparably well at both end-diastole and peak systolic ejection (representing minimum and maximum average motion of the ventricle respectively), so we did not consider the time point in the cycle independently in this work. One image from each acquired cycle was segmented with the number of cycles obtained for each animal varying from 3 to 6. The images were chosen to represent a cross section of time points in the cardiac cycle.

To exploit the temporal continuity in myocardial speckle, the search window for the coefficients of the predictor is chosen such that encompasses distances comparable to the correlation length of the speckle. For our data this is typically on the order of 30 voxels in the axial direction and 4 voxels in each angular direction. This window is chosen by visual inspection and is fixed for all experiments. To further reduce model complexity, the number of coefficients is decimated by a factor of two in each dimension, so each coefficient represents multiple voxels.

3.2 Goals

The experiments had two goals. The first was to demonstrate that the multiframe conditional model produced more accurate segmentations than the previous approach while preserving its robustness to certain image inhomogeneities, such as moderate dropout. The second was to establish how the solutions varied based on how many time points were included in the analysis. To this end, the new

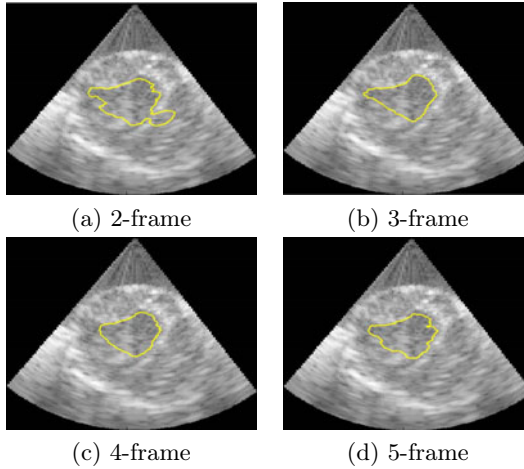


Fig. 7. Slices from 3D images showing typical segmentations using 2,3,4, and 5 frames

mSLiP algorithm was implemented with 2, 3, 4, and 5 frames from the cardiac sequence and compared to manual tracings of B-mode data and automated B-mode boundary detection using standard (Chan and Vese-based) level sets [1]. Manual tracings were performed by one expert and thus do not represent a true gold standard because expert contours typically vary. The Chan-Vese method is chosen for comparison because it assumes intensity is homogeneous over each region and only assumes a smoothness prior for the boundary. It is inherently related to our algorithm as one-frame mSLiP reduces to the Chan-Vese model. While comparison with the Chan-Vese method clearly shows the added value of the proposed method, it would not typically be used in practice for segmentation in echocardiography.

3.3 Advantages of the Proposed Method

Examples of the performance of the algorithms are shown on slices from the 3D data sets in figures 4, 5, 6, and 7. A typical segmentation surface is shown in Fig. 4. While the results are generated in the frustum coordinate system on analytic images, they are overlaid on the corresponding B-mode images that have been scan-line converted into Cartesian coordinates for ease of visual inspection and to present the anatomy in a familiar aspect ratio. Fig. 5 demonstrates typical cases where the Chan-Vese contours leak out through a dropout in signal intensity and mSLiP segmentations adhere to the endocardium. In all cases, the mSLiP segmentation adheres to the proper boundary. While the effect on segmentation quality is significant, these images show cases where the contour leaks out through relatively minor signal dropout. Fig. 6 demonstrates the effect of more significant dropout on the segmentation. Fig. 7 includes a typical example that

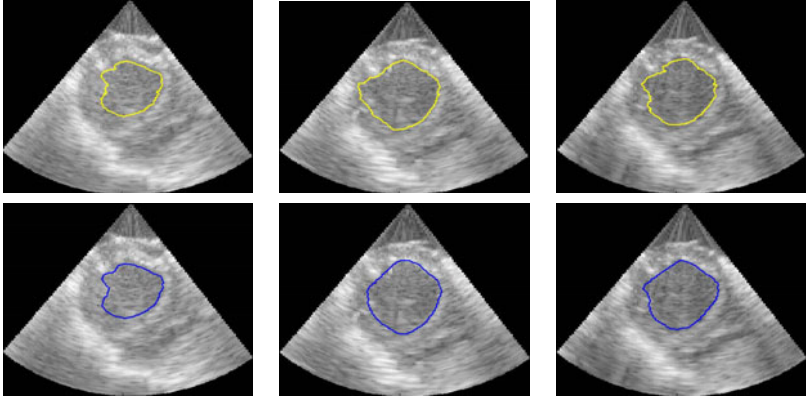


Fig. 8. Slices from 3D images showing 4-frame mSLiP segmentations (yellow) and manual tracings (blue)

demonstrates how the segmentation results vary based on the amount of temporal information used in the analysis. In this data set, the 2-frame mSLiP still leaks through dropout, but the inclusion of more frames in the predictor prevents this. Also note that the 4-frame result is superior to the 5-frame segmentation.

3.4 Validation and Results

The automated segmentations generated by our algorithm and the Chan-Vese approach were compared to manual tracings. The algorithms were quantitatively evaluated based on the following three segmentation quality indices: 1) Hausdorff distance (HD); 2) mean absolute distance (MAD), and the Dice coefficient. So that the contour distance measures (HD and MAD) were anatomically meaningful, all results were scan-line converted into Cartesian coordinates prior to quantitative evaluation. For visual evaluation, typical slices from segmentation surfaces along with manual tracings are shown in Fig. 8. The results of this analysis are shown in figures 9, 10, and 11. Of particular interest for this work is the improvement in the Dice coefficient as it represents clinically relevant volume correspondence. Note that the improvement in the Dice coefficient shows diminishing returns. This is the result of the motion of the muscle as time points further from the frame being segmented are used. Similar improvements can be observed in both HD and MAD. It should be noted that, based on the confidence intervals shown in the plot, the improvement in HD is not necessarily of any significance, while the improvement in MAD tracks with the improvement in the Dice Coefficient, which can be expected since MAD represents the average distance from the correct contour and thus typically improves as volume correspondence improves.

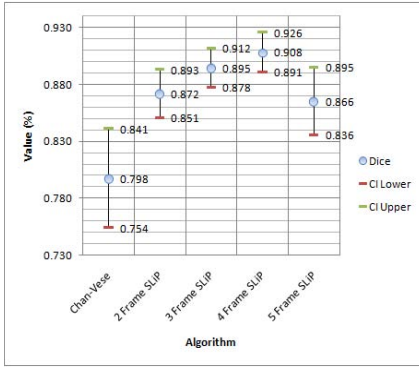


Fig. 9. Dice Coefficient

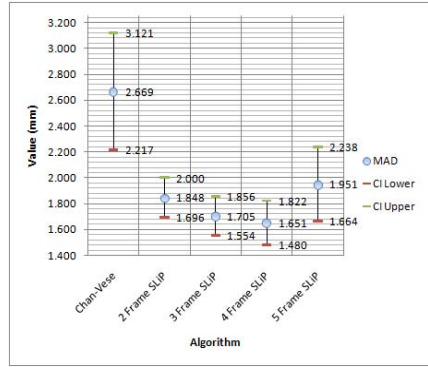


Fig. 10. Mean Absolute Distance

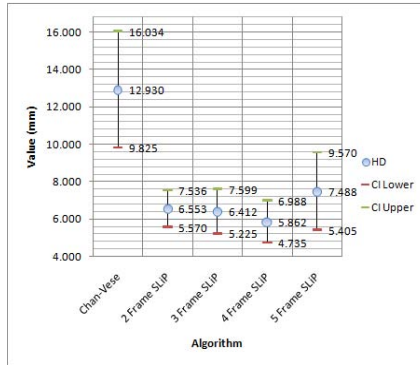


Fig. 11. Hausdorff Distance

4 Discussion and Conclusions

We have proposed a conditional probability model for segmentation from RF echocardiography that jointly models the probability of each region based on relating the frame of interest to a subsequent sequence of frames and the complex intensities of the frame being segmented. Frames are related by means of a spatio-temporal linear predictor whose coefficients are chosen in a least squares sense based on complex speckle values. The algorithm relies solely on the signal rather than *a priori* knowledge of expected shape. The proposed method also utilizes temporal coherence in the data without computing expensive tracking parameters prior to segmentation. Finally, the mSLiP segmentation feature does not make a piecewise homogeneous assumption for the image and thus does not leak through boundaries that have relatively poor contrast (i.e., moderate dropout).

This approach is intended as a preprocessing step in RF phase-based speckle tracking. Because the motion of the blood is irregular, performing speckle tracking on the blood pool provides no meaningful information. The mSLiP

objective function segments based on the same concept, so it is well suited for providing a geometric constraint for speckle tracking.

The key innovation of this work is the use of correlations among an ensemble of frames towards the segmentation of a single frame. This leads to a better estimation of predictor parameters and provides more context for the segmenter when there is prominent signal dropout. The results showed a significant improvement in segmentation accuracy, but that including more temporal data in the analysis only improves the results within a small range of time points. The number of time points must be chosen such that the motion of the ventricle over the sequence is small.

Future work will focus on schemes to include more temporal data without incurring diminishing returns in segmentation results. It is also of interest to see if these methods can be extended to capture the epicardium, which is currently complicated by a difficulty in discriminating the boundary between the left ventricle and the liver (the liver also demonstrates a strong temporal coherence).

References

1. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. on Imag. Proc.* 10(2), 266–277 (2001)
2. Dydenko, I., Friboulet, D., Gorce, J.M., D’hooge, J., Bijmens, B., Magnin, I.E.: Towards ultrasound cardiac image segmentation based on the radiofrequency signal. *Med. Imag. Anal.* 7(3), 353–367 (2003)
3. Langton, C.: Hilbert transform, analytic signal, and the complex envelope. *Sig. Proc. and Sim News* (1999)
4. Lubinski, M.A., Emelianov, S.Y., O’Donnell, M.: Speckle tracking methods for ultrasonic elasticity imaging using short-time correlation. *IEEE Trans. Ultra Ferro Freq. Cont.* 46(1), 82–96 (1999), <http://dx.doi.org/10.1109/58.741427>
5. Nillesen, M.M., Lopata, R.G.P., Huisman, H.J., Thijssen, J.M., Kapusta, L., de Korte, C.L.: 3D cardiac segmentation using temporal correlation of radio frequency ultrasound data. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 927–934. Springer, Heidelberg (2009)
6. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: a survey. *IEEE Trans. on Med. Imag.* 25(8), 987–1010 (2006)
7. Pearlman, P.C., Tagare, H.D., Sinusas, A.J., Duncan, J.S.: 3D radio frequency ultrasound cardiac segmentation using a linear predictor. In: Jiang, T., Navab, N., Plum, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 502–509. Springer, Heidelberg (2010)
8. Qian, X., Tagare, H.D.: Overcoming dropout while segmenting cardiac ultrasound images. In: *Proc. ISBI*, pp. 105–108 (2006)
9. Shung, K.K., Thieme, G.A.: *Ultrasonic scattering in biological tissues*. CRC Press, Boca Raton (1993)
10. Tao, Z., Tagare, H.D., Beaty, J.D.: Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images. *IEEE Transactions on Medical Imaging* 25(11), 1483–1491 (2006)
11. Yan, P., Jia, C.X., Sinusas, A., Thiele, K., O’Donnell, M., Duncan, J.S.: Lv segmentation through the analysis of radio frequency ultrasonic images. *Proc. IPMI 20*, 233–244 (2007)

Optimal Graph Based Segmentation Using Flow Lines with Application to Airway Wall Segmentation

Jens Petersen¹, Mads Nielsen¹, Pechin Lo¹, Zaigham Saghir²,
Asger Dirksen², and Marleen de Bruijne^{1,3}

¹ Department of Computer Science, University of Copenhagen, Denmark

² Department of Respiratory Medicine, Gentofte Hospital, Denmark

³ Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands

Abstract. This paper introduces a novel optimal graph construction method that is applicable to multi-dimensional, multi-surface segmentation problems. Such problems are often solved by refining an initial coarse surface within the space given by graph columns. Conventional columns are not well suited for surfaces with high curvature or complex shapes but the proposed columns, based on properly generated flow lines, which are non-intersecting, guarantee solutions that do not self-intersect and are better able to handle such surfaces.

The method is applied to segment human airway walls in computed tomography images. Comparison with manual annotations on 649 cross-sectional images from 15 different subjects shows significantly smaller contour distances and larger area of overlap than are obtained with recently published graph based methods.

Airway abnormality measurements obtained with the method on 480 scan pairs from a lung cancer screening trial are reproducible and correlate significantly with lung function.

Keywords: Segmentation, graph, flow lines, airway walls.

1 Introduction

Graph based methods similar to Wu and Chen's [10] polynomial time solutions to optimal net surface problems using minimum cut algorithms have seen a growing use within medical image segmentation in the last couple of years. The methods can find the globally optimal solution for multiple surfaces in multiple dimensions given surface cost functions, and various geometric constraints [4,5,9,11].

In order to use these methods, the segmentation problem needs to be transformed from image space to some discretized graph space defined by a set of columns. Each column is associated with a point on the sought surface and represents the set of possible solutions, or positions, it can take. The suitability of the graph space depends on how well plausible solutions in image space can be represented. For instance if a graph column does not cross the sought surface,

then this surface can not be represented in graph space. Similarly, admissible solutions in graph space should represent valid surfaces in image space. If this is not the case the found solutions can for instance have self-intersections. It is also important that the transformation allows for a meaningful representation of the surface cost functions and geometric constraints, such as for instance surface smoothness.

There have been methods using simple mathematical transformations, such as [49], in which images of tubular airway segments were unfolded using polar transforms in two or three-dimensions. The graph columns were oriented perpendicular to the resulting straight contours or terrain like surfaces allowing for an easy representation of surface smoothness constraints. However such approaches only work with very simple shapes.

In [5] a rough initial segmentation was used to define the transform. The columns were placed at surface points and oriented along the surface normals. By limiting their length to the minimum distance to the medial axes of the initial segmentation, problems with intersecting columns and thus possible self-intersecting surfaces were avoided. The approach can however result in columns that are too short to reach the desired solution, as shown in Fig. 1(a). In such areas the curvature of the surfaces through the columns also often does not follow the curvature of the sought surfaces, which can make it harder to construct good segmentation cost functions.

Yin et al. [11] suggested columns inspired by the non-intersecting property of electric lines of force. The columns were constructed by placing electrical charges at surface points of the initial segmentation and tracing the electric lines of force within the field inward and outward. The method is computationally intractable for large scale problems, as every surface point charge influences the computation

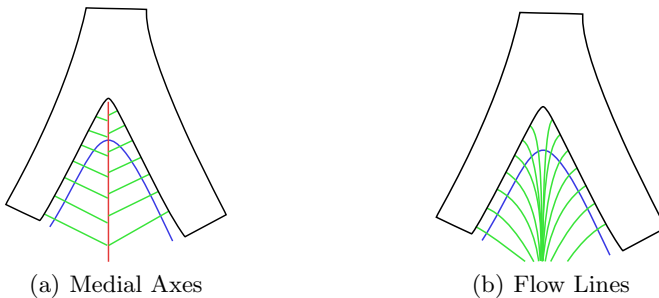


Fig. 1. Figure 1(a) illustrates the fishbone like structure of surface normal direction columns (*green*) based on the distance to the medial axis (*red*) in areas of the initial segmentation (*black*) with high curvature. Notice that the four inner-most columns do not cross the sought surface border (*blue*), which means that the segmented surface will be wrong in these positions. Figure 1(b) shows the advantage of columns based on flow lines (*green*). Notice that all columns cross the sought surface, and the curvature of the surfaces through the columns are more similar to the curvature of the sought surface.

of every electric line of force. Furthermore the electric lines of force can behave erratically if the initial segmentation contains small scale errors or noise.

We suggest to combine a regularization of the initial segmentation with a curvilinear transformation using an implementation which is computationally tractable for large scale problems. The columns are defined by greatest ascent and descent flow lines, calculated from the convolution of the initial segmentation with a C^1 continuous integrable filter. Such flow lines are uniquely defined and also non-intersecting, and fast approximations can often be computed by limiting the extent of the convolution kernels. Furthermore, regularization terms are naturally incorporated into the framework, making it easier to deal with noise and small errors in the initial segmentation. Figure [1\(b\)](#) illustrates the concept.

We apply the method to the problem of segmenting the human airway walls in three dimensions in computed tomography images, to allow quantification of abnormalities such as airway thickening and bronchiectasis. This is a dual surface problem, consisting of an inner and an outer surface boundary, between the airway lumen and the wall and between the lung parenchyma and the wall respectively, where the many bifurcations form regions of high curvature that would cause problems for conventional graph construction approaches.

2 Method

2.1 Initial Segmentation

We will assume the existence of a coarse initial segmentation, a single object given by the voxels in the set S , whose surface should be roughly similar to the surfaces we are looking for. For our application we used the airway tree segmentation algorithm described in [7](#), which returns a binary three-dimensional segmentation of the airway lumen.

The initial segmentation needs to be converted to a mesh. To this end, we used vertices at the center of each surface face and the neighborhood given by the face edge neighbors, such that each vertex has 4 neighbors. We will denote the vertices in this mesh with V_B , and represent the neighborhood with an edge set E_B , where $(i, j) \in E_B$ denotes that the vertices $i, j \in V_B$ are neighbors. The resolution of the mesh used in the experiments was $0.5 \text{ mm} \times 0.5 \text{ mm} \times 0.5 \text{ mm}$.

2.2 Flow Lines

A flow line in a vector field is tangent to the field at each point, and if it is defined in terms of some scalar potential, it will follow the direction with the greatest rate of change of this potential. Consider the potentials E defined by the convolution:

$$E(\mathbf{x}) = \int Q(\hat{\mathbf{x}})R(\hat{\mathbf{x}} - \mathbf{x})d\hat{\mathbf{x}} , \quad (1)$$

where \mathbf{x} is the position to be evaluated and Q is an indicator function for the initial segmentation:

$$Q(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S \\ 0 & \text{if } \mathbf{x} \notin S \end{cases} .$$

In this work we experimented with generating the potentials from two different types of filters. The first are of the form:

$$R(\mathbf{x}) = \frac{1}{\alpha + |\mathbf{x}|^2} , \quad (2)$$

where $\alpha > 0$ is a regularization constant, which makes R well defined for all \mathbf{x} . Increasing the value of α has the added effect of smoothing the result, which is useful if the initial segmentation contains a lot of noise. Notice that when $\alpha \rightarrow 0$, E becomes equal to the electric potential times a constant arising from a 'charge density' given by Q . This option therefore is similar to the method introduced in [11], but rather than having a discrete set of surface point 'charges', which introduce local singularities, it is defined everywhere and thus allows us to trace the flow lines consistently through the surface. We will refer to this as the potential kernel/filter.

The second type of flow line is given by the Gaussian kernel, that is:

$$R(\mathbf{x}) = ce^{-|\mathbf{x}|^2/(2\sigma^2)} , \quad (3)$$

which is separable, unlike the potential filter. The convolution operation thus becomes much less expensive.

The flow lines are traced inward in the gradient direction and outward in the negative gradient direction, sampling the columns at regular arc length intervals. This was done using a sampling interval of 0.5 mm. At some point the gradient flattens to the point where we can no longer trace the column due to numerical issues or due to the size of the convolution kernel, resulting in a column with a finite number of inner and outer column points.

2.3 Graph

We use an optimal graph construction technique similar to Ishikawa's [3], which has slightly fewer edges than the Wu and Chen method [10]. It should be noted that neither of these papers deal with columns of varying lengths, however this can easily be dealt with similar to how the *out of bounds edges* were handled by Ishikawa [3], see (7) and Fig. 2(b).

We have a column V_i traced from each mesh vertex $i \in V_B$, with the following set of points: $\{i_{-I_i}, i_{-(I_i-1)}, \dots, i_0, \dots, i_{O_i}\}$, where I_i and O_i correspond to the number of inner and outer column points relative to $i_0 = i$, and need to construct a maximum flow graph $G = (V, E)$, with vertices V and edges E , to find the set of surfaces M . We thus construct columns of vertices $V_i^m = \{i_k^m \mid i_k \in V_i\}$ in V , where $m \in M$ such that:

$$V = \bigcup_{i \in V_B, m \in M} V_i^m \cup \{s, t\} . \quad (4)$$

Here s and t denote the source and sink nodes respectively. In the case of airway wall segmentation, $M = \{0, 1\}$ would denote the fact that there is an inner and outer surface sub-graph.

Let $(v \xrightarrow{c} u)$ denote a directed edge from vertex v to vertex u with capacity c . We define a cost function $w(i_k^m) \geq 0$, mapping a node with index $k \in \{-I_i, 1 - I_i, \dots, 0, \dots, O_i\}$ in a column V_i to the inverse likelihood that it is part of the surface m , see Sect. 2.4. Such a data term can be implemented by the following edges:

$$E_{\text{data}} = \left\{ \left\{ (i_k^m \xrightarrow{w(i_k^m)} i_{k+1}^m) \mid i_k^m, i_{k+1}^m \in V_i^m \right\} \cup \left\{ (i_{O_i}^m \xrightarrow{w(i_{O_i}^m)} t), (s \xrightarrow{\infty} i_{I_i}^m) \right\} \mid i \in V_B, m \in M \right\}. \quad (5)$$

Since each column is a direct line of flow from the source to the sink, it must be cut at least once. However in some degenerate cases, multiple cuts might exist in each column. This is something to be avoided, as it could change the topology of the surface. A solution is to add infinite cost edges opposite to the data term edges:

$$E_{\infty} = \left\{ (i_k^m \xrightarrow{\infty} i_{k-1}^m) \mid i \in V_B, m \in M, i_{k-1}^m, i_k^m \in V_i^m \right\}. \quad (6)$$

An example of these edges is given in Fig. 2(a).

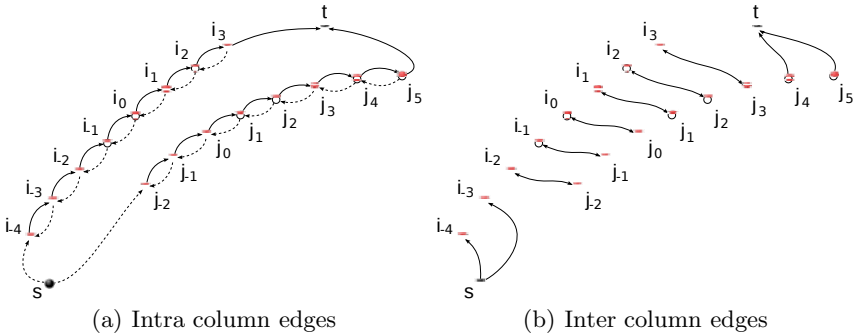


Fig. 2. Two example neighboring columns V_i and V_j , with $I_i = 4$ and $I_j = 2$ inner column points and $O_i = 3$ and $O_j = 5$ outer column points (note m superscript left out for clarity). The data term edges are given by the black arrows and the infinite cost edges are dotted in Fig. 2(a). The inter column edges generated from a linear edge penalty function in Fig. 2(b) in black. If the edge penalty is of the form: $f_{i,j}(x) = p^m x$, the (bi-)directional capacity of these inter column edges are then given by p^m . Notice that when combining the inter and intra column edges the vertex j_5 has two edges connecting it to the sink node. This is just for illustrative purposes, in practice the capacities of these are summed and implemented with a single edge. The inter column edge from the s to i_{-4} can be left out completely.

Let $f_{i,j,m,n}(|k-l|)$ define a convex non-decreasing function describing the inverse likelihood that both vertex $i_k \in V_i$ and $j_l \in V_j$ are part of the surfaces $m, n \in M$ respectively. Note m can be equal to n and i can be equal to j , but not simultaneously. This can be thought of as an edge penalty, since it attributes a penalty of including the 'edge' (i_k^m, j_l^n) in the solution and can be used to implement surface smoothness and separation penalties, see (9). Both Ishikawa (3) and Wu and Chen (10) show how such edge penalties can be implemented in the case where the columns have equal length. The following inter column edges extend the implementation to the case where both the number of inner and outer column points can vary:

$$E_{\text{inter}} = \left\{ \left\{ (i_k^m \xrightarrow{\Delta_{i,j,m,n}(k-l)} j_l^n) \mid i_k^m \in V_i^m, j_l^n \in V_j^n \right\} \cup \right. \\ \left. \left\{ (s \xrightarrow{\Delta_{i,j,m,n}(k-l)} j_l^n) \mid j_l^n \in V_j^n, k \in L(j, i) \right\} \cup \right. \\ \left. \left\{ (i_k^m \xrightarrow{\Delta_{i,j,m,n}(k-l)} t) \mid i_k^m \in V_i^m, l \in U(i, j) \right\} \right. \\ \left. \mid i, j \in V_B, m, n \in M \right\}, \quad (7)$$

where L and U describe the needed edge endpoints that are out of bounds:

$$L(i, j) = \{k \mid I_i > I_j, k \in \{-I_i, 1 - I_i, \dots, -I_j - 1\}\} \\ U(i, j) = \{k \mid O_i > O_j, k \in \{O_i, O_i - 1, \dots, O_j + 1\}\},$$

and Δ the capacity of the edges calculated from the edge penalty function:

$$\Delta_{i,j,m,n}(x) = \begin{cases} 0 & \text{if } x < 0 \\ f_{i,j,m,n}(1) - f_{i,j,m,n}(0) & \text{if } x = 0 \\ (f_{i,j,m,n}(x+1) - f_{i,j,m,n}(x)) - & \text{if } x > 0 \\ (f_{i,j,m,n}(x) - f_{i,j,m,n}(x-1)) & \end{cases}. \quad (8)$$

We used the following edge penalty functions in our experiments:

$$f_{i,j,m,n}(x) = \begin{cases} p^m x & \text{if } m = n \text{ and } (i, j) \in E_B \\ qx & \text{if } m \neq n \text{ and } i = j \\ 0 & \text{else} \end{cases}. \quad (9)$$

p^m is the smoothness penalty, defining the cost of each index the solution varies between neighboring columns in the same surface m . q is the separation penalty, defining the cost inherent in the solution for each index the surfaces are separated in each column. The advantage of using linear edge penalty functions is that far fewer edges are needed to implement the penalty as $\Delta_{i,j,m,n}(x) = 0$ for all $x \neq 0$. An illustration of these edges is given in Fig. 2(b). The total edge set E in the maximum flow graph is given by:

$$E = E_{\text{data}} \cup E_{\infty} \cup E_{\text{inter}}. \quad (10)$$

Note the edge penalties described in this Sect. are soft. It is possible to add hard constraints as well, by adding infinite cost edges between the columns. We refer to (4,10) for examples of this.

We used the algorithm described in (1) to find the minimum cut.

2.4 Cost Functions

In the case of airway walls, the columns will usually start inside the air-filled lumen area, which has low density, move through the airway wall where the density rises, and finally end up in the lung parenchyma where the density falls again. The CT intensity directly reflects this density change. A common way to find such boundaries is to use weightings of the first and second order derivatives of the intensity along the columns [4,5,9]. In this way the positions of the found surfaces can be adjusted relative to the 'real' surfaces. For the experiments described in this paper we used the cost functions of Petersen et al. [9], with inner and outer surface derivative weights given by $\gamma^0, \gamma^1 \in [0, 1]$. The derivatives are obtained using central differences from cubic interpolated values.

3 Experiments and Results

3.1 Data

The material used comes from the Danish lung cancer screening trial [8]. The images were obtained using a Multi Detector CT scanner with a low dose (120 kV and 40 mAs), reconstructed using a hard kernel with a resolution of approximately $0.78 \text{ mm} \times 0.78 \text{ mm} \times 1 \text{ mm}$.

For evaluation, we used manual annotations in 649 two-dimensional cross-sectional images with a resolution of $0.5 \text{ mm} \times 0.5 \text{ mm}$ extracted at random positions perpendicular to and centered on the airways in 15 scans. Some of these images contain more than one airway branch. Such other branches were marked and excluded from the analysis. The images were split randomly in a training and a test data set, consisting of 329 and 319 images from 8 and 7 subjects.

To evaluate suitability of the segmentations for deriving measures of airway morphology as a sign of Chronic Obstructive Pulmonary Disease (COPD), 480 subjects for whom repeated scans and spirometry within a two year period were available, were randomly selected to evaluate reproducibility of measures of airway morphology as well as their correlation with lung function.

3.2 Comparison to Existing Approaches

Experiments were conducted with a recently published two-dimensional method [9] and three three-dimensional methods based on different ways of constructing the columns. One was a method using straight columns $S_{k,\tau}$, as described in [5], the medial axes and normals were determined using the method of reference [2] using k neighbors and an error tolerance of τ respectively. We refer to the original article for a definition of these parameters. Normals with too large errors were replaced by a nearest neighbor normal. The other two methods used the proposed flow line columns calculated from the potential and Gaussian kernels, denoted P_α and G_σ respectively.

Let M_l , M_a and X denote the manually marked lumen, airway and excluded parts respectively. Similarly let A_l and A_a denote the parts segmented by the

algorithms. We then evaluate the correctness of a segmentation in the cross-sections using the relative area of overlap outside the excluded area, Φ as follows:

$$\Phi(M_l, M_a, A_l, A_a, X) = \frac{|(M_l \cap A_l)/X|}{|M_l/X| + |A_l/X|} + \frac{|(M_a \cap A_a)/X|}{|M_a/X| + |A_a/X|}. \quad (11)$$

Define the contour $\mathcal{C}(A)$ of an area A as the set of pixels belonging to A , where at least one of the pixels in the standard 4-neighborhood is not part of A . The average contour distance was then defined by:

$$\Psi(M_l, M_a, A_l, A_a, X) = \text{mean}_{x \in \mathcal{C}(A_l/X)} d(\mathcal{C}(M_l/X), x), \quad (12)$$

where $d(A, x)$ defines the minimum euclidean distance between the point x and the set A .

Parameter Tuning. The methods have inner and outer smoothness penalties, inner and outer cost function derivative weightings and separation penalties, denoted p^m, γ^m and q where $m \in M$ respectively. Optimal values for these parameters for each of the methods were obtained by searching the parameter space on the training data set using an iterative algorithm. In each iteration a parameter is searched by probing left and right search interval limits. If no improvement in the mean value of Φ is detected the intervals are halved around the current best guess and the process is repeated. This continues until a better guess is found or the difference between the left and right search intervals becomes less than some threshold. The search then proceeds with the next parameter in a loop with reset left and right search interval limits until no more parameters get updated. In order to avoid getting stuck in a local minima, the algorithm was repeated a number of times using random values as initial guesses for the parameters.

Since parameter optimization is time consuming, a small set of pilot experiments were performed on the training set to determine suitable values of k, τ, α and σ . The optimal value of τ was found to be about 2.5, meaning roughly 3% of the normals were discarded for having an error above the tolerance. The segmentation results improved with increasing k , flattening out at about 64. k is inversely related to the resolution of the medial axis and thus directly related to the length of the columns and number of self-intersections. The fact that the relative high value of 64 works best seems to indicate that the straight column method suffers from too short columns. The experiments with α indicated that the potential kernel did not require additional regularization, so we chose a value of $\alpha = 10^{-4}$ mm as this would have practically no regularizing effect, yet still allowed the filter to be well defined everywhere. A similar conclusion holds for σ , however as the Gaussian tends to zero much faster than the potential kernel, the choice of σ seems to be a trade-off between too short columns or too much regularization. A value of 0.45 mm provided the best results for our case.

Results. Running the proposed method on an image from our data usually takes less than 10 minutes (Intel Xeon 1.60 GHz using no parallelization). Figure [3](#)

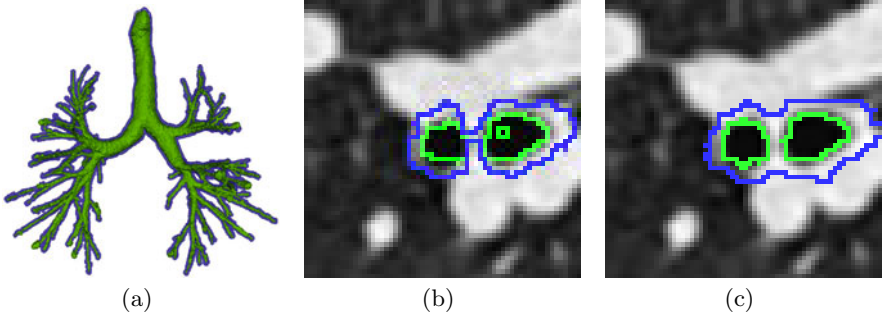


Fig. 3. Figure 3(a), 3(b) and 3(c) show the inner (green) and outer (blue) airway wall surface of a segmentation result in three dimensions obtained using $G_{\sigma=0.45}$ and two cross-sections near a bifurcation obtained with $S_{k=64,\tau=2.5}$ and $G_{\sigma=0.45}$ respectively. Notice how the outer surface cannot be correctly found using $S_{k=64,\tau=2.5}$ in the region with high curvature between the airway branches and there is even a hole in the segmentation, probably due to self-intersecting surfaces.

shows a segmentation result in three dimensions and cross-sections illustrating results of $G_{\sigma=0.45}$ and $S_{k=64,\tau=2.5}$.

Table 1 shows the results of the comparisons with the manual annotations in the test data set, for each of the investigated methods using the optimal parameters. $G_{\sigma=0.45}$ achieved the best result when measured with both metrics, whereas $S_{k=64,\tau=2.5}$ and $2D$ were the worst in terms of the relative area of overlap and average contour distance respectively. $G_{\sigma=0.45}$ was significantly better than any of the other three-dimensional methods and also significantly better than $2D$ using Ψ ($p < 0.05$) and shows the smallest variance in the quality of the results of all the methods ($p < 0.0001$). Results were compared using a paired-sample t -test and a two-sample F -test respectively.

Table 1. The results of different methods and kernels on the test data set. Mean \pm standard deviation of (11) and (12). The best result marked with a bold font.

	$2D$	$S_{k=64,\tau=2.5}$	$P_{\alpha=10^{-4}}$	$G_{\sigma=0.45}$
Φ	0.884 ± 0.079	0.865 ± 0.086	0.880 ± 0.079	0.890 ± 0.059
Ψ (mm)	0.115 ± 0.176	0.105 ± 0.132	0.113 ± 0.177	0.092 ± 0.102

3.3 Tree Extraction and Measurements

Airway centerlines and branch generations were extracted from the airway tree with a front propagation method, as described in [6]. Airway morphology was quantified using the Inner Volume (IV) and the Wall Volume Percentage ($WV\%$), which are three-dimensional extensions to commonly used measures of airway abnormality [9]. The segmentations were grouped in generations by

assigning each voxel to the generation of the nearest centerline point. Let W denote the area classified as belonging to a specific generation, then the measures were computed in individual generations as follows:

$$IV = |W \cap A_l|, \quad (13)$$

$$WV\% = 100 \times WV / (IV + WV)\%, \quad (14)$$

where $WV = |W \cap A_a / A_l|$. Branches of generation 1 to 8 were included in the measurements. The trachea, defined as generation 0, was excluded.

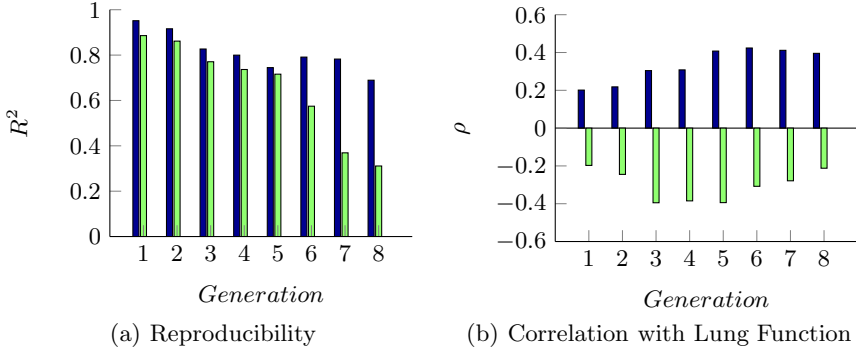


Fig. 4. Figure 4(a) shows the reproducibility of the measures, IV (blue) and $WV\%$ (green), in repeated scans quantified as R^2 in branch generations 1 to 8. Figure 4(b) shows statistically significant correlation with lung function in the same generations of the same measures.

The change in the measures during the roughly one year period between the repeated scans is assumed to be mostly due to measurement variability. This is a fair assumption given the relatively slow development of COPD and was backed up by paired-sample t -tests, showing no significant difference in the means of the measures in the same generation in repeated scans using a Bonferroni adjusted significance level of $p = 0.05$. It is thus possible to quantify the reproducibility of the measures, with the coefficient of determination, calculated from Pearson product moment correlation coefficients of the measures in repeated scans, see Fig. 4(a).

IV is the most reproducible measurement, which is to be expected given the high contrast ratio between the air-filled lumen and high density wall. The reproducibility of $WV\%$, which depend on the segmentation of the outer surface, show however that it is also determined with enough precision to allow reproducible measurements, up to roughly generation 6.

Figure 4(b) shows Spearman's correlation coefficients ρ of the measures and lung function measured by Forced Expiratory Volume in one second as a percentage of the predicted value (FEV1 (% predicted)). IV is positively correlated, indicating luminal narrowing with COPD, whereas $WV\%$ is negatively correlated, indicating wall thickening, which is in agreement with current knowledge of the disease process in COPD. These results indicate the method can be used to measure abnormalities caused by COPD up to at least generation 8.

4 Discussion and Conclusion

In this paper we performed extensive parameter tuning for all methods to allow for an as objective comparison as possible. However, it is our experience that results are not very sensitive to the settings of these parameters and suitable settings can be obtained by something as simple as a few trial and error runs using manual inspection of the segmentation results.

The results indicate the importance of choosing a good convolution kernel. For our specific application the Gaussian performed better than the potential, which we think is mainly due to it tending to zero much faster, limiting long range effects. More experiments will be needed to investigate which kernels are suitable to which applications.

The proposed segmentation method is not able to extend the initial segmentation beyond the lengths of the graph columns, and is thus not able to make up for larger errors, such as missing branches. This can explain a large part of the decline in reproducibility with generations, seen in Fig. 4(a). For instance investigations on the amount of segmented branches compared to the theoretical maximum, assuming a binary tree structure, revealed that almost all the branches were segmented in generations 5, whereas the number had dropped to about 50% in generation 6. Measurements conducted in corresponding branches, as opposed to generations, might thus still be reproducible after generation 6 and 8.

To conclude, a new graph construction technique applicable to multi-dimensional multi-surface segmentation problems was proposed. The method runs in polynomial time and is able to penalize for non-smoothness and separation of the found surfaces. The results are guaranteed to not self-intersect and are robust in regions with large curvature.

We applied the method to the problem of segmenting human airway walls in CT images and results were shown to be significantly more accurate than those of recently published two- and three-dimensional methods. Large scale evaluations on 480 images from a lung cancer screening trial, showed good reproducibility of the obtained airway abnormality measures and a significant correlation with lung function.

Acknowledgements. This work is partly funded by the Netherlands Organisation for Scientific Research (NWO), and AstraZeneca, Sweden.

References

1. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(9), 1124–1137 (2004)
2. Dey, T.K., Sun, J.: Normal and Feature Estimation from Noisy Point Clouds. In: *Proceedings of the 26th International Conference on Foundations of Software Technology and Theoretical Computer Science*, pp. 21–32 (2006)

3. Ishikawa, H.: Exact Optimization for Markov Random Fields with Convex Priors. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(10), 1333–1336 (2003)
4. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal Surface Segmentation in Volumetric Images - A Graph-Theoretic Approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(1), 119–134 (2006)
5. Liu, X., Chen, D.Z., Wu, X., Sonka, M.: Optimal Graph-Based Segmentation of 3D Pulmonary Airway and Vascular Trees Across Bifurcations. In: *First International Workshop on Pulmonary Image Analysis*, pp. 103–111 (2008)
6. Lo, P., van Ginneken, B., Reinhardt, J.M., de Bruijne, M.: Extraction of Airways from CT (EXACT 2009). In: *The Second International Workshop on Pulmonary Image Analysis*, pp. 175–189 (2009)
7. Lo, P., Sporning, J., Pedersen, J.J.H., de Bruijne, M.: Airway tree extraction with locally optimal paths. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 51–58. Springer, Heidelberg (2009)
8. Pedersen, J.H., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B.G., Døssing, M., Mortensen, J., Richter, K., Clementsen, P., Seersholm, N.: The danish randomized lung cancer CT screening trial—overall design and results of the prevalence round. *J. Thorac. Oncol.* 4(5), 608–614 (2009)
9. Petersen, J., Lo, P., Nielsen, M., Edula, G., Ashraf, H., Dirksen, A., Bruijne, M.d.: Quantitative Analysis of Airway Abnormalities in CT. In: Karssemeijer, N., Summers, R.M. (eds.) *Medical Imaging 2010: Computer-Aided Diagnosis*. Proceedings of SPIE, vol. 7624 (2010)
10. Wu, X., Chen, D.Z.: Optimal net surface problems with applications. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) *ICALP 2002*. LNCS, vol. 2380, pp. 1029–1042. Springer, Heidelberg (2002)
11. Yin, Y., Song, Q., Sonka, M.: Electric field theory motivated graph construction for optimal medical image segmentation. In: Torsello, A., Escolano, F., Brun, L. (eds.) *GbrPR 2009*. LNCS, vol. 5534, pp. 334–342. Springer, Heidelberg (2009)

Surface–Region Context in Optimal Multi-object Graph-Based Segmentation: Robust Delineation of Pulmonary Tumors^{*}

Qi Song¹, Mingqing Chen¹, Junjie Bai¹, Milan Sonka^{1,2,3}, and Xiaodong Wu^{1,2}

¹ Department of Electrical & Computer Engineering

² Department of Radiation Oncology

³ Department of Ophthalmology & Visual Sciences, University of Iowa,
Iowa City, IA 52242, USA

{qi-song,xiaodong-wu,milan-sonka}@uiowa.edu

Abstract. Multi-object segmentation with mutual interaction is a challenging task in medical image analysis. We report a novel solution to a segmentation problem, in which target objects of arbitrary shape mutually interact with terrain-like surfaces, which widely exists in the medical imaging field. The approach incorporates context information used during simultaneous segmentation of multiple objects. The object–surface interaction information is encoded by adding weighted inter-graph arcs to our graph model. A globally optimal solution is achieved by solving a *single* maximum flow problem in a low-order polynomial time. The performance of the method was evaluated in robust delineation of lung tumors in megavoltage cone-beam CT images in comparison with an expert-defined independent standard. The evaluation showed that our method generated highly accurate tumor segmentations. Compared with the conventional graph-cut method, our new approach provided significantly better results ($p < 0.001$). The Dice coefficient obtained by the conventional graph-cut approach (0.76 ± 0.10) was improved to 0.84 ± 0.05 when employing our new method for pulmonary tumor segmentation.

1 Introduction

Medical image segmentation allows analyzing medical image data in a quantitative manner, which plays a vital role in numerous biomedical applications. Recently, graph-based methods with a global energy optimization property have attracted considerable attentions (e.g., [12,74]). Despite their widespread use for image segmentation, they may have problem in the presence of multiple target objects with weak boundaries, similar intensity information and serious mutual interaction among each other [10].

To solve the problem of multi-object segmentation with mutual interaction, Boykov *et al.* [1] developed a multi-region framework based on a graph cut

^{*} This work was supported in part by NSF grants CCF-0830402 and CCF-0844765, and NIH grants R01-EB004640 and K25-CA123112.

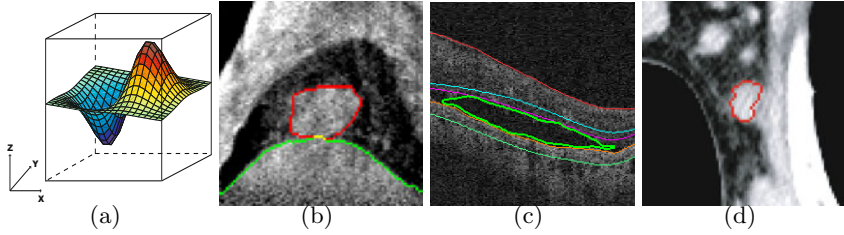


Fig. 1. (a) The terrain-like surface intersects each (x, y) -column exactly one time. (b) Example slices of a lung tumor (red) in megavoltage cone-beam CT. The tumor is attached/adjacent to the diaphragm (green). (c) Example slices of a fluid-filled region (green) in retinal optical coherence tomography (OCT). The fluid region is surrounded by intraretinal layers. (d) Example slices of a lymph node (red) in X-ray CT data. The lymph node is connected to surrounding soft tissues.

method. An interaction term is added into the energy function, which incorporated the geometric interactive relationship between different regions. Their method is topologically flexible and shares some elegance with the level set methods. Li *et al.* [4] proposed an approach called “graph search” for globally optimal segmentation of multiple *terrain-like surfaces* (Fig. 1(a)), which was successfully extended to a wide spectrum of applications [13, 10, 3, 2]. For each target surface a corresponding geometric sub-graph is constructed. The relative position information between pairs of surfaces is enforced by adding weighted arcs between corresponding sub-graphs. Multiple surfaces can be detected simultaneously in a globally optimal fashion by solving a single maximum flow problem in a transformed graph. Compared with the graph cut approach, this method requires no human interaction and provides easier way to incorporate shape prior information [9]. The major limitation of Li *et al.*’s family of the graph search method lies in the fact that it is non-trivial to deal with the region of complicated topology without an approximate segmentation of the region’s boundary surface.

In this paper, we consider a special structure consisting of mutually interacting terrain-like surfaces and regions of “arbitrary shape”. Here “arbitrary shape” means that the target region itself does not have any specific preferred shape. Mutual interaction exists between the pairs of the terrain-like surfaces and the regions in a way that their topological information and relative positions are usually known beforehand. Incorporating these interrelations into the segmentation plays a significant role for accurate segmentation when target surfaces and regions lack clear edge and have similar intensity distribution. Fig. 1(b), (c) and (d) show some typical medical imaging examples.

The main idea is to make use of the advantages from both the graph cut method [12] and Li *et al.*’s graph search method [4]. For regions of arbitrary shape we construct a corresponding sub-graph following the graph cut method making use of its shape flexibility. For terrain-like surfaces the sub-graph is built based on the graph search approach, which requires no initial seeds and provides good shape control. The key is to introduce an additional interaction term in the

energy function, which can be implemented by adding inter-graph arcs between the pairs of different types of graphs, enforcing known interacting information between the target terrain-like surfaces and the regions of arbitrary shape. Then a globally optimal solution can be obtained by solving a *single* maximum flow problem, yielding the simultaneous segmentation of surfaces and regions.

Our paper is organized as follows. In Section 2 we describe our proposed framework in details. In Sections 3–5 a quantitative and comparative performance evaluation over a representative number of experiments with Mega-Voltage Cone Beam CT (MVCBCT) images for lung tumor is performed. Section 6 presents concluding remarks.

2 Method

To simplify the presentation and to ease the understanding of our method, let us first consider the task of detecting one terrain-like surface and one region of arbitrary topology with mutual interaction between each other. Note that the same principles used for this illustration are directly applicable to multiple pairs of surfaces and regions with interactions between those two kinds of targets.

2.1 Energy Function

Consider a volumetric image $\mathcal{I}(X, Y, Z)$ of size $X \times Y \times Z$. For each (x, y) pair the voxel subset $\{v(x, y, z) | 0 \leq z < Z\}$ forms a column parallel to the z-axis, denoted by $p(x, y)$. Each column has a set of neighborhoods for a certain neighbor setting \mathcal{N}_c [4]. A *terrain-like surface* of particular interest, denoted as S_T , is the surface that intersects each column $p(x, y)$ at exactly one voxel. It can be defined as a function $S_T(x, y)$, mapping each (x, y) pair to its z-value. The target region of arbitrary shape, denoted as R_A , includes all the voxels v inside the region. Fig. 2(a) shows one example 2-D slice from a 3-D image.

To explain the employed cost function, let us start with the energy terms used for detecting the terrain-like surface, which are similar in form to those described in [9]. Suppose an edge-based cost c_v is assigned to each voxel $v \in I$, which is inversely related to the likelihood that the desired surface S_T indeed contains this voxel. For any pair of neighboring columns $(p, q) \in \mathcal{N}$ a convex function penalizing the surface shape change of S_T on p and q is expressed as $f_{p,q}(S_T(p) - S_T(q))$. Then the energy term E_{gs} takes the form

$$E_{gs}(S_T) = \sum_{v \in S_T} c_v + \sum_{(p,q) \in \mathcal{N}_c} f_{p,q}(S_T(p) - S_T(q)). \quad (1)$$

For segmentation of the target region R_A we employ the well-known binary graph cut energy [12, 11]. Let l denote the binary variables assigned for each voxel, indexed as l_v over voxels $v \in I$. In our notation, $l_v = 1$ means that v belongs to the target region R_A and $l_v = 0$ means that v belongs to the background. The graph cut energy E_{gc} is expressed as

$$E_{gc}(R_A) = \sum_{v \in I} D_v(l_v) + \sum_{(v_i, v_j) \in \mathcal{N}_v} V_{i,j}(l_i, l_j), \quad (2)$$

where D_v is the data term measuring how well the label l_v fits the voxel v given the image data. \mathcal{N}_v defines the neighboring relationship between voxels, and the boundary energy term $V_{i,j}(l_i, l_j)$ is the penalty of assigning the neighboring voxels (v_i, v_j) to labels l_i and l_j , respectively.

As mentioned above, incorporation of known interrelations between terrain-like surface S_T and object R_A plays a key role for accurate segmentation. To enforce a priori object–surface interaction information, we add an interaction term $E_{interaction}(S_T, R_A)$ to our energy function and the energy function takes the form

$$E(S) = E_{gs}(S_T) + E_{gc}(R_A) + E_{interaction}(S_T, R_A). \quad (3)$$

Our objective is to find the optimal set $S = \{S_T, R_A\}$ such that the above energy is minimized.

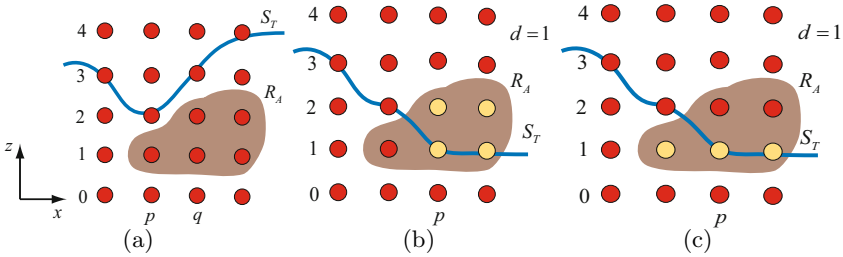


Fig. 2. An example 2-D slice from a 3-D image. (a) The terrain-like surface S_T and the region of arbitrary shape R_A , $S_T(p) = 2$, $S_T(q) = 3$. (b) Incorporation of the interaction constraint that the region R_A tends to be positioned “lower” than the terrain-like surface S_T . For voxels $v \in R_A$ and $S_T(p) - z(v) < d$ (yellow voxels), a penalty w_v is given; d is set as 1. (c) The interaction constraint that R_A tends to be positioned “higher” than S_T . For voxels $v \in R_A$ ($l_v = 1$) and $z(v) - S_T(p) < d$ (yellow voxels), penalty w_v is given; d is set as 1.

2.2 Incorporation of Geometric Interactions

In this section, we specify the geometric interactions incorporated in our framework between the target surface S_T and the region R_A , and show how to enforce as the interaction energy term $E_{interaction}(S_T, R_A)$. We start with the case that the region R_A tends to be lower than the terrain-like surface S_T with a given distance d . For any violating voxels in R_A , a penalty is given. More specifically, let $z(v)$ denote the z coordinate for voxel v . $S_T(p)$ is the z -value for the surface S_T on column p , representing the “height” of the surface on that column. Then for any voxel $v \in p$, if $v \in R_A$ and $S_T(p) - z(v) < d$, a penalty w_v is given (Fig. 2(b)). Our interaction energy term takes the form

$$E_{interaction}(S_T, R_A) = \sum_{p \in \mathcal{I}} \sum_{\substack{v \in p \\ S_T(p) - z(v) < d}} w_v l_v. \quad (4)$$

For the constraint that R_A is a priori expected to be positioned “higher” than the terrain-like surface S_T , a similar formulation is employed. For any voxel $v \in p$, if $v \in R_A(l_v = 1)$ and $z(v) - S_T(p) < d$, a penalty w_v is given (Fig. 2(c)). Our interaction energy term then takes the form

$$E_{interaction}(S_T, R_A) = \sum_{p \in \mathcal{I}} \sum_{\substack{v \in p \\ z(v) - S_T(p) < d}} w_v l_v. \quad (5)$$

2.3 Graph Construction

Our graph transformation scheme formulates the energy minimization problem as a single computation of maximum flow in the graph.

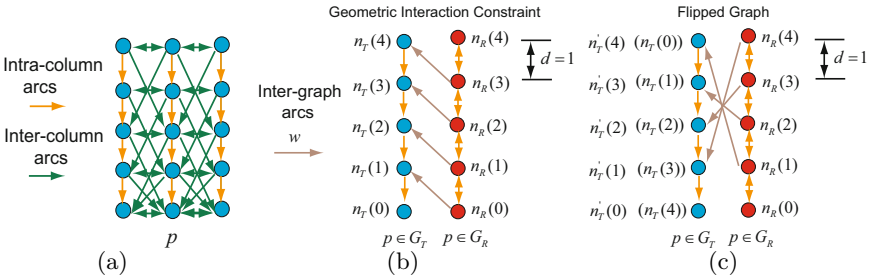


Fig. 3. (a) Graph construction of $G_T(N_T, A_T)$ for the graph search energy term $E_{gs}(S_T)$. Two types of arcs are introduced: intra-column arcs (yellow arrows) and inter-column arcs (green arrows). (b) Incorporation of geometric interaction constraints by adding inter-graph arcs (brown arrows). The region R_A tends to be lower than the terrain-like surface S_T with a distance of $d = 1$; $n_T(i)$ and $n_R(i)$ correspond to the same voxel in original image. (c) The flipped graph for constraints that R_A tends to be higher than the surface S_T with a distance of $d = 1$. $n'_T(i)$ and $n_R(Z-i-1)$ correspond to the same voxel in the original image.

For the graph cut energy term $E_{gc}(S_T)$ a sub-graph $G_R(N_R, A_R)$ is built using the method described in [12]. Every voxel $v \in \mathcal{I}$ has a corresponding node $n_R \in N_R$. Two additional nodes, the source (object) s and the sink (background) t , are added. Each node n_R has one t -link to each of the source and sink, which enforces the data-term energy. Each pair of neighboring nodes is connected by an n -link, which encodes the boundary energy term. The minimum-cost s - t cut divides the graph G_R into two parts: All nodes belonging to the target object are included in the source set and all background nodes are in the sink set.

For the graph search energy term $E_{gs}(S_T)$ a sub-graph $G_T(N_T, A_T)$ is constructed, which follows the method described in [11][10]. Every node $n_T(x, y, z) \in N_T$ corresponds to exactly one voxel $v(x, y, z) \in \mathcal{I}$. The positions of nodes reflect the positions of corresponding voxels in the image domain. Two types of arcs are added to the graph: (1) The intra-column arcs with $+\infty$ weight serve to enforce

the monotonicity of the target surface S_T ; and (2) the inter-column arcs incorporate the shape-prior penalties $f_{p,q}$ between the neighboring columns p and q . Each node is assigned a weight w_n such that the total weight of a closed set in the graph G_T equals to the edge-cost term in E_{gs} . A typical graph G_T is shown in Fig. 3(a). Then, as in [11,10], each node n_T is connected to either the source s by the arc with weight $-w_n$ if $w_n < 0$ or the sink t by the arc with weight w_n if $w_n > 0$. Note that the source s and the sink t are the same nodes used in G_R for the implementation of the graph cut energy term. Using this construction, we merge the two sub-graphs G_R and G_T as a single s - t graph G . The original energy minimization can be achieved by solving a maximum flow problem in the graph. The target surface S_T can be defined by the minimum-cost s - t cut in the graph. All nodes in G_T above surface S_T belong to the sink set and all nodes on or below S_T belong to the source set of the cut [11].

To incorporate geometric interaction constraints, additional inter-graph arcs are added between two sub-graphs. We begin with the case that region R_A tends to be lower than the terrain-like surface S_T with a given distance d . If $n_R(x_1, y_1, z_1)$ in the subgraph G_R belongs to the source set (labeled as “object”) and $n_T(x_1, y_1, z_1 + d)$ in the subgraph G_T belongs to the sink set, which indicates that $S_T(x_1, y_1) - z_1 < d$, a penalty w contributes to the objective energy function. That can be enforced by adding a directed arc with a weight of w from each node $n_R(x, y, z)$ to $n_T(x, y, z + d)$, as shown in Fig. 3(b).

To enforce the constraint that R_A tends to be higher than S_T with distance d , a “flip” operation is involved. A transformed graph $G'_T(N'_T, A'_T)$ is constructed, in which a node $n'_T(x, y, z)$ corresponds to a voxel $v(x, y, Z - z - 1)$ in the image \mathcal{I} . The interaction penalty is given by adding a directed arc with weight w from $n_R(x, y, z)$ to $n'_T(x, y, Z - z + d - 1)$, which is associated with the voxel $v(x, y, z - d)$ in original image. Fig. 3(c) shows the flipped graph.

Once the graph is constructed, a globally optimal solution can be found by solving a *single* maximum flow problem, which minimizes the total energy $E(S)$ in a low-order polynomial time.

3 Application for Lung Tumor Segmentation on MVCBCT Image

In this section, we exemplify the application of our method on lung tumor segmentation from MVCBCT images. MVCBCT is a promising technique used in clinic for daily imaging of patients [8] for lung tumor radiotherapy. Successful segmentation of lung tumors from the respiratory correlated 3-D images reconstructed from the projection data of MVCBCT scans can provide important information of tumor motion and volume changes, which allows better delineation of lung tumors for radiation therapy [6,5].

Here we mainly focus on the segmentation of the primary lung tumor from the reconstructed MVCBCT images, which is a very challenging work. First, the quality of the MVCBCT images is poor. Serious noise interference exists. Second, the lung tumor is frequently located next to the lung surface, the adjacent tissues

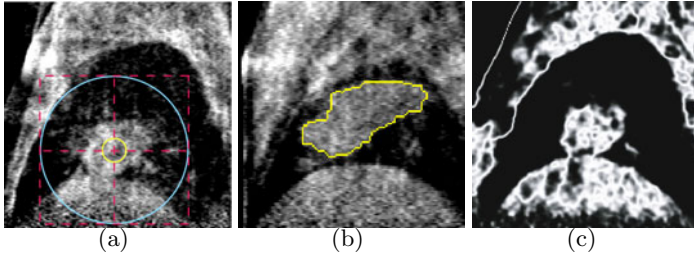


Fig. 4. (a) One typical example of the initialization step in sagittal view. The yellow sphere lies inside the tumor and the blue sphere completely contains the tumor. The cross point indicates the given center point. (b) The tumor (yellow) lies adjacent to both the diaphragm and the upper boundary of the lung. (c) The cost image for data term of the graph cut energy. The higher intensity value indicates the larger probability that the voxel belongs to the tumor.

have a similar intensity profile, and no clear boundary exists in-between. To overcome those difficulties, our new approach was employed for simultaneous segmentation of the lung primary tumor and the adjacent boundary.

3.1 Initialization

As the initialization step, one center point and two radii were required as the user input, from which two spheres were generated. The smaller one was required to be completely inside the tumor and the larger one was required to be completely outside the tumor. Fig. 4(a) shows a typical example. The segmentation was then conducted on the bounding box area of the larger sphere.

3.2 Simultaneous Segmentation of Lung Boundary and Tumor

The graph is constructed using the method described in Section 2.3, which contains two sub-graphs: G_T for lung boundary detection and G_R for tumor segmentation. To construct G_T , the user also need to interactively define the direction from which the tumor is adjacent to the lung boundary. In certain cases the tumor may be adjacent to the lung surface from two directions, as shown in Fig. 4(b). In this situation, we segment the tumor and two boundary surfaces together. Three sub-graphs are constructed accordingly: G_{T1} and G_{T2} for detection of lung boundaries from two directions; G_R for tumor segmentation. The geometric interactions are enforced between two pairs of graphs: (G_{T1}, G_R) and (G_{T2}, G_R) .

Cost design for G_T : For lung boundary detection a gradient-based cost function was employed for edge-based costs. The negative magnitude of the gradient of the image \mathcal{I} was computed at each voxel: $c(x, y, z) = -|\nabla\mathcal{I}|$. Note that the intensities inside the lung are generally lower than surrounding tissues, a Sobel kernel was used to favor a bright-to-dark or dark-to-bright transition depending

on the direction of the target surface. A second order shape prior penalty was employed with the form: $f(h) = \beta \cdot h^2$, where β was a constant parameter. The shape prior penalty penalized the change of the surface topology: $h = S_T(p) - S_T(q)$ between neighboring columns p and q .

Cost design for G_R : For tumor detection the data term D_v for voxel v was designed as follows: For all voxels lying inside of the smaller circle, which belong to the object, $D_v(l_v = 1) = 0$, $D_v(l_v = 0) = +\infty$. Similarly, for all voxels outside the larger circle $D_v(l_v = 1) = +\infty$, $D_v(l_v = 0) = 0$. For all other voxels the intensity distribution for tumor followed Gaussian distribution. The mean intensity value \bar{i} and the standard deviation value σ were obtained from all the voxels inside of the smaller circle. For voxel v with intensity i_v the data term was given as

$$D_v(l_v = 1) = -\log Pr(i_v | l_v = 1) \propto \frac{(i_v - \bar{i})^2}{\sigma^2}, \quad (6)$$

$$D_v(l_v = 0) = -\log(1 - Pr(i_v | l_v = 1)) \propto -\log(1 - \exp(-\frac{(i_v - \bar{i})^2}{\sigma^2})). \quad (7)$$

A typical cost image for data term is shown in Fig. 4(c).

For the boundary penalty a gradient-based cost function was used with the similar form as described in [12]

$$w_{ij} = -\log(1 - \exp(\frac{-|\nabla I|^2(i, j)}{2\sigma_g^2})), \quad (8)$$

where w_{ij} corresponds to the weight of the n-link arc between neighboring voxels v_i and v_j ; $|\nabla I|^2(i, j)$ denotes the squared gradient magnitude.

Cost design for interaction constraint: For the geometric interaction constraint we required that the minimum distance between the target tumor and the lung boundary be at least one. A hard constraint is enforced with $d = 1$ and $w_v = +\infty$.

4 Experimental Methods

The performance evaluation of the reported method was carried in 20 MVCBCT scans obtained from three patients with non-small cell lung cancer. Each set of patient MVCBCT scans was acquired longitudinally over eight weeks of radiation therapy treatment. For each scan two volumetric images were reconstructed, one for full-exhalation phase and one for full-inhalation phase, resulting in 40 images [5]. The size of the reconstructed images was $128 \times 128 \times 128$ voxels with cubic voxel sizes of $1.07 \times 1.07 \times 1.07 \text{ mm}^3$. Out of the 40 datasets, 2 were rejected for poor image quality by experts prior to any work reported here [5] and our experiments were conducted on the remaining 38 datasets.

Surfaces of lung tumors were obtained by expert manual tracing as reported in [5] and served as the independent standard for assessing segmentation correctness of our approach. The employed values of the above-described segmentation method parameters were selected empirically. The same parameter values of $\beta = 5$, $d = 1$, and $\sigma_g = 10$ were applied to all analyzed datasets.

The segmentation performance was assessed using the Dice similarity coefficient (DSC). $DSC = 2|V_m \cap V_c|/(|V_m| + |V_c|)$, where V_m denotes the volume of the independent standard and V_c denotes the volume of the computer-determined object. All DSC values were computed in 3-D.

To determine the performance of our novel surface-and-region segmentation approach in comparison with a conventional approach of solely using a graph cut method to detect the tumor without simultaneously segmenting associated lung boundary surfaces, these two approaches were applied to all 38 MVCBCT images with identical spherical initialization (Fig. 4) and the obtained DSC performance indices were compared for the two methods. Statistical significance of the observed differences was determined using Student *t*-test for which *p* value of 0.05 was considered significant.

5 Results

Our simultaneous surface–region segmentation method as well as Boykov’s conventional graph-cut segmentation method were applied to all 38 MVCBCT images for which the independent standard was available. Our approach achieved tumor segmentation correctness characterized by $DSC = 0.840 \pm 0.049$ while the conventional approach yielded statistically significantly lower segmentation

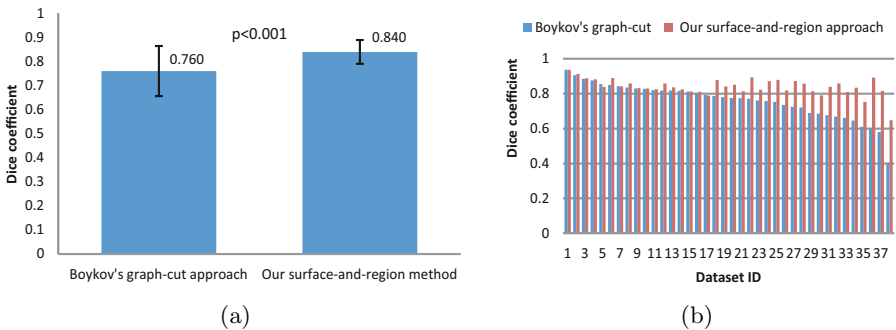


Fig. 5. Quantitative and comparative performance evaluation based on computed DSC in 38 volumetric MVCBCT images. (a) Overall tumor segmentation performance of Boykov’s graph cut approach and our new surface-and-region method. (b) Dice coefficients for all 38 datasets ordered according to the performance of the conventional graph-cut approach (blue). Note the uneven character of the segmentation performance for the conventional method compared to the highly-robust segmentation performance of the new approach.

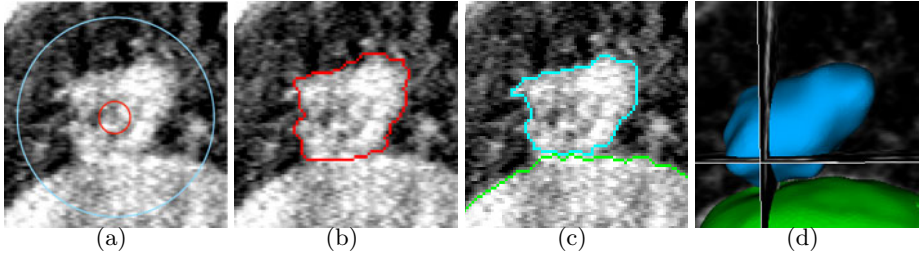


Fig. 6. Typical tumor segmentation examples. (a) one 2-D slice of 3-D MVCBCT image with outlines of spherical initialization. (b) Manual segmentation of the lung tumor – independent standard. (c) Simultaneous region-and-surface segmentation of the diaphragm (green) and the lung tumor (blue) using our new approach showing excellent segmentation performance – $DSC = 0.878$. (d) The 3-D representation of the diaphragm (green) and the tumor (blue).

performance of $DSC = 0.760 \pm 0.104$ ($p < 0.001$). Fig. 5(a) displays these overall results graphically as $Mean \pm stdev$.

Fig. 5(b) shows the pairwise performance comparisons for all 38 datasets, ordered according to the performance of the conventional approach. Fig. 5(b) thus clearly demonstrates that the performance of the conventional method is very uneven in the analyzed data set. In contrary, our new approach not only shows an overall improvement of the segmentation, it also demonstrates that the segmentation performance is very similar for all tumors in the entire set of the 38 analyzed images, thus showing a markedly higher robustness of the new approach resulting from the incorporation of image-based surface context.

Fig. 6 shows a typical outcome of our new tumor segmentation method and gives visual comparison with the independent standard. Fig. 7 shows tumor

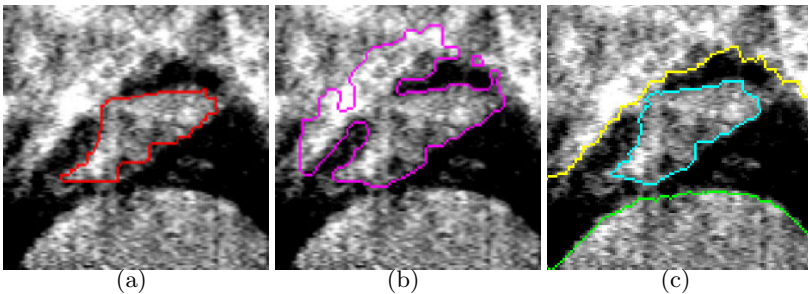


Fig. 7. Performance comparison in a difficult image. (a) Independent standard obtained by manual segmentation and shown in one 2-D slice of the 3-D volume. (b) Tumor segmentation failure resulting from the conventional graph cut method – $DSC = 0.689$. (c) Tumor segmentation obtained using our new method – segmented lung surfaces are shown in yellow and green, tumor surface is in blue – $DSC = 0.814$.

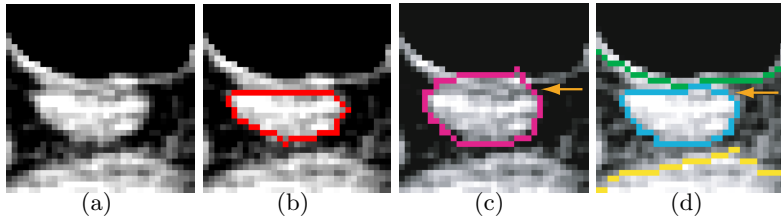


Fig. 8. Feasibility examples of lymph node segmentation in X-ray CT data. (a) One 2-D slice from 3-D CT image. (b) Manual segmentation of lymph node boundaries. (c) Segmentation failure resulting from the conventional graph-cut method. (d) Simultaneous region–surface segmentation of the lymph node (blue) and surrounding tissues (green and yellow) using our new approach showing excellent segmentation performance. The improvement is indicated by the arrow.

segmentation results obtained using the two compared methods in a difficult image, in which the tumor is closely adjacent to the lung surface from two directions. Fig. 7(b) shows a segmentation failure of the conventional approach while Figs. 7(c,d) show correctly segmented tumor using our new method.

Current non-optimized implementation requires about 5 minutes on a Linux workstation (3 GHz, 32 GB memory). The run time decreases to less than 10 s when working with downsampled images for which voxel sizes doubled along each of the x , y , and z directions with only a limited segmentation performance penalty ($DSC_2 = 0.826 \pm 0.046$).

6 Discussion and Conclusion

We presented a novel graph-based framework incorporating surface-region context information to solve the segmentation problem of a special structure having target objects of arbitrary topology mutually interacting with terrain-like surfaces, which widely exists in medical imaging fields. The proposed approach was successfully applied to lung tumor segmentation in MVCBCT. The result shows the power of our algorithm. To demonstrate the diverse applications of our framework, we also applied our method to lymph node segmentation in CT image. Fig. 8 shows an illustrative result.

References

1. Delong, A., Boykov, Y.: Globally optimal segmentation of multi-region objects. In: ICCV (2009)
2. Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., Wu, X.: Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 245–270. Springer, Heidelberg (2011), (this volume)

3. Han, D., Wu, X., Sonka, M.: Optimal multiple surfaces searching for video/image resizing - a graph-theoretic approach. In: ICCV (2009)
4. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images - a graph-theoretic approach. *IEEE TPAMI* 28(1), 119–134 (2006)
5. Chen, M., Siochi, R.A.: A clinical feasibility study on respiratory sorted megavoltage cone beam CT. In: International Workshop on Pulmonary Image Analysis (2010)
6. Chen, M., Siochi, R.A.: Diaphragm motion quantification in megavoltage cone-beam CT projection images. *Medical Physics* 37, 2312–2320 (2010)
7. Schaap, M., Neefjes, L., Metz, C., van der Giessen, A., Weustink, A., Mollet, N., Wentzel, J., van Walsum, T., Niessen, W.: Coronary lumen segmentation using graph cuts and robust kernel regression. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 528–539. Springer, Heidelberg (2009)
8. Morin, O., Gillis, A., Chen, J., Aubin, M., Bucci, M., Roach, M., Pouliot, J.: Megavoltage cone-beam CT: system description and clinical applications. *Medical Dosimetry* 31(1), 51–61 (2006)
9. Song, Q., Wu, X., Liu, Y., Haeker, M., Sonka, M.: Simultaneous searching of globally optimal interacting surfaces with shape priors. In: CVPR (2010)
10. Song, Q., Liu, Y., Liu, Y., Saha, P.K., Sonka, M., Wu, X.: Graph search with appearance and shape information for 3-D prostate and bladder segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 172–180. Springer, Heidelberg (2010)
11. Wu, X., Chen, D.Z.: Optimal net surface problems with applications. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) ICALP 2002. LNCS, vol. 2380, pp. 1029–1042. Springer, Heidelberg (2002)
12. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient N-D image segmentation. *IJCV* 70(2), 109–131 (2006)
13. Yin, Y., Zhang, X., Williams, R., Wu, X., Anderson, D., Sonka, M.: LOGISMOS - Layered Optimal Graph Image Segmentation of Multiple Objects and Surfaces: Cartilage segmentation in the knee joints. *IEEE Trans. Medical Imaging* 29(12), 2023 (2010)

Optimal Weights for Multi-atlas Label Fusion

Hongzhi Wang, Jung Wook Suh, John Pluta,
Murat Altinay, and Paul Yushkevich*

PICSL, Department of Radiology, University of Pennsylvania

Abstract. Multi-atlas based segmentation has been applied widely in medical image analysis. For label fusion, previous studies show that image similarity-based local weighting techniques produce the most accurate results. However, these methods ignore the correlations between results produced by different atlases. Furthermore, they rely on pre-selected weighting models and ad hoc methods to choose model parameters. We propose a novel label fusion method to address these limitations. Our formulation directly aims at reducing the expectation of the combined error and can be efficiently solved in a closed form. In our hippocampus segmentation experiment, our method significantly outperforms similarity-based local weighting. Using 20 atlases, we produce results with 0.898 ± 0.019 Dice overlap to manual labelings for controls.

1 Introduction

Atlas-based segmentation is motivated by the observation that segmentation strongly correlates with image appearance. A target image can be segmented by referring to labeled images that have similar image appearance. After warping an atlas's reference image to the target image via deformable registration, one can directly transfer labels from the atlas to the target image. As an extension, multi-atlas based segmentation makes use of more than one atlas to compensate for potential errors imposed by using any single atlas.

Errors produced by atlas-based segmentation can be attributed to dissimilarity in anatomy and/or appearance between the atlas and the target image. Recent research has been focusing on addressing this problem. For instance, research has been done on optimally constructing a single atlas from training data that is the most representative of a population [8]. Constructing multiple representative atlases from training data has been considered as well, and usually works better than single-atlas based approaches. Multi-atlas construction can be done either by constructing one representative atlas for each mode obtained from clustering training images [1] or by simply selecting the most relevant atlases for the target image on the fly [13]. Either way, one needs to combine the segmentation results obtained by referring to different atlases to produce the

* The authors thank the anonymous reviewers for their critical and constructive comments. This work was supported by the Penn-Pfizer Alliance grant 10295 (PY) and the NIH grants K25 AG027785 (PY) and R01 AG037376 (PY).

final solution. In this regard, image similarity-based local weighting has been shown to be the most accurate label fusion strategy [2114].

For label fusion, similarity-based local weighting techniques assign higher weights to atlases that have more similar appearance to the target image. These methods require a pre-selected weighting model to transfer local appearance similarity into non-negative weights. The optimal parameter of the weighting model usually needs to be determined in an ad hoc fashion through experimental validation. More important, the correlations between the results produced by different atlases are completely ignored. As a result, these methods cannot produce optimal solutions when the results are correlated, e.g. instead of producing random errors, different atlases tend to select the same wrong label.

In this paper, we propose a novel label fusion approach to automatically determine the optimal weights. Our key idea is that to minimize errors in the combined result, assigning weights to atlases should explicitly consider the correlations between results produced by different atlases with respect to the target image. Under this formulation, the optimal label fusion weights can be efficiently computed from the covariance matrix in a closed form. To estimate the correlations between atlases, we follow the basic assumption behind atlas-based segmentation and estimate label correlations from local appearance correlations between the atlases. We apply our method to segment the hippocampus from MRI and show significant improvements over similarity-based label fusion with local weighting.

2 Label Fusion Based Multi-atlas Segmentation

In this section, we briefly review previous label fusion methods. Let T_F be a target image and $A^1 = (A_F^1, A_S^1), \dots, A^n = (A_F^n, A_S^n)$ be n registered atlases. A_F^i and A_S^i denote the i th warped atlas image and the corresponding warped manual segmentation. Each A_S^i is a candidate segmentation for the target image. Label fusion is the process combining these candidate segmentations to produce the final segmentation. For example, the majority voting method [6,9] simply counts the votes for each label from all registered atlases and chooses the label receiving the most votes. The final segmentation \hat{T}_S is produced by:

$$\hat{T}_S(x) = \operatorname{argmax}_{l \in \{1 \dots L\}} p_x(l) \quad (1)$$

where l indexes through labels and L is the number of labels. x indexes through image voxels. $p_x(l)$ is the votes for label l at x , given by:

$$p_x(l) = \sum_{i=1}^n \frac{1}{n} p(l|A^i, x) \quad (2)$$

where $p(l|A^i, x)$ is the posterior probability that A^i votes for label l at x , with $\sum_{l \in \{1, \dots, L\}} p(l|A^i, x) = 1$. Typically, for deterministic atlases that have one unique label for every location, $p(l|A^i, x)$ is 1 if $l = A_S^i(x)$ and 0 otherwise.

Continuous label posterior probabilities can be used as well especially when probabilistic atlases are involved. Even for deterministic atlases, continuous label posterior probabilities still can be derived, see [14] for some examples.

Majority voting makes a strong assumption that different atlases produce equally accurate segmentations for the target image. Since atlas-based segmentation uses example-based knowledge representations, the segmentation accuracy produced by an atlas depends on the appearance similarity between the warped atlas image and the target image. To improve label fusion accuracy, recent work focuses on developing segmentation quality estimations based on local appearance similarity. For instance, the votes received by label l can be estimated by:

$$p_x(l) = \sum_{i=1}^n w^i(x) p(l|A^i, x) \quad (3)$$

$w^i(x)$ is a local weight assigned to the i_{th} atlas, with $\sum_{i=1}^n w^i(x) = 1$. The weights are determined based on the quality of segmentation produced by each atlas such that more accurate segmentations play more important roles in the final decision. One way to estimate the weight is based on local image similarity under the assumption that images with similar appearance are more likely to have similar segmentations. When the summed squared distance (SSD) and a Gaussian weighting model are used [14], the weights can be estimated by:

$$w^i(x) = \frac{1}{Z(x)} \exp\left(-\sum_{y \in \mathcal{N}(x)} [A_F^i(y) - T_F(y)]^2 / \sigma\right) \quad (4)$$

where $\mathcal{N}(x)$ defines a neighborhood around x and $Z(x) = \sum_{i=1}^n w^i(x)$ is a normalization constant. In our experiment, we use a $(2r + 1) \times (2r + 1) \times (2r + 1)$ cube-shaped neighborhood specified by the radius r . Since segmentation quality usually is nonuniform over the entire image, the estimation is applied based on local appearance dissimilarity. The inverse distance weighting has been applied as well [2]:

$$w^i(x) = \frac{1}{Z(x)} \left[\sum_{y \in \mathcal{N}(x)} (A_F^i(y) - T_F(y))^2 \right]^{-\beta} \quad (5)$$

where σ and β are model parameters controlling the weight distribution. Experimental validations usually are required to choose the optimal parameters. Furthermore, the correlations between atlases are not considered in the weight estimation. Next, we introduce a method that does not have these limitations.

3 Estimating Optimal Weights through Covariance Matrix

The vote produced by any single atlas for a label l can be modeled as the true label distribution, $p(l|T_F, x)$, plus some random errors, i.e.:

$$p(l|A^i, x) = p(l|T_F, x) + \epsilon(A^i, x) \quad (6)$$

Averaging over all segmentations produced by the same error distribution, the error produced by A^i at x can be quantified by:

$$E \left[(p(l|A^i, x) - p(l|T_F, x))^2 \right] = E [\epsilon(A^i, x)^2] \quad (7)$$

After combining results from multiple atlases, the error can be quantified by:

$$E \left[(p_x(l) - p(l|T_F, x))^2 \right] = \sum_{i=1}^n \sum_{j=1}^n w^i(x) w^j(x) M_x(i, j) \quad (8)$$

where $p_x(l)$ is given by (3) and M_x is the covariance matrix with:

$$M_x(i, j) = E \left[(p(l|A^i, x) - p(l|T_F, x)) (p(l|A^j, x) - p(l|T_F, x)) \right] \quad (9)$$

$M_x(i, i)$ quantifies the errors produced by i_{th} atlas and $M_x(i, j)$ estimates the correlation between two atlases w.r.t. the target image when $i \neq j$. Positive correlations indicate that the corresponding atlases tend to make similar errors, e.g., they tend to vote for the same wrong label. Negative correlations indicate that the corresponding atlases tend to make opposite errors. To facilitate our analysis, we rewrite (8) in matrix format as follows:

$$E \left[(p_x(l) - p(l|T_F, x))^2 \right] = W_x^t M_x W_x \quad (10)$$

where $W_x = [w^1(x); \dots; w^n(x)]$ and t stands for transpose. For optimal label fusion, the weights should be selected s.t. the combined error is minimized, i.e.,

$$W_x^* = \underset{W_x}{\operatorname{argmin}} W_x^t M_x W_x \quad \text{subject to} \quad \sum_{i=1}^n W_x(i) = 1 \quad (11)$$

The optimal weights can be solved via applying Lagrange multipliers, provided the covariance matrix M_x . The solution is:

$$W_x = \frac{M_x^{-1} \mathbf{1}_n}{\mathbf{1}_n^t M_x^{-1} \mathbf{1}_n} \quad (12)$$

where $\mathbf{1}_n = [1; 1; \dots; 1]$ is a vector of size n . When M_x is not full rank, the weights can be reliably estimated using quadratic programming optimization [11].

In fact, previous segmentation quality-based local weighting approaches can be derived from (12) by ignoring the correlations between atlases, i.e., setting $M_x(i, j) = 0$ for $i \neq j$. The main difference is that the weights computed by our method can be either positive or negative, while the weights used by segmentation quality-based weighting are non-negative. When the segmentations produced by different atlases are positively correlated, applying negative weights to some of the atlases allows to cancel out the common errors shared by these negatively weighted atlases and other positively weighted atlases, which may result in smaller combined errors (see (8)).

3.1 Estimating Correlations between Atlases

Since the true label distribution $p(l|T_F, x)$ is unknown, we seek approximations to estimate the correlations between atlases w.r.t. the target image. To simplify the estimation problem, we consider binary label posterior probabilities for the target image, i.e. $p(l|T_F, x) \in \{0, 1\}$. Under this constraint, correlations between atlases are non-negative because:

$$[p(l|A^i, x) - p(l|T_F, x)] [p(l|A^j, x) - p(l|T_F, x)] \geq 0 \quad \text{for } 1 \leq i, j \leq n \quad (13)$$

Following previous image similarity-based local weighting methods, we estimate label errors by local image dissimilarities as follows:

$$|p(l|A^i, x) - p(l|T_F, x)| \sim |A_F^i(x) - T_F(x)| \quad (14)$$

For continuous label posterior probabilities $p(l|A^i, x) \in [0, 1]$, $|A_F^i(x) - T_F(x)|$ estimates the expected error produced by the atlas. When binary label posterior probabilities are used for atlases, i.e. $p(l|A^i, x) \in \{0, 1\}$, $|A_F^i(x) - T_F(x)|$ corresponds the probability that the atlas makes an error. Note that we use a linear function to model the relationship between segmentation labels and image intensities. Typically, the real appearance-label relationship is more complicated than linear correlations. However, as we show below, using such a simple linear model already produces excellent label fusion results.

To further simplify the estimation, we assume that each atlas produces segmentations for the target image independently. Hence,

$$M_x(i, j) = |p(l|A^i, x) - p(l|T_F, x)| |p(l|A^j, x) - p(l|T_F, x)| \quad (15)$$

$$\sim |A_F^i(x) - T_F(x)| |A_F^j(x) - T_F(x)| \quad (16)$$

Since the segmentation quality produced by an atlas usually varies smoothly over spatial locations, for more robustness, we estimate the correlation at x using samples from its neighborhood by:

$$M_x(i, j) \sim \sum_{y \in \mathcal{N}(x)} |A_F^i(y) - T_F(y)| |A_F^j(y) - T_F(y)| \quad (17)$$

Note that when $i = j$, the label error produced by any single atlas is estimated by the commonly used summed squared distance over local image intensities as $E [(p(l|A^i, x) - p(l|T_F, x))^2] \sim \sum_{y \in \mathcal{N}(x)} |A_F^i(y) - T_F(y)|^2$.

The image similarity-based covariance estimation captures the atlas correlations produced by the actual registrations. However, local image similarity is not always a reliable estimator for registration errors, therefore may not always be reliable for estimating error correlations. To address this problem, one may incorporate empirical covariances estimated from training data to complement the similarity-based estimation. To make the comparison with previous image similarity-based label fusion techniques more straightforward, we focus on using the image similarity-based covariance estimation in this paper and leave the exploration of empirical covariance estimation to future work.

3.2 Toy Examples

In this section, we demonstrate the usage of our method with two toy examples. In the first example, suppose that for a target image, atlases A^1 and A^2 produce segmentations with similar qualities at location x , but their results are uncorrelated, with the covariance matrix $M_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The optimal weights computed by (12) are $W_x^* = [0.5; 0.5]$, which are the same as the solution produced by segmentation quality-based weighting. Now suppose that another atlas A^3 , which is identical to A^1 , is added to the atlas library. Obviously, A^1 and A^3 are strongly correlated because they produce identical label errors. Ignoring such correlations, quality-based weighting assigns equal weights to each of the three atlases. Hence, the final result is biased by the atlas that is repeatedly applied. However, given $M_x(1,3)=1$ and $M_x(2,3)=0$, our method assigns the weights, $W_x^* = [0.25; 0.5; 0.25]$. The bias caused by using A^1 twice is corrected.

Fig. 1 shows applying our method to another toy example. In this binary segmentation problem, we have five atlases. The pairwise correlations between the results produced by the atlases are all positive. Note that A^1 and A^4 have the largest combined inter-atlas correlations, indicating that they contain the most

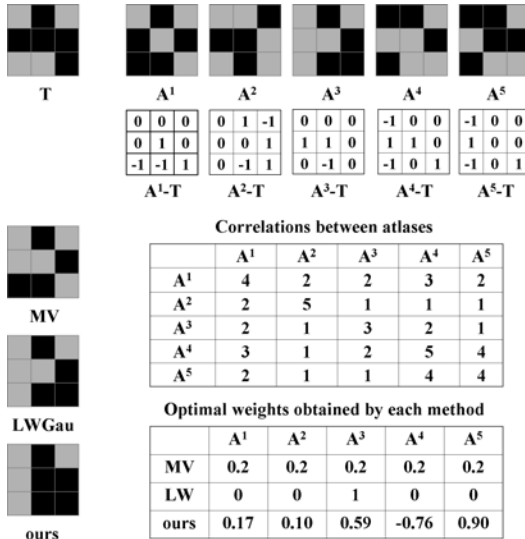


Fig. 1. Illustration of label fusion on a toy example. The target segmentation, T , is shown on the top left corner, followed by five warped atlases, A^1 to A^5 . For simplicity, each atlas produces binary votes and we assume that the images have the same appearance patterns as the segmentations. The voting errors produced by each atlas are shown in the matrix underneath it. For similarity-based label fusion with Gaussian weighting (LWGau) and our method, the atlas weights computed for the center pixel are also used for other pixels in this example.

common errors produced by all atlases. As a result, majority voting produces a result biased towards atlas A^1 and A^4 . Due to the strong correlations, similarity-based label fusion with Gaussian weighting (LWGau) reduces to single-atlas segmentation, i.e. only the most similar atlas, A^3 , has a non-zero weight. To compensate the overall bias towards A^4 , only A^4 receives a negative weight by our method to cancel out the consistent errors among all atlases.

3.3 Remediating Registration Errors by Local Searching

As recently shown in [5], the performance of atlas-based segmentation can be moderately improved by applying a local searching technique. This method also uses local image similarity as an indicator of registration accuracy and remedies registration errors by searching for the correspondence, that gives the most similar appearance matching, within a small neighborhood around the registered correspondence in each warped atlas.

Note that the goal of image-based registration is to correspond the most similar image patches between the registered images. However, the correspondence obtained from registration may not give the maximal similarity between all corresponding regions. For instance, deformable image registration algorithms usually need to balance the image matching constraint and the regularization prior on deformation fields. A global regularization constraint on the deformation fields is necessary to clarify the ambiguous appearance-label relationship arising from employing small image patches for matching. However, enforcing a global regularization constraint on the deformation fields may compromise the local image matching constraint. In such cases, the correspondence that maximizes the appearance similarity between the warped atlas and the target image may be within a small neighborhood of the registered correspondence.

Motivated by this observation, instead of using the original registered correspondence, we remedy registration errors by searching for the correspondence, that gives the most similar appearance matching, within a small neighborhood centered around the registered correspondence in each atlas. The locally searched optimal correspondence is:

$$x^i = \operatorname{argmin}_{x' \in \mathcal{N}'(x)} [A_F^i(\mathcal{N}(x')) - T_F(\mathcal{N}(x))]^2 \quad (18)$$

x^i is the location from i_{th} atlas with the best image matching for location x in the target image within the local neighborhood $\mathcal{N}'(x)$. Again, we use a cubic neighborhood definition, specified by a radius r_s . Note that \mathcal{N}' and \mathcal{N} may represent different neighborhoods and they are the only free parameters in our method. Instead of the registered corresponding patch $A^i(\mathcal{N}(x))$, we apply the searched patch $A^i(\mathcal{N}(x^i))$ to produce the fused label at x for the target image, i.e. (3) becomes $p_x(l) = \sum_{i=1}^n w^i(x^i)p(l|A^i, x^i)$.

To search for the most similar image patches, larger searching windows are more desirable. However, using larger searching windows more severely compromises the regularization constraint on the deformation fields, which complicates the appearance-label relationship on local patches. As a result, the linear

appearance-label function (14) becomes less accurate. It is reasonable to expect an optimal searching range that balances these two factors.

4 Experiments

In this section, we apply our method to segment the hippocampus using T1-weighted magnetic resonance imaging (MRI). The hippocampus plays an important role in memory function. Macroscopic changes in brain anatomy, detected and quantified by MRI, consistently have been shown to be highly predictive of AD pathology and highly sensitive to AD progression [15]. Accordingly, automatic hippocampus segmentation from MRI has been widely studied.

We use the data in the Alzheimer’s Disease Neuroimaging Initiative (ADNI, www.loni.ucla.edu/ADNI). Our study is conducted using 3 T MRI and only includes data from mild cognitive impairment (MCI) patients and controls. Overall, the data set contains 139 images (57 controls and 82 MCI patients). The images are acquired sagittally, with 1×1 mm in-plane resolution and 1.2 mm slice thickness. To obtain reference segmentation, we first apply a landmark-guided atlas-based segmentation method [12] to produce the initial segmentation for each image. Each fully-labeled hippocampus is manually edited by a trained human rater following a previously validated protocol [7].

For cross-validation evaluation, we randomly select 20 images to be the atlases and another 20 images for testing. Image guided registration is performed by the Symmetric Normalization (SyN) algorithm implemented by ANTS [3] between each pair of the atlas reference image and the test image. The cross-validation experiment is repeated 10 times. In each cross-validation experiment, a different set of atlases and testing images are randomly selected from the ADNI dataset. The results reported are averaged over the 10 experiments.

We focus on comparing with similarity-based local weighting methods, which are shown to be the most accurate label fusion methods in recent experimental studies, e.g. [2,14]. We use majority voting (MV) and the STAPLE algorithm [16] to define the baseline performance. For each method, we use binary label posteriors obtained from the deterministic atlases. For similarity-based label fusion, we apply Gaussian weighting (4) (LWGau) and inverse distance weighting (5) (LWInv).

Our method has two parameters, r for the local appearance window used in similarity-based covariance estimation, r_s for the local searching window used in remedying registration errors. For each cross-validation experiment, the parameters are optimized by evaluating a range of values ($r \in \{1, 2, 3\}$; $r_s \in \{0, 1, 2, 3\}$) using the atlases in a leave-one-out cross-validation strategy. We measure the average overlap between the automatic segmentation of each atlas obtained via the remaining atlases and the reference segmentation of that atlas, and find the optimal parameters that maximize this average overlap. Similarly, The optimal local searching window and local appearance window are determined for LWGau and LWInv as well. In addition, the optimal model parameters are also determined for LWGau and LWInv, with the searching range $\sigma \in [0.05, 0.1, \dots, 1]$ and $\beta \in [0.5, 1, \dots, 10]$, respectively.

For robust image matching, instead of using the raw image intensities, we normalize the intensity vector obtained from each local image intensity patch s.t. the normalized vector has zero mean and unit variance. To reduce the noise effect, we spatially smooth the weights computed by each method for each atlas. We use mean filter smoothing with the smoothing window \mathcal{N} , the same neighborhood used for local appearance patches.

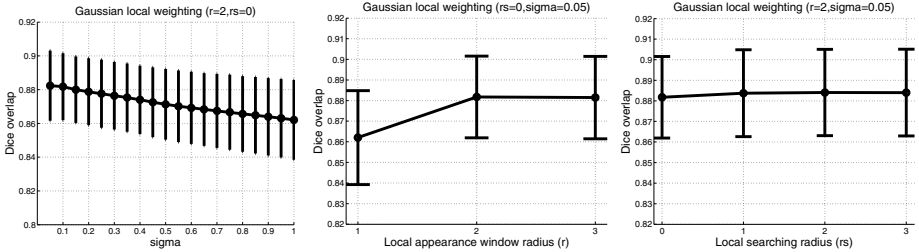


Fig. 2. Visualizing some of the parameter selection experiments for LWGau using leave-one-out on the atlases for the first cross-validation experiment. The figures show the performance of LWGau with respect to the Gaussian weighting function (left), local appearance window (middle) and local searching window (right), respectively when the other two parameters are fixed (the fixed parameters are shown in the figure’s title).

Fig. 2 shows some parameter selection experiments for LWGau in the first cross-validation experiment. The results are quantified in terms of Dice overlaps between automatic and manual segmentations of the atlases. The Dice overlap between two regions, A and B , measures the volume consistency as $\frac{2|A \cap B|}{|A| + |B|}$. For this cross-validation experiment, the selected parameters for LWGau are $\sigma = 0.05$, $r=2$, $r_s = 2$. Note that local searching only slightly improves the performance for LWGau. Similar results are observed for LWInv as well.

For our method, when the covariance matrix M_x is not full rank, we use the quadratic programming optimization tool *quadprog* in MATLAB (version R2009b) to estimate the weights. Fig. 3 shows the performance of our method when applied on the atlases in a leave-one-out fashion in the first cross-validation experiment. Without using local searching, our method already outperforms LWGau as shown in Fig. 2. Comparing to LWGau, local searching yields more improvement for our method. Overall, our method produces $\sim 1\%$ Dice improvement over LWGau and LWInv in this cross-validation experiment.

Using the appearance window with $r = 1$, our method performs significantly worse than using larger appearance windows. This indicates that the estimated error covariance using too small appearance windows are not reliable enough. When small appearance window with $r = 1$ is applied, our method performs comparably to the competing methods, but when larger appearance windows are used, our method significantly outperforms the competing methods. Note that applying larger appearance windows yields smoother local appearance similarity variations, therefore results in smoother local weights for label fusion. When

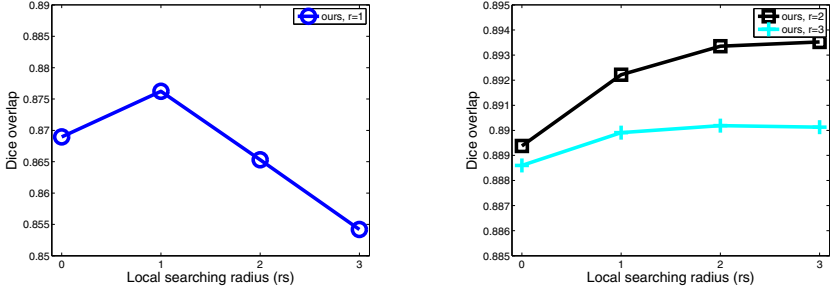


Fig. 3. Leave-one-out performance by our method on the atlases for the first cross-validation experiment when different appearance and searching windows are used. Since the results produced using appearance window with $r = 1$ are significantly worse than using larger appearance windows, for better visualization, we separately plot the results using $r = 1$ on the left. Note that the best results produced by our method is about 1% better than those produced by LWGau in Fig. 2.

Table 1. The performance in terms of Dice overlap produced by each method

method	left	right
MV	0.836 ± 0.084	0.829 ± 0.069
STAPLE	0.846 ± 0.086	0.841 ± 0.086
LWGau	0.886 ± 0.027	0.875 ± 0.030
LWInv	0.885 ± 0.027	0.873 ± 0.030
ours	0.894 ± 0.024	0.885 ± 0.026

large appearance windows with $r > 2$ are applied, the performance drops as r increases for all three methods. Hence, over-smoothing the local weights reduces the label fusion accuracy.

In terms of average number mislabeled voxels, LWGau and LWInv produce 369 and 372 mislabeled voxels for each hippocampus, respectively. By contrast, our method produces 352 mislabeled voxels. Table 1 shows the results in terms of Dice overlap produced by each method. Overall, LWGau and LWInv produce similar results. Both significantly outperform majority voting and the STAPLE algorithm. Our method outperforms similarity-based local weighting approaches by $\sim 1\%$ of Dice overlap. Since the average intra-rater segmentation overlap is 0.90, our method reduces the performance gap between MALF segmentation and intra-rater segmentation from $\sim 2\%$ Dice overlap to $\sim 1\%$ Dice overlap, a 50% improvement. Our improvement is statistically significant, with $p < 0.00001$ on the paired Student’s t-test for each cross-validation experiment. Fig. 4 shows some results produced by LWGau and our method.

Comparing to the state of the art in hippocampus segmentation. As pointed out in 4, direct comparisons of quantitative segmentation results across publications are difficult and not always fair due to the inconsistency in the underlying segmentation protocol, the imaging protocol, and the patient population. However,

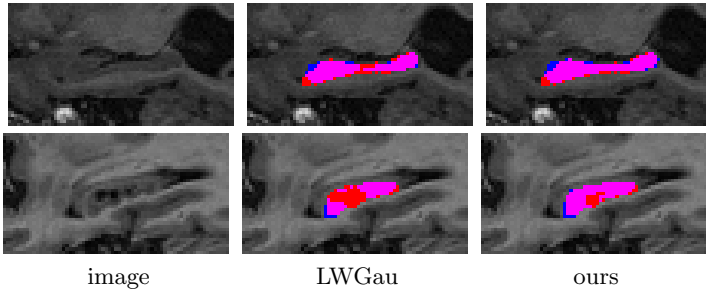


Fig. 4. Sagittal views of segmentations produced by LWGau and our method. Red: manual; Blue: automatic; Pink: overlap between manual and automatic segmentation.

the comparisons carried out below indicate the highly competitive performance achieved by our label fusion technique.

[45,10] present the highest published hippocampus segmentation results produced by MALF. All these methods are based on label fusion with similarity-based local weighting. The experiments in [45] are conducted in a leave-one-out strategy on a data set containing 80 control subjects. They report average Dice overlaps of 0.887 and 0.884, respectively. For controls, we produce Dice overlap of 0.898 ± 0.019 , more than 1% Dice overlap improvement. [10] uses a template library of 55 atlases. However, for each atlas, both the original atlas and its flipped mirror image are used. Hence, [10] effectively uses 110 atlases for label fusion. [10] reports results in Jaccard index ($JI(A, B) = \frac{|A \cap B|}{|A \cup B|}$) for the left side hippocampus of 10 controls, 0.80 ± 0.03 , and 10 MCI patients, 0.81 ± 0.04 . Our results for the left side hippocampus are 0.823 ± 0.031 for controls and 0.798 ± 0.041 for MCI patients. Overall, our results for controls are better than the state of the art and our results for MCI patients are slightly worse, but we use significantly fewer atlases than [45,10].

5 Conclusions

We proposed a novel method to derive optimal weights for label fusion. Unlike previous label fusion techniques, our method automatically computes weights by explicitly considering the error correlations between atlases. To estimate the correlations between atlases, we use a linear appearance-label model. In our experiment, our method significantly outperformed the state of the art label fusion technique, the similarity-based local weighting methods. Our hippocampus segmentation results also compare favorably to the state of the art in published work, even though we used significantly fewer atlases.

References

1. Allasonniere, S., Amit, Y., Troune, A.: Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B* 69(1), 3–29 (2007)

2. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de-Solorzano, C.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Tran. Medical Imaging* 28(8), 1266–1277 (2009)
3. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12, 26–41 (2008)
4. Collins, D., Pruessner, J.: Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *NeuroImage* 52(4), 1355–1366 (2010)
5. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Nonlocal patch-based label fusion for hippocampus segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010. LNCS*, vol. 6363, pp. 129–136. Springer, Heidelberg (2010)
6. Hansen, L.K., Salamon, P.: Neural network ensembles. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 12(10), 993–1001 (1990)
7. Hasboun, D., Chantome, M., Zouaoui, A., Sahel, M., Deladoeuille, M., Sourour, N., Duymes, M., Baulac, M., Marsault, C., Dormont, D.: MR determination of hippocampal volume: Comparison of three methods. *Am. J. Neuroradiol.* 17, 1091–1098 (1996)
8. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphism atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
9. Kittler, J.: Combining classifiers: A theoretical framework. *Pattern Analysis and Application* 1, 18–27 (1998)
10. Leung, K., Barnes, J., Ridgway, G., Bartlett, J., Clarkson, M., Macdonald, K., Schuff, N., Fox, N., Ourselin, S.: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer’s Disease. *NeuroImage* 51, 1345–1359 (2010)
11. Murty, K.G.: *Linear Complementarity, Linear and Nonlinear Programming*. Helderman-Verlag (1988)
12. Pluta, J., Avants, B., Glynn, S., Awate, S., Gee, J., Detre, J.: Appearance and incomplete label matching for diffeomorphic template based hippocampus segmentation. *Hippocampus* 19, 565–571 (2009)
13. 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)
14. Sabuncu, M., Yeo, B., Leemput, K.V., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Trans. on Medical Imaging* 29(10), 1714–1720 (2010)
15. Scahill, R., Schott, J., Stevens, J., Fox, M.R.N.: Mapping the evolution of regional atrophy in Alzheimer’s Disease: unbiased analysis of fluidregistered serial MRI. *Proc. Natl. Acad. Sci. U. S. A.* 99(7), 4703–4707 (2002)
16. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. on Medical Imaging* 23(7), 903–921 (2004)

Characterizing Spatially Varying Performance to Improve Multi-atlas Multi-label Segmentation

Andrew J. Asman¹ and Bennett A. Landman^{1,2,3}

¹Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

²Biomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

³Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA 21218

{andrew.j.asman,bennett.landman}@vanderbilt.edu

Abstract. Segmentation of medical images has become critical to building understanding of biological structure-functional relationships. Atlas registration and label transfer provide a fully-automated approach for deriving segmentations given atlas training data. When multiple atlases are used, statistical label fusion techniques have been shown to dramatically improve segmentation accuracy. However, these techniques have had limited success with complex structures and atlases with varying similarity to the target data. Previous approaches have parameterized raters by a single confusion matrix, so that spatially varying performance for a single rater is neglected. Herein, we reformulate the statistical fusion model to describe raters by regional confusion matrices so that co-registered atlas labels can be fused in an optimal, spatially varying manner, which leads to an improved label fusion estimation with heterogeneous atlases. The advantages of this approach are characterized in a simulation and an empirical whole-brain labeling task.

Keywords: Simultaneous truth and performance level estimation (STAPLE), Statistical fusion, Classifier fusion, Rater performance, Automated segmentation.

1 Introduction

Knowledge of the connections and relationships between biological structures and function is essential to scientific and clinical interpretation of medical images. Segmentation plays a pivotal role in building understanding of these relationships as it enables association of quantitative and functional observations with structural counterparts. Fully automated segmentation methods are important tools for achieving robust, high-throughput analysis, yet imaging and anatomical variability render these challenging objectives. The gold standard approach is to have a human expert manually label each image; yet manual segmentation (labeling) suffers from intra- and inter-rater reliability concerns. It is natural to have multiple experts (i.e., raters) perform the segmentation task so that the output of each individual can be combined (“fused”) to form an estimate of the “ground truth.” Manual labeling can be extraordinarily time consuming and cost prohibitive, so alternative approaches have been developed to efficiently derive multiple possible segmentations. In both manual and

automated cases, a central challenge becomes finding the optimal method to fuse the possible segmentations so that a consistent, reliable and accurate estimate of the true segmentation can be obtained.

A practical and simple approach to fusion is using majority vote, where the “ground truth” estimate is obtained by declaring the label for each voxel that was reported most frequently [1, 2]. This approach, however, 1) does not guarantee a unique majority label, 2) does not provide information about the likelihood of the estimated segmentation and 3) does not provide any information about rater performance. The Simultaneous Truth and Performance Level Estimation approaches (aka, STAPLE) were recently presented to optimally estimate observations based on rater reliability [3, 4]. When raters are collectively unbiased and independent, this algorithm increases the accuracy of a single labeling by probabilistically fusing multiple less accurate delineations, e.g., following the general theory of statistical boosting.

Multi-atlas segmentation through non-rigid registration and label transfer represents a fully automated method for performing segmentation through label fusion based on existing label sets (i.e., the atlases). Using a database of manually segmented atlases, this segmentation technique has been shown to be reasonably robust and accurate. As with the manual labeling method, the problem becomes determining the optimal method to combine the observed segmentations. Statistical fusion techniques have been widely used in multi-atlas labeling [5-8]. However, when applied to multiple, intricate labels, voting techniques have been shown to dramatically outperform statistical fusion techniques [5, 9, 10]. Additionally, applications have largely focused on specific anatomical features with relatively few labels [7, 8].

Optimality of the statistical fusion approaches hinges on the validity of the labeling process model (e.g., the stochastic model of how a rater makes a mistake). Existing STAPLE approaches have used spatially invariant models — the probability distribution of error does not change voxel-by-voxel. Intuitively (and empirically – see Figure 1), there should exist certain regions where the registration (and label transfer) is more accurate than in other regions. Furthermore, the pattern of spatial agreement

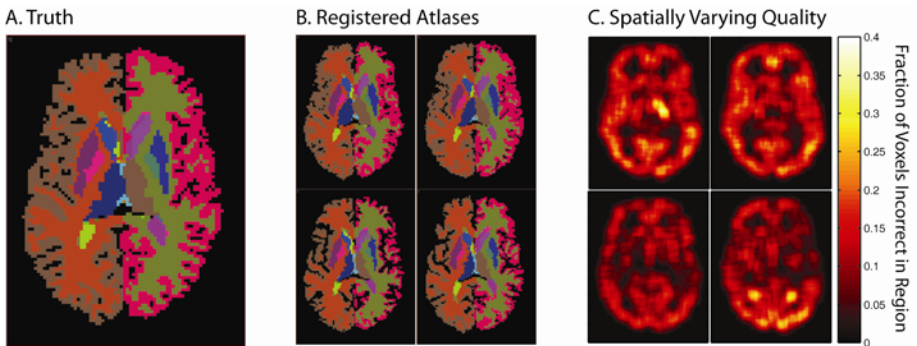


Fig. 1. The spatial quality variation exhibited by registered atlases. A representative slice from the true labels (manually drawn) of a target brain is presented in (A). Four example observations of this slice can be seen in (B). The quality of the observations seen in (B) is compared to the true labels seen in (A) to construct the spatial quality variation heat maps presented in (C). Note that the spatial variation is independent of the actual labels of the brain.

will vary on an atlas-by-atlas basis. Traditionally, single measures of rater performance that describe the quality of the registered atlases are an amalgamation of many sub-regions of varying accuracy. In the context of voting, regional and intensity based fusion techniques have emerged [7, 9]. However, other than simply ignoring consensus voxels [4], spatially varying characteristic have not been derived for the statistical fusion approach. Herein, we generalize the statistical fusion algorithm (STAPLE) to seamlessly account for spatially varying performance (Spatial STAPLE) by the inclusion of regional performance level estimations (confusion matrices).

This manuscript is organized as follows. Section II details the extension of traditional STAPLE theory to include spatially varying performance levels. Section III illustrates the advantages of Spatial STAPLE on a simulation example and in an empirical task of whole-brain labeling. Finally, Section IV provides brief concluding remarks.

2 Theory

The following presentation of theory for Spatial STAPLE closely follows the approach of Warfield, et al [3].

2.1 Problem Definition

Consider an image of N voxels with the task of determining the correct label for each voxel in that image. Consider a collection of R raters (registered atlases) that provide an observed delineation of all N voxels exactly once. The set of labels, \mathbf{L} , represents the set of possible values that a rater can assign to all N voxels. Let \mathbf{D} be an $N \times R$ matrix describing the labeling decisions of all R raters at all N voxels where $D_{ij} \in \{0, 1, \dots, L - 1\}$. Let \mathbf{T} be a vector of N elements that represents the hidden true segmentation for all voxels, where $T_i \in \{0, 1, \dots, L - 1\}$.

A characterization of the R raters performance at each voxel is characterized by $\boldsymbol{\theta}$. Each element of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_{jm}$, describes an $L \times L$ confusion matrix and is defined for rater j and region B_m , where B_m is a vector that indicates the voxels over which $\boldsymbol{\theta}_{jm}$ is defined. Let M be the total number of regions, and let the union of all regions equal the full volume, (i.e. $\cup_{m=1}^M B_m = \{1, \dots, N\}$) and all regions are mutually disjoint (i.e. $B_i \cap B_j = \emptyset, \forall i \neq j$). Let the complete data be (\mathbf{D}, \mathbf{T}) and let the probability mass function of the complete data be $f(\mathbf{D}, \mathbf{T} | \boldsymbol{\theta})$.

Figure 2 illustrates the regional confusion matrix approach where an observation is divided into quadrants. Each quadrant is described by a different confusion matrix. One of the observed quadrants is of significantly higher quality than the other quadrants. Traditional STAPLE would fail to recognize this phenomenon, while Spatial STAPLE is capable of detecting this regional quality variation.

2.2 Spatial STAPLE Algorithm

The goal of the Spatial STAPLE algorithm is to accurately estimate the true segmentation using the R raters segmentation decisions and the current estimate of the rater performance level parameters. The estimated performance level parameters will be selected such that they maximize the complete data log likelihood function

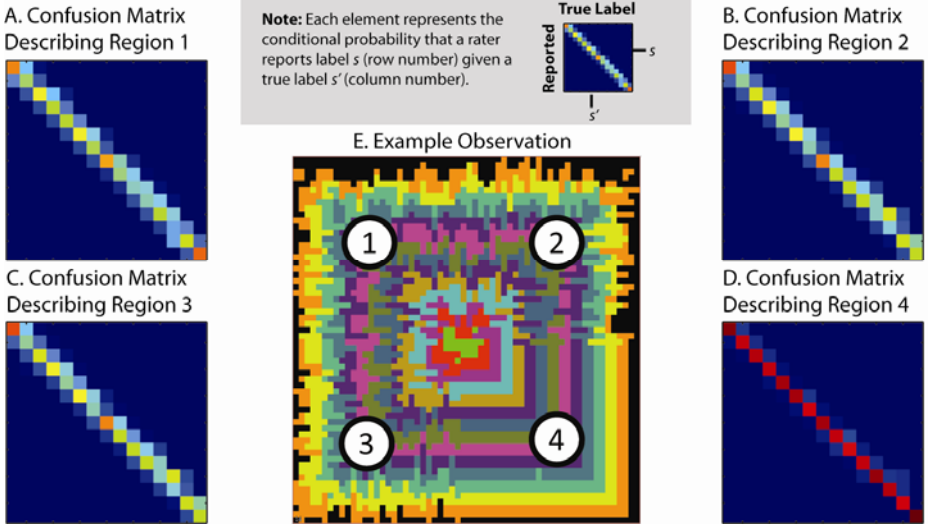


Fig. 2. A visual representation of the Spatial STAPLE algorithm. The images in (A – D) represent varying confusion matrices for the various regions of the presented observation. The confusion matrix presented in (D) is of significantly higher quality than the other confusion matrices as represented by the fact that it is nearly a diagonal matrix. The observation can be seen in (E) where the regions corresponding to each confusion matrix is specified.

$$\hat{\theta} = \arg \max_{\theta} \ln f(\mathbf{D}, \mathbf{T} | \theta). \quad (1)$$

It is assumed that the segmentation decisions are all conditionally independent given the true segmentation and the performance level parameters, that is $(D_{ij} | T_i \theta_{jm}) \perp (D_{ij'} | T_i \theta_{j'm}) \forall j \neq j'$. This model expresses the assumption that the raters derive their segmentations of the same image independently from one another and that the quality of the segmentation by each rater is captured by the estimation of the performance level parameters.

Our solution to the expectation maximization (E-M) algorithm used to solve (1) is now presented. The complete data used to solve this E-M algorithm is the observed data, \mathbf{D} , and the true segmentation, \mathbf{T} . The true segmentation is regarded as the missing or hidden data, and is unobservable. Let θ_{jm} be the covariance, or confusion matrix associated with rater j and region B_m and let

$$\theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1R} \\ \theta_{21} & \theta_{22} & \dots & \theta_{2R} \\ \dots & \dots & \dots & \dots \\ \theta_{M1} & \theta_{M2} & \dots & \theta_{MR} \end{bmatrix} \quad (2)$$

be the complete set of unknown parameters for the R segmentations and the M disjoint subsets of the full set of voxels.

Following the notation proposed by Warfield, et al [3], the estimation of the true segmentation (E-step) is represented using common E-M notation. Let \mathbf{W} be an $L \times N$ matrix, where each element W_{si} represents the probability that the true label associated with voxel i is label s . The solution for \mathbf{W} on iteration k is

$$\begin{aligned} W_{si}^{(k)} &\equiv f(T_i = s | \mathbf{D}_i, \boldsymbol{\theta}_{jm}^{(k-1)}) \\ &= \frac{f(T_i = s) \prod_j f(D_{ij} = n | T_i = s, \boldsymbol{\theta}_{jm}^{(k-1)})}{\sum_{s'} f(T_i = s') \prod_j f(D_{ij} = n | T_i = s', \boldsymbol{\theta}_{jm}^{(k-1)})} \\ &= \frac{f(T_i = s) \prod_j \theta_{jmns}^{(k-1)}}{\sum_{s'} f(T_i = s') \prod_j \theta_{jmns'}^{(k-1)}} \end{aligned} \quad (3)$$

where the value of m is selected such that $i \in B_m$ and $\theta_{jmns}^{(k-1)}$ represents the probability that rater j observed label n given that the true label is s in region B_m . The prior probability, $f(T_i = s)$, can be either a global prior or a voxelwise prior. In the case of this paper $f(T_i = s)$ is a global prior that represents an *a priori* estimate of the fraction of voxels in the true segmentation that have label value s .

The estimate of the performance level parameters (M-step) is presented in (4). The main difference between the representation seen in (4) and the traditional STAPLE representation [3] is the fact that $\boldsymbol{\theta}$ is an $R \times M$ set of confusion matrices, where each rater has M confusion matrices associated with it, each defined over a region space that is a subset of the full voxel set.

$$\begin{aligned} \boldsymbol{\theta}_{jm}^{(k)} &= \arg \max_{\boldsymbol{\theta}_{jm}} \sum_{i \in B_m} E[\ln f(D_{ij} | T_i, \boldsymbol{\theta}_{jm}) | \mathbf{D}, \boldsymbol{\theta}_{jm}^{(k-1)}] \\ &= \arg \max_{\boldsymbol{\theta}_{jm}} \sum_{i \in B_m} \sum_s W_{si}^{(k)} \ln f(D_{ij} | T_i = s, \boldsymbol{\theta}_{jm}). \end{aligned} \quad (4)$$

Using the estimate of $\boldsymbol{\theta}_{jm}^{(k-1)}$ for $f(D_{ij} | T_i = s, \boldsymbol{\theta}_{jm})$ and the constraint that each column of the confusion matrix must be normalized, we obtain the final result

$$\theta_{jms's}^{(k)} = \frac{\sum_{i \in B_m} I(D_{ij} = s') W_{si}^{(k-1)}}{\sum_{i \in B_m} W_{si}^{(k-1)}} \quad (5)$$

where $\theta_{jms's}^{(k)}$ is the probability that rater j reports label s' when the true label is s at voxel i , and I is the indicator function. This quantity is determined over the voxel set defined by B_m .

2.3 Sliding Windows and Biasing Priors

As evidenced in Figure 1, the spatial quality of a given rater or registered atlas can vary dramatically in a relatively small region. Thus, the number of voxels contained within set B_m should be relatively small. However, as B_m is diminished to account for this phenomenon, the ability to accurately characterize rater performance is dramatically hampered due to the limited number of degrees of freedom when estimating $\theta_{jm}^{(k)}$. To compensate for this problem, we introduce the idea of using a given region, B_m , as a sliding window with significant amount of overlap with other regions. Due to the overlap of the sliding windows, a given voxel i may be an element in multiple region sets. As a result, we use nearest neighbor interpolation to determine the appropriate $\theta_{jm}^{(k)}$ for a given voxel i .

With the overlap of the sliding windows, the estimations at a given $\theta_{jm}^{(k)}$ should be relatively smooth and stable. For example, if a given label is only observed in region B_m a handful of times, the estimation of a rater's quality at observing this label will be limited. As a result, we introduce the idea of using a whole-image estimate of performance level parameters from STAPLE for regularization. Within a maximum *a posteriori* approach, we would use an informed parametric prior based upon this estimate. Nevertheless, parametric characterization of the manifold of confusion matrices is involved and not strictly necessary. Rather, for computational and stability concerns, we introduce an implicit prior in the following form. Let $\theta_j^{(0)}$ be the confusion matrix associated with rater j estimated from the STAPLE algorithm. The performance level parameters can then be calculated as

$$\theta_{jms's}^{(k)} = \frac{\sigma \theta_{js's}^{(0)} + \sum_{i \in B_m} I(D_{ij} = s') W_{si}^{(k-1)}}{\sigma \sum_s \theta_{js's}^{(0)} + \sum_{i \in B_m} W_{si}^{(k-1)}} \quad (6)$$

where σ is a scale factor that is dependent upon the size of the region (sliding window) B_m , the number of voxels, and the number of raters. Our empirically derived expression for this scale factor is

$$\sigma = \kappa \frac{N}{N_w \ln R} \quad (7)$$

where κ is an arbitrary constant, N_w is the number of voxels per window B_m , R is the total number of raters and N is the total number of voxels. Note that σ is designed such that the value of κ should be close to unity for a given experiment.

2.4 Initialization and Convergence

The initialization strategies, convergence detection strategies and data-adaptive prior initialization strategies are essentially equivalent to the traditional STAPLE approach [3]. As with the STAPLE approach, each element of θ , θ_{jm} , is initialized to approximately 0.99 along the diagonal of the confusion matrix. The off-diagonal elements are filled with normally distributed values around 0.01. The standard normalization criterion for each column of θ_{jm} is maintained. Convergence is detected when the

average change in the on-diagonal values in θ is less than a constant (herein, 0.001). The data-adaptive prior, $f(T_i = s)$, present in the true segmentation estimation was initialized as a global prior such that each element represents the fraction of the total observed voxels ($N \times R$) that were equal to label s . Lastly, from Eq. 7, a value of $\kappa = 1$ was used for all experiments presented.

3 Methods and Results

All simulations and experiments were implemented in MATLAB (Mathworks, Natick, MA). All studies were run on a 64-bit quad-core 3.07 GHz desktop computer with 13 GB of RAM, running Ubuntu 10.04.

3.1 Spatially Varying Rater Quality Simulation

First, we consider a simulation in which raters exhibited spatially varying performance levels when segmenting a 3D volume ($80 \times 60 \times 60$) consisting of 15 embedded

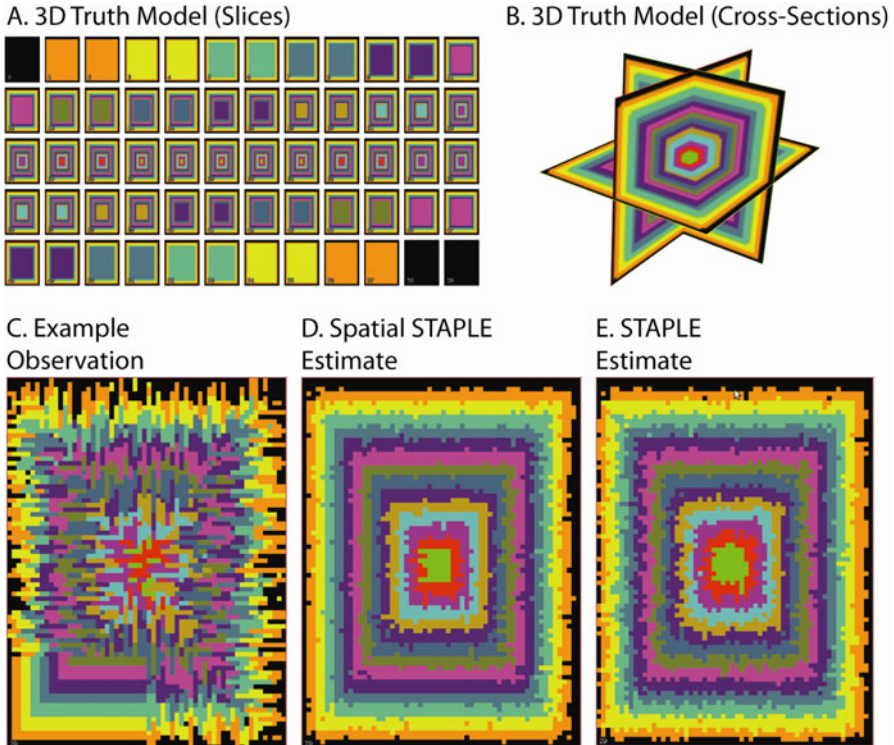


Fig. 3. Spatially varying rater quality simulation. The 3D truth model used in the simulation can be seen in both (A) and (B). (A) shows each of the individual slices of the model and (B) shows the three main cross-sections. (C) presents an example observation of the truth model. Representative estimates from both Spatial STAPLE (D) and STAPLE (E) are shown using 10 raters.

cubes (Figure 3). The volume was divided into 16 different equally sized regions. A total of 16 raters observed the volume where each rater was “perfect” in one of the 16 regions while exhibiting boundary-shift error behaviors in the remaining 15 regions. In the boundary-shift error regions, the level of boundary shift for each boundary point was chosen from a Gaussian distribution with zero mean and a variance specific to each rater. Note that when all 16 raters are fused, there exists a single rater that is perfect at all voxels in the volume.

Spatial STAPLE used a sliding window of $10 \times 10 \times 10$ voxels for a total of 1600 performance level estimations evenly distributed throughout the volume for each rater. Ten Monte Carlo iterations were used to assess the accuracy of STAPLE, Spatial STAPLE and majority vote for fusing the observations between 3 and 16 raters. The sensitivity of the biasing prior was assessed for 8 raters.

Both STAPLE and Spatial STAPLE were superior to majority vote for all numbers of raters fused (Figure 4A). For low numbers of raters the STAPLE and Spatial STAPLE estimates were of approximately equivalent quality. However, as the number of volumes increases, the Spatial STAPLE estimate becomes far superior to the STAPLE estimate.

For high values of κ (Eq. 7) the accuracy of Spatial STAPLE converges to the STAPLE result (e.g., increased impact of prior), while very low values of κ resulted in poor segmentations due to limited sample size in small sliding windows (Figure 4B). Nevertheless, a wide range of κ , the biasing prior resulted in estimates that were significantly superior to the STAPLE estimates.

3.2 Multi-atlas, Multi-label Empirical Experiment

Second, we examine an empirical application of label fusion to multi-atlas labeling for extraction of the cortical gray matter using atlases with six tissue labels. Fifteen

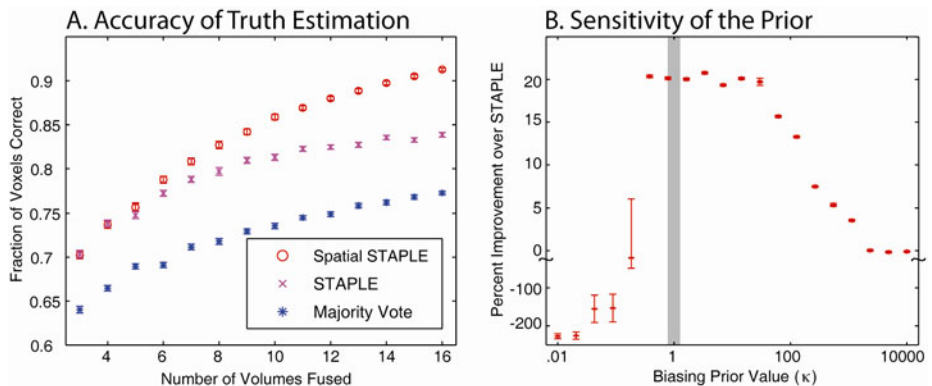


Fig. 4. Simulation results for spatial varying performance. The results from the 10 Monte Carlo iteration simulation can be seen in (A). It is evident that both STAPLE and Spatial STAPLE outperform majority vote for all numbers of raters. For increasing numbers of raters Spatial STAPLE dramatically outperforms STAPLE. The sensitivity of the implicit prior can be seen in (B). Note that for high values, the Spatial STAPLE estimate converges to the STAPLE estimate. For low values, the Spatial STAPLE estimate is unstable and results in estimations that are dramatically worse than STAPLE.

atlases were registered to a collection of 24 target atlases using the Vectorized Adaptive Bases Registration Algorithm (VABRA) [11] and corresponding whole-brain segmentations were correspondingly deformed to match each target. All segmentations were obtained from the Open Access Series of Imaging Studies (OASIS) [12]. The registered labels were then used as observations for fusion with STAPLE, Spatial STAPLE and majority vote. The Dice Similarity Coefficient (DSC) [13] was used to compare label volumes. A window size of $10 \times 10 \times 10$ voxels was used with 8000 performance level estimations for each rater rectilinearly distributed throughout the volume. Spatial STAPLE provides improvement over STAPLE in certain regions of the brain, particularly the insula (Figure 5). As the number of volumes fused increases, a consistent improvement in label agreement is evident.

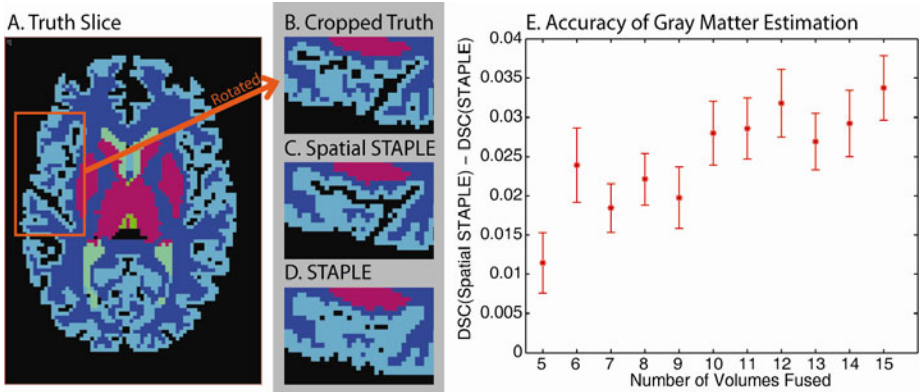


Fig. 5. Results from an empirical experiment using 6 labels. A representative truth slice from the 6-label model can be seen in (A). A cropped and rotated region (A) is presented in (B). The estimates seen in (C) and (D) represent the output from Spatial STAPLE (C) and STAPLE (D) using 8 volumes. The accuracy of the gray matter estimation (in terms of the difference in DSC values) for varying numbers of volumes can be seen in (E). For reference, the average STAPLE DSC was approximately 0.8.

Third, the performance of all three methods was characterized for an intricate whole-brain labeling with 41 labels otherwise using the same procedure as above. Spatial STAPLE exhibits significant improvements over STAPLE estimate (Figure 6). For several of the labels, particularly the gray matter and some smaller labels in the mid-brain, Spatial STAPLE shows visible improvement over the majority vote estimate (Figure 6). For more than 5 registered atlases, both Spatial STAPLE and majority vote are superior to STAPLE. The median performance of majority vote and Spatial STAPLE are very similar as can be seen in Figure 6F; however, Spatial STAPLE results in a non-trivial improvement over majority vote for a large number of outlier cases.

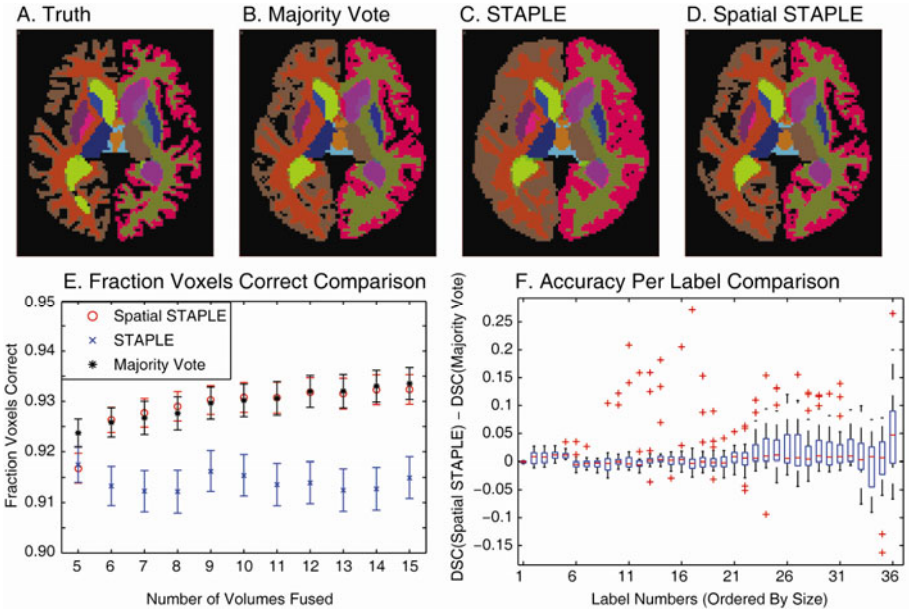


Fig. 6. Results from the empirical experiment using 41 labels. A representative slice from the truth model can be seen in (A). The majority vote, STAPLE and Spatial STAPLE estimates for this slice using 8 volumes can be seen in (B), (C) and (D), respectively. A comparison of the fraction of voxels correct when fusing 5 to 15 volumes to all 24 target atlases can be seen in (E). A per label comparison between Spatial STAPLE and majority vote can be seen in (F). Only the 36 labels that were consistent between the 24 target atlases are shown. Note that both Spatial STAPLE and majority vote outperform STAPLE for increasing numbers of volumes in (E). Lastly, note that there are a large number of outliers for which Spatial STAPLE outperforms majority vote in (F).

4 Discussion and Conclusion

Spatial STAPLE accounts for spatial quality variation through seamless integration of multiple confusion matrices with the established STAPLE theory. This algorithm presents a dramatic change in how we can view the implementation of statistical fusion algorithms.

With only a single confusion matrix, one is limited to cases where each rater is consistent throughout an entire volume. Our approach provides a conceptually simple and computationally tractable method to account for spatially varying performance characteristics. As presented, highly irregular performance level variations (as observed in Figure 1) can be accounted for and utilized to perform local weighting of atlas labels, and, in turn, improve estimations of the true segmentation and understanding of rater performance. The introduction of a data informed prior regularizes the estimation process and enables localized performance estimates without substantially decreasing precision. Nevertheless, important topics of future research persist, such as a more thorough formulation of the biasing prior and a characterization of the

sensitivity of the algorithm to the initialization of the performance level parameters. Additionally, this algorithm will have to be compared to more sophisticated voting based fusion algorithms [5, 9] to more completely assess performance.

A subtle, but important, result of this work is that by accounting for spatially varying rater performance in these demonstrations, the often-observed reliability gap between statistical fusion and majority vote has been closed or overcome. Both the use of local performance characterization and the introduction of data-informed priors could be used with other recent innovations in statistical fusion (such as new data types or rater models). The degree of correlation and impact of data informed priors can be readily controlled through hierarchical or regional derivations. As presented, the framework estimates and exploits spatially varying structure, but the performance gains from this approach suggest that this proposed framework could also be used to estimate the structure of potential correlations in rater performance and, perhaps, exploit these characteristics in the design of statistical fusion paradigm.

Acknowledgments

This work was supported in part by NIH/NINDS 1R01NS056307 and NIH/NINDS 1R21NS064534.

References

1. Warfield, S., Dengler, J., Zaers, J., Guttman, C.R., Wells III, W.M., Ettinger, G.J., Hiller, J., Kikinis, R.: Automatic identification of gray matter structures from MRI to improve the segmentation of white matter lesions. *J. Image Guid Surg.* 1, 326–338 (1995)
2. Kikinis, R., Shenton, M.E., Gerig, G., Martin, J., Anderson, M., Metcalf, D., Guttman, C.R., McCarley, R.W., Lorenson, W., Cline, H., et al.: Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *J. Magn. Reson. Imaging* 2, 619–629 (1992)
3. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging* 23, 903–921 (2004)
4. Rohlfing, T., Russakoff, D.B., Maurer, C.R.: Performance-Based Classifier Combination in Atlas-Based Image Segmentation Using Expectation-Maximization Parameter Estimation. *IEEE Transactions on Medical Imaging* 23, 983–994 (2004)
5. Langerak, T., van der Heide, U., Kotte, A., Viergever, M., van Vulpen, M., Pluim, J.: Label Fusion in Atlas-Based Segmentation Using a Selective and Iterative Method for Performance Level Estimation (SIMPLE). *IEEE Transactions on Medical Imaging* 29, 2000–2008 (2010)
6. Lotjonen, J.M., Wolz, R., Koikkalainen, J.R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D.: Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49, 2352–2365 (2010)
7. 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 Transactions on Medical Imaging* 28, 1000–1010 (2009)

8. Depa, M., Sabuncu, M.R., Holmvang, G., Nezafat, R., Schmidt, E.J., Golland, P.: Robust Atlas-Based Segmentation of Highly Variable Anatomy: Left Atrium Segmentation. In: Camara, O., Pop, M., Rhode, K., Sermesant, M., Smith, N., Young, A. (eds.) STACOM 2010. LNCS, vol. 6364, pp. 85–94. Springer, Heidelberg (2010)
9. Sabuncu, M., Yeo, B., Van Leemput, K., Fischl, B., Golland, P.: A Generative Model for Image Segmentation Based on Label Fusion. *IEEE Transactions on Medical Imaging* 29, 1714–1729 (2010)
10. Kittler, J., Alkoot, F.: Sum versus vote fusion in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 110–115 (2003)
11. Rohde, G.K., Aldroubi, A., Dawant, B.M.: The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging* 22, 1470–1479 (2003)
12. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, non-demented, and demented older adults. *Journal of Cognitive Neuroscience* 19, 1498–1507 (2007)
13. Dice, L.: Measures of the amount of ecologic association between species. *Ecology* 26, 297–302 (1945)

Discovering Dense and Consistent Landmarks in the Brain

Dajiang Zhu¹, Degang Zhang^{1,2}, Carlos Faraco^{3,4}, Kaiming Li^{1,2}, Fan Deng¹,
Hanbo Chen¹, Xi Jiang¹, Lei Guo², L. Stephen Miller^{4,5}, Tianming Liu^{1,4}

¹University of Georgia, Dept. of Computer Science,
415 Boyd GSRC Building, Athens, GA, USA
dajiang.zhu@gmail.com

²Northwestern Polytechnical University, Shaanxi, China

³University of Georgia, Biomedical Health Sciences Inst., Athens, GA, USA

⁴University of Georgia, BioImaging Research Center, Athens, GA, USA

⁵University of Georgia, Dept. of Psychology, Athens, GA, USA

Abstract. The lack of consistent and reliable functionally meaningful landmarks in the brain has significantly hampered the advancement of brain imaging studies. In this paper, we use white matter fiber connectivity patterns, obtained from diffusion tensor imaging (DTI) data, as predictors of brain function, and to discover a dense, reliable and consistent map of brain landmarks within and across individuals. The general principles and our strategies are as follows. 1) Each brain landmark should have consistent structural fiber connectivity pattern across a group of subjects. We will quantitatively measure the similarity of the fiber bundles emanating from the corresponding landmarks via a novel trace-map approach, and then optimize the locations of these landmarks by maximizing the group-wise consistency of the shape patterns of emanating fiber bundles. 2) The landmark map should be dense and distributed all over major functional brain regions. We will initialize a dense and regular grid map of approximately 2000 landmarks that cover the whole brains in different subjects via linear brain image registration. 3) The dense map of brain landmarks should be reproducible and predictable in different datasets of various subject populations. The approaches and results in the above two steps are evaluated and validated via reproducibility studies. The dense map of brain landmarks can be reliably and accurately replicated in a new DTI dataset such that the landmark map can be used as a predictive model. Our experiments show promising results, and a subset of the discovered landmarks are validated via task-based fMRI.

Keywords: Brain landmark, diffusion tensor imaging, fiber shape, optimization.

1 Introduction

The lack of consistent and reliable functionally meaningful landmarks in the brain significantly hampered the advancement of brain imaging studies. It is widely called for to investigate approaches to mapping dense brain landmarks that represent the common brain architecture at the individual level, while automatically establishing

correspondences across subjects [1, 2]. In the brain science literature, anatomic landmarks such as sulci and gyri on the cerebral cortex have been commonly used to differentiate and identify brain regions [1]. However, the spatial resolutions and functional specificities of brain landmarks at the gyral or sulcal scales are limited. In particular, each sulcus or gyrus may contain multiple functionally inhomogeneous sub-units [3] and thus using gyral or sulcal region as brain landmark has limitations.

Typically, functional landmarks can be identified via task-based fMRI, which has been widely accepted as a readily available, reliable approach to identify functionally-specialized brain regions [4]. However, in many application scenarios, there is no task-based fMRI data available. For instance, it is challenging to conduct task-based fMRI studies and acquire high-quality data for elderly or children participants [5-7]. Additionally, the human brain is composed of many functional networks such as working memory, vision, auditory, language, motor, attention, emotion, and default mode systems [3], and acquisition of extensive task-based fMRI data for mapping multiple brain networks is very time-consuming and sometime an impossible task. Instead, a typical diffusion tensor imaging (DTI) scan needs less than ten minutes, is much less demanding, and is widely available. Therefore, we are strongly motivated to identify and discover functionally meaningful landmarks based only on DTI data.

Interestingly, a variety of recent studies [8-11] indicated that white matter connectivity patterns are good predictors of functional landmarks. That is, white matter connectivity patterns are quite consistent across brains for the same functional landmark. This is similar to the “connectional fingerprint” concept presented in [9, 10]. Essentially, each brain’s cytoarchitectonic area has a unique set of extrinsic inputs and outputs, called the “connectional fingerprint” [9], and this is crucial in determining the functions that each brain area performs [9]. The close relationship between structural connectivity pattern and brain function has been reported in other studies in the literature [8-11]. Therefore, in this paper, we propose to use consistent white matter fiber connectivity patterns as predictors of brain function, and to discover a dense, reliable and consistent map of brain landmarks based on DTI data within and across individuals.

Our computational framework is composed of the three phases shown in Fig. 1, and described as follows: 1) A dense, regular grid map of landmarks is initialized on a randomly selected template subject, and then warped to other individual subjects via a linear registration algorithm. As a result, approximately 2000 brain landmarks are initialized with rough correspondences on 15 subjects’ DTI images. 2) The locations of each corresponding grid landmark are optimized within a group via the maximization of group-wise consistency of structural fiber connection patterns. Based on the premise in the above paragraph, the landmarks with consistent fiber connection patterns across subjects are regarded as the discovered landmarks. 3) A subset of the discovered landmarks is validated via a task-based fMRI study, which activated sixteen functionally meaningful regions in the working memory network.

The major contributions of this paper are twofold. 1) We present a novel framework for optimizing the locations of the initialized brain landmarks via group-wise consistency maximization of the landmarks’ structural connectivity patterns. In particular, we developed a novel approach for quantitative measurement of the consistency of a landmark’s structural connectivity pattern by projecting the fiber curves onto a standard spherical space, which we refer to as trace-map. Our results indicate that the

structural connectivity patterns for each individual landmark in different brains are quite consistent after the optimization. Our validation results using task-based fMRI demonstrate the functional validity of our DTI-based landmarks. These results provide direct evidence to support the view that the structural connectivity pattern of a landmark is a good predictor of brain function [8-11]. 2) We present a novel framework that allows us to identify a dense map of brain landmarks (over 300) with much finer granularity and significantly better functional homogeneity than the traditional Brodmann atlas. Importantly, our reproducibility studies demonstrate that the discovered dense map of brain landmarks is reproducible and predictable in separate subject populations. The landmark map prediction framework, together with the discovered landmarks within our datasets, provides a novel solution to the challenging problem of structured and individualized representations of the human brain.

2 Methods

2.1 Overview

Figure 1 summarizes the flowchart of our computational framework for dense and consistent landmark discovery. The framework is composed of three major steps, including initialization of the dense landmark, optimization of each initialized landmark within a group of subjects, and identification and validation of consistent landmarks. The details of each step will be presented in the following sections.

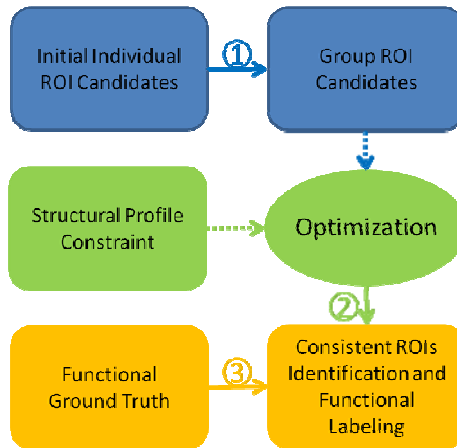


Fig. 1. The flowchart of the landmark discovery framework. (1) A dense, regular grid map of landmarks is initialized on a randomly selected template subject, and then warped to other individual subjects via a linear registration algorithm. As a result, 2056 landmarks are initialized with rough correspondences across fifteen subjects. (2) The locations of each corresponding grid landmark in different subjects are optimized within a group via the maximization of group-wise consistency of structural fiber connection patterns. (3) A subset of optimized landmarks is validated based on a task-based fMRI study which consistently activated sixteen functional regions of working memory system.

2.2 Dataset, Preprocessing and Initialization

The multimodal fMRI and DTI dataset was obtained from our recent study in [12]. Briefly, fMRI data was acquired for each subject for 2 runs. Each run included three block types: a modified version of the operation-span (OSPAN) task, an Arithmetic task and a Baseline. DTI data were also acquired. The scans were performed on a 3T GE Signa HDx scanner. Acquisition parameters were as follows: fMRI: 64×64 matrix, 4mm slice thickness, 220mm FOV, 30 slices, TR=1.5s, TE=25ms, ASSET=2; DTI: 128×128 matrix, 2mm slice thickness, 256mm FOV, 60 slices, TR=15100ms, TE=variable, ASSET=2, 3 B0 images, 30 optimized gradient directions, b-value=1000; all aligned to the AC-PC line.

Each participant’s fMRI data was pre-processed and analyzed, using the OSPAN – Arithmetic contrast, through FSL’s FEAT. In total, we identified 16 activated ROIs. DTI pre-processing included skull removal, motion correction and eddy current correction. Fiber tracking was performed using MEDINRIA (FA threshold=0.2; minimum fiber length=20). Brain tissue segmentation was conducted on DTI data using a method similar to [13] and the cortical surface was reconstructed using the marching cubes algorithm. In this paper, the DTI space was used as the standard space in which we reported the results. Because DTI and fMRI sequences are both echo planar imaging (EPI) sequences, their distortions tend to be similar. So the misalignment between DTI and fMRI images is much less than that between T1 and fMRI images [14]. Co-Registration between fMRI and DTI data was performed using FSL FLIRT. The tracked fibers and activated ROIs were then mapped onto the cortical surface.

We randomly selected one case from 15 subjects as the template and generated a dense, regular map of 3D grid points within the boundary box of the reconstructed cortical surface. The intersection locations between the grid map and the cortical surface are used as the initial landmarks, as shown in Fig. 2a. As a result, we generated 2056 landmarks on the template subject. Then, we registered this grid map of landmarks to other subjects using the linear registration algorithm FSL FLIRT (Fig. 2b). This linear warping initialized the dense grid map of landmarks and established their rough correspondences.

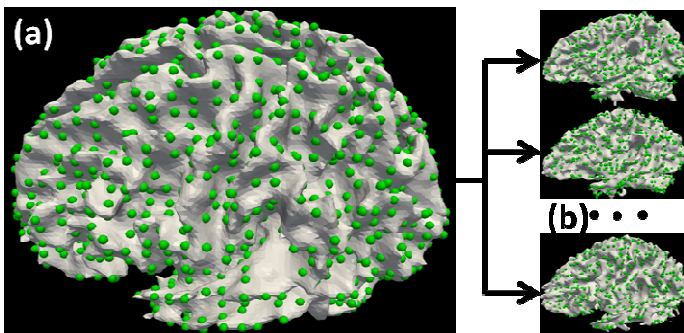


Fig. 2. Illustration of landmark initialization among a group of subjects. (a) We generated a dense regular grid map on a randomly selected template. (b) We registered this grid map to other subjects using linear registration algorithm. The green bubbles are the landmarks.

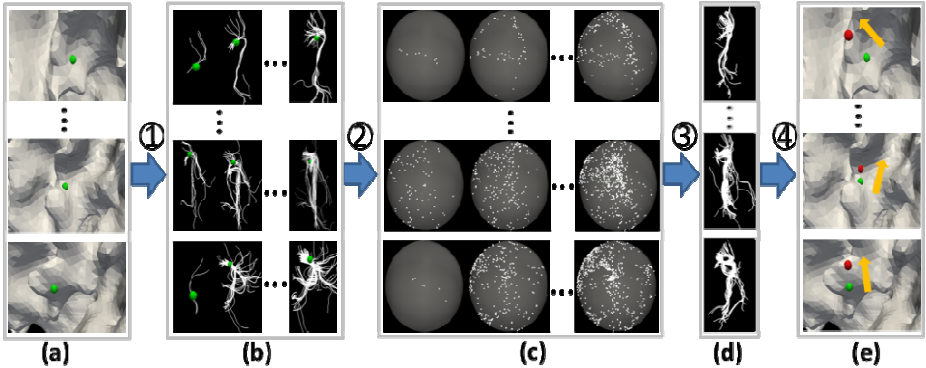


Fig. 3. The workflow of our landmark optimization algorithm. (a) The corresponding initialized landmarks of a group of subjects. The green bubbles are the landmarks. (b) A group of fiber bundles extracted from the neighborhood of the landmark. (c) Trace-maps corresponding to each fiber bundle. (d) The optimized fiber bundle of each subject. (e) The movements of the landmarks from initial locations (green) to the optimized locations (red). (1) Extracting fiber bundles from different locations close to the initial landmark. (2) Transforming the fiber bundles to trace-maps. (3) Finding the group of fiber bundles which make the group variance the least. (4) Finding the optimized location of initial landmark (red bubble).

2.3 Optimization of Landmark Locations

The algorithmic pipeline we developed for individual landmark location optimization is outlined in Fig. 3. It is noted that in this pipeline, each landmark will be optimized separately. First, we extracted white matter fiber bundles emanating from small regions around the neighborhood of the initial landmark. The centers of these small regions will be determined by the vertices of the cortical surface mesh, and each small region will serve as the candidate for landmark location optimization. Here, one green bubble in Fig. 3a represents one initial landmark and its correspondence across the different subjects. Fig. 3b shows examples of the fiber bundles we extracted. Then, we projected the fiber bundles to the standard trace-maps (details in section 2.3.1) as shown in Fig. 3c, and calculated the distance between any pair of trace-maps in different subjects within the group. Finally, we perform a whole-space search to find one group of fiber bundles (Fig. 3d) which gives the least group-wise variance. Fig. 3e shows examples of the optimized locations (red bubble) and the landmark movements (yellow arrow).

2.3.1 Trace-Map Model

A variety of algorithms have been developed for white matter fiber shape analysis in the literature [e.g., 15, 16]. However, it is still a challenging problem to quantitatively describe the shapes of fiber bundles and measure their similarities. In this paper, we propose a novel method by which to describe fiber bundles and have dubbed it “trace-map model.” First, we divide each fiber curve into segments which are composed of a collection of points (Fig. 4a). Then, we perform a Principal Component Analysis (PCA) to find the principal direction of one of the segments, which is then represented

as a vector, as shown in Fig. 4a. Next, we translate the vectors to the origin of a global spherical coordinate system and shoot from the origin to a unit sphere whose center is the origin (Fig. 4c). In this way, we can have a trace point on the sphere, and then perform the same procedure on the segments of all fibers in each bundle (Fig. 4c). Fig. 4d shows two examples. The first row is a U-shape fiber bundle and its corresponding trace-map. The second row is a line-shape case.

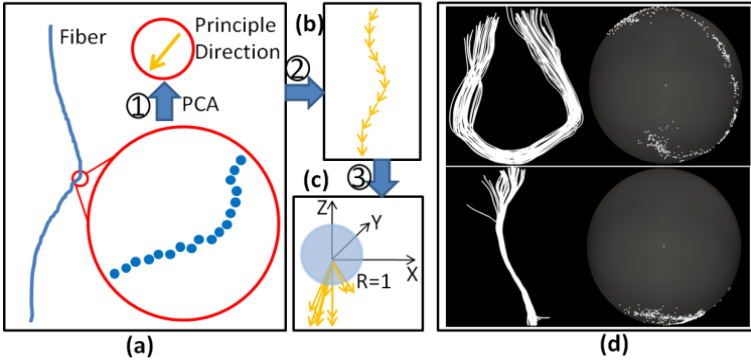


Fig. 4. (a) Calculation of the principal direction for one segment of each fiber. (b) Each segment could be represented by a series of vectors. (c) After translation to the origin of a global coordinate system, each vector shoots to a unit sphere whose center is the origin. (d) Two examples of fiber bundles and their trace-maps. The top row is a U-shape fiber bundle example and the bottom row is a line-shape one. For both cases, the left are fiber bundles and right are their trace-map representations.

There are two issues to be noted here. One is that we must make sure all subjects' brains are aligned. In our implementation, we calculate the principal direction of each brain using PCA, which we use to align different brains. Thus, fiber bundles with similar shapes but different orientations can be differentiated by different trace-point distributions on the standard sphere surface. The second issue is that we need to identify one of the two ends of the fiber bundles as a starting point. Since each fiber bundle is extracted from a small region on the cortical surface, we select the end that is closer to the center of the region as the starting point. This is very important to ensure that trace-maps of one fiber are consistent across different optimization procedures. The proposed trace-map model has the following advantages. 1) It is an effective way to present a fiber bundle. Essentially, the trace-map model transforms a fiber bundle to a set of points distributed on the surface of a unit sphere. It projects a complex fiber's geometric features into point distribution patterns on a standard space, in which different fiber bundles from different subjects can thus be compared quantitatively. The trace-map patterns reflect the accumulation of the strength of the fiber bundle in different directions. To a certain extent, it is similar to the idea of inflating the convoluted cortical surface onto a standard sphere in FreeSurfer [17]. 2) The trace-map model is not sensitive to small changes in the composition of a fiber bundle. This is a very important property when we perform between subjects comparisons because we

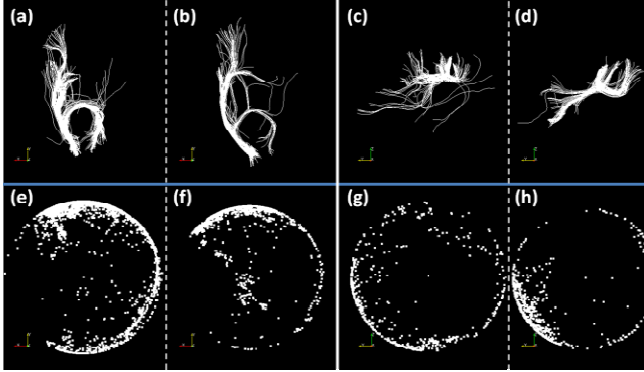


Fig. 5. (a), (b) and (c), (d) are two pairs of similar fiber bundles. (e)-(h) are their respective trace-maps.

want to determine whether the fiber bundles have similar overall shapes. 3) The trace-map model has the property of rotation invariance, because the coordinate system is already aligned before the projection.

2.3.2 Fiber Bundle Comparison Based on Trace-Maps

Our rationale for comparing fiber bundles through trace-maps is that similar fiber bundles have similar overall trace-map patterns. Fig. 5 shows 4 examples. Fig. 5a and 5b is a pair of fiber bundles that appear similar upon visual inspection. We can see that their trace-maps, Fig. 5e and 5f, are also similar. Fig. 5c and 5d show another pair of similar fiber bundles. Their trace-maps are shown in Fig. 5g and 5h. We can clearly see the similar patterns of the point distributions in the trace-maps.

After representing the fiber bundle by the trace-map model, the bundles can be compared by defining the distances between their corresponding trace-maps, as shown in Fig. 6. We built a standard sphere coordinate system as showed in Fig. 6a and set up the sample points on the standard sphere surface by adjusting angle Φ and θ . The step of angle change is $\pi/6$. Hence, we totally have 144 sample points as showed in Fig. 6b. For each trace-map, we can calculate the point density at the location of certain sample point. In other words, we can use a histogram vector of 144 dimensions to represent a trace-map. Each dimension in this vector is the point density information of a specific sample point. As a result, the vector can reflect the point distribution of a trace-map uniquely. The point density $\text{den}(P_i)$ is defined as :

$$\text{den}(P_i) = n_i/N \quad (1)$$

where n_i is the number of points in the trace-map whose center is P_i with radius d . In this paper, $d=0.3$. N is total number of points in the trace-map. As shown in Fig. 6b, we calculate the point density within the range of the yellow circle. The distance of two trace-maps is defined as:

$$D(T, T') = \frac{\sum_{i=1}^n (T_i - T'_i)^2}{n} \quad (2)$$

where T and T' are two vectors representing different trace-maps. T_i and T'_i are the i th attributes of the vector T and T' . n is the number of sample points, and in this paper n equals 144. Note that the point density here is normalized so that we do not require that the numbers of points in different trace-maps are equal.

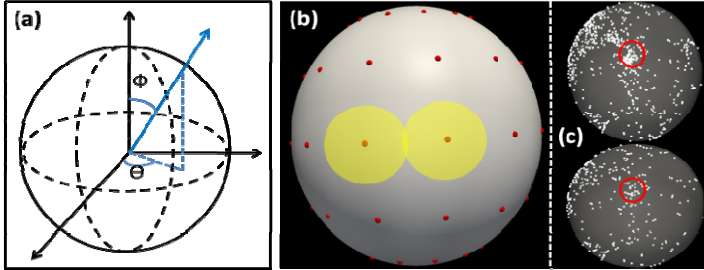


Fig. 6. (a) A sphere coordinate system for finding the sample points. We totally have 144 sample points by adjusting angle Φ and θ . (b) A sphere with 144 sample points. (c) Two trace-maps. The two red circles belong to the same sample point and will be compared based on the point density information within red circles.

2.3.3 Optimization of Landmark Locations

We formulate the problem of optimization of landmark locations as an energy minimization problem, which aims to maximize the consistency of structural connectivity patterns across a group of subjects. By searching the whole-space of landmarks locations, we can find an optimal combination of new landmark locations that ensure the fiber bundles from different subjects have the least group variance. Mathematically, the energy function we want to minimize is defined as:

$$E(S_1, S_2, \dots, S_m) = \sum E(S_k, S_l), k \neq l \text{ and } k, l=1,2,\dots,m \quad (3)$$

$S_1 \dots S_m$ are m subjects. We let $E(S_k, S_l) = D(T_k, T_l)$, and rewrite the Eq. (3) as below:

$$E(S_1, S_2, \dots, S_m) = \frac{\sum \sum_{i=1}^n (T_{ki} - T_{li})^2}{n}, k \neq l \text{ and } k, l=1,2,\dots,m \quad (4)$$

For any two subjects S_k and S_l , we transformed them to the corresponding vector format, T_k and T_l , of trace-maps. T_{ki} and T_{li} are the i th attributes of T_k and T_l respectively.

In our implementation, for each landmark of the subject, we examined around 30 locations and extracted their corresponding emanating fiber bundles as the candidates for optimization. Then, we transformed the fiber bundles to trace-maps. After representing them as vectors, we calculated the distance between any pair of them from different subjects. Thus, we can do search in the whole space of landmark location combinations to find the optimal one which has the least variance of fiber bundles shapes within the group.

3 Experimental Results

Our experimental results include three parts. First, we performed quantitative measurements of the consistencies of fiber shape patterns before and after optimization, and conducted a reproducibility study of the optimization framework in separate groups of subjects. Then, we performed a comparison of optimized landmarks with the activated brain regions detected from a working memory task-based fMRI study [12]. Last, we will examine the influence of different initial locations of the landmarks on our optimization results.

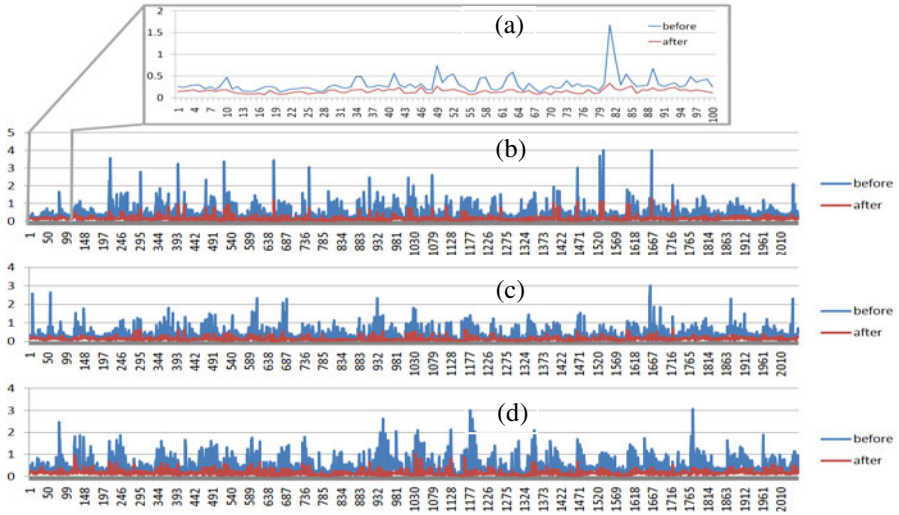


Fig. 7. Comparison of group distance of each landmark before and after optimization. The horizontal axis shows the landmark index. The vertical axis is the group trace-map distance. We randomly divided 15 subjects to three groups (5 for each group) and did the optimization separately. (b), (c) and (d) show the results of three groups. (a) is an enlarged view of a part of (b).

3.1 Consistency of Fiber Shape Patterns and Reproducibility Study

We randomly divided the 15 subjects into three equal groups. We then performed the landmark optimization framework in Section 2 on each subject group separately. The group-wise trace-map distances for each landmark measured by Eq. (2) in three groups before (blue curve) and after (red curve) optimization are shown in Fig. 7b-7d. We can see that after optimization, the group distance decreased significantly, indicating that the fiber shape patterns are very similar for corresponding landmarks across subjects. The average decrease of the group trace-map distance was 64.4%. After a random, visual examination of hundreds of landmarks across the three groups, we found that most of the optimized landmarks have quite similar fiber shape patterns. Fig. 10 (in next section) shows examples of the fiber bundles from the three groups before and after optimization. It is striking that even though the landmarks in the three groups of subjects were optimized *separately*, their converged landmark

results have quite similar fiber connection patterns. This means that the optimization results for the linearly initialized landmarks are quite *reproducible* across subjects, providing direct evidence to the existence of a common brain architecture and support to the connectional fingerprint concept [9].

As a visualization, Fig. 8 shows the movement trajectories of all 2056 landmarks before (blue end) and after (red end) optimization. As shown by red arrows, we can see that some of initial landmarks will move to certain common convergence areas consistently. These cortical areas have more consistent structural connection patterns, and provide the common attraction basins for landmarks initialized within their neighborhood. We will revisit this point in Section 3.3.

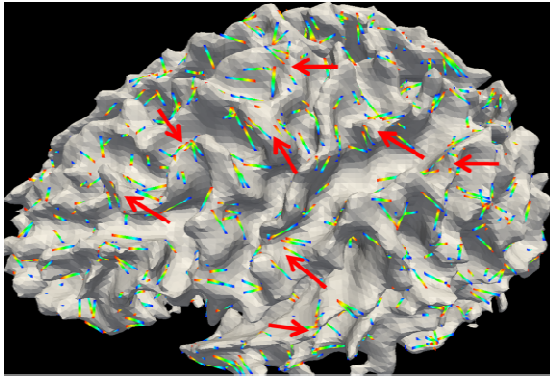


Fig. 8. The trajectories of movements of all 2056 landmarks before (blue end) and after (red end) optimization

3.2 Validation Using Task-Based fMRI

To evaluate the functional consistencies of the optimized landmark locations, we used working memory task-based fMRI data that provided benchmark brain regions of functional consistency to examine the results in Section 3.1. Since the working memory fMRI data provided 16 consistently activated brain regions, we searched the neighborhoods of these functional regions and found that, across different subjects, some optimized landmarks fell into the neighborhoods consistently. This result indicated that optimized locations of landmarks within the working memory regions have consistent functional meanings. As examples, Fig. 9 shows two working memory regions (ROI #10 and ROI #11). Each sub-figure shows a subject and the enlarged view of the neighborhoods (purple circle). Yellow bubbles represent the working memory ROIs. Bubbles with other colors represent optimized landmarks. In particular, we examined two corresponding optimized landmark (red and green bubbles) for functional ROI #10. We can see that both red (grid index=480) and green bubble (grid index=827) fall into the working memory regions consistently across different subjects. In particular, Fig. 10 shows the fiber connection patterns of these two landmarks before and after optimization for the 15 subjects. The brown and purple bubbles on the fiber bundles represent the landmarks before and after optimization, respectively. From the figure, it can be seen that after optimization the

structural connectivity patterns are much more consistent, suggesting the success of our method.

We examined the neighborhoods of all of the 16 working memory ROIs, and found that for each ROI at least 3 landmarks converged within their neighborhood. This result suggests that consistent functional brain regions identified via task-based fMRI tend to be the convergence centers of our landmark optimization procedure, which indicates that there is a deep-rooted regularity of human brain architecture.

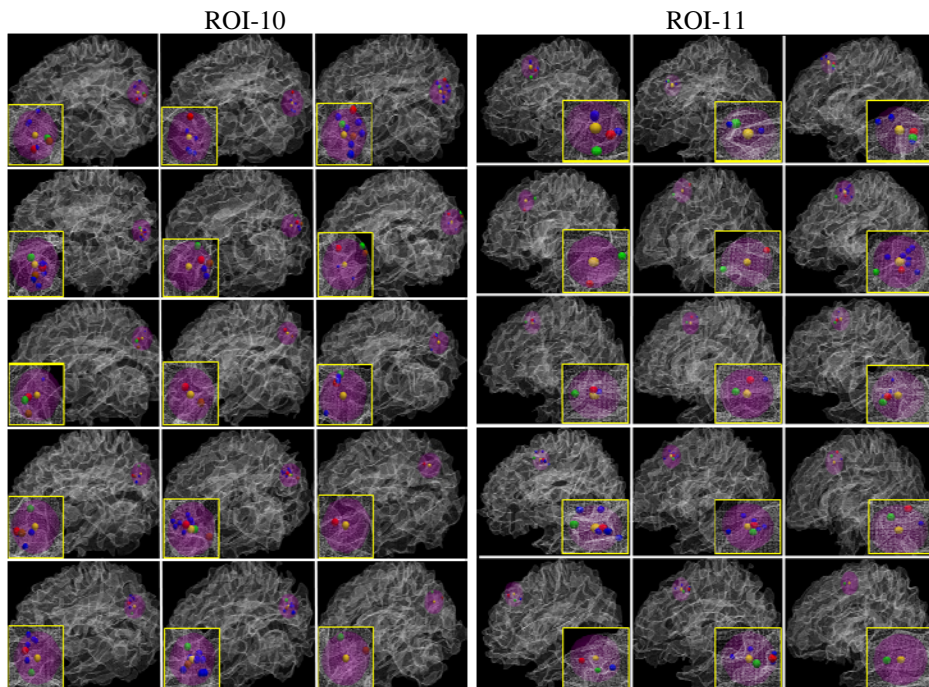


Fig. 9. Illustration of detected functional ROIs via fMRI and optimized landmarks by our methods. Each sub figure showed a subject and the enlarged view of the neighborhood (purple circle). Yellow bubbles represent the ROI. Blue bubbles represent the noise landmarks. Bubbles with other colors represent consistent landmarks. Within one ROI, the bubbles with same colors indicate they have same grid index.

3.3 Sensitivity to Initial Landmark Locations

In our landmark discovery framework, the initialization of landmarks is achieved by linear registration. Here, we examine the sensitivity of the landmark optimization procedure to different initial locations. To do that, we randomly moved the initial landmarks in a range of 3 to 4mm. Then, we repeated our pipeline in Section 2 on two initializations and the results are shown in Fig. 11. Fig. 11a represents a subject before the optimization. Red and green bubbles represent two different initializations. Fig. 11b represents the results after the optimization. We can see that most landmarks

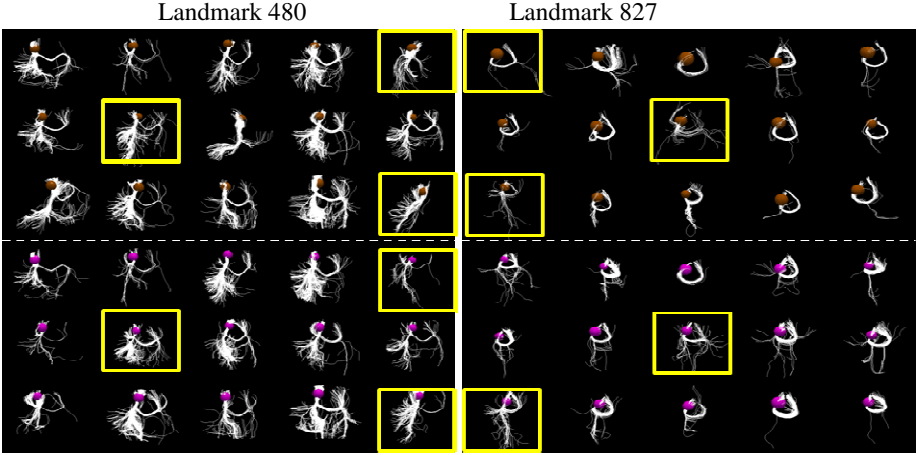


Fig. 10. Optimization results of 15 subjects and 2 optimized landmarks. Each sub figure represents a subject. The fiber bundles with brown bubbles are before the optimization and purple bubbles are after the optimization. Highlighted areas depict landmarks with obvious improvement. For landmark 480, group distances before and after optimization are 0.15 and 0.06. For landmark 827, group distances before and after optimization are 0.10 and 0.07.

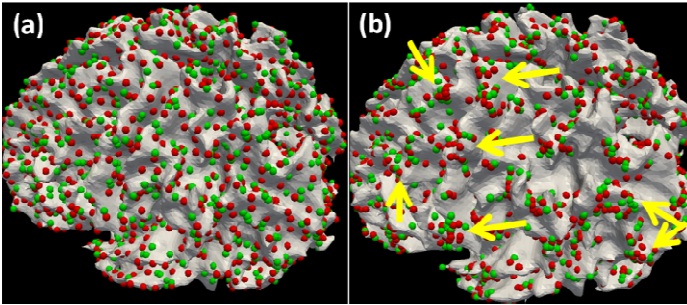


Fig. 11. (a) shows initial landmarks before the optimization. (b) shows optimized landmarks. Red and green bubbles represent landmarks by two different initializations.

moved to certain converging regions as highlighted by the yellow arrows. For quantitatively evaluating the average distance between two different initializations, we defined $D(S, S')$ as the distance between two points distribution S, S' :

$$D(S, S') = \frac{\sum_{i=1}^m Dis_{min}(S_i, S') + \sum_{j=1}^n Dis_{min}(S'_j, S)}{2} \tag{5}$$

where $Dis_{min}(S_i, S')$ is the minimal distance between the i th point in $S(S_i)$ and S' , $Dis_{min}(S'_j, S)$ is the minimal distance between the j th point in $S'(S'_j)$ and S . We selected 5 subjects to calculate the average distance between two initializations and the distance between their optimization results. The average distances before and after

the optimization are 3.35mm and 1.82mm, respectively. This result proves that even different initializations will lead to relatively consistent convergence.

Lastly, the Affinity Propagation (AP) clustering algorithm [18] was applied on the optimized landmarks based on their Euclidean distances and we can detect over 300 reliable clusters across 15 subjects, and each cluster is considered as a reliable brain landmark in this paper. This implies that landmarks with different initializations still converged to consistent cortical regions.

4 Discussion and Conclusion

In this paper, a novel framework for the discovery of dense brain landmarks was presented. Extensive experiments demonstrated the promise of this framework and a validation study via task-based fMRI was provided for a subset of the discovered landmarks. In the near future, we will conduct extensive task-based fMRI studies, in addition to the working memory task-based fMRI in this paper, to functionally label this dense map of brain landmarks. Other future studies may then use the discovered dense map of brain landmarks as the next-generation brain atlas. Importantly, this atlas provides much finer granularity and better functional homogeneity than the Brodmann brain atlas that has been widely used in the brain science field for over 100 years. Also, the dense landmark map can be used as a general reference coordinate system for functional brain mapping via fMRI. That is, with the availability of the dense landmark map, it will be possible to report functional activity detected through fMRI in relation to the identified landmarks. Since the dense landmarks are consistent across different brains with intrinsically established correspondences, the comparison and pooling of results from different fMRI studies become much more effective and efficient. Finally, the dense landmark map will have significant and wide applications in clinical neuroscience. In general, clinical diagnosis, therapy and follow up of numerous brain diseases entail accurate localization of brain landmarks. With the availability of the dense landmark map, morphological, connectional and functional alterations in brain diseases can be reported in relation to the predicted landmarks and be compared with those corresponding measurements in the brains of normal controls. This capability will facilitate numerous significant applications in clinical neuroscience and brain diseases in the future.

References

1. Van Essen, D.C., Dierker, D.L.: Surface-Based and Probabilistic Atlases of Primate Cerebral Cortex. *Neuron*. 56, 209–225 (2007)
2. Derrfuss, J., Mar, R.A.: Lost in localization: The need for a universal coordinate database. *NeuroImage* 48, 1–7 (2009)
3. Ashburner, J., Friston, K., Penny, W.: *Human Brain Function*. Academic Press, London (2004)
4. Friston, K.J.: Modalities, modes, and models in functional neuroimaging. *Science* 326(5951), 399–403 (2009)

5. Jack Jr., C.R., Bernstein, M.A., Borowski, B.J., Gunter, J.L., Fox, N.C., Thompson, P.M., Schuff, N., Krueger, G., Killiany, R.J., Decarli, C.S., Dale, A.M., Carmichael, O.W., Tosun, D., Weiner, M.W.: Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement.* 6(3), 212–220 (2010)
6. Li, Z., Santhanam, P., Coles, C.D., Lynch, M.E., Hamann, S., Peltier, S., Hu, X.: Increased "Default Mode" Activity in Adolescents Prenatally Exposed to Cocaine. *Human Brain Mapping* 32, 759–770 (2010)
7. Epstein, J.N., Casey, B.J., Tonev, S.T., Davidson, M., Reiss, A.L., Garrett, A., Hinshaw, S.P., Greenhill, L.L., Vitolo, A., Kotler, L.A., Jarrett, M.A., Spicer, J.: Assessment and Prevention of Head Motion During Imaging of Patients with Attention Deficit Hyperactivity Disorder. *Psychiatry Res.* 155(1), 75–82 (2007)
8. Li, K., Guo, L., Faraco, C., Zhu, D., Deng, F., Zhang, T., Jiang, X., Zhang, D., Chen, H., Hu, X., Miller, S., Liu, T.: Individualized ROI Optimization via Maximization of Groupwise Consistency of Structural and Functional Profiles. In: *Neural Information Processing Systems, NIPS* (2010)
9. Passingham, R.E., Stephan, K.E., Kotter, R.: The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.* 3(8), 606–616 (2002)
10. Brett, M., Johnsrude, I.S., Owen, A.M.: The problem of functional localization in the human brain. *Nat. Rev. Neurosci.* 3(3), 243–249 (2002)
11. 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. *PNAS* 106(6), 2035–2040 (2009)
12. Faraco, C., Unsworth, N., Langley, J., Terry, D., Li, K., Zhang, D., Liu, T., Miller, S.: Complex span tasks and hippocampal recruitment during working memory. *NeuroImage* 55, 773–787 (2011)
13. Liu, T., Li, H., Wong, K., Tarokh, A., Guo, L., Wong, S.T.C.: Brain Tissue Segmentation Based on DTI Data. *NeuroImage* 38(1), 114–123 (2007)
14. Li, K., Guo, L., Li, G., Nei, J., Faraco, C., Zhao, Q., Miller, S., Liu, T.: Cortical surface based identification of brain networks using high spatial resolution resting state FMRI data. In: *International Symposium on Biomedical Imaging, ISBI* (2010)
15. O'Donnell, L.J., Kubicki, M., Shenton, M.E., Dreusicke, M.H., Grimson, W.E.L., Westin, C.F.: A method for clustering white matter fiber tracts. *American Journal of Neuroradiology* 27, 1032–1036 (2006)
16. Brun, A., Knutsson, H., Park, H.-J., Shenton, M.E., Westin, C.-F.: Clustering fiber traces using normalized cuts. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3216, pp. 368–375. Springer, Heidelberg (2004)
17. FreeSurfer, <http://surfer.nmr.mgh.harvard.edu/>
18. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* 315, 972–976 (2007)

Auto-alignment of Knee MR Scout Scans through Redundant, Adaptive and Hierarchical Anatomy Detection

Yiqiang Zhan, Maneesh Dewan, and Xiang Sean Zhou

Siemens Medical Solutions, USA

Abstract. 3D knee magnetic resonance (MR) scout scan is an emerging imaging sequence that facilitates technicians in aligning the imaging planes of diagnostic high resolution MR scans. In this paper, we propose a method to automate this process with the goal of improving the accuracy, robustness and speed of the workflow. To tackle the various challenges coming from MR knee scout scans, our auto-alignment method is built upon a redundant, adaptive and hierarchical anatomy detection system. More specifically, we learn 1) a *hierarchical redundant* set of anatomy detectors, and 2) ensemble of group-wise spatial configurations across different anatomies, from training data. These learned statistics are integrated into a comprehensive objective function optimized using an expectation-maximization (EM) framework. The optimization provides a new framework for *hierarchical* detection and *adaptive* selection of anatomy primitives to derive optimal alignment. Being extensively validated on 744 clinical datasets, our method achieves high accuracy (sub-voxel alignment error), robustness (to severe diseases or imaging artifacts) and fast speed (~ 5 secs for 10 alignments).

Keywords: MR, Scout scans, Knee, Alignment, Detection, Learning.

1 Introduction

Magnetic resonance imaging (MRI) has been successfully used to diagnose various knee diseases ranging from acute knee injuries to chronic knee dysfunction, e.g., ligament tears, patella dislocation, meniscal pathology and arthropathies, etc [8]. Owing to the inherent imaging physics and speed limitations of MR, the diagnostic MR images have high resolution within slice and low resolution across slices (c.f. Fig. 1 (a)). Hence, the diagnostic ability and quality of MR images highly depends on the accuracy of positioning of the MR slice group (imaging planes), such that the anatomy of interest lies in the imaging plane of high resolution. Recently, a new MR sequence, 3D knee scout scan, has been introduced to improve the quality of MR knee workflow. A 3D knee scout scan has low but **isotropic** image resolutions. Although it is not of diagnostic quality, it provides 3D context for accurate positioning of the high-resolution (high-res) diagnostic slice groups. In the current workflow, the technicians will position

and adjust (rigidly align) the high-res slice groups in accordance with the relevant anatomies using the MR scout scans. The high-res MR knee scans are then done using these manually aligned slice groups. For example, as shown in Fig. 1 (b), for a high-res menisci scan, the transversal imaging slice group should be parallel to the meniscus plane and the transversal in-plane rotation should be aligned to the line connecting the two lower edges of condyle (the yellow dashed line). Since the manual alignment is time consuming and often not reproducible for follow-up studies, an automatic alignment method is desired to increase the speed, accuracy and reproducibility of the MR knee workflow.

However, automatic alignment of MR knee scout scans is not trivial due to the following challenges coming from MR knee scout scans: (1) **Complex transformation:** Compared to other alignment problems, knee transformation is much more complex, including 3D rotation, scaling, and articulation (Fig. 3(c)(e)). This not only increases the dimensions of the parameter space, but also induces additional appearance variations of the same anatomy. (2) **Large variations of intensity and contrast:** Different knee coils and magnetic fields are used in various clinical sites, which causes the intensities of the same anatomy and contrast between different anatomies to vary a lot across MR scout scans. (3) **Abnormality/Variations in knee scout scans:** Besides variations introduced from patient ages and sizes ((Fig. 3(c)(d)), knee diseases, implants and imaging artifacts dramatically change the appearance of anatomies (Fig. 3(f)(g)). Moreover, owing to patient constraints or technician errors, sometimes the key anatomy used for alignment might be out of the field of view.(Fig. 3(a)(b)). (4) **Runtime efficiency requirements:** Auto-align algorithm must have high runtime efficiency as its key target is to speed up the MR knee workflow.

To handle all the aforementioned challenges simultaneously, we propose an iterative alignment method that uses the learned local appearance characteristics and spatial configurations of anatomy primitives from a large set of training dataset. In principle, our method belongs to feature-based alignment/registration category. However, our method has three distinctive properties, which are particularly designed to tackle the challenges of knee MR scout scan alignment.

- **Redundancy:** The learned local appearance cues of distinctive anatomy primitives can not always be reliably detected, especially when they have dramatic appearance changes or are completely out of FOV (*Challenge 2,3*). Therefore, we employ “more than enough” anatomy primitives to derive alignment. For

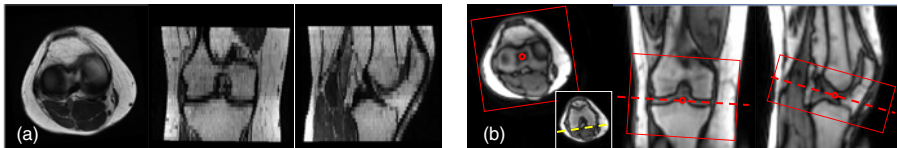


Fig. 1. (a) A typical diagnostic MR knee transversal scan. (b) A MR knee scout scan. Rectangle boxes: coverage of MR slice group. Circles: imaging centers. Yellow dashed line (in the small thumbnail) defines the transversal in-plane rotation.

example, in theory, three non-collinear points can exclusively determine a rigid alignment. In our method, a far greater number of anatomies are used to highly increase the robustness of our approach with respect to the various abnormalities in MR scout scans.

- **Adaptivity:** “Redundant” anatomy primitive detectors are “adaptively” used in different stages of alignment. The anatomy primitives with high transformation-invariant appearance are used in the initial stage, while the ones with larger transformation-sensitive appearances are employed in the later stage. “Adaptivity” makes our method efficient without losing robustness (*Challenge 3,4*).

- **Hierarchy:** We build a hierarchy of detectors with different transformation sensitivities i.e. for each anatomy primitive, we learn detectors with different levels of alignment. At run-time, detectors less sensitive to transformation are used first to estimate an approximate but robust transformation. The detectors more sensitive to transformation (more discriminative) are run at a later stage, and hence increase the accuracy of alignment. By using this “hierarchy” strategy, appearance variations resulting from complex transformation (*Challenge 1*) are gradually removed and accurate detection and alignment can be finally achieved.

The above strategies are realized through a comprehensive objective function. The three terms in this objective function correspond to anatomy primitive-based alignment, anatomy detection using appearance characteristics and spatial constraints of the relative positions of anatomy primitives. An expectation-maximization (EM) framework is employed to optimize the objective function.

2 Related Work

As MR knee scout alignment is a new application in MR workflow, to our best knowledge, there are very few studies on this problem so far [1][5]. Although these pioneer studies open the window for automation of the MR knee workflow, they still have errors around 3-5mm, due to the aforementioned challenges in MR knee scout. At the same time, research on related problems, e.g., brain alignment and face alignment, has been conducted for more than a decade. In this section, we will review these methods and discuss if there are any off-the-shelf methods to solve our problem.

Due to various applications in neuroscience, brain alignment/registration has been extensively studied. In general, brain registration aims to maximize the similarity between two brain images within a specific transformation space. The main core of different methods lies in the definition of the similarity functions, which can be classified into voxel-based [12][3] and feature-based similarities [6][9]. In voxel-based registration methods, information from the entire image is employed to establish a voxel-wise corresponding map across different images. However, in knee scout scan alignment, the alignment of slice group usually needs a reference to only a part of the image and/or a few anatomy primitives. Global image information thus becomes unnecessary and sometimes will decrease alignment accuracy. For example, for knee menisci alignment (see Fig. 1(b)), by using a voxel-based similarity function, it is almost impossible to be immune to

the deformation/artifacts in any regions of scout scans(see Fig. 3 (f)) as these areas contribute to similarities as well. Although a pre-segmentation can help to mask out the relevant anatomies for registration, segmentation itself might not be robust to severe abnormality, either. Feature-based similarity function is more suitable to our problem as it provides the capability to focus on local information. However, in most studies [9], features are defined by low level salient structures, e.g., boundary points, crest lines. As these features do not bear exclusive anatomy definitions, the matching of features can not guarantee the matching of corresponding anatomies. (Note that corresponding anatomies might not always appear as salient structures due to disease and imaging artifacts.) [6] started to use real anatomy landmarks for feature-based registration. However, this method is not fully automatic as its landmark detection requires manual operations. In addition, it is worth noting that a large number of deformable registration methods can not be used in our application, as MR scout scan have to be rigidly aligned and the conversion from deformation fields to rigid transformation is still an open problem.

Face alignment is another related research topic originating from the areas of face detection and recognition, which aims to remove appearance variations coming from different poses to increase the accuracy of face recognition. In [10], face detectors for different poses are exclusively performed and the pose corresponding to the strongest detector response is regarded as the aligned pose. Different machine learning technologies were also applied to solve this problem, e.g., vector boosting [4], etc. In [7], face alignment and detection are tackled altogether through learning a convolution network. Although these methods achieve tremendous success in face alignment, they can not be directly borrowed to our application due to the inherent differences between face and knee alignment. First, knee transformation is much more complex than face. While face transformation is usually modeled as a 2D rotation, knee transformation involves 3D rotation, scaling and articulate transformation. With the higher transformation space/dimension, some existing face alignment methods become computationally infeasible. Second, knee and face alignment have different accuracy requirements. Errors of face alignment are usually allowed to be up to 10 degrees. In contrast, as the accuracy of knee alignment is very critical to the subsequent high-res diagnostic scans, the accuracy of an acceptable knee alignment is much higher (<3mm translation error, <2 degree rotation error).

3 Methods

Formulation: In our method, a MR scout scan is aligned to a canonical space based on a set of automatically detected anatomical primitives/landmarks. A classic landmark-based alignment problem is usually formulated as a least square problem:

$$\min_T \| T\mathbf{m} - \mathbf{p} \|^2 \quad (1)$$

where \mathbf{p} denotes the detected landmarks in a MR scout scan and \mathbf{m} denotes corresponding landmarks in the model space (Both \mathbf{p} and \mathbf{m} are vectors con-

catenated by 3D coordinates of landmarks.) T is the optimal transformation that brings the scout scan to the canonical model space. Since the alignment and detection are highly dependent on each other, in order to estimate them jointly, Eq. 1 should be extended to incorporate landmark detection as:

$$\max_{T, \mathbf{p}} e^{-\|T\mathbf{m} - \mathbf{p}\|^2} + D(\mathbf{p} | I; T) + S(\mathbf{p}) \quad (2)$$

The first additional term, $D(\mathbf{p} | I; T)$, denotes the likelihood of the landmark locations, given image (scout scan) I and transformation T . This term can be derived from anatomy detectors which capture the appearance characteristics of landmarks. It is worth noting that D depends not only on image I but also on transformation T , which is in accordance to the fact that appearance characteristics are influenced by transformations as well. The second additional term $S(\mathbf{p})$ indicates the likelihood of the landmark locations in terms of their spatial relations. In other words, this term incorporates high-level spatial statistics across different anatomies and is exploited to make the alignment robust to erroneous detections. (In Eq. 2, alignment residual has an exponential format to have the same value range as the second and third term.) The current objective function lacks “adaptivity”, which is critical to achieve robust alignment, especially when abnormality exists (diseases, artifacts or bad positioning). Hence, Eq. 2 is extended by inducing a selection matrix \mathbb{C} as:

$$\max_{T, \mathbf{p}, \mathbb{C}} e^{-\|T\mathbb{C}\mathbf{m} - \mathbb{C}\mathbf{p}\|^2} + D(\mathbb{C}\mathbf{p} | I; T) + S(\mathbb{C}\mathbf{p}) \quad (3)$$

Here, \mathbb{C} is a diagonal matrix whose diagonal elements can be either 1 or 0. It indicates whether an anatomy landmark is used for alignment. By inducing \mathbb{C} , three additional properties are introduced into the alignment framework. First, \mathbf{p} can be defined to contain more anatomy landmarks beyond what is “necessary”. For example, although three non-collinear landmarks is enough to derive a rigid transformation, our system exploits much more landmarks (redundancy) to achieve robustness. Second, given different values of \mathbb{C} , the redundant landmarks can be adaptively selected based on appearance (detection confidence) or spatial locations (spatial statistics). Third, during the optimization process, \mathbb{C} is updated dynamically such that detectors with different transformation sensitivities are used hierarchically. Overall, the additional \mathbb{C} in Eq. 3 facilitates the alignment system to be more robust and accurate. Next, we will introduce details of Eq. 3 as well as its optimization method.

Learning-based Anatomy Detection: $D(\mathbb{C}\mathbf{p} | I; T)$ in Eq. 3 indicates the likelihood of the existence of landmarks, which is in fact an anatomy detection problem. Inspired by Viola’s work on face detection [11], we use a learning-based method that starts from a huge set of *elementary* features. A classifier with feature selection capability is then employed to select and combine these elementary features to characterize complex anatomy appearance.

In training, we first annotate landmarks in a set of MR scout scans as positive samples. The voxels “off” of these landmarks are considered as negative samples.

A large number of elementary features are then extracted in the neighborhood of each training sample. Our elementary features are generated by a set of mother functions, $\{H_l(\mathbf{x})\}$, which consists of one or more 3D rectangle functions with different polarities. By scaling the mother functions and convoluting them with the original image, a set of spatial-frequency spaces are constructed. For any voxel $\mathbf{x}_0 \in \mathfrak{R}^3$, its feature vector $\mathfrak{F}(\mathbf{x}_0)$ is finally obtained by sampling these spatial-frequency spaces in the neighborhood of \mathbf{x}_0 as Eq. 4.

$$\mathfrak{F}(\mathbf{x}_0) = \bigcup_l \{F_l(\mathbf{x}_i, s_j) | \mathbf{x}_i \in \mathbb{N}(\mathbf{x}_0), s_{min} < s_j < s_{max}\} \quad (4)$$

where $F_l(\mathbf{x}, s) = H_l(s\mathbf{x}) * I(\mathbf{x})$ denotes the spatial-frequency spaces constructed by the mother functions. These feature vectors are fed into *Adaboost* algorithm [11] to learn appearance signatures of anatomy primitives. The output of the learned classifier $\mathcal{A}(\mathfrak{F}(\mathbf{x}))$ indicates the existence likelihood of a landmark at \mathbf{x} .

Our method has its hallmark in the alignment of training images, which is different from a direct extension of [11]. Since the elementary features are not rotation-invariant, spatial transformations of images will change feature values. Hence, learning appearance characteristics becomes more complicated. One can envision two extreme solutions. First one is to remove this kind of feature variations through alignment of training images. The flip side is that these detectors learned from “perfectly” aligned training samples will be fragile to spatial transformations at run-time, since they never “saw” feature variations resulting from spatial transformations. Another way is to learn local appearance detectors in presence of these transformation variations. This makes the detectors robust to transformation, though the cost one has to pay is detection accuracy. From another perspective, the appearance features of different anatomies exhibit different sensitivity to spatial transformations. As shown in Fig. 2, local appearances around the center of tibia intercondylar eminence (CTIE) is more sensitive to the rotation than that of the center of lateral meniscus (CLM), i.e., with random perturbation, the overlaid images are fuzzier around CTIE (within red rectangles). Thus, all these factors should also be considered in the training framework for generating effective learned detectors.

Therefore, we propose to train hierarchy of detectors using images with different levels of alignment. We first manually align all training images to a common canonical space. Random rigid perturbations are then applied to these aligned training images for feature extraction. Assume the alignment matrix of a training image I_i is T_i , after applying random rigid transformation $r_i \in \mathcal{R}$ (where \mathcal{R} defines the random perturbation range), the construction of spatial-frequency spaces is re-formulated as:

$$F_l(\mathbf{x}, s; T_i, r_i) = H_l(s\mathbf{x}) * (r_i \circ T_i \circ I_i(\mathbf{x})) \quad (5)$$

where \circ is the rigid transformation operator. Elementary features are then sampled in $F_l(\mathbf{x}, s; T_i, r_i)$ using Eq. 4 and used to train a detector.

For the same anatomy, by changing the random perturbation range \mathcal{R} , a set of detectors with different transformation sensitivities can be trained as

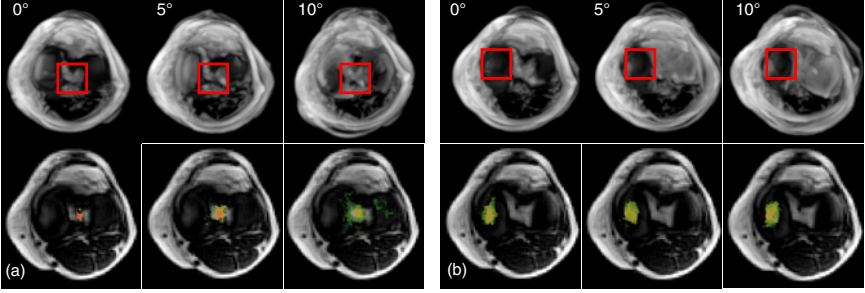


Fig. 2. Training samples with different levels of alignment and their corresponding response maps with respect to different anatomies. (a) CTIE (b)CLM. (Top row) Overlay of 20 training samples using different levels of alignment. (Bottom row) Overlay of response maps from detectors trained with images with different levels of alignment. (From left to right) 0, 5, 10 degrees random rotation after alignment, respectively.

$\{\mathcal{A}(\mathfrak{F}(p); \mathcal{R}_j) | j = 1 \cdots N\}$, where \mathcal{R}_j indicates different random perturbation ranges and $\mathcal{R}_1 \subset \mathcal{R}_2 \subset \cdots \subset \mathcal{R}_N$. Finally, $D(\mathbb{C}\mathbf{p} | I; T)$ in Eq. 3 is defined as the integration of the responses given by these detectors across all anatomy primitives as Eq. 6.

$$D(\mathbb{C}\mathbf{p} | I; T) = \sum_i c_{ii} \sum_j \mathcal{A}_i(\mathfrak{F}(p_i); \mathcal{R}_j) / (\sum_i c_{ii} \sum_j \mathbf{1}) \quad (6)$$

Here, $\mathcal{A}_i(\cdot; \mathcal{R}_j)$ denotes the detector corresponding to the i th anatomy primitive trained from samples with random perturbation range \mathcal{R}_j . c_{ii} is a diagonal element of selection matrix \mathbb{C} , corresponding to the i th anatomy primitive.

Fig 2 (bottom row) shows two interesting characteristics of detectors trained with different levels of alignment. *First*, detectors trained with looser alignment (larger perturbation) show a flatter response map, while detectors trained with a tighter alignment (less perturbation) are more discriminative and provide a peaky response map. *Second*, different anatomies show different sensitivities to spatial transformations. As shown in Fig 2 (a), responses from CLM detectors display similar patterns with variations in the random perturbation range. On the contrary, responses from CTIE detectors (Fig 2 (b)) change from a clear peaky pattern to a noisy flatter pattern with increase in perturbation range. These two characteristics indicate the necessity of a “hierarchical” detection schema to achieve both robustness and accuracy, which is realized in our iterative optimization framework.

Ensemble of Group-wise Spatial Configurations: Imaging artifacts, severe diseases and variations of field of views (FOV) can change local appearance dramatically causing erroneous detections. Hence, it is important to prune these erroneous detections by exploiting spatial correlations between different anatomy primitives (role of $S(\mathbb{C}\mathbf{p})$ in Eq. 3).

Different from the active shape model that learns global shape statistics, we propose to learn group-wise spatial configurations for each anatomy primitive p_i . Assume $U(\mathbf{p} \setminus p_i)$ is a subset of anatomy primitives which does not contain p_i , i.e., $U(\mathbf{p} \setminus p_i) \subset \{\mathbf{p} \setminus p_i\}$. The group-wise spatial configuration between p_i and $U(\mathbf{p} \setminus p_i)$ is modeled as a conditional probability following Gaussian distribution:

$$s(p_i | U(\mathbf{p} \setminus p_i)) = \frac{1}{(2\pi)^{(3/2)} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(p_i - \mu)^T \Sigma^{-1} (p_i - \mu)\right) \quad (7)$$

where μ and Σ are two statistical coefficients that are learned as follows.

We employ a linear model to capture the spatial correlation between p_i and $U(\mathbf{p} \setminus p_i)$, i.e., $p_i = \mathcal{C} \cdot \mathbb{U}$, where \mathbb{U} is a vector concatenated by $\{p_j | p_j \in U(\mathbf{p} \setminus p_i)\}$. Given a set of training samples, the linear correlation matrix \mathcal{C} can be learned by solving a least squares problem. Furthermore, μ and Σ are calculated as:

$$\begin{aligned} \mu &= E[\mathcal{C} \cdot \mathbb{U}] \\ \Sigma &= E[(\mathcal{C} \cdot \mathbb{U} - \mu)(\mathcal{C} \cdot \mathbb{U} - \mu)^T] \end{aligned} \quad (8)$$

In this way, each anatomy primitive p_i receives a set of conditional probabilities from different subsets of $\{U_j(\mathbf{p} \setminus p_i) | j = 1 \dots M\}$ defined as Eq. 7. It is worth noting that a correctly detected p_i might also receive a lower conditional probability from a subset $U(\mathbf{p} \setminus p_i)$ that includes erroneous detections. On the other hand, in most cases, the conditional probabilities received by an erroneously detected p_i are *all* low. (The only exception is that the erroneous detections happen to construct a correct spatial configuration, which very rarely happens.) Therefore, the maximum value of all conditional probabilities received by p_i is used to measure the ‘‘eligibility’’ of the spatial location of the i th anatomy locations. $S(\mathbb{C}\mathbf{p})$ in Eq. 3 is thus defined as:

$$S(\mathbb{C}\mathbf{p}) = \sum_i c_{ii} \max_{|U_j| \leq K} s(p_i | U_j(\mathbf{p} \setminus p_i)) / \sum_i c_{ii} \quad (9)$$

where K is used ($K = 3$ in this study) to constrain the cardinality of U_j .

The advantage of our spatial correlation modeling lies in two aspects. First, instead of learning global spatial statistics, we learn spatial correlations within small groups of anatomies, which are assembled in a ‘‘democratic’’ way at run-time. It makes our method robust to missing or gross detection failures. Second, by constraining the cardinality of U_j and using a linear model, our spatial model will not overfit erroneous detections. In addition, the run-time efficiency is also guaranteed.

Adaptive and Hierarchical Alignment: Since the objective function after incorporating Eq. 6 and 9 becomes high-dimensional, non-linear and non-convex, it is difficult to optimize by traditional approaches. Recall Eq. 3, it has three sets of optimization variables T , \mathbf{p} and \mathbb{C} , where T is the transformation parameters to be estimated and \mathbf{p} and \mathbb{C} are two sets of latent variables. This structure is similar to the maximum-likelihood estimation problem which is often tackled

using expectation-maximization (EM) algorithm. Thereby, our approach optimizes \mathbf{p} , \mathbb{C} and T iteratively following the idea of EM algorithm. The initial T_0 is set as an identity matrix. Afterwards, “expectation stage” (E-stage) and “maximization stage” (M-stage) are performed alternatively until convergence.

E-stage: During E-stage, we aim to optimize the two latent variables \mathbf{p} and \mathbb{C} . \mathbf{p} is first optimized by maximizing $D(\mathbb{C}\mathbf{p} | I; T)$ in Eq. 3. It is realized by invoking learned anatomy detectors in the image with current transformation, i.e., $T(I)$, to generate response maps corresponding to different anatomies. The maximum positions of these response maps are considered as intermediate optimal \mathbf{p} . By setting $c_{ii} = 0$ for p_i with low response and $c_{ii} = 1$ otherwise, we obtain the initial \mathbb{C} , which is further optimized according to Eq. 9. This optimization is performed by iteratively “peeling away” detections that “violate” learned spatial statistics. For each iteration, we calculate the minimum value of $\max_{|U_j| \leq K} s(p_i | U_j(\mathbf{p} \setminus p_i))$ across all p_i with $c_{ii} = 1$. If this value is less than a pre-set threshold, the corresponding p_i will be considered as an erroneous detection and its corresponding c_{ii} is set as zero. This process is repeated until \mathbb{C} converges, i.e., no more landmarks are removed as erroneous detection. After each E-stage, we get intermediate optimization results \mathbf{p} and \mathbb{C} , which will be used in M-stage.

M-stage: In M-stage, the intermediately optimized \mathbf{p} and \mathbb{C} are fixed. The rigid transformation T is then optimized using the close-form solution presented in 2. The derived transformation T will be considered as intermediately optimized transformation and used in the next E-stage.

The three hallmarks of our approach, redundancy, adaptivity and hierarchy are indeed realized in this optimization framework. The implementation of “*redundancy*” is straightforward by including more than enough anatomy primitives in \mathbf{p} . Although some of them might be erroneously detected, they can be effectively pruned through the optimization of \mathbb{C} in E-stage. “*Adaptivity*” and “*hierarchical*” are realized in the iterative optimization procedure. As introduced before, anatomies have appearance characteristics with varying sensitivity to spatial transformation. In the initial iterations, the lower sensitive anatomies (e.g. CLM, c.f. Fig. 2) with high detector responses are more likely to be selected. In the later iterations, as the image alignment improves, transformation-sensitive anatomies start to show higher detector responses and become highly likely to be selected. In this way, different anatomies are “adaptively” selected in the optimization. From another perspective, each anatomy has a hierarchy of detectors with different sensitivities to spatial transformations. In the initial stages, when the image is far from the well aligned position, detectors with looser alignment (more perturbation) exhibit higher responses and dominate $D(\mathbb{C}\mathbf{p} | I; T)$. With the iterations, as the image becomes well aligned, detectors with more rigorous alignment start to generate higher responses as well and thereby contribute more to $D(\mathbb{C}\mathbf{p} | I; T)$. Hence, detectors with different levels of alignment are “hierarchically” employed to gradually remove the appearance variations resulting from transformation. These hallmarks thus make our algorithm highly accurate and robust.

4 Experiments

In clinical workflow, our auto-alignment is employed to generate various high-res slice groups relevant to different anatomy parts, e.g., meniscus, patella, cruciate ligaments, etc. Due to space limitation, in this section, we focus on experimental results of knee meniscus slice group, which is the most frequently used for high-res MR imaging. Our method is general and achieves similar performance on other slice boxes. As shown in Fig. 1 (b), the alignment of knee meniscus slice group is defined by the meniscus plane (norm of the slices) and the lower edge of the femur condyle in transversal view (slice in-plane rotation).

Our experimental data includes 924 knee scout scans (180 for training and 744 for testing) with isotropic resolution $2mm$. These datasets come from different clinical sites and were generated by different types of Siemens MR Scanners (Avanto, Verio, etc.) with different knee coils (KN, K15, EX, etc.), which induces additional appearance variations and makes the auto-alignment even more challenging. We validate our method in two aspects, robustness and accuracy.

Robustness: A clinically feasible auto-alignment algorithm has to be robust to the variations not only due to age, gender, demography, but also coming from disease and imaging artifacts. In this study, we test our method on 744 knee scout scans. The alignment results are visually checked and evaluated by experienced experts as “perfect”, “acceptable” (clinically accepted but can be improved) and “gross failure” (obvious alignment errors). Table 1 shows the comparison of the performance of our method with other three methods. **Seg:** an in-house benchmark system that is mainly based on heuristic segmentation and detection. **Reg:** a mutual information-based registration method and **Det:** our method without using learned spatial configurations which can be considered as a direct extension of Viola’s face detection method on our application. As shown in table 1, **Det** is better than **Seg** and **Reg** as local information helps in robustness. In our method, the robustness is further increased by incorporating ensemble of group-wise spatial configurations. Some representative challenging cases are shown in Fig. 3. Note that even in these challenging cases our algorithm does the knee meniscus alignment very well, i.e. the central transversal slice in the aligned volume passes through the menisci and the condyles are aligned (small thumbnail image in each subfigure).

Accuracy: We also perform quantitative evaluation of the accuracy of our method. In this study, we randomly pick 50 representative clinical cases covering

Table 1. Robustness comparisons of different auto-slice-positioning methods on knee meniscus slice group. GF: Gross Failure. AC: Acceptable. PF: Perfect.

	Seg	Reg	Det	Proposed Method
GF	95 (12.7%)	78 (10.5%)	20 (2.7%)	1 (0.1%)
AC	58 (7.8%)	80 (9.4%)	11 (1.5%)	7 (1.0%)
PF	591 (79.4%)	596 (80.1%)	713 (95.8%)	736 (98.9%)

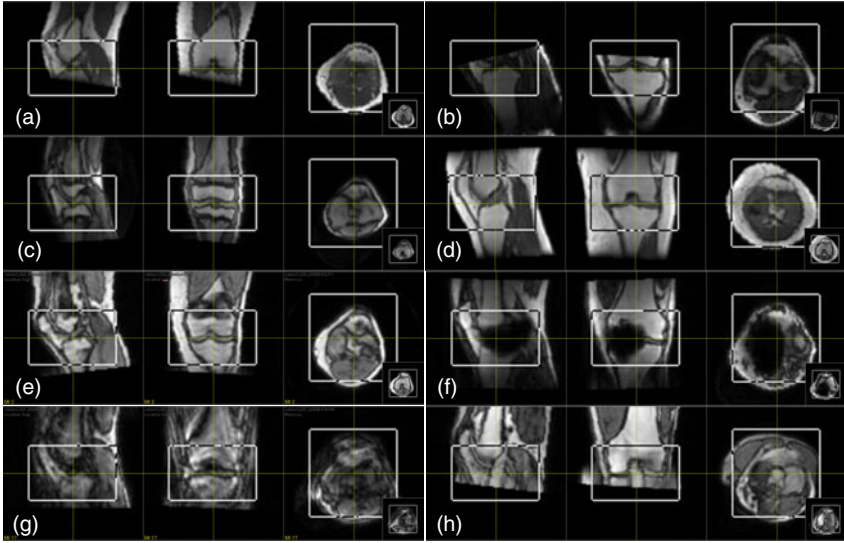


Fig. 3. Robust alignment of our algorithm on representative challenging cases. Each subfigures show the 3 orthogonal MPRs of the aligned volume at the center of the knee meniscus slice box: (a) tibia out of FOV, (b) femur out of FOV, (c) a pediatric patient, (d) an old patient, (e) a patient with bone disease, (f) a patient with metal implant, (g) motion artifacts and (h) folding artifacts.

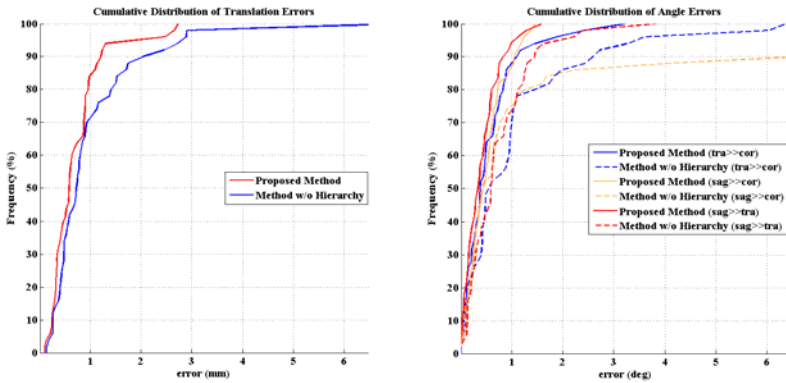


Fig. 4. Cumulative distribution of alignment errors using different alignment methods

different genders, ages and knee sizes. These 50 cases are manually aligned and used as ground truth to derive the quantitative errors. For comparison, we also implemented a method which is similar to the proposed one but without hierarchical detection and alignment, i.e., \mathbb{C} in Eq. 3 is omitted. (We did not compare with **Seg**, **Reg** and **Det**, as gross failure cases make their accuracy numbers

much worse.) Our method has the root mean square error 0.93mm(translation), 0.83° (tra >> cor), 0.54° (sag >> tra), 0.63° (sag >> cor). Our method actually achieves sub-voxel accuracy. The method without hierarchical detection and alignment has larger errors 1.52mm (translation), 1.72° (tra >> cor), 1.02° (sag >> tra), 3.30° (sag >> cor), respectively. More detailed quantitative errors, cumulative error distributions of these two methods, are shown in Fig. 4.

Runtime efficiency: Running on a computer with Dual-Core AMD Opteron (tm) Processor 2212 (2GHz), our method is able to generate 10 alignment for different knee anatomies in 5 seconds. It satisfies the requirement of MR knee workflow.

5 Conclusions

In this paper, we proposed an auto-alignment method for MR knee scout scans, which exploits redundant, adaptive and hierarchical strategies to detect anatomy primitives and align scout scans accordingly. Our method is highly robust, accurate and fast with extensive demonstration on 700+ dataset. It paves the way to automate the positioning of slice boxes for high-res MR knee scanning and improve the MR knee workflow efficacy. It comes as no surprise that the proposed method has been integrated into the latest Siemens MR Scanners.

Acknowledgments. The authors would like to thank Martin Harder for a lot of valuable discussions and his data collection efforts.

References

1. Bystrov, D., Pekara, V.: Automated planning of mri scans of knee joints (2007)
2. Eggert, D.W.: Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications* 9, 272–290 (1997)
3. Glocker, B., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. *Medical Image Analysis* 12, 731–741 (2008)
4. Huang, C., Ai, H., Li, Y., Lao, S.: Vector boosting for rotation invariant multi-view face detection. In: *ICCV* (2005)
5. Jolly, M., Guehring, J.: Automatic Femur Segmentation and Condyle Line Detection in 3D MR Scans for alignment of High Resolution MR. In: *ISBI* (2010)
6. Joshi, A., Shattuck, D., Thompson, P., Leahy, R.: Surface-constrained volumetric brain registration using harmonic mappings. *IEEE TMI* 26, 1657–1669 (2007)
7. Osadchy, M., Cun, Y., Miller, M.: Synergistic face detection and pose estimation with energy-based models. *J. of Machine Learning Research* 8, 1197–1215 (2007)
8. Ostlere, S.: Imaging the knee. *Imaging* 15, 217–241 (2007)
9. Pennec, X., Ayache, N., Thirion, J.P.: Landmark-based registration using features identified through differential geometry. Academic Press, London (2008)
10. Schneiderman, H., Kanade, T.: A statistical method for 3d object detection applied to faces and cars. In: *CVPR* (2000)
11. Viola, P., Jones, M.J.: Robust real-time face detection. *International Journal of Computer Vision* 57, 137–154 (2004)
12. Wells, W., Viola, P., Atsumid, H., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis* 1, 35–51 (1996)

Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis

Sandy Durrleman, Marcel Prastawa, Guido Gerig, and Sarang Joshi

SCI Institute, University of Utah, 72 S. Central Campus Dr.,
UT-84112 Salt Lake City

Abstract. In this paper, we propose a novel approach for intensity based atlas construction from a population of anatomical images, that estimates not only a template representative image but also a common optimal parameterization of the anatomical variations evident in the population. First, we introduce a discrete parameterization of large diffeomorphic deformations based on a finite set of control points, so that deformations are characterized by a low dimensional geometric descriptor. Second, we optimally estimate the position of the control points in the template image domain. As a consequence, control points move to where they are needed most to capture the geometric variability evident in the population. Third, the optimal number of control points is estimated by using a $\log -L^1$ sparsity penalty. The estimation of the template image, the template-to-subject mappings and their optimal parameterization is done via a single gradient descent optimization, and at the same computational cost as independent template-to-subject registrations. We present results that show that the anatomical variability of the population can be encoded efficiently with these compact and adapted geometric descriptors.

1 Introduction

Fundamental to Computational Anatomy is the estimation of the template and the template-to-subject mappings that characterize anatomical variability in a population. For statistical analysis of a set of images it is crucial to efficiently parametrize the anatomical variability. One approach consists of extracting features of the mappings, like the Jacobian matrix of the deformation [3], or its determinant [9]. More comprehensive approaches rely on the statistical analysis of the displacement fields of the voxels grid by using a log-Euclidean technique [2], or on the Riemannian characterization of the group of diffeomorphisms parametrized via the continuous initial momenta map in the LDDMM setting [14].

The intrinsic problem with such approaches is that the fundamental anatomical variability is parametrized in an infinite dimensional space (practically on the order the size of the imaging modality), which does not reflect the intrinsic dimensionality of the anatomical variations in a finite database of images. Indeed, the estimated deformations are usually constrained to be spatially smooth, thus preventing every voxel from moving independently. Moreover, the need for deformation is not equally distributed over the domain, since intensity-based registrations

are mostly driven by the level sets of the template image. As a consequence, we hypothesize that the anatomical variability can be characterized by a compact geometric parametrization of much smaller dimension. It has already been shown in [7] that smooth vector fields parameterizing diffeomorphisms can be efficiently approximated by a small number of momenta via a “matching pursuit” technique, where momenta stands for vectors attached to control points.

In this paper we propose a control point parameterization of large deformation diffeomorphisms following [10] to drive template-to-subject image registrations. This differs from LDDMM image registration, for which the deformation is parameterized by a continuous map of momenta that are always parallel to the image gradient [12]. Here, we propose to use a finite set of momenta, which are not constrained in their direction. Control points techniques have been widely used for small deformation transformations [8], its use for large deformation matching of images is challenging. In [13], diffeomorphisms were built by a composition of small B-splines transforms without a comprehensive variational formulation. In [1], diffeomorphisms were characterized via a finite set of initial momenta located at the vertices of a “texture mesh”, but no attempt was made to estimate an optimal mesh describing a whole population of images. The inherent difficulty is to find an efficient way to transport information back and forth from source to target. Indeed, control points flow from source to target (via the deformation ϕ), whereas the variability in the population is in the target image domain and hence needs to be pulled back to the source to build the template ($I_{\text{src}} \circ \phi^{-1}$). We solve this issue via a new formulation of the dynamical system which drives the LDDMM registration, borrowed from optimal control theory. It enables us to easily transport points and vectors via simple integration of ODE. One of the striking results of this formulation is that the optimal positions of the control points in the template space can be found *at no additional computational cost*. We demonstrate that the control points are naturally attracted by the contours of the template image, and a regularity term optimizes their spatial distribution to avoid redundancy in the parameterization. The number of control points determines the number of degrees of freedom of the deformations and therefore the accuracy of the template-to-subject deformations and the sharpness of the atlas. To optimize the number of control points for a given atlas sharpness, we use a log L^1 -penalty term on the set of initial momenta, in the spirit of the in vogue sparse statistical methods. Our results show that this prior enables to prune the set of control points without sacrificing the description of the anatomical variability. This sparse and adapted parameterization of the variability seems therefore adapted to the statistical analysis of a collection of images.

We follow the now well established paradigm for atlas estimation by first defining pairwise image matching and then using it for population analysis.

2 Image Matching with Discrete Parameterization of Deformations

Image matching term. Let I_{src} and I_{tar} two images to be matched and ϕ a diffeomorphism of the image domain. The source image is deformed into I_1 via

$I_{\text{src}} \circ \phi^{-1}$. Let $\mathbf{y} = (y_1, \dots, y_M)$ be the concatenation of the location of the voxels in the target image. The deformed image I_1 linearly interpolates the gray levels of the neighboring voxels around positions $\phi^{-1}(y_k)$. We denote $\mathbf{y}(0) = \phi^{-1}(\mathbf{y})$.

The deformation estimation is driven by the L^2 norm between the images:

$$\|I_{\text{src}} \circ \phi^{-1} - I_{\text{tar}}\|^2 = \sum_{k=1}^M (I_{\text{src}}(y_k(0)) - I_{\text{tar}}(y_k))^2 = A(\mathbf{y}(0)), \tag{1}$$

which depends only on the positions $\mathbf{y}(0)$, since the target voxel positions \mathbf{y} are fixed on a regular lattice.

Non-linear diffeomorphic deformations. A standard way to construct non-linear diffeomorphisms is to integrate infinitesimal transformation parametrized via a time-varying vector field $v_t(x)$ over the time interval $[0, 1]$: $\dot{\phi}_t(x) = v_t(\phi_t(x))$, with initial condition $\phi_0(x) = x$. Under the conditions detailed in [4] and satisfied here, the resulting $(\phi_t)_{t \in [0,1]}$ is a flow of diffeomorphisms (for each time $t \in [0, 1]$, ϕ_t is a diffeomorphic deformation).

Let $\mathbf{c}_0 = \{c_1, \dots, c_N\}$ be a finite set of control points. These points move in space according to the deformation: $c_i(t) = \phi_t(c_i)$. By analogy with landmark matching [10], we parametrize the instantaneous velocity field $v_t(x)$ by a set of *time-varying* vectors $\boldsymbol{\alpha}(t) = \{\alpha_1(t), \dots, \alpha_N(t)\}$ as:

$$v_t(x) = \sum_{i=1}^N K(x, c_i(t))\alpha_i(t), \tag{2}$$

where K is an interpolating kernel, assumed hence forth without loss of generality, to be Gaussian: $K(x, y) = \exp(-|x - y|^2/\sigma^2)$.

The positions of the control points $c_i(t)$ depend on the velocities and therefore on the vectors $\boldsymbol{\alpha}(t)$. They satisfy a set of N coupled ODEs: $\dot{c}_i(t) = \dot{\phi}_t(c_i) = v_t(c_i) = \sum_{j=1}^N K(c_i(t), c_j(t))\alpha_j(t)$, with initial condition at $t = 0$: $c_i(0) = c_i$. In a matrix form, this can be written as: $\dot{\mathbf{c}}(t) = f(\mathbf{c}(t), \boldsymbol{\alpha}(t))$ with $\mathbf{c}(0) = \mathbf{c}_0$.

Once these positions have been estimated, the motion of any point x_0 is computed by solving the ODE: $\dot{x}(t) = v_t(x)$, $x(0) = x_0$. It follows that the flow of diffeomorphisms is entirely determined by the time-varying vectors $\boldsymbol{\alpha}(t)$ and the initial control points \mathbf{c}_0 .

In particular, the points $\phi_1^{-1}(y_k)$ in the data term are computed by flowing the positions y_k from $t = 1$ (target space) back to time $t = 0$ (source space): one integrates backward the ODE: $\dot{y}_k(t) = v_t(y_k) = \sum_{i=1}^N K(y_k(t), c_i(t))\alpha_i(t)$ with the final condition $y_k(1) = y_k$. This equation in matrix form becomes $\dot{\mathbf{y}} = g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))$, $\mathbf{y}(1) = \mathbf{y}$. The solution at time $t = 0$ is $\mathbf{y}(0) = \phi_1^{-1}(\mathbf{y})$.

One defines the regularity of the deformation as its total kinetic energy: $\int_0^1 \|v_t\|_V^2 dt = \int_0^1 \sum_{i=1}^N \sum_{j=1}^N \alpha_i(t)^t K(c_i(t), c_j(t))\alpha_j(t) dt$, using the Sobolev norm of the velocity field associated with the kernel K . We write it as $\int_0^1 L(\mathbf{c}(t), \boldsymbol{\alpha}(t)) dt$.

Criterion minimization. Now, the matching criterion can be written as:

$$E(\mathbf{c}_0, \boldsymbol{\alpha}(t)) = A(\mathbf{y}(0)) + \gamma \int_0^1 L(\mathbf{c}(t), \boldsymbol{\alpha}(t)) dt \tag{3}$$

subject that:

$$\begin{aligned}\dot{\mathbf{c}}(t) &= f(\mathbf{c}(t), \boldsymbol{\alpha}(t)) & \mathbf{c}(0) &= \mathbf{c}_0 \\ \dot{\mathbf{y}}(t) &= g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t)) & \mathbf{y}(1) &= \mathbf{y}\end{aligned}\quad (4)$$

where γ is the scalar trade-off between the regularity of the deformation and the fidelity to data. The minimization of this criterion with respect to the time-varying vectors $\boldsymbol{\alpha}(t)$ leads to the source-to-target deformation parameterized by the control points \mathbf{c}_0 . The minimization with respect to the \mathbf{c}_0 leads to the optimal positions of the control points which parameterize the best matching possible. Both optimizations will be done via a single gradient descent.

It has been shown that the regularity term for $\gamma > 0$ ensures that the flow of diffeomorphisms builds a geodesic path in a certain group of diffeomorphisms [4]. This means that at the minimum the flow of diffeomorphisms is entirely determined by the initial momenta at time $t = 0$: $(\mathbf{c}_0, \boldsymbol{\alpha}(0))$.

We optimize the criterion via a gradient descent. The initial conditions are $\boldsymbol{\alpha}(t) = 0$ for all t (which gives $\phi_t(x) = x$, i.e. no deformation) and the control points \mathbf{c}_0 are given on a regular lattice. A variation of the momenta $\delta\boldsymbol{\alpha}(t)$ and of the control points initial position $\delta\mathbf{c}_0$ induces a variation of the whole path of the control points $\delta\mathbf{c}(t)$ from source ($t = 0$) to target ($t = 1$). Then, this changes the path of the target voxels flowing back to the source: it leads to a variation of the positions $\delta\mathbf{y}(0)$ and hence a variation of the criterion. We show in Appendix A that the gradient of the criterion with respect to these two variables is:

$$\begin{cases} \nabla_{\boldsymbol{\alpha}} E(t) = 2\gamma\boldsymbol{\alpha}(t) + \boldsymbol{\eta}^c(t) + \tilde{\boldsymbol{\eta}}^y(t) \\ \nabla_{\mathbf{c}_0} E = \boldsymbol{\eta}^c(0) \end{cases}\quad (5)$$

where $\boldsymbol{\eta}^c$ and $\tilde{\boldsymbol{\eta}}^y$ are $3N$ dimensional vectors, which satisfy one forward and one backward integral equations:

$$\begin{aligned}\eta_p^y(t) &= -\nabla_{y_p(0)} A + \int_0^t \sum_{q=1}^N \frac{2}{\sigma^2} K(y_p(s), c_q(s)) \alpha_q(s)^t \eta_p^y(s) (y_p(s) - c_q(s)) ds \\ \eta_i^c(t) &= -\int_t^1 \sum_{j=1}^N \frac{2}{\sigma^2} K(c_i(s), c_j(s)) \left(\alpha_j(s)^t \eta_i^c(s) + \alpha_i(s)^t \eta_j^c(s) + 2\gamma \alpha_i(s)^t \alpha_j(s) \right) (c_i(s) - c_j(s)) \\ &\quad - \sum_{k=1}^M \frac{2}{\sigma^2} K(c_i(s), y_k(s)) \alpha_i(s)^t \eta_k^y(s) (c_i(s) - y_k(s)) ds \\ \nabla_{y_k(0)} A &= 2 \left(I_0(y_k(0)) - I_1(y_k) \right) \nabla_{y_k(0)} I_0\end{aligned}$$

and $\tilde{\boldsymbol{\eta}}^y$ is the solution of the set of N linear equations for all t :

$$\sum_{j=1}^N K(c_i(t), c_j(t)) \tilde{\eta}_j^y(t) = \sum_{k=1}^M K(c_i(t), y_k(t)) \eta_k^y(t)\quad (6)$$

The auxiliary variable $\boldsymbol{\eta}^y$ is computed forward in time: it transports the usual image force $\nabla_{y_k(0)} A$ from the source ($t = 0$) to the target space ($t = 1$). Once at time $t = 1$, this variable is used as a source term in the ODE satisfied by the

Algorithm 1. Image Matching with finite-dimensional parameterization

-
- 1: $\alpha_i(t) \leftarrow 0$ for all $i = 1, \dots, N$ and all t
 - 2: $\mathbf{c}_0 \leftarrow$ initial positions of control points (input)
 - 3: **repeat** {Gradient descent}
 - 4: {Compute path of control points (forward integration)}
 - 5: $c_i(t) = c_i(0) + \int_0^t \sum_{j=1}^N K(c_i(s), c_j(s)) \alpha_j(s) ds$
 - 6: {Compute deformed source image (backward integration)}
 - 7: $y_k(t) = y_k(1) - \int_t^1 \sum_{j=1}^N K(y_k(s), c_j(s)) \alpha_j(s) ds$
 - 8: {Compute gradient of source image}
 - 9: $\nabla_{y_k(0)} E = 2 \left(I_0(y_k(0)) - I_1(y_k) \right) \nabla_{y_k(0)} I_0$
 - 10: {Compute auxiliary variable η^y (forward integration)}
 - 11: $\eta_p^y(t) = -\nabla_{y_p(0)} E - \int_0^t \sum_{q=1}^N \alpha_q^t(s) \eta_p^y(s) \nabla_1 K(y_p(s), c_q(s)) ds$
 - 12: {Compute auxiliary variable η^c (backward integration)}
 - 13: $\eta_i^c(t) = \int_t^1 \sum_{j=1}^N \left(\alpha_j(s)^t \eta_i^c(s) + \alpha_i(s)^t \eta_j^c(s) + 2\gamma \alpha_i(s)^t \alpha_j(s) \right) \nabla_1 K(c_i(s), c_j(s)) +$
 $\sum_{k=1}^M \alpha_i(s)^t \eta_k^y(s) \nabla_2 K(y_k(s), c_i(s)) ds$
 - 14: {Solve the linear system}
 - 15: $\sum_{j=1}^N K(c_i(t), c_j(t)) \tilde{\eta}_j^y(t) = \sum_{k=1}^M K(c_i(t), y_k(t)) \eta_k^y(t)$ ($\tilde{\eta}^y$ is of dimension $3N$)
 - 16: {Compute gradient}
 - 17: $\nabla_i J(t) = 2\gamma \alpha_i(t) + \eta_i^c(t) + \tilde{\eta}_i^y(t)$
 - 18: {Update time-varying momenta}
 - 19: $\alpha_i(t) \leftarrow \alpha_i(t) - \varepsilon \nabla_{\alpha_i} J(t)$
 - 20: {Update initial positions of control points}
 - 21: $c_i(0) \leftarrow c_i(0) - \varepsilon \eta_i^c(0)$
 - 22: **until** Convergence
-

variable η^c , which is integrated backward in time. At time $t = 0$, this variable is used to update the position of the control points in the source image domain, at no additional cost. The overall gradient descent is summarized in Algorithm [1](#).

Remark 1 (Interpretation in the small deformation setting). To better understand these equations, we linearize this model in time. Then, the flow of diffeomorphisms is reduced to the transform: $\phi(x) = x + v(x)$ parameterized by the fixed momenta $(\mathbf{c}_0, \alpha(0))$. For small deformations, the inverse is approximated by $\phi^{-1}(y_k) = y_k - v(y_k)$. The matching criterion becomes:

$$E(\mathbf{c}_0, \alpha) = \left\| I_{\text{src}} \circ \phi^{-1} - I_{\text{tar}} \right\|^2 + \gamma \|v\|_V^2, \quad (7)$$

whose gradient can be computed straightforwardly as:

$$\begin{aligned} \frac{1}{2} \nabla_{\alpha_i} E &= - \sum_{k=1}^M K(c_i, y_k) (I_{\text{src}}(y_k - v(y_k)) - I_{\text{tar}}(y_k)) \nabla_{y_k - v(y_k)} I_{\text{src}} + \gamma \sum_{j=1}^N K(c_i, c_j) \alpha_j \\ \frac{1}{2} \nabla_{c_i} E &= \sum_{k=1}^M \frac{2}{\sigma^2} K(c_i, y_k) (I_{\text{src}}(y_k - v(y_k)) - I_{\text{tar}}(y_k)) (\nabla_{y_k - v(y_k)} I_{\text{src}})^t \alpha_i (c_i - y_k) \\ &\quad - \gamma \sum_{j=1}^N \frac{2}{\sigma^2} K(c_i, c_j) \alpha_i^t \alpha_j (c_i - c_j) \end{aligned}$$

This is exactly the linearization of the gradient (5), at order 0 for the first equation and at order 1 for the second one (the zeroth order vanishing).

The first equation consists of two terms: the first one is the convolution of the usual image force at the control points, the second one is a regularizer of the estimated momenta, which can be seen as a low-pass filter on the momenta. The second equation is the update rule for the control points positions. The first term shows that they are attracted by the voxels where the gradient of the image is large (i.e. the contours), provided that the momenta α_i pushes in the ‘right’ direction, that of the image force (making the dot product positive). The second term is a repulsion term which moves away two control points which carry momenta pointing in the same direction, thus limiting the redundancy of the parameterization at the scale of the kernel σ .

3 Atlas Estimation with Common Sparse Parameterization of Deformations

3.1 Joint Optimization of the Template Image and the Deformations

Given a set of N_s images, one wants to find a template image I_0 , the optimal set of control points \mathbf{c}_0 in the image domain and the optimal set of N_s time-varying vectors $\alpha_s(t)$, which drive the registration of the template to every image. As a consequence, the criterion to be minimized is given by:

$$E(I_0, \mathbf{c}_0, \alpha_1(t), \dots, \alpha_{N_s}(t)) = \sum_{s=1}^{N_s} \left\{ A_s(\mathbf{y}_s(0)) + \gamma \int_0^1 L(c_s(t), \alpha_s(t)) ds \right\} \quad (8)$$

subject that:

$$\begin{aligned} \dot{\mathbf{c}}_s(t) &= f(\mathbf{c}_s(t), \alpha_s(t)) & \mathbf{c}_s(0) &= \mathbf{c}_0 \\ \dot{\mathbf{y}}_s(t) &= g(\mathbf{y}_s(t), \mathbf{c}_s(t), \alpha_s(t)) & \mathbf{y}_s(1) &= \mathbf{y} \\ A_s(\mathbf{y}_s(0)) &= \|I_0(\mathbf{y}_s(0)) - I_s\|^2 \end{aligned} \quad (9)$$

where we notice that the initial and final condition of the ODEs are shared among the population. This criterion has the same form as in (3). Denoting E_s each term of the sum, the gradient with respect to \mathbf{c}_0 and the $\alpha_s(t)$ is given as:

$$\nabla_{\alpha_s(t)} E = \nabla_{\alpha_s(t)} E_s \quad \nabla_{\mathbf{c}_0} E = \sum_{s=1}^{N_s} \nabla_{\mathbf{c}_0} E_s \quad (10)$$

This means that the weighting vectors are computed for N_s parallel registrations by integrating the ODEs of (5). After the back and forth integrations, the auxiliary variables $\eta_s^c(0)$ are averaged over the subjects to update the common parameterization of the deformations \mathbf{c}_0 .

The gradient with respect to the template image I_0 is the sum of the gradient $\nabla_{I_0} A_s(\mathbf{y}_s(0))$. The value $I_0(y_k(0))$ is computed as $\sum_{p \in \mathcal{N}(y_k(0))} \rho_p(y_k(0)) I_0(\pi_p(y_k(0)))$, where $\mathcal{N}(y)$ denotes the set of 8 neighboring voxels $\pi_p(y)$ of the position y and $\rho_p(y)$ the weights corresponding to the trilinear interpolation. Let

R_s be the residual image $I_0(\mathbf{y}(0)) - I_s$. Then a variation δI_0 of the template image leads to:

$$\begin{aligned} \frac{1}{2}\delta A(\mathbf{y}(0)) &= \sum_i R_s(y_i) \sum_{p \in \mathcal{N}(y_i(0))} \rho_p(y_i(0)) \delta I_0(\pi_p(y_i(0))) \\ &= \sum_j \left(\sum_{\{i; \exists k, \pi_k(y_i(0))=y_j\}} \rho_k(y_i(0)) R_s(y_i) \right) \delta I_0(y_j) = \sum_j \nabla_{I_0} A_s(y_j) \delta I_0(y_j) \end{aligned}$$

where one multiplies the gray levels for each voxel in the source domain instead of the target domain. This shows that the gradient of A is the partial volume projection [6] of the residual image. This is computed by flowing the voxel y_k back in the source domain and distributing its gray level at the 8 neighboring voxels with the same weight as for a trilinear interpolation. The accumulation of such values for every voxel and every subject gives the gradient.

3.2 Sparsity Enforced by $\log -L^1$ Penalty Term

The number of control points determines the number of degrees of freedom of the deformations. The more control points, the more accurate the registrations, the sharper the template image. Consequently, the optimization of the previous criterion tends to use as many control points as possible. However, as we will see in our experiments, from a certain point, adding more momenta only marginally increase the atlas sharpness at the cost of adding much more noise in the description of the variability. To find the optimal number of degrees of freedom for an accurate description of the variability, we use a numerically stabilized $\log -L^1$ penalty motivated by [5] :

$$E = \sum_{s=1}^{N_s} \left\{ A_s(\mathbf{y}_s(0)) + \gamma \int_0^1 L(c_s(t), \boldsymbol{\alpha}_s(t)) ds + \gamma_{\text{sp}} \sum_{i=1}^N \log^c (\|\boldsymbol{\alpha}_i^s(0)\|) \right\} \quad (11)$$

where N denotes the total number of control points, $\boldsymbol{\alpha}_i^s$ the i th initial momentum vector of the s th subject. For numerical stability we truncate the log function near 0 via: $\log^c(x) = \max(\log(x), \log(c)) - \log(c)$ a positive penalty. This penalty function has almost no effect on large momenta, whereas it enforces the small momenta to converge to zero. The constant c is a threshold to avoid numerical instability, set typically at the voxel size. Every 5 iterations of the gradient descent, control points with an initial momenta smaller than c are pruned.

This penalty induces only a small change in the algorithm. The quantity $\gamma_{\text{sp}} \boldsymbol{\alpha}_i^s / \|\boldsymbol{\alpha}_i^s\|^2$ is added to each $\nabla_{\boldsymbol{\alpha}_i} E^s$ as soon as $\|\boldsymbol{\alpha}_i^s\| > c$.

4 Experiments

Our method is independent of the dimension of the images. In this section, we focus on 2D images for a better visualization and understanding of the results.

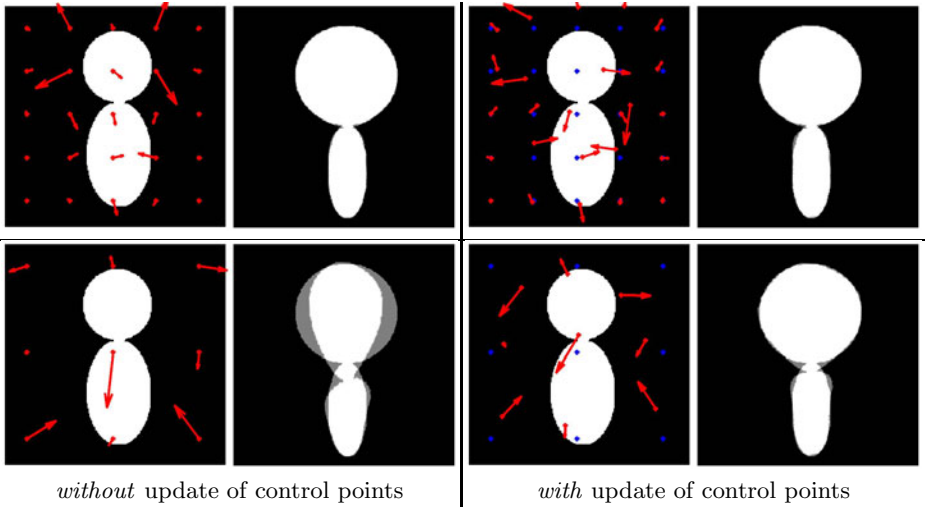


Fig. 1. Synthetic image matching with 25 (top) and 9 (bottom) control points. On the left of each panel: the source image with the initial momenta (red arrows). On the right the superimposition of the deformed source and target image. First row shows that a *discrete* parameterization is sufficient for a perfect matching. Second row shows that moving the control points to their optimal positions gives a much better representation of the shape differences for a fixed number of parameters. $\sigma = 50$ voxels, $\gamma = 10^{-2}$.

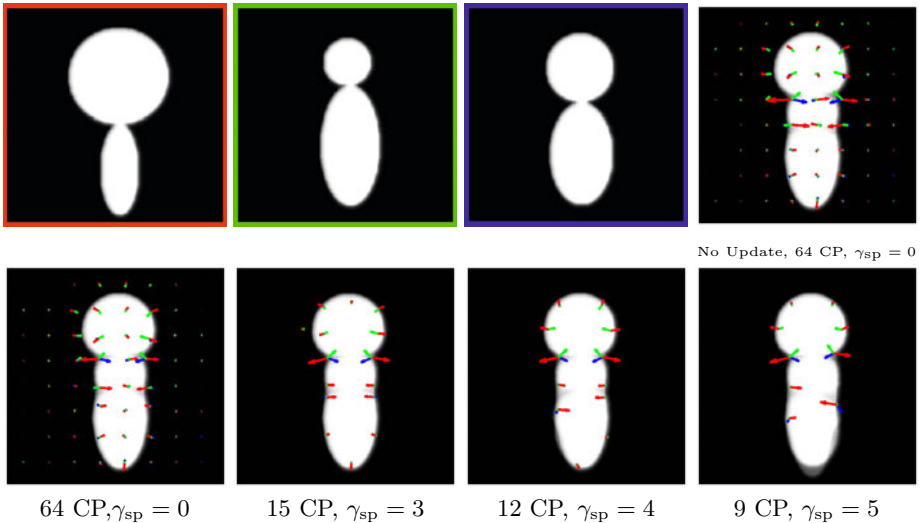


Fig. 2. Atlas from 3 images (top left). Template and initial momenta shown without (top right, $A = 163$) and with (bottom left, $A = 127$) update of control points. The sparsity term drastically reduces the number of control points, for a comparable atlas sharpness: data term $A = 134$ and 206 for $\gamma_{sp} = 3$ and 4 . For $\gamma_{sp} = 5$, the atlas sharpness worsens ($A = 393$), as noticeable at the bottom of the image. $\gamma = 0.1$, $\sigma = 30$ voxels.

In Fig. 1, we show the matching of two synthetic 2D images of size 256×256 . It shows that a *discrete* parameterization (with the maximum number of degrees of freedom: one control point every σ) enables a perfect matching, as would do a parameterization by a *continuous* momenta map as in [4]. Using much fewer momenta, the matching is less accurate. But, moving the control points at their optimal position near the contours drastically increases the matching accuracy.

In Fig. 2, we construct an atlas from 3 synthetic images. The sparsity prior enables to select the most important momenta and to give a compact representation of the shape variability. Fig. 3 shows that the initial number of control points can be divided by 5 without sacrificing much of the atlas sharpness.

In Fig. 4, we show the motion of the control points to the contours of the source image during a registration between MRI slices of size 176×256 . In Fig. 5, we construct an atlas from 5 of such images. The sparsity prior shows that the main variations in the population are located at the skull, near the ventricles and the main sulci, like the frontal sulcus. These structures are indeed the most salient and variable at the scale of analysis: $\sigma = 5$ voxels.

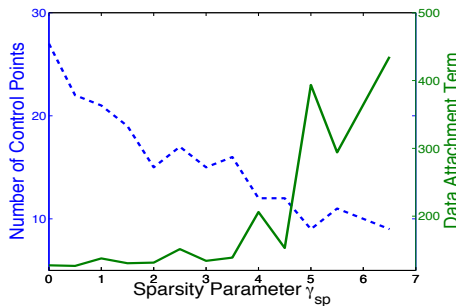


Fig. 3. Graph of the the number of control points and the fidelity-to-data term with respect to the sparsity parameter γ_{sp} with data of Fig. 2. There is a whole range $\gamma_{sp} \in]0, 4]$ for which the number of control points can be divided by 2.4 to 5.3 (from 27 to 12 compared to the initial 64 control points) for an almost constant atlas sharpness.

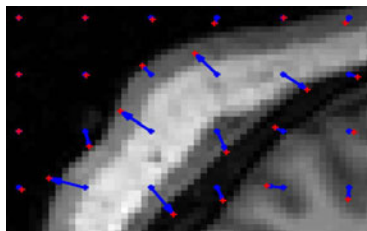


Fig. 4. Registration of brain MRI slices: close-up on the left posterior area of the source image. Initial control points in blue moved to the positions shown in red. They moved toward the contours of the image. They do not move in absence of image force in homogeneous areas. $\sigma = 5$ voxels and $\gamma = 10^{-3}$.

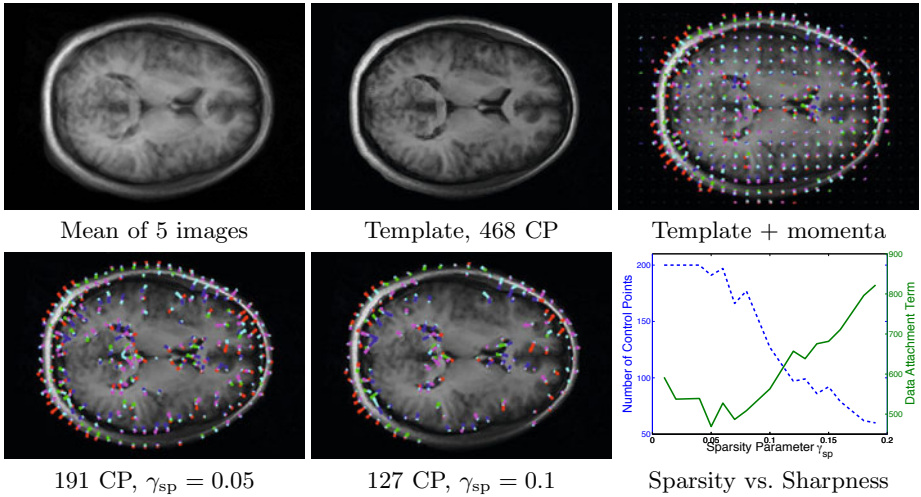


Fig. 5. Atlas from 5 brain images. Top row: estimated atlas with update of control points but without sparsity enforced. Bottom row: with enforced sparsity, the control points focus on the skull, the ventricles and the major sulci. The graph shows that we can achieve an equivalent description of the variability with 40.8% to 27.1% of the original 468 momenta, for $\gamma_{sp} \in]0, 0.1]$. $\sigma = 5$, $\gamma = 5 \cdot 10^{-3}$.

5 Discussion and Conclusion

In this paper, we present a new method for parameterizing large and dense image deformations via a discrete set of control points. Given a set of images, we estimate the template image, the template-to-subject deformations and their parameterization via the optimal placement of the control points and the optimal number of them according to a sparse prior. The whole estimation is posed as a single optimization problem and is solved by a single gradient descent. This is more controllable and more efficient than usual alternated minimizations. The algorithm requires only a solution of ODEs and linear systems; no heuristic rule is used to update the control points. By contrast, a death/birth procedure of control points could have been investigated, but at the cost of more heuristic and arbitrary priors. Our results show that the anatomical variability can be efficiently described by small number of well-placed momenta. We expect to show in the future that these new parameterizations substantially improve the statistical analysis of sets of 3D anatomical images. Future work will extend this framework to include the construction of geodesic diffeomorphisms by integration of Hamiltonian systems, as initiated in [14, 11]. Future work will also focus on the automatic estimation of the best trade-offs between atlas sharpness, sparsity and regularity of the deformations by adding priors on γ and γ_{sp} .

Acknowledgments. This work was supported by NIH grants: NIBIB (5R01 EB007 688), NCRRR (P41 RR023953), ACE-IBIS (RO1 HD055741), and NA-MIC (U54 EB005149).

References

1. Allassonnière, S., Trouvé, A., Younes, L.: Geodesic shooting and diffeomorphic matching via textured meshes. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) EMMCVPR 2005. LNCS, vol. 3757, pp. 365–381. Springer, Heidelberg (2005)
2. 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)
3. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: Identifying global anatomical differences: deformation-based morphometry. *Human Brain Mapping* 6(5-6), 348–357 (1998)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61, 139–157 (2005)
5. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted L^1 minimization. *Journal of Fourier Analysis and Applications* 14(5), 877–905 (2008)
6. Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Thèse de sciences (phd thesis), Université de Nice-Sophia Antipolis (March 2010)
7. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models of sets of curves and surfaces based on currents. *Med. Im. Anal.* 13(5), 793–808 (2009)
8. Glasbey, C.A., Mardia, K.V.: A penalised likelihood approach to image warping. *Journal of the Royal Statistical Society, Series B* 63, 465–492 (2001)
9. Gogtay, N., Lu, A., Leow, A.D., Klunder, A.D., Lee, A.D., Chavez, A., Greenstein, D., Giedd, J.N., Toga, A.W., Rapoport, J.L., Thompson, P.M.: 3D growth pattern abnormalities visualized in childhood-onset schizophrenia using tensor-based morphometry. *Proc. Natl. Acad. Sci.* 105(41), 15979–15984 (2008)
10. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. *IEEE Transaction on Image Processing* 9(8), 1357–1370 (2000)
11. Marsland, S., McLachlan, R.: A hamiltonian particle method for diffeomorphic image registration. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 396–407. Springer, Heidelberg (2007)
12. Miller, M., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision* 24(2), 209–228 (2006)
13. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic Registration Using B-Splines. In: Larsen, R., Nielsen, M., Sparring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
14. Singh, N., Fletcher, P., Preston, J., Ha, L., King, R., Marron, J., Wiener, M., Joshi, S.: Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 529–537. Springer, Heidelberg (2010)

A Differentiation of the Criterion

A variation of the momenta $\delta\alpha(t)$ and the initial position of the control points $\delta\mathbf{c}_0$ induces a variation of the path of the control points $\delta\mathbf{c}(t)$ and then of the voxel positions $\delta\mathbf{y}(t)$. This induces the variation of the criterion:

$$\delta E = \nabla_{\mathbf{y}(0)} A^t \delta \mathbf{y}(0) + \gamma \int_0^1 (\partial_1 L(t) \delta \mathbf{c}(t) + \partial_2 L(t) \delta \alpha(t)) dt \quad (12)$$

The variations $\delta\mathbf{c}(t)$ and $\delta\mathbf{y}(t)$ satisfy the linearized ODEs:

$$\begin{aligned}\dot{\delta\mathbf{c}}(t) &= \partial_1 f(\mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\mathbf{c}(t) + \partial_2 f(\mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\boldsymbol{\alpha}(t) \\ \dot{\delta\mathbf{y}}(t) &= \partial_1 g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\mathbf{y}(t) + \partial_2 g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\mathbf{c}(t) + \partial_3 g(\mathbf{y}(t), \mathbf{c}(t), \boldsymbol{\alpha}(t))\delta\boldsymbol{\alpha}(t)\end{aligned}$$

with $\delta\mathbf{c}(0) = \delta\mathbf{c}_0$ and $\delta\mathbf{y}(1) = 0$. Let $R_{st} = \exp(\int_s^t \partial_2 f(u)du)$ and $V_{st} = \exp(\int_s^t \partial_1 g(u)du)$. The solution of these linear ODEs with source terms are:

$$\delta\mathbf{c}(t) = R_{0t}\delta\mathbf{c}_0 + \int_0^t R_{st}\partial_2 f(s)\delta\boldsymbol{\alpha}(s)ds \text{ and } \delta\mathbf{y}(t) = -\int_t^1 V_{st}\left(\partial_2 g(s)\delta\mathbf{c}(s) + \partial_3 g(s)\delta\boldsymbol{\alpha}(s)\right)ds$$

Plugging these equations into (12) and using Fubini's theorem leads to:

$$\begin{aligned}\delta E &= \int_0^1 \left(\gamma\partial_1 L(t) + \nabla_{\mathbf{y}(0)} A^t V_{t0}\partial_2 g(t)\right)R_{0t}dt\delta\mathbf{c}_0 \\ &+ \int_0^1 \left(\gamma\partial_2 L(t) + \int_t^1 \gamma\partial_1 L(s)R_{ts}\partial_2 f(t)ds + \nabla_{\mathbf{y}(0)} A^t \left(V_{t0}\partial_3 g(t) + \int_t^1 V_{s0}\partial_2 g(s)R_{ts}\partial_2 f(t)ds\right)\right)\delta\boldsymbol{\alpha}(t)dt\end{aligned}$$

Therefore the gradient of E with respect to the L^2 metric is given as:

$$\nabla_{\boldsymbol{\alpha}} E(t) = \gamma\partial_2 L(t)^t + \partial_3 g(t)^t \boldsymbol{\eta}^y(t) + \partial_2 f(t)^t \boldsymbol{\eta}^c(t) \text{ and } \nabla_{\mathbf{c}_0} E = \boldsymbol{\eta}^c(0) \quad (13)$$

where we have denoted:

$$\boldsymbol{\eta}^y(t) = V_{t0}^t \nabla_{\mathbf{y}(0)} A \text{ and } \boldsymbol{\eta}^c(t) = \int_t^1 R_{ts}^t \left(\gamma\partial_1 L(s)^t + \partial_2 g(s)^t \boldsymbol{\eta}^y(s)\right)ds \quad (14)$$

Since $V_{t0} = \text{id} - \int_0^t V_{s0}\partial_1 g(s)ds$ and $R_{ts} = \text{id} + \int_t^s R_{us}\partial_1 f(u)du$, we have:

$$\begin{aligned}\boldsymbol{\eta}^y(t) &= -\nabla_{\mathbf{y}(0)} A - \int_0^t \partial_1 g(s)^t \boldsymbol{\eta}^y(s)ds \\ \boldsymbol{\eta}^c(t) &= \int_t^1 \left(\gamma\partial_1 L(s)^t + \partial_2 g(s)^t \boldsymbol{\eta}^y(s) + \partial_1 f(s)^t \boldsymbol{\eta}^c(s)\right)ds\end{aligned} \quad (15)$$

Given the definitions of f, g and L , the L^2 -gradient can be written as:

$$\nabla_{\boldsymbol{\alpha}}^{L^2} E(t) = 2\gamma\mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\alpha}(t) + \mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))\boldsymbol{\eta}^c(t) + \mathbf{k}(\mathbf{y}(t), \mathbf{c}(t))\boldsymbol{\eta}^y(t) \quad (16)$$

where $\mathbf{k}(\mathbf{y}, \mathbf{c})$ denotes the $3M$ -by- $3N$ block matrix whose (i, j) th-block is $K(y_i, c_j)$.

The Sobolev gradient associated to the metric $\mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))$ is therefore:

$$\nabla_{\boldsymbol{\alpha}}^{\text{Sob}} E(t) = 2\gamma\boldsymbol{\alpha}(t) + \boldsymbol{\eta}^c(t) + \mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))^{-1}\mathbf{k}(\mathbf{y}(t), \mathbf{c}(t))\boldsymbol{\eta}^y(t) \quad (17)$$

which requires to solve a linear system. If control points gets closer than σ , then the matrix $\mathbf{k}(\mathbf{c}(t), \mathbf{c}(t))$ is badly conditioned [6]. In this case, we adjust σ in this matrix to the minimal distance between control points and use the gradient $\mathbf{k}^{\text{adj}}(\mathbf{c}(t), \mathbf{c}(t))^{-1}\nabla^{L^2} E$, which interpolates between the L^2 gradient ($\sigma_{\text{adj}} \rightarrow 0$) and the Sobolev gradient ($\sigma_{\text{adj}} \rightarrow \sigma$).

Learning an Atlas of a Cognitive Process in Its Functional Geometry

Georg Langs^{1,3}, Danial Lashkari¹, Andrew Sweet¹, Yanmei Tie²,
Laura Rigolo², Alexandra J. Golby², and Polina Golland¹

¹ Computer Science and Artificial Intelligence Lab,
Massachusetts Institute of Technology, Cambridge, MA, USA
{[langsg](mailto:langsg@csail.mit.edu),[daniel](mailto:daniel@csail.mit.edu),[sweet](mailto:sweet@csail.mit.edu),[polina](mailto:polina@csail.mit.edu)}@csail.mit.edu

² Department of Neurosurgery, Brigham and Women's Hospital,
Harvard Medical School, Boston, MA, USA
{[ytie](mailto:ytie@bwh.harvard.edu),[lrigolo](mailto:lrigolo@bwh.harvard.edu),[agolby](mailto:agolby@bwh.harvard.edu)}@bwh.harvard.edu

³ Computational Image Analysis and Radiology Lab, Department of Radiology,
Medical University of Vienna, Vienna, Austria

Abstract. In this paper we construct an atlas that captures functional characteristics of a cognitive process from a population of individuals. The functional connectivity is encoded in a low-dimensional embedding space derived from a diffusion process on a graph that represents correlations of fMRI time courses. The atlas is represented by a common prior distribution for the embedded fMRI signals of all subjects. The atlas is not directly coupled to the anatomical space, and can represent functional networks that are variable in their spatial distribution. We derive an algorithm for fitting this generative model to the observed data in a population. Our results in a language fMRI study demonstrate that the method identifies coherent and functionally equivalent regions across subjects.

1 Introduction

The functional architecture of the cerebral cortex consists of regions and networks of regions that become active during specific tasks or at rest when the brain is suspected to engage in activities such as memory encoding [1]. The functional networks vary spatially across individuals due to natural variability [15], developmental processes in early childhood [9] or adulthood [4], or pathology [5]. Reorganization can appear over remarkably short periods of even few days [4]. The relationship between the structure of functional networks and their spatial distribution is not well understood.

The traditional brain imaging paradigm in most functional MRI (fMRI) studies treats functional activity as a feature of a position within the anatomical coordinate frame. The anatomical variability in a population is typically mitigated by smoothing and non-rigid registration of the anatomical data, and the corresponding normalization of functional signals into a *stereotactic* space. The remaining spatial variability of functional regions is ignored. An alternative approach is

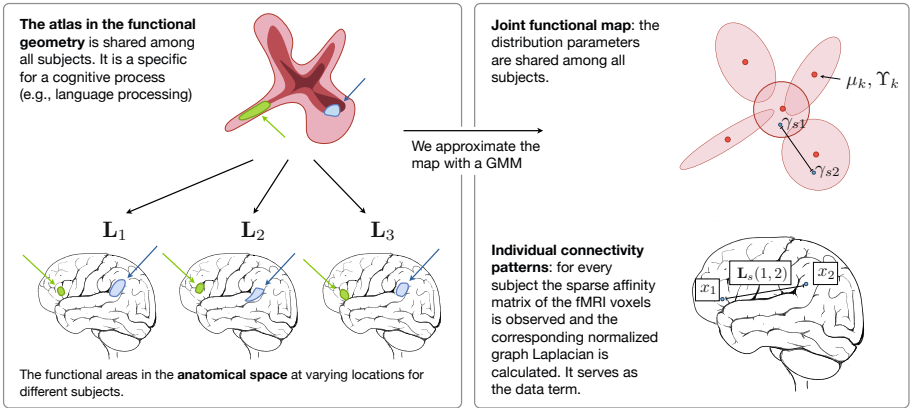


Fig. 1. Joint functional geometry scheme

to localize functional regions of interest (fROIs) in individuals or groups [15] as a precursor to analysis, and subsequently study the responses in the resulting small number of fROIs.

Both approaches limit the range of questions that can be formulated on the fMRI observations. For example, the spatial normalization framework cannot express or account for spatial variability within the population since it assumes perfect spatial correspondences when detecting networks by averaging over multiple subjects. In contrast, the fROI approach is based on detection results for each subject, which can be infeasible if the activation is weak and cannot be distinguished from noise in individual subjects without averaging over the group.

We propose a different approach to characterize functional networks in a population of individuals. We do not assume a tight coupling between anatomical location and function, but view functional signals as the basis of a descriptive map that represents the global connectivity pattern during a specific cognitive process. We develop a representation of those networks based on manifold learning techniques and show how we can learn an *atlas* from a population of subjects performing the same task. Our main assumption is that the connectivity pattern associated with a functional process is consistent across individuals. Accordingly, we construct a generative model (the atlas) for these connectivity patterns that describes the common structures within the population.

The clinical goal of this work is to provide additional evidence for localization of functional areas. A robust localization approach is important for neurosurgical planning if individual activations are weak or reorganization has happened due to pathologies such as tumor growth. Furthermore the method provides a basis for understanding the mechanisms underlying formation and reorganization in the cerebral system.

Related work. A spectral embedding [18] represents data points in a map that reflects a large set of pair-wise affinity values in the Euclidean space.

Diffusion maps establish a metric based on the concept of diffusion processes on a graph [2]. A probabilistic interpretation of diffusion maps has recently been proposed [13]. Previously demonstrated spectral methods in application to fMRI analysis mapped voxels into a space that captured joint functional characteristics of brain regions [10]. This approach represents the magnitude of co-activation by the density in the embedding. Functionally homogeneous units have been shown to form clusters in the embedding in a study of parceled resting-state fMRI data [17]. In [7] multidimensional scaling was employed to retrieve a low dimensional representation of positron emission tomography (PET) signals in a set of activated regions. In an approach closely related to the method proposed in this paper [11], an embedding of fMRI signals was used to match corresponding functional regions across different subjects. Recently a probabilistic generative model that connects the embedding coordinates with a similarity matrix has been demonstrated in [14].

2 Generative Model of Functional Connectivity

We start by reviewing the original diffusion maps formulation. We then derive a probabilistic likelihood model for the data based on this mapping and use the model to link diffusion maps of functional connectivity across subjects.

2.1 Diffusion Distances, Diffusion Maps, and fMRI Time Courses

Given an fMRI sequence $\mathbf{I} \in \mathbb{R}^{T \times N}$ that contains N voxels, each characterized by an fMRI signal over T time points, we calculate matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ that assigns a non-negative symmetric weight to each pair of voxels (i, j)

$$\mathbf{W}(i, j) = e^{\frac{\langle \mathbf{I}_i, \mathbf{I}_j \rangle}{\epsilon}}, \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the correlation coefficient of the time courses \mathbf{I}_i and \mathbf{I}_j , and ϵ is the weight decay. We define a graph whose vertices correspond to voxels and whose edge weights are determined by \mathbf{W} [2,10]. In practice, we discard all edges that have a weight below a chosen threshold. This construction yields a sparse graph which is then transformed into a Markov chain. We define the Markov transition matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$, where \mathbf{D} is a diagonal matrix such that $d_i = D(i, i) = \sum_j w(i, j)$ is the degree of node i . By interpreting the entries $\mathbf{P}(i, j)$ as transition probabilities, we can define the diffusion distance

$$D_t(i, j) = \sum_{i'=1, \dots, N} \frac{(\mathbf{P}^t(i, i') - \mathbf{P}^t(j, i'))^2}{\phi(i')} \quad \text{where} \quad \phi(i) = \frac{d_i}{\sum_{i'} d_{i'}}. \quad (2)$$

The distance is determined by the probability of traveling between vertices i and j by taking all paths of at most t steps. The transition probabilities are based on the functional connectivity of node pairs; the diffusion distance integrates the connectivity values over possible paths that connect two points and defines a geometry that captures the entirety of the connectivity structure. This distance

is characterized by the operator \mathbf{P}^t , the t^{th} power of the transition matrix. The value of the distance $D_t(i, j)$ is low if there is a large number of paths of at most length t steps with high transition probabilities between the nodes i and j .

The diffusion map coordinates $\mathbf{\Gamma} = [\gamma_1, \gamma_2, \dots, \gamma_N]^T$ yield a low dimensional embedding of the signal such that the resulting pairwise distances approximate diffusion distances, i.e., $\|\gamma_i - \gamma_j\|^2 \approx D_t(i, j)$ [13]. They are derived from the right eigenvectors of the transition matrix. In Appendix A we show that a diffusion map can be viewed as a solution to a least-squares problem. Specifically, we define a symmetric matrix $\mathbf{A} = \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$, and let \mathbf{L} be the normalized graph Laplacian

$$\mathbf{L} = \mathbf{D}^{-1/2} \mathbf{A}^{2t} \mathbf{D}^{-1/2}. \quad (3)$$

The embedding coordinates are then found as follows:

$$\mathbf{\Gamma}^* = \underset{\mathbf{\Gamma} \in \mathbb{R}^{N \times L}}{\operatorname{argmin}} \sum_{i,j} d_i d_j (\mathbf{L}(i, j) - \gamma_i^T \gamma_j)^2, \quad (4)$$

where L is the dimensionality of the embedding. To simplify notation, we omit t for \mathbf{L} and $\mathbf{\Gamma}$ in the derivations, assuming that all the results are derived for a fixed, known diffusion time.

2.2 A Generative Model for Diffusion Maps across Subjects

The goal of the generative model is to explain jointly the distribution of pairwise functional affinities of voxels across all subjects. We use latent variables $\mathbf{\Gamma} = \{\mathbf{\Gamma}_s\}_{s=1}^S$ to represent the diffusion map coordinates for S subjects indexed by $s \in \{1, \dots, S\}$. We can interpret Eq. (4) as maximization of a Gaussian likelihood model. We let γ_{si} denote the embedding coordinates of voxel i in subject s and let \mathbf{L}_s be the normalized graph Laplacian for subject s . We further assume that elements of \mathbf{L}_s are conditionally independent given the embedding coordinates:

$$p(\mathbf{L}_s(i, j) | \gamma_{si}, \gamma_{sj}) = \mathcal{N}\left(\mathbf{L}_s(i, j); \gamma_{si}^T \gamma_{sj}, \frac{\sigma_s^2}{d_{si} d_{sj}}\right), \quad (5)$$

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2 .

Note that the variance depends on the degrees d_i, d_j , which is technically a problem since these quantities depend on the data \mathbf{W} . We find that in practice, the method works well and leave the development of rigorous probability models for diffusion maps as an interesting future direction.

In the absence of a prior distribution on $\mathbf{\Gamma}_s$, fitting this model to the data yields results similar to the conventional diffusion maps for each subject independently from the rest of the population.

The goal of this paper is to define an atlas that represents a population-wide structure of functional connectivities in the space of diffusion maps. To capture this common structure, we define a shared prior distribution on the embedding coordinates $\mathbf{\Gamma}_s$ for all subjects, and expect the embedded vectors to be in correspondence across subjects [11]. Here, we assume that the common distribution in the embedding space is a mixture of K Gaussian components.

We let $z_{si} \in \{1, \dots, K\}$ be the component assignment for voxel i in subject s and obtain the prior on the embedding coordinates of voxel i in subject s :

$$p(\gamma_{si} | z_{si} = k; \boldsymbol{\mu}, \boldsymbol{\Theta}) = \mathcal{N}(\gamma_{si}; \mu_k, \boldsymbol{\Theta}_k), \quad (6)$$

where μ_k and $\boldsymbol{\Theta}_k$ are the center and covariance matrices for component k . We let the component assignments be independently distributed according to the weights of different components, i.e.,

$$p(z_{si} = k) = \pi_k. \quad (7)$$

Together, Eqs. 5 to Eqs. 7 the joint distribution of the embeddings $\boldsymbol{\Gamma}$, the component assignments \mathbf{z} , and the observed affinities $\mathbf{L} = \{\mathbf{L}_s\}_{s=1}^S$. The distribution is parameterized by component centers $\{\mu_k\}$, covariance matrices $\{\boldsymbol{\Theta}_k\}$, and weights $\{\pi_k\}$.

By adding the group prior over diffusion maps, we constrain the resulting subject maps to be aligned across subjects and further encourage them to resemble the population-wide structures characterized by the mixture model (Fig. 1). The mixture model acts as a population atlas in the embedding space.

3 Atlas Learning and Inference

We employ the variational EM algorithm [8] to estimate the parameters of our model from the observed data. We approximate the posterior distribution of latent variables $p(\boldsymbol{\Gamma}, \mathbf{z} | \mathbf{L})$ with a product distribution of the form

$$q(\boldsymbol{\Gamma}, \mathbf{z}) = \prod_{s,i} q(\gamma_{si})q(z_{si}). \quad (8)$$

The problem is then formulated as the minimization of the Gibbs free energy

$$\mathcal{F} = \mathbb{E}_q [\ln q(\boldsymbol{\Gamma}, \mathbf{z}) - \ln p(\boldsymbol{\Gamma}, \mathbf{z}, \mathbf{L}; \boldsymbol{\mu}, \boldsymbol{\Theta}, \boldsymbol{\pi})], \quad (9)$$

where \mathbb{E}_q indicates the expected value operator with respect to distribution $q(\cdot)$. We derive coordinate descent update rules that, given an initialization of all latent variables and parameters, find a local minimum of the cost function in Eq. 9. Appendix B presents the derivation of the update rules.

3.1 Initialization

The algorithm requires initial estimates of the latent variables and model parameters. Initialization affects convergence and the quality of the final solution. Here, we describe a simple initialization scheme that matches closely the structure of our model and enhances the performance of the algorithm.

In general, the relationship between the diffusion map coordinates $\boldsymbol{\Gamma}$ and the corresponding symmetric matrix \mathbf{L} is defined up to an arbitrary orthonormal matrix \mathbf{Q} since $(\boldsymbol{\Gamma}\mathbf{Q})(\boldsymbol{\Gamma}\mathbf{Q})^T = \boldsymbol{\Gamma}\mathbf{Q}\mathbf{Q}^T\boldsymbol{\Gamma}^T = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \mathbf{L}$. In order to define an atlas of the functional connectivity across all subjects, we seek matrix \mathbf{Q}_s for each

subject s such that the maps $\{\Gamma_s \mathbf{Q}_s\}_{s=1}^S$ are aligned in a common coordinate frame. Consider aligning the diffusion map Γ_s of subject s to the diffusion map Γ_r of reference subject r . Similar to the construction of the diffusion map, we compute the inter-subject affinities between the fMRI signals of subjects s and r using Eq. (II) and only keep those with a correlation above the threshold. This step produces a set of M node pairs $\{(i_m, j_m)\}_{m=1}^M$, characterized by affinities $\{w_m\}_{m=1}^M$. The initialization should ensure that nodes with similar fMRI signals are close in the common embedding space. Therefore, we choose matrix \mathbf{Q} that minimizes the weighted Euclidean distance between pairs of corresponding embedding coordinates

$$\mathbf{Q}_{sr}^* = \underset{\mathbf{Q}}{\operatorname{argmin}} \left[\sum_{m=1}^M w_m \|\mathbf{Q} \gamma_{si_m} - \gamma_{rj_m}\|_{L_2}^2 \right]. \quad (10)$$

We define matrices $\Gamma_{s_m} = [\gamma_{si_1}, \dots, \gamma_{si_M}]^T$ and $\Gamma_{r_m} = [\gamma_{rj_1}, \dots, \gamma_{rj_M}]^T$. It can be shown that $\mathbf{Q}_{sr}^* = \mathbf{V}\mathbf{U}^T$, where we use the singular value decomposition $\mathbf{U}\Sigma\mathbf{V}^T = \Gamma_{s_m}^T \operatorname{diag}(\mathbf{w}_m) \Gamma_{r_m}$ [16].

We find initial estimates of $\{\mu_k, \Theta_k, \pi_k\}_{k=1}^K$ by fitting a K component Gaussian mixture model to the initial estimates of the atlas embedding coordinates $\{\Gamma_s \mathbf{Q}_{sr}^*\}_{s=1}^S$ for a randomly chosen reference subject r .

4 Experiments and Results

Data. We demonstrate the method on a set of six healthy control subjects. The fMRI data was acquired using a 3T GE Signa system (TR=2s, TE=40ms, flip angle=90°, slice gap=0mm, FOV=25.6cm, volume size of $128 \times 128 \times 27$ voxels, voxel size of $2 \times 2 \times 4$ mm³). The language task (antonym generation) block design was 5min 10s long, starting with a 10s pre-stimulus period. Eight task and seven rest blocks, 20s each, alternated in the design. For each subject, an anatomical T1 MRI scan was acquired and registered to the functional data. Grey matter was segmented with FSL [19] on the T1 data. The grey matter labels were transferred to the co-registered fMRI volumes, and computation was restricted to grey matter.

Evaluation. We construct a joint functional diffusion map for all six subjects. For the results presented in this paper, we set the dimensionality of the diffusion map to be $L = 20$ and choose a diffusion time $t = 2$ that satisfies $(\lambda_L/\lambda_1)^t < 0.2$ for all subjects. To facilitate computation we only keep nodes for which the degree is above a certain threshold. In the experiments reported here we choose a threshold of 100. For the EM algorithm, we fix a value of $\sigma_s = 10^2 N_s^{-1} \sum_i d_{si}$ for the first 10 iterations, then allow this parameter to update for the remaining iterations according to the rule defined in Appendix B. In our experiments, an initial value of σ_s specifically proportional to 10^2 allows the algorithm to achieve the lowest Gibbs free energy.

We hypothesize that working in the embedding space should allow us to more robustly capture the functional structure common to all subjects. In order to

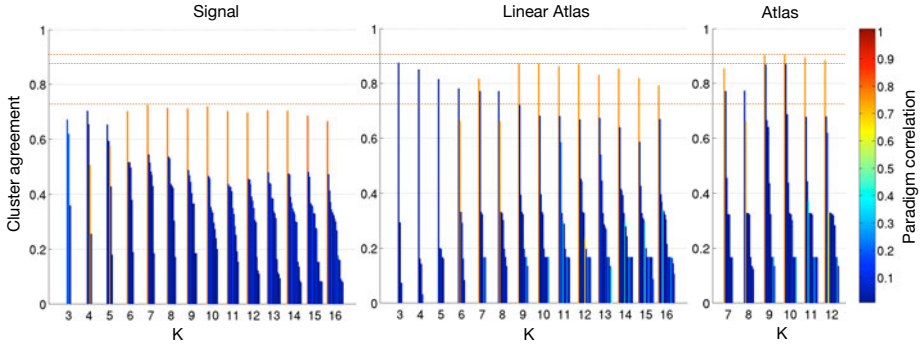


Fig. 2. Mean cluster Dice scores for clustering in *Signal*, *Linear Atlas*, and *Atlas*. For each number of clusters K , we report the mean Dice score across subjects for each cluster. Color illustrates correlation of the cluster average fMRI signal with the paradigm signal.

validate this, we compare the consistency of clustering structures found in the space of fMRI time courses (*Signal*), a low-dimensional ($L=20$) PCA embedding of these time courses (*PCA-Signal*), and the low-dimensional ($L=20$) embedding proposed in this paper. We report results for the initial alignment (*Linear-Atlas*) and the result of learning the joint atlas (*Atlas*).

We first apply clustering to signals from individual subjects separately to find subject-specific cluster assignments. We then apply clustering to signals combined from *all* subjects to construct the corresponding group-wise cluster assignments. Since our group atlas for the lower-dimensional space is based on a mixture model, we also choose a mixture-model for clustering in the *Signal* and *PCA-Signal* spaces. In both cases, each component in the mixture is an isotropic von Mises-Fisher distribution, defined on a hyper-sphere after centering and normalization of the fMRI signals to unit variance [12].

Likewise, we cluster the diffusion map coordinates $\mathbf{\Gamma}_s$ separately in each subject to obtain subject-specific assignments. We cluster the diffusion map coordinates of all subjects aligned to the first subject, $\{\mathbf{\Gamma}_s \mathbf{Q}_{(s,1)}\}_s$ for the *Linear-Atlas* and in $\{\mathbf{\Gamma}_s^*\}_s$ for *Atlas* to obtain group-wise clustering assignments. Analyzing the consistency of clustering labels across methods evaluates how the population structure captures the individual embeddings. For the diffusion maps, Euclidean distance is a meaningful metric; we therefore use a mixture model with Gaussian components that share the same isotropic variance.

Since clusters are labeled arbitrarily in each result, we match group and subject-specific cluster labels by solving a bipartite graph matching problem. Here, we find a one-to-one label correspondence that maximizes voxel overlap between pairs of clusters, similar to the method used in [12]. After matching the labels, we use the Dice score [3] to measure the consistency between group and subject-specific assignments for each cluster.

4.1 Results

Fig. 2 reports the consistency of clusters between group-level and subject-specific assignments, measured in terms of Dice score averaged across subjects. To illustrate the temporal nature of the clusters, the colors of the bars indicate the correlation of the average fMRI signal in the cluster with the fMRI language paradigm convolved with the hemodynamic response function. Note that the paradigm was not used at any point during the generation of the maps or clusters. The cluster with the highest paradigm correlation is the most consistent cluster over a large range of cluster numbers. The highest Dice score (0.725) for *Signal* clustering is achieved, with similar values for larger numbers of clusters. Although the plot is not shown here, clustering in the *PCA-Signal* space exhibits no noticeable improvement. Initial alignment of the diffusion maps into the *Linear-Atlas* substantially increases the Dice score of the highest ranked clusters for all K , with a maximum value of 0.876. The variational EM algorithm performed using a range of reasonable cluster numbers and further improves the cluster agreement for the top ranked clusters (0.905).

Fig. 3 shows the networks of the subjects that correspond to the top ranked atlas cluster ($K=10$), together with the corresponding average fMRI signal. The paradigm is recovered very well, and for most subjects the cluster network plausibly spans visual, motor, and language areas.

Fig. 4 compares the location and average signal of the top ranked of 10 clusters for *Signal* and *Atlas* clustering in a single subject. While both recover parts of the paradigm, the clustering in the diffusion map atlas exhibits more consistency between the group and the subject levels. Additionally, the *Signal* clusters suffer from a relatively high dispersion across the entire cortex. This is not the

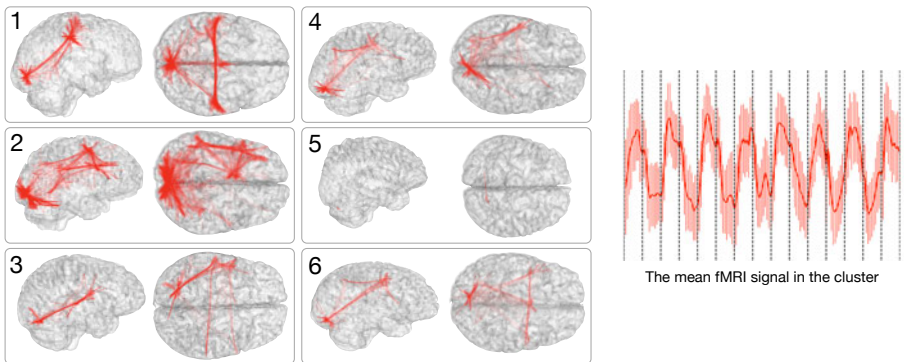


Fig. 3. A cluster in the joint map corresponds to a network in each subject. Here we illustrate a network in 6 subjects that corresponds to one cluster in the atlas, and the mean fMRI signal of this cluster. The 8 block stimulus in this language study was not used by the analysis, but was recovered by the algorithm and corresponding networks were identified across all subjects. They typically span the visual cortex, the language areas (Wernicke and Broca), and the motor areas in some cases.

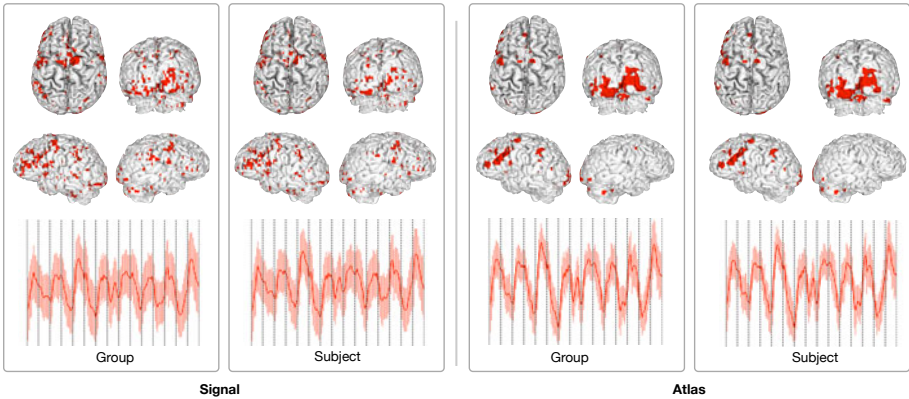


Fig. 4. Most consistent cluster in the *Signal* space and the *Atlas* shown in the anatomical space. For each method, we show the group-wise (left) and subject-specific (right) assignment. Also shown is the average and standard deviation of the cluster fMRI signal. Results for subject 2.

case for the diffusion map atlas. In summary, these results demonstrate that the representation of fMRI time courses in the low dimensional space of diffusion maps better captures the functional connectivity structure across subjects. Not only are clustering assignments more consistent, but the anatomical characteristics of these clusters are also more plausible. Furthermore, our results using the variational EM algorithm suggest that the probabilistic population model further improves the consistency across the population, and consolidates the distribution in the embedding space.

5 Conclusion

We propose a method to learn an atlas of the functional connectivity structure that emerges during a cognitive process from a group of individuals. The atlas is a group-wise generative model that describes the fMRI responses of all subjects in the embedding space. The embedding space is a low dimensional representation of fMRI time courses that encodes the functional connectivity patterns within each subject. Results from a fMRI language experiment indicate that the diffusion map framework captures the connectivity structure reliably, and leads to valid correspondences across subjects. Future work will focus on the application of the framework to the study of reorganization processes.

Acknowledgements. This work was funded in part by the NSF IIS/CRCNS 0904625 grant, the NSF CAREER 0642971 grant, the NIH NCCR NAC P41-RR13218, NIH NIBIB NIMIC U54-EB005149, NIH U41RR019703, and NIH P01CA067165 grants, the Brain Science Foundation, the Klarman Family Foundation, and EU (FP7/2007-2013) n°257528 (KHRESMOI).

References

1. Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L.: The brain's default network: anatomy, function, and relevance to disease. *Ann. N Y Acad. Sci.* 1124, 1–38 (2008)
2. Coifman, R.R., Lafon, S.: Diffusion maps. *App. Comp. Harm. An.* 21, 5–30 (2006)
3. Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* 26(3), 297–302 (1945)
4. Elbert, T., Rockstroh, B.: Reorganization of human cerebral cortex: the range of changes following use and injury. *Neuroscientist* 10(2), 129–141 (2004)
5. Elkana, O., Frost, R., Kramer, U., Ben-Bashat, D., Hendler, T., Schmidt, D., Schweiger, A.: Cerebral reorganization as a function of linguistic recovery in children: An fmri study. *Cortex* (December 2009)
6. Friedland, S., Torokhti, A.: Generalized rank-constrained matrix approximations. Arxiv preprint math/0603674 (2006)
7. Friston, K., Frith, C., Fletcher, P., Liddle, P., Frackowiak, R.: Functional topography: multidimensional scaling and functional connectivity in the brain. *Cerebral Cortex* 6(2), 156 (1996)
8. Jaakkola, T.: Tutorial on variational approximation methods. In: *Advanced Mean Field Methods: Theory and Practice*, pp. 129–159 (2000)
9. Kuhl, P.K.: Brain mechanisms in early language acquisition. *Neuron* 67(5), 713–727 (2010)
10. Langs, G., Samaras, D., Paragios, N., Honorio, J., Alia-Klein, N., Tomasi, D., Volkow, N.D., Goldstein, R.Z.: Task-specific functional brain geometry from model maps. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 925–933. Springer, Heidelberg (2008)
11. Langs, G., Tie, Y., Rigolo, L., Golby, A., Golland, P.: Functional geometry alignment and localization of brain areas. In: Lafferty, J., Williams, C.K.I., Shawe-Taylor, J., Zemel, R., Culotta, A. (eds.) *Advances in Neural Information Processing Systems*, vol. 23, pp. 1225–1233 (2010)
12. Lashkari, D., Vul, E., Kanwisher, N., Golland, P.: Discovering structure in the space of fMRI selectivity profiles. *Neuroimage* 50(3), 1085–1098 (2010)
13. Nadler, B., Lafon, S., Coifman, R., Kevrekidis, I.: Diffusion maps—a probabilistic interpretation for spectral embedding and clustering algorithms. In: *Principal Manifolds for Data Visualization and Dimension Reduction*, pp. 238–260 (2007)
14. Rosales, R., Frey, B.: Learning generative models of affinity matrices. In: *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI 2003)*. pp. 485–492 (2003)
15. Saxe, R., Brett, M., Kanwisher, N.: Divide and conquer: a defense of functional localizers. *Neuroimage* 30(4), 1088–1096 (2006)
16. Scott, G., Longuet-Higgins, H.: An algorithm for associating the features of two images. *Proceedings: Biological Sciences* 244(1309), 21–26 (1991)
17. Thirion, B., Dodel, S., Poline, J.B.: Detection of signal synchronizations in resting-state fmri datasets. *Neuroimage* 29(1), 321–327 (2006)
18. Von Luxburg, U.: A tutorial on spectral clustering. *Statistics and Computing* 17(4), 395–416 (2007)
19. 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. *Neuroimage* 45(suppl. 1), S173–S186 (2009)

A Diffusion Map Coordinates

In the standard diffusion map analysis, the embedding coordinates $\mathbf{\Gamma}$ for a L -dimensional space are obtained via the first L eigenvectors of matrix $\mathbf{A} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ [13]. Here we show that we can represent the embedding as a solution of a least-squares problem formulated directly on the similarity matrix \mathbf{W} .

Formally, $\mathbf{\Gamma} = \mathbf{D}^{-1/2}\mathbf{V}_{1:L}\mathbf{\Lambda}_L^t$, where $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ is the eigenvector decomposition of matrix \mathbf{A} , t is the diffusion time, and subscripts indicate that we select the first L eigenvectors. Matrix $\tilde{\mathbf{A}} = \mathbf{V}_{1:L}\mathbf{\Lambda}_L\mathbf{V}_{1:L}^T$ is a low-rank approximation of matrix \mathbf{A} that is quite accurate if the remaining eigenvalues are much smaller than the sum of the first L eigenvalues. We define

$$\begin{aligned} \mathbf{L} &= \mathbf{D}^{-1/2}\mathbf{A}^{2t}\mathbf{D}^{-1/2} \approx \mathbf{D}^{-1/2}\tilde{\mathbf{A}}^{2t}\mathbf{D}^{-1/2} = \mathbf{D}^{-1/2}(\mathbf{V}_{1:L}\mathbf{\Lambda}_L\mathbf{V}_{1:L}^T)^{2t}\mathbf{D}^{-1/2} \\ &= \mathbf{D}^{-1/2}\mathbf{V}_{1:L}\mathbf{\Lambda}_L^t\mathbf{\Lambda}_L^t\mathbf{V}_{1:L}^T\mathbf{D}^{-1/2} = \mathbf{\Gamma}\mathbf{\Gamma}^T \end{aligned} \quad (11)$$

and use a generalization of the Eckart-Young theorem [6] to formulate the eigen decomposition as an optimization problem:

$$\mathbf{\Gamma}^* = \underset{\mathbf{\Gamma} \in \mathbb{R}^{N \times L}}{\operatorname{argmin}} \|\mathbf{A}^2 - \mathbf{D}^{1/2}\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{D}^{1/2}\|_F^2 = \underset{\mathbf{\Gamma} \in \mathbb{R}^{N \times L}}{\operatorname{argmin}} \sum_{i,j} d_i d_j (\mathbf{L}_t(i,j) - \gamma_i^T \gamma_j)^2, \quad (12)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

B Variational EM Update Rules

We use a natural choice of a multinomial distribution for cluster membership $q(z_{si} = k)$ for $s \in \{1, \dots, S\}$, $i \in \{1, \dots, N_s\}$, and a Gaussian distribution for the embedding coordinates $q(\gamma_{si}) = \mathcal{N}(\gamma_{si}; \mathbb{E}[\gamma_{si}], \operatorname{diag}(\mathbb{V}[\gamma_{si}]))$, parameterized by its mean $\mathbb{E}[\gamma_{si}]$ and component-wise variance $\mathbb{V}[\gamma_{si}]$.

E-Step. We determine the parameter values of the approximating probability distribution $q(\cdot)$ that minimize the Gibbs free energy in Eq. (9) by evaluating the expectation, differentiating with respect to each parameter and setting the derivatives to zero. This yields

$$q(z_{si} = k) \propto \frac{\pi_k}{|\mathbf{\Theta}_k|^{1/2}} \exp \left\{ -\frac{1}{2} \left((\mathbb{E}[\gamma_{si}] - \mu_k)^T \mathbf{\Theta}_k^{-1} (\mathbb{E}[\gamma_{si}] - \mu_k) + \operatorname{trace}(\operatorname{diag}(\mathbb{V}[\gamma_{si}])\mathbf{\Theta}_k^{-1}) \right) \right\}, \quad \text{s.t.} \quad \sum_k q(z_{si} = k) = 1,$$

$$\mathbb{V}[\gamma_{si}(l)] = \left(\sum_k q(z_{si} = k) \mathbf{\Theta}_k^{-1}(l, l) + \frac{d_{si}}{\sigma_s^2} \sum_{j \neq i} d_{sj} (\mathbb{E}[\gamma_{sj}(l)]^2 + \mathbb{V}[\gamma_{sj}(l)]) \right)^{-1},$$

$$\begin{aligned} \mathbb{E}[\gamma_{si}(l)] = & \mathbb{V}[\gamma_{si}(l)] \left[\sum_k q(z_{si} = k) \left[\Theta_k^{-1}(l, l) \mu_k(l) - \sum_{l' \neq l} \Theta_k^{-1}(l, l') (\mathbb{E}[\gamma_{si}(l') - \mu_k(l')]) \right] \right. \\ & \left. + \frac{d_{si}}{\sigma_s^2} \sum_{j \neq i} d_{sj} \left[\mathbf{L}_s(i, j) \mathbb{E}[\gamma_{sj}(l)] - \mathbb{E}[\gamma_{sj}(l)] \sum_{l' \neq l} \mathbb{E}[\gamma_{si}(l')] \mathbb{E}[\gamma_{sj}(l')] \right] \right]. \end{aligned}$$

Rather than solve the coupled system of equations above, we iteratively update each parameter of the distribution $q(\cdot)$ while fixing all the other parameters.

M-Step. We now find the parameter values similar to the standard EM algorithm, but using the approximating distribution $q(\cdot)$ to evaluate the expectation. Specifically, we find

$$\pi_k = \frac{1}{\sum_s N_s} \sum_{s,i} q(z_i = k), \quad \mu_k = \sum_{s,i} \frac{q(z_{si} = k)}{\sum_{s',i'} q(z_{s'i'} = k)} \mathbb{E}[\gamma_{si}], \quad (13)$$

$$\Theta_k = \sum_{s,i} \frac{q(z_{si} = k)}{\sum_{s',i'} q(z_{s'i'} = k)} [(\mathbb{E}[\gamma_{si}] - \mu_k) (\mathbb{E}[\gamma_{si}] - \mu_k)^T + \text{diag}(\mathbb{V}[\gamma_{si}])], \quad (14)$$

$$\sigma_s^2 = \frac{2}{N_s(N_s - 1)} \sum_{i,j \neq i} d_{si} d_{sj} \mathbb{E}[(\mathbf{L}_s(i, j) - \gamma_{si}^T \gamma_{sj})^2]. \quad (15)$$

Parameterization-Invariant Shape Statistics and Probabilistic Classification of Anatomical Surfaces

Sebastian Kurtek¹, Eric Klassen², Zhaohua Ding³,
Malcolm J. Avison³, and Anuj Srivastava¹

¹ Department of Statistics, Florida State University, Tallahassee, FL

² Department of Mathematics, Florida State University, Tallahassee, FL

³ Institute of Imaging Science, Vanderbilt University, Nashville, TN

Abstract. We consider the task of computing shape statistics and classification of 3D anatomical structures (as continuous, parameterized surfaces). This requires a Riemannian metric that allows re-parameterizations of surfaces by isometries, and computations of geodesics. This allows computing Karcher means and covariances of surfaces, which involves optimal re-parameterizations of surfaces and results in a superior alignment of geometric features across surfaces. The resulting means and covariances are better representatives of the original data and lead to parsimonious shape models. These two moments specify a normal probability model on shape classes, which are used for classifying test shapes into control and disease groups. We demonstrate the success of this model through improved random sampling and a higher classification performance. We study brain structures and present classification results for Attention Deficit Hyperactivity Disorder. Using the mean and covariance structure of the data, we are able to attain an 88% classification rate.

Keywords: Riemannian framework, parameterization invariance, shape statistics and models, classification, anatomical structures, ADHD.

1 Introduction

Shape is an important feature of anatomical objects and can be immensely useful in characterizing objects for the purpose of monitoring and characterization of a subject's health. Studying shapes of 3D anatomical structures in the brain is of particular interest because many diseases can potentially be linked to alterations of these shapes. In this paper we are focused on shape analysis of parametrized surfaces of anatomical objects, using a *Riemannian framework that allows comparison, matching, deformation, averaging, modeling, and classification* of observed shapes.

There have been many different representations of surfaces. Several groups have proposed methods for studying the shapes of surfaces by embedding them in volumes and deforming these volumes [9], [11]. While these methods are both prominent and pioneering in medical image analysis, they are typically computationally expensive. An alternative approach is based on manually-generated landmarks under the Kendall shape theory [7] and active shape models [5]. Others study 3D shape variabilities using level sets [16], curvature flows [10], or point cloud matching via the iterative closest

point algorithm [1]. Also, there has been remarkable success in the use of medial representations for shape analysis, especially in medical image analysis, see e.g. [2], [8].

However, the most natural representation for studying shapes of 3D objects seems to be using their boundary. In case of parameterized surfaces, there is an additional issue of handling the parameterization variability. Some papers, e.g. those using SPHARM [3] or SPHARM-PDM [19], tackle this problem by choosing a fixed parameterization. A large set of papers in the literature treat parameterization (or registration) as a pre-processing step [12], [4], [6]. In other words, they take a set of surfaces and use some energy function, such as the entropy or the minimum description length to register points across surfaces. Once the surfaces are registered, they are compared using standard procedures. There are several fundamental problems with this approach. Firstly, due to a registration procedure based on ensembles, the distance between any two shapes is dependent on the other shapes in the ensemble. Secondly, the registration and the comparison steps are typically disjoint and under different metrics. This certainly lacks the formalism needed to define proper distances and leads to sub-optimal registrations.

To the best of our knowledge, there are very few techniques in the literature on a Riemannian shape analysis of parameterized surfaces that can provide geodesic paths and be invariant to re-parameterization. To elaborate on this approach, let f_1 and f_2 denote two surfaces; f_1 and f_2 are elements of an appropriate space \mathcal{F} , which is made precise later, and let $\langle\langle \cdot, \cdot \rangle\rangle$ be the chosen Riemannian metric on \mathcal{F} . Then, under certain conditions, the geodesic distance between shapes of f_1 and f_2 will be given by:

$$\min_{\gamma, O} \left(\min_{\substack{F: [0, 1] \rightarrow \mathcal{F} \\ F(0) = f_1, F(1) = O(f_2 \circ \gamma)}} \left(\int_0^1 \langle\langle F_t(t), F_t(t) \rangle\rangle^{(1/2)} dt \right) \right). \quad (1)$$

(This assumes that translation and scaling variability has already been removed.) Here $F(t)$ is a parameterized path in \mathcal{F} , and the quantity $L[F] = \int_0^1 \langle\langle F_t(t), F_t(t) \rangle\rangle^{(1/2)} dt$ denotes the length of F . The minimization inside the brackets, thus, denotes the problem of finding a geodesic path (locally the shortest path) between the surfaces f_1 and $O(f_2 \circ \gamma)$, where O and γ stand for an arbitrary rotation and re-parameterization of f_2 , respectively. The minimization outside the bracket seeks the optimal rotation and re-parameterization of the second surface so as to best match it with the first surface. In simple words, the outside optimization solves the registration problem while the inside optimization solves for an optimal deformation (geodesic) and a formal distance (geodesic distance) between shapes. Thus, an important strength of this approach is that the **registration and comparison are solved jointly rather than sequentially**. Another strength is that this framework can be easily extended to different types of surfaces.

The rest of this paper is organized as follows. In Section 2, we present the framework and some examples of computing geodesics between toy surfaces, simple 3D objects and anatomical surfaces in the brain. In Section 3, we give a methodology for calculating shape statistics of surfaces such as the Karcher mean and the covariance. Finally, in Section 4, we report ADHD classification results using different techniques.

2 Novel Riemannian Framework

We will assume that the surfaces of interest are closed, i.e. no boundaries, and do not have holes. We will represent a surface with its embedding $f : \mathbb{S}^2 \rightarrow \mathbb{R}^3$. Let the set of all such parameterized surfaces be $\mathcal{F} = \{f : \mathbb{S}^2 \mapsto \mathbb{R}^3 \mid \int_{\mathbb{S}^2} \|f(s)\|^2 ds < \infty \text{ and } f \text{ is smooth}\}$, where ds is the standard Lebesgue measure on \mathbb{S}^2 . We choose the natural Riemannian structure in the tangent space, $T_f(\mathcal{F})$: for any two elements $m_1, m_2 \in T_f(\mathcal{F})$, define an inner product: $\langle m_1, m_2 \rangle = \int_{\mathbb{S}^2} \langle m_1(s), m_2(s) \rangle ds$, where the inner product inside the integral is the standard Euclidean product. The resulting \mathbb{L}^2 distance between any two points $f_1, f_2 \in \mathcal{F}$ is $(\int_{\mathbb{S}^2} \|f_1(s) - f_2(s)\|^2 ds)^{1/2}$. One can represent surfaces as elements of \mathcal{F} as stated here and use the \mathbb{L}^2 distance to compare shapes of surfaces. Although this framework is commonly used, it is not suitable for analyzing shapes of surfaces as it is not invariant to re-parameterizations. Let Γ be the set of all diffeomorphisms of \mathbb{S}^2 . This set will act as the re-parametrization group for surfaces, with its action given by composition: for a $\gamma \in \Gamma$, $f \in \mathcal{F}$, the re-parameterized surface is given by $f \circ \gamma$. It is easy to see that $\|f_1 \circ \gamma - f_2 \circ \gamma\| \neq \|f_1 - f_2\|$ in general, and that is a problem in using \mathbb{L}^2 distances on \mathcal{F} for shape analysis.

To define a new Riemannian metric, we first present a representation of surfaces first introduced in [13], [14]:

Definition 1. Define the mapping $Q : \mathcal{F} \rightarrow \mathbb{L}^2$ as $Q(f)(s) = \sqrt{\|a(s)\|} f(s)$, where $\|a(s)\| = \|f_x(s) \times f_y(s)\|$ is the area multiplication factor of f at $s = (x, y) \in \mathbb{S}^2$.

Here $\|\cdot\|$ denotes the standard 2-norm of a vector in \mathbb{R}^3 . The factor $\|a(s)\|$ is the ratio of infinitesimal areas of the surface at $f(s)$ and the domain at s . For any $f \in \mathcal{F}$, we will refer to $q(s) \equiv Q(f)(s)$ as the q -map of f . Since \mathcal{F} is a set of smooth surfaces, the set of all q -maps is a subset of \mathbb{L}^2 . The action of Γ on \mathbb{L}^2 , the space of q -maps, is given by $(q, \gamma) = \sqrt{J_\gamma}(q \circ \gamma)$. An important fact about the map Q is that if we re-parameterize a surface by γ and then obtain its q -map (Definition 1), or if we obtain its q -map (Definition 1) and then act by γ , the result will be the same.

We choose the natural \mathbb{L}^2 metric on the space of q -maps. That is, for any two elements $w_1, w_2 \in T_q(\mathbb{L}^2)$, define an inner product: $\langle w_1, w_2 \rangle = \int_{\mathbb{S}^2} \langle w_1(s), w_2(s) \rangle ds$. The Riemannian metric that we will use on \mathcal{F} is the pullback of the \mathbb{L}^2 metric from the space of q -maps. For this purpose, we first derive the differential of Q at f , denoted by $Q_{*,f}$. This is a linear mapping between tangent spaces $T_f(\mathcal{F})$ and \mathbb{L}^2 . For a tangent vector $v \in T_f(\mathcal{F})$, the mapping $Q_{*,f}$ is given by $Q_{*,f}(v) = \frac{1}{2\|a\|^{3/2}}(a \cdot a_v)f + \sqrt{\|a\|} v$. We use this differential of Q to define a Riemannian metric on \mathcal{F} as follows.

Definition 2. For any $f \in \mathcal{F}$ and any $v_1, v_2 \in T_f(\mathcal{F})$, define the inner product $\langle\langle v_1, v_2 \rangle\rangle_f \equiv \langle Q_{*,f}(v_1), Q_{*,f}(v_2) \rangle$, where the inner product on the right side is the standard inner product in \mathbb{L}^2 .

Substituting for $Q_{*,f}$, we obtain an expression for $\langle\langle v_1, v_2 \rangle\rangle_f$:

$$\left\langle \frac{1}{4\|a\|^3} (a \cdot a_{v_1})f, (a \cdot a_{v_2})f \right\rangle + \left\langle \frac{1}{2\|a\|} [(a \cdot a_{v_2})v_1 + (a \cdot a_{v_1})v_2], f \right\rangle + \langle \|a\|v_1, v_2 \rangle.$$

An important property of this Riemannian metric is that the action of Γ on \mathcal{F} is by isometries. With this induced metric, \mathcal{F} becomes a Riemannian manifold and we want to compute geodesic distances between two points, say f_1 and f_2 , in \mathcal{F} .

2.1 Pre-shape and Shape Space

Shape analysis of surfaces can be made invariant to certain global transformations by normalizing. The translation of surfaces is easily taken care of by centering: $f_{centered}(s) = f(s) - \frac{\int_{S^2} f(s) \|a(s)\| ds}{\int_{S^2} \|a(s)\| ds}$. Scaling can be removed by re-scaling all surfaces to have unit area, $f_{scaled}(s) = \frac{f(s)}{\sqrt{\int_{S^2} \|a(s)\| ds}}$. With a slight abuse of notation, we define the space of normalized surfaces as \mathcal{F} . \mathcal{F} forms the pre-shape space in our analysis. The remaining groups – rotation and re-parameterization – are dealt with differently, by removing them algebraically from the representation space. The rotation group $SO(3)$ acts on \mathcal{F} , $SO(3) \times \mathcal{F} \rightarrow \mathcal{F}$ according to $(O, f) = Of$ and the re-parameterization group Γ acts on \mathcal{F} , $\mathcal{F} \times \Gamma \rightarrow \mathcal{F}$ with $(f, \gamma) = (f \circ \gamma)$. Since the actions of $SO(3)$ and Γ on \mathcal{F} commute we can define an action of the product of the groups on \mathcal{F} . The orbit of a surface f is given by $[f] = \{O(f \circ \gamma) | O \in SO(3), \gamma \in \Gamma\}$ and the set of closures of all orbits is defined to be \mathcal{S} .

The next step is to define geodesic paths in \mathcal{F} and \mathcal{S} . We start with the case of \mathcal{F} ; the geodesic distance between any two points $f_1, f_2 \in \mathcal{F}$ is given by

$$d_{\mathcal{F}}(f_1, f_2) = \min_{\substack{F : [0, 1] \rightarrow \mathcal{F} \\ F(0) = f_1, F(1) = f_2}} \left(\int_0^1 \langle \langle F_t(t), F_t(t) \rangle \rangle^{(1/2)} dt \right).$$

We will use a path- straightening approach for solving this problem. Once we have an algorithm for finding geodesics in \mathcal{F} , we can obtain geodesics and geodesic lengths in \mathcal{S} by solving an additional minimization problem over $SO(3) \times \Gamma$, as stated in Eqn. [11](#).

2.2 Geodesics in the Pre-shape Space \mathcal{F}

Here we address the problem of finding geodesics between surfaces f_1 and f_2 in \mathcal{F} using a path-straightening approach [\[15\]](#). A similar approach for geodesics on shape spaces of closed curves was used in [\[18\]](#). The basic idea here is to connect f_1 and f_2 by any initial path, e.g. using a straight line under the \mathbb{L}^2 metric, and then iteratively “straighten” it until it becomes a geodesic. This update is performed using the gradient of an energy function as described next.

Let $F : [0, 1] \rightarrow \mathcal{F}$ denote a path in \mathcal{F} . The energy of the path F under the induced metric is defined to be:

$$E[F] = \int_0^1 \int_x \int_y \left[\frac{1}{4\|A\|^3} (A \cdot A_t)^2 (F \cdot F) + \frac{1}{\|A\|} (A \cdot A_t) (F_t \cdot F) + \|A\| (F_t \cdot F_t) \right] dy dx dt.$$

In this expression we have suppressed the argument t for all of the quantities. Also, we use $A(t)$ to imply $a(F(t))$. It is well known that a critical point of E is a geodesic path

in \mathcal{F} . To find a critical point, we are going to use the gradient ∇E_F which, in turn, is approximated using directional derivatives, $\nabla E_F(G)$, where $G \in \mathcal{G}$ is a perturbation of the path F . The expression for $\nabla E_F(G)$ is available analytically. Here \mathcal{G} denotes the set of all possible perturbations of F . We start with an initial path F and iteratively update it in the direction of ∇E until we arrive at the critical point F^* , which is the desired geodesic.

2.3 Geodesics in Shape Space \mathcal{S}

Now, we consider the problem of finding geodesics between surfaces in \mathcal{S} . This requires solving an additional optimization over the product group $SO(3) \times \Gamma$ by iterating between the following two steps.

1. **Rotation:** We can use a gradient approach for this optimization, but instead we will use an efficient albeit approximate technique based on Procrustes analysis. For a fixed $\gamma \in \Gamma$, the minimization over $SO(3)$ is performed as follows. Compute the 3×3 matrix $C = \int_{\mathbb{S}^2} f_1(s) f_2(s)^T ds$. Then, using the singular value decomposition $C = U \Sigma V^T$, we can define the optimal rotation as $O^* = UV^T$ (if the determinant of C is negative, the last column of V^T changes sign).
2. **Re-Parameterization:** In order to solve the optimization problem over Γ in Eqn. [11](#) we will use a gradient approach. Although this approach has an obvious limitation of converging to a local solution, it is still general enough to be applicable to general cost functions. Additionally, we have tried to circumvent the issue of a local solution by taking multiple initializations. This is similar to the gradient approach taken in [\[13\]](#), [\[14\]](#); the difference lies in the cost function used for optimization. In earlier papers, we address a problem of the type $\min_{\gamma \in \Gamma} \|q_1 - (q_2, \gamma)\|^2$, where q_1 and q_2 are q -maps of f_1 and f_2 and here we minimize a cost function of the type $d_{\mathcal{F}}(f_1, f_2 \circ \gamma)^2$, but the approach remains the same.

Computational Cost: We used the Matlab environment on an Intel Core 2 Duo CPU (2.66GHz, Windows XP). When we sample the surfaces with 2500 points and use 1400 basis elements for the path straightening optimization, the average computational cost for computing a geodesic in \mathcal{F} (10 iterations) is 95s and in \mathcal{S} is 490s.

In Figure [11](#) we display geodesics in \mathcal{F} and \mathcal{S} for toy surfaces with peaks at different locations. Along the geodesic in the shape space, the peaks smoothly move to the locations of the peaks on f_2 . On the other hand, along the geodesic in the pre-shape space, the peaks on f_1 are contracted while the peaks on f_2 are created. The difference

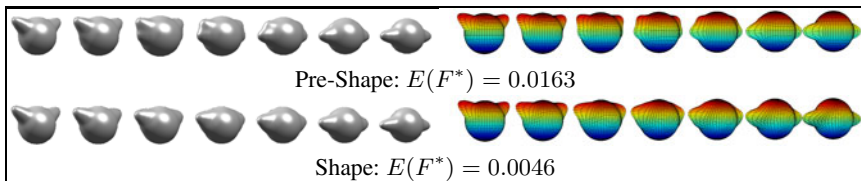


Fig. 1. Geodesic comparison for shapes of surfaces with dual peaks

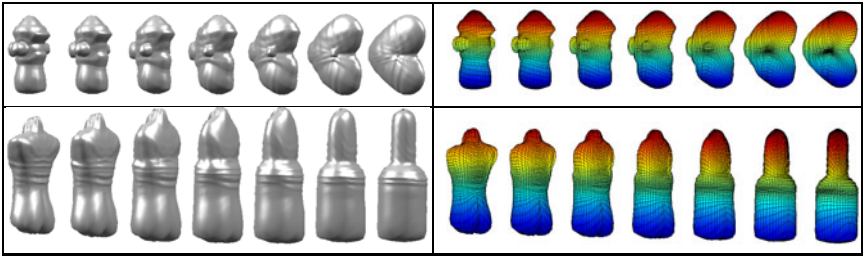


Fig. 2. Geodesics for shapes of a hydrant and a heart (top) and a torso and a bottle (bottom)

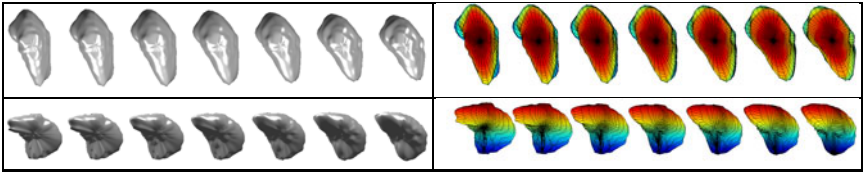


Fig. 3. Geodesics for left putamen and pallidus (top), and left caudate and right putamen (bottom). Left sides show rendered surfaces while the right sides show polygon meshes.

in these geodesics is due to improved matching in \mathcal{S} . It is important to note that the decrease in the geodesic path energy from \mathcal{F} to \mathcal{S} is significant. In Figure 2, we present geodesics in \mathcal{S} for the shapes of some simple 3D objects. In each example, the geodesic path is smooth and natural due to improved feature matching. In Figure 3, we present geodesics in the shape space between anatomical structures in the brain (left putamen and left pallidus, left caudate and right putamen). In the first example, f_1 and f_2 are quite similar and thus the evolution along the geodesic path is more subtle than in the previous examples. In the second example, the deformation is clearly visible.

3 Shape Statistics of Surfaces

In this section we will present tools for computing two important shape statistics – the Karcher mean and the covariance – for a set of surfaces.

3.1 Karcher Mean of Surfaces

To our knowledge there is no natural embedding of the shape space \mathcal{S} inside a Hilbert space. Thus, we cannot pursue the idea of computing extrinsic statistics for surfaces and will use intrinsic statistics. For obtaining an intrinsic sample mean for a set of surfaces, we use the Karcher mean. For the given surfaces $\{f_1, f_2, \dots, f_n\} \in \mathcal{F}$, the sample Karcher mean is given by: $\bar{f} = \operatorname{argmin}_{[f] \in \mathcal{S}} \sum_{i=1}^n d([f], [f_i])^2$. Here d denotes the length of the geodesic in the shape space between $[f]$ and $[f_i]$. A gradient-based approach for finding the Karcher mean is given in [7] and is not repeated here. We use that algorithm for finding the Karcher mean shape of a set of surfaces.

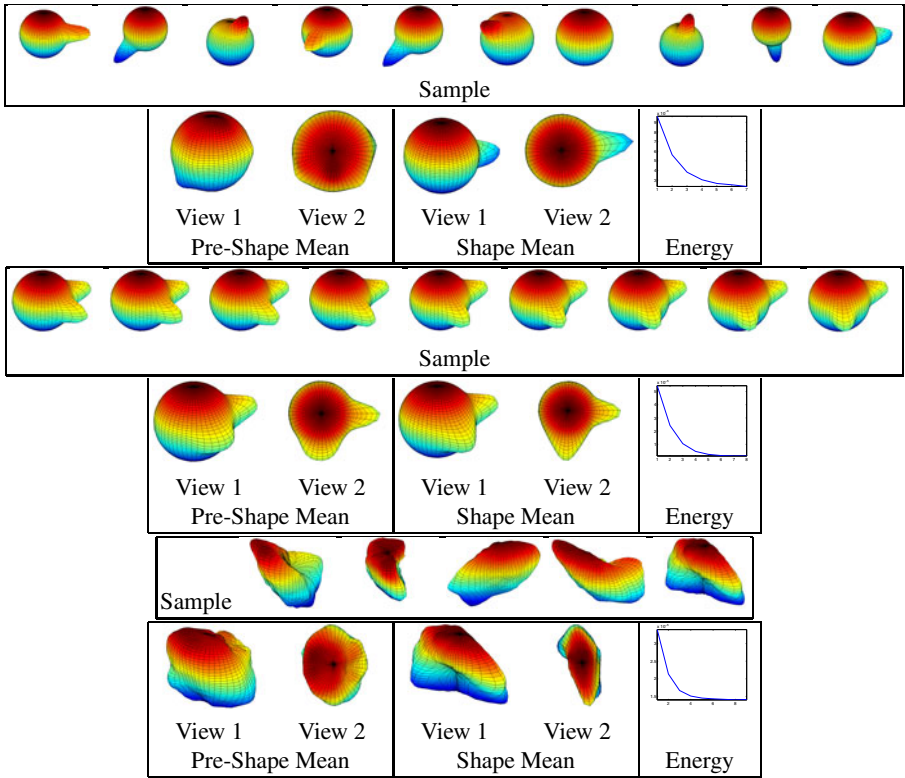


Fig. 4. Mean computation for a sample of surfaces with one peak (top), dual peaks (middle), and left putamen surfaces with random grid placement (bottom)

Next, we present some examples of Karcher mean shapes using toy objects and anatomical surfaces. For comparison, we also display $\tilde{f} = (1/n) \sum_{i=1}^n f_i$, i.e. without any rotational or re-parameterizational alignment. For each example we show the decrease in the gradient of the cost function during the computation of the Karcher mean. In the top part of Figure 4 we present means for ten unimodal surfaces with random peak placements on a sphere. The \tilde{f} surface has ten very small peaks at the locations of the peaks in the sample. On the other hand, the mean in \mathcal{S} has one peak, which is of the same size as all of the peaks in the sample. In this simple example one can clearly see the effect of feature preservation due to rotational and re-parameterizational alignment. In the middle part of Figure 4 we present mean shapes of nine surfaces with dual peaks. We note that \tilde{f} has one peak aligned (at the location of the common peak in the sample) and one very wide and small peak. The mean in \mathcal{S} has two peaks due to a crisp alignment of peaks and thus is a much better representative of the sample. In the bottom part of Figure 4 we present results of mean computations for five left putamen surfaces with different parameterizations (different coordinate systems on the same surface). \tilde{f} does not have the shape of a left putamen. On the other hand, the \bar{f} surface has the original correct shape. The optimization over the re-parameterization group plays a very important role in this example.

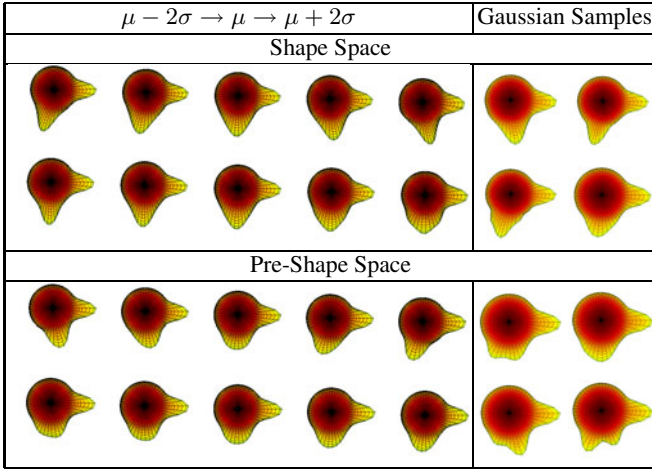


Fig. 5. Principal directions of variation and random samples from a Gaussian model

3.2 Estimation of the Karcher Covariance

Once the sample Karcher mean has been computed, the evaluation of the Karcher covariance is performed as follows. This task is difficult because: (1) although \mathcal{F} is a vector space, we are using a non-standard metric $\langle\langle \cdot, \cdot \rangle\rangle$ and (ii) the shape space of interest is actually a quotient space of \mathcal{F} . To compute the covariance, we first find the shooting vectors from the mean \bar{f} to the shape orbits of each of the given surfaces $[f_i]$. That is, let $\nu_i = \dot{F}^*(0)$, where $F^*(0) = \bar{f}$ and $F^*(1) = O_i^*(f_i \circ \gamma_i^*)$, $i = 1, 2, \dots, n$. We then perform principal component analysis by applying the Gram-Schmidt procedure (under the chosen metric $\langle\langle \cdot, \cdot \rangle\rangle$), to generate an orthonormal basis $\{B_j | j = 1, \dots, k\}$, $k \leq n$, of the observed $\{\nu_i\}$ in the vector space $T_{\bar{f}}(\mathcal{F})$. We project each of the vectors ν_i onto this orthonormal basis using $\nu_i \approx \sum_{j=1}^k c_{i,j} B_j$, where $c_{i,j} = \langle\langle \nu_i, B_j \rangle\rangle$. Now each original surface can simply be represented using the coefficient vector $c_i = \{c_{i,j}\}$. Then, the covariance matrix can be computed in the coefficient space using $K = (1/(n - 1)) \sum_{i=1}^n c_i c_i^T \in \mathbb{R}^{k \times k}$. We can use the SVD of K to determine the principal directions of variation in the given data. For example, if $u \in \mathbb{R}^k$ corresponds to a principal singular vector of K , then the corresponding tangent vector in $T_{\bar{f}}(\mathcal{S})$ is given by $\sum_{j=1}^k u_j B_j$. Hence, a concatenation of mappings $c \mapsto \sum_{j=1}^k c_j B_j \mapsto f = \exp_{\bar{f}}(\sum_{j=1}^k u_j B_j)$ provides a transformation of a coefficient vector c into a shape f . We can impose a shape model on a class by imposing a Gaussian model on its coefficients $c \sim N(0, K)$, where K is the sample covariance matrix of that class.

We used the data and means presented in the middle example of Figure 4 to calculate the principal directions of variation in the sample and to draw some random samples from a Gaussian model mentioned above for both the pre-shape and shape spaces. The results are displayed in Figure 5. Some of the samples drawn from the model in the pre-shape space are invalid as they have three rather than two peaks. On the other hand, in the shape space, all of the samples have two sharp peaks. This is due to improved

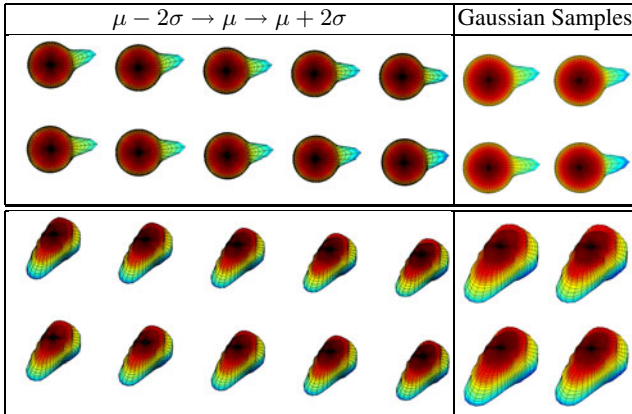


Fig. 6. Principal directions of variation and random samples in \mathcal{S} from a Gaussian distribution for surfaces with one peak (top) and left pallidus surfaces (bottom)

feature matching in \mathcal{S} . In the shape space, the principal directions of variation simply reflect the placement of the peaks on each surface in the data. Next, we computed the principal directions of variation and some random samples based on a Gaussian model for the data in the top example of Figure 4. We show the results in the top portion of Figure 6. In the last example, presented in the bottom portion of Figure 6, we used 15 left pallidus structures. We note that our model is compact and all of the random samples are valid structures.

4 ADHD Classification

In this section we present disease classification results for persons with Attention Deficit Hyperactivity Disorder and controls. The clinical data sets used in this work were T1 weighted brain magnetic resonance images of young adults of ages between 18 and 21. These subjects were recruited from the Detroit Fetal Alcohol and Drug Exposure Cohort. Among the 34 subjects studied, 19 were diagnosed with ADHD and the remaining 15 were controls (non-ADHD). We consider six different anatomical structures for classification (L. and R. Pallidus, L. and R. Putamen, L. and R. Parietal Lobe). We chose these structures based on previous literature that suggests that they are major players in ADHD development [13], [17], [20].

We perform disease classification using several different distance functions. In the third row of Table 1, we report the leave-one-out nearest neighbor (LOO NN) classification rate using all pairwise single structure distances (called d_1) between q -maps [13]. It is very expensive to compute the entire distance matrix and thus we would like to be able to perform classification using statistical summaries such as the mean and covariance. In addition, the covariance structure of the sample may provide useful information for disease detection. Thus, for each of the six anatomical structures, we first compute the mean structure for the disease and control groups using the framework introduced in

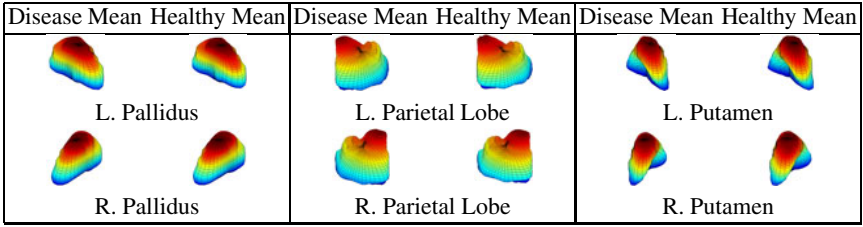


Fig. 7. Means for six anatomical structures in the ADHD study

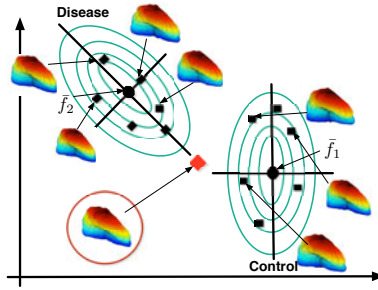


Fig. 8. Role of covariance in classification

Table 1. ADHD classification rate using five different measures

Distance Type	Structure Side	Pallidus		Parietal Lobe		Putamen	
		Left	Right	Left	Right	Left	Right
LOO NN distance	ICP(%)	67.6	55.9	61.8	67.6	61.8	47.1
LOO NN distance	SPHARM-PDM(%)	44.1	52.9	-	-	50.0	55.9
LOO NN shape distance	d_1 (%)	76.5	61.8	70.6	67.6	82.4	67.6
Distance to mean	d_2 (%)	70.6	64.7	67.6	67.6	70.6	64.7
Covariance-adjusted distance	d_3 (%)	88.2	67.6	67.6	70.6	82.4	82.4

this paper. In Figure 7 we display the means for all six anatomical structures for ADHD cases and controls. It is not easy to discern differences between the two means for each structure by looking at them. First, we classify each subject as a case or control based on the geodesic distance to the disease and control means (called d_2). We also form a covariance for each of the samples (six structures, two disease groups) in a leave-one-out manner using the orthonormal basis coefficients described in the previous section. Given this information, we can compute a covariance-adjusted distance to the mean for each subject using the underlying covariance structure as $d_3 \equiv \frac{1}{2}c'K_c^{-1}c + \frac{\|\nu^\perp\|}{\epsilon}$, where $\nu^\perp = \nu - \sum_{j=1}^k c_j B_j$ and $\epsilon = \min(\text{diag}(K))$. Once again we perform nearest neighbor classification under d_3 . For comparison purposes, we also report classification rates using the ICP (iterative closest point) algorithm and SPHARM-PDM under the

leave-one-out nearest neighbor classifier. SPHARM-PDM failed to provide a parameterization for the parietal lobe data and thus those classification rates are not reported. The results are presented in Table 1. The classification rates using (d_1) are better than the ones using distances to the means only. Nonetheless, the decrease in classification rates using the mean only (d_2) is not large in general and in the case of the right pallidus it actually increases. The best performance is attained when we use the mean and the covariance structure of the given data. In fact, using this measure allows us to improve the classification rates for most of the six structures, with the left pallidus performing best at 88.2%. We also see great improvement in the right putamen classification rate. In Figure 8, we show the benefits of using the covariance structure in disease classification. Using the mean only distances (d_2) , the test shape (in red) is closer to the control group, but after including the covariance structure in the distance calculation (d_3) , this test shape is closer to the disease group. In addition, all of our distance measures outperform standard surface comparison techniques such as ICP or SPHARM-PDM.

5 Conclusion

We have presented a Riemannian framework for parameterization-invariant statistical analysis of surfaces. Most previous methods view registration as a pre-processing step, while one of our main contributions is to involve the registration directly in our framework. We have used a novel Riemannian metric for comparison of surfaces and computation of the first two statistical moments, the Karcher mean and Karcher covariance. We utilize this metric to demonstrate the computation of geodesics between real 3D objects, toy surfaces and anatomical shapes. In addition, we showcase the benefits of our method by comparing statistical models developed within our framework to those that do not involve parameterization invariance. Finally, we use the computed statistical summaries for disease classification in a previously explored ADHD study, where we are able to significantly improve the rates by involving both mean and covariance information. We find that the left pallidus structure provides the highest disease classification rate of 88.2%.

Acknowledgements. This research was supported by AFOSR FA9550-06-1-0324, ONR N00014-09-1-0664, NSF DMS-0915003 (AS) and NIH/NIDA R21-DA021034 (MJA).

References

1. Almhdie, A., Léger, C., Deriche, M., Lédée, R.: 3D registration using a new implementation of the ICP algorithm based on a comprehensive lookup matrix: Application to medical imaging. *Pattern Recognition Letters* 28(12), 1523–1533 (2007)
2. Bouix, S., Pruessner, J.C., Collins, D.L., Siddiqi, K.: Hippocampal shape analysis using medial surfaces. *Neuroimage* 25, 1077–1089 (2001)
3. Brechbühler, C., Gerig, G., Kübler, O.: Parameterization of closed surfaces for 3D shape description. *Computer Vision and Image Understanding* 61(2), 154–170 (1995)
4. Cates, J., Meyer, M., Fletcher, P., Whitaker, R.: Entropy-based particle systems for shape correspondence. In: *MICCAI Mathematical Foundations of Computational Anatomy*, pp. 90–99 (2006)

5. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
6. Davies, R., Twining, C., Cootes, T., Taylor, C.: Building 3-d statistical shape models by direct optimization. *IEEE Medical Imaging* 29(4), 961–981 (2010)
7. Dryden, I.L., Mardia, K.: *Statistical Shape Analysis*. John Wiley & Son, Chichester (1998)
8. Gorczowski, K., Styner, M., Jeong, J., Marron, J., Piven, J., Hazlett, H., Pizer, S., Gerig, G.: Multi-object analysis of volume, pose, and shape using statistical discrimination. *IEEE Pattern Analysis and Machine Intelligence* 32(4), 652–666 (2010)
9. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics* LVI(4), 617–694 (1998)
10. Gu, X., Wang, S., Kim, J., Zeng, Y., Wang, Y., Qin, H., Samaras, D.: Ricci flow for 3D shape analysis. In: *IEEE International Conference on Computer Vision* (2007)
11. Joshi, S., Miller, M., Grenander, U.: On the geometry and shape of brain sub-manifolds. *Pattern Recognition and Artificial Intelligence* 11, 1317–1343 (1997)
12. van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D.: A survey on shape correspondence. *Eurographics State-of-the-Art Report* (2010)
13. Kurtek, S., Klassen, E., Ding, Z., Jacobson, S., Jacobson, J., Avison, M., Srivastava, A.: Parameterization-invariant shape comparisons of anatomical surfaces. *IEEE Medical Imaging* 30(3), 849–858 (2011)
14. Kurtek, S., Klassen, E., Ding, Z., Srivastava, A.: A novel Riemannian framework for shape analysis of 3D objects. In: *IEEE Computer Vision and Pattern Recognition*, pp. 1625–1632 (2010)
15. Kurtek, S., Klassen, E., Gore, J., Ding, Z., Srivastava, A.: Elastic geodesic paths in shape space of parametrized surfaces. *IEEE Pattern Analysis and Machine Intelligence In Review* (2010)
16. Malladi, R., Sethian, J., Vemuri, B.: A fast level set based algorithm for topology-independent shape modeling. *J. of Math. Imaging and Vision* 6, 269–290 (1996)
17. Qiu, A., Crocetti, D., Adler, M., Mahone, E.M., Denckla, M.B., Miller, M.I., Mostofsky, S.H.: Basal ganglia volume and shape in children with attention deficit hyperactivity disorder. *Am. J. Psychiatry* 166(1), 74–82 (2009)
18. Srivastava, A., Klassen, E., Joshi, S.H., Jermyn, I.H.: Shape analysis of elastic curves in Euclidean spaces. *IEEE Pattern Analysis and Machine Intelligence* 99(PrePrints) (2010)
19. Styner, M., Oguz, I., Xu, S., Brechbuhler, C., Pantazis, D., Levitt, J., Shenton, M., Gerig, G.: Framework for the statistical shape analysis of brain structures using SPHARM-PDM. In: *MICCAI Open Science Workshop* (2006)
20. Teicher, M., Anderson, C., Polcari, A., Glod, C., Mass, L., Renshaw, P.: Functional deficits in basal ganglia of children with attention-deficit/hyperactivity disorder shown with functional magnetic resonance imagery relaxometry. *Nature Medicine* 6(4), 470–473 (2000)

On the Extraction of Topologically Correct Thickness Measurements Using Khalimsky's Cubic Complex

M. Jorge Cardoso, Matthew J. Clarkson, Marc Modat,
and Sebastien Ourselin

Centre for Medical Image Computing (CMIC), University College London, UK

Abstract. The extraction of thickness measurements from shapes with spherical topology has been an active area of research in medical imaging. Measuring the thickness of structures from automatic probabilistic segmentations is normally hindered by the presence of noise, partial volume (PV) effects and the limited resolution of medical images. Also, the complexity of certain shapes, like the highly convoluted and PV corrupted cerebral cortex, results in topologically inconsistent measurements. In this paper we explore the use of Khalimsky's cubic complex for the extraction of topologically correct thickness measurements from probabilistic or fuzzy segmentations without explicit parametrisation of the edge. A sequence of element collapse operations is used to correct the topology of the segmentation. The Laplace equation is then solved between multiple equipotential lines and the thickness measured with an ordered upwind differencing method using an anisotropic grid with the probabilistic segmentation as a speed function. Experiments performed on digital phantoms show that the proposed method obtains topologically correct thickness measurements with an increase in accuracy when compared to two well established techniques. Furthermore, quantitative analysis on brain MRI data showed that the proposed algorithm is able to retrieve expected group differences between the cortical thickness of AD patients and controls with high statistical significance.

Keywords: Khalimsky, PDE, Laplacian, thickness.

1 Introduction

The extraction of 3D consistent measurements of thickness from anatomical structures is an important post processing step in medical image analysis. For example, measurements of myocardial thickness are an important component for the diagnosis of cardiac disease and for risk stratification, and are of high prognostic value in patients with nonischemic cardiomyopathy and infiltrative diseases. In addition, changes in the thickness of the cerebral cortex are also of interest in various diseases such as Alzheimer's [10] and Huntington's [14] disease. The automatic extraction of thickness measurements from anatomical structures has the potential to provide a biomarker for diagnosis and disease progression.

However, the reliable extraction of sub-voxel accurate measurements of thickness from probabilistic segmentations is still an unsolved problem.

Thickness estimation methods can be separated into surface-based and voxel-based techniques. Surface based methods [4,8] typically fit a triangulated mesh to the region of interest, making them computationally expensive, especially if a requirement is to maintain topological constraints. Additionally, the parametrisation of the surface can be complex and curvature constraints can bias the thickness measurements [15]. The thickness measurement can also be biased by the choice of the smoothness parameters as these will impact the ability of the algorithm to fit a mesh to highly convoluted structures [15].

Voxel-based methods on the other side extract the value of thickness directly from the voxel grid and are computationally very efficient, however, their accuracy is critically dependant on the image resolution and the quality of the segmentation. Furthermore, topological problems might occur in highly convoluted areas because of the limited knowledge about spatial information. Overall, voxel-based methods can be clustered into 3 subgroups: mathematical morphology [11], partial differential equation [5,6,16,13,1] and line integral based [2]. Mathematical morphology based methods use a combination of skeletonisation and region growing techniques in order to calculate the minimal Euclidean distance between points. These methods work on a binarised version of the segmentation thus having their accuracy limited by the voxel size. Partial differential equation (PDE) based methods solve the Laplace equation between the inner and outer surfaces as if they were charged conductors, resulting in isopotential electric field lines between them. The thickness is then equal to the sum of the lengths of the normals to these isolines. The normals to the isolines that go from the inner to the outer surfaces are guaranteed to be diffeomorphic, as they are bijective and never cross. However, this method requires both cortical surfaces to be homotopic to a sphere in order to avoid saddle points in the Laplacian field and collisions on the advancing front. Even though this is a necessary condition, the Laplace equation based methods are being used without enforcing these topological constraints [1]. Finally, line integral based methods calculate thickness of the structure of interest by finding the direction that minimises the line integrals over its probabilistic segmentation at each position of the 3D volume. The accuracy of this method is very high for simple shapes because it works on the probabilistic segmentation and not on a binarised grid, however, the choice of stopping criteria becomes critical in highly convoluted areas like the cerebral cortex.

The aim of this work is to combine the features of all the above voxel-based methods in a unified, fully automated **K**halimsky based **T**hickness **E**stimation algorithm, called **KTE**, that is topologically correct and partial-volume aware. We use the properties of the Khalimsky grid and an iterative set of element collapse operations to correct the topology of the segmentation. We then extend the single Laplacian field method to a multi-stage Laplacian in order to encompass the partial-volume containing areas and we modify the classic PDE

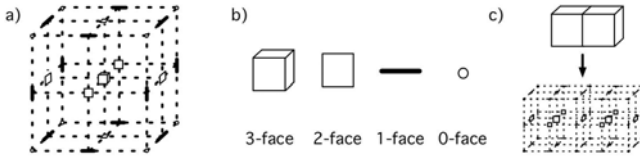


Fig. 1. Example of: a) a Cubic Complex, b) its m -face and c) the object in the cubical space associated with neighbouring voxels

proposed by [16] to have a spatially varying speed function that is dependent on the probabilistic segmentation.

2 Method

2.1 Topology Preserving and the Khalimsky's Cubic Complex

Topology-preserving operations are used in many image analysis applications in order to transform an object while leaving its topological characteristics unchanged. A typical topology-preserving transformation based on simple point deletion can be described as follows. Let the object be a set X of points in the voxel grid, and a subset K of X be a constraint set. At each iteration, choose a simple point x_i in X_0 but that is not in K according to a priority function (e.g. euclidean distance) and set $X_{i+1} = X_i \setminus x_i$. Iterate until no other point can be removed. This process results in a homotopic skeleton of X constrained by K . Notwithstanding the simplicity, thinning algorithms that work in the voxel space have some well-described problems regarding the minimality of the set X_n and the existence of lumps [12,3]. Abstract complexes, like the Khalimsky space [7] have been promoted, in particular by V. Kovalevsky [9], to provide a sound topological basis for image analysis. Intuitively, a cubic complex can be seen as a space where every voxel is represented by a structure composed of a set of elements having various dimensions (e.g. cubes, squares, edges, vertices) put together according to some rules. For illustrative purposes see Fig. 1. Let \mathbb{Z}^3 denote the original 3D cartesian voxel grid and \mathbb{F}^3 the set composed of all m -faces of \mathbb{Z}^3 for all $m = 0$ to 3. An m -face of \mathbb{Z}^n is called a point if $m = 0$, a line $m = 1$, a square if $m = 2$ and a cube if $m = 3$. Please refer to [12,3] for a complete formal description of the cubic complex framework. For our application, and as a voxel is equivalent to an elementary cube, an easy correspondence can be made between the voxel classical view and the framework of cubical complexes. In the rest of the paper, we use the term voxel to mean a 3-cell. This abstract space provides a sound basis for digital topology and topology correction but also, due to the dimensionality of its elements, an interesting framework for thickness measurement extraction. Within this framework, one is able to connect different areas with infinitesimally thin surfaces, correcting the topology of the segmentations with minimal influence on the thickness estimation step.

2.2 Collapse Operation and Topology Correction

Collapse operation: The collapse operation is a well-known operation that preserves topology. Let F be a complex in \mathbb{F}^3 and let $f \in F$. If g is a *proper face* of F and f is the only face of F which *strictly includes* g , then the pair (f, g) is considered a free pair of F . Here, the dimension of the face g is $\dim(f) - 1$. As the pair (f, g) is a free pair, it can be removed without any change in the topology of F by an operation called *elementary collapse* defined as $F \setminus \{f, g\}$. This operation is the \mathbb{F}^3 equivalent to the removal of a simple point in \mathbb{Z}^3 .

This operation can then be used to collapse a shape into another while maintaining the same topology. In a more general form, let the operation that collapses the set F onto the constraint set K , with a priority function P be denoted by $\text{Col}(F, K, P)$, where a lower value of the priority function corresponds to a higher priority.

Topology correction of spherical objects: Starting from a probabilistic segmentation of a spherical object, let S_{In} , S_{Out} and S_{Obj} be probabilistic segmentations of the internal area, external area and the object of interest. Thus, S_{In_i} , S_{Out_i} and S_{Obj_i} represents the probability for voxel i in \mathbb{Z}^3 to belong to each corresponding class. These probabilistic segmentations can be converted from \mathbb{Z}^3 to \mathbb{F}^3 by setting the value at each m -face equal to the average of their $m + 1$ surrounding faces, for all $m \in 0, 1, 2$ and the value at the 3-face as the original value for the corresponding voxel in \mathbb{Z}^3 . For example, the value at the 0-face, $S_{Obj_{(i,m)}}$ with $m = 0$ will be equal to the average of its surrounding 1-face. This is equivalent to an ordered linear interpolation.

The topology correction for spherical objects will be done in four steps: in step 1, starting from an object with a Euler characteristic of 1, in this case a box bounding the object in \mathbb{F}^3 defined as $F_{Box_{(i,m)}}$, set $F_{In_{pv}} = \text{Col}(F_{Box}, K_{In_{pv}}, S_{In})$ with the constraint set $K_{In_{pv}} = (S_{In} > t)^-$ as the topological closure (represented by the $-$ operation) of all the elements of $S_{In_{(i,m)}}$ above a certain threshold t . This constraint set contains all the voxels that have more than t probability of belonging to the inner area. On an ideal segmentation, the value of threshold t would be 0, however, due to noise, t is set to 0.05 for probabilistic segmentation on standard clinical datasets. This operation is the Khalimsky equivalent of shrink wrapping the constraint set without a change in topology, resulting in a set that has all the voxels that have a probability above t of belonging to the inner area while maintaining its initial Euler characteristic of 1.

In step 2, set $F_{In_{pure}} = \text{Col}(F_{In_{pv}}, K_{In_{pure}}, S_{In})$ with $K_{In_{pure}} = (S_{In} > (1 - t))^-$ as the constraint set. This will continue the collapsing of the previous structure $F_{In_{pv}}$ ordered by the probability of belonging to the inner area S_{In} and will result in a set that contains all the voxels that have a probability above $(1 - t)$ of belonging to the inner area while maintaining its initial Euler characteristic of 1.

In order to get a topologically corrected $F_{Out_{pv}}$, the set $F_{In_{pure}}$ has to be inverted. However, in a cubic complex space, the inversion of a complex is not a complex. A topological closure could be used to transform the inverted set

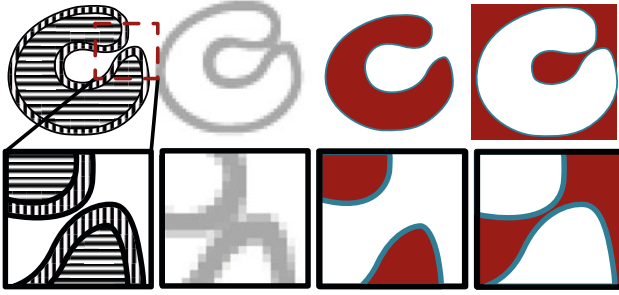


Fig. 2. From left to right: An example object with the outside in white, the inside in horizontal lines and the object of interest with vertical lines; A rasterised version of the object with PV effect (notice the erroneously connected structure due to PV); the result of the bounding box collapse after stage 1 in blue and after stage 2 in red; the result of the collapse of the inverse of the topology corrected pure inside area after stage 3 in blue and after stage 4 in red;

into a complex, but this operation would change the topological characteristics of the inverted set. Under the assumption of smoothness for the probabilities S and consequently smoothness of the edge of $F_{In_{pure}}$, one can remove all the *incomplete* m -face from the complex and still leave the topological characteristics unchanged. This operation consists of removing all the external m -faces that are not limited by all their n -face elements, with $n < m$, in an ordered manner from $m = 3$ to 0. For example, a 2-face (square) will be removed if any of its surrounding 1-faces (lines) and 0-faces (points) do not belong to the inverted set. This operation can be described as an opening on the incomplete m -faces of the set.

Let the inverted set of $F_{In_{pure}}$ according to the above described method be represented by $F_{In_{pure}}^{-1}$. Then set $F_{Out_{pv}} = Col(F_{In_{pure}}^{-1}, K_{Out_{pv}}, S_{Out})$ with the constraint set $K_{Out_{pv}} = (S_{In} > t)^-$. This will collapse the inverted set $F_{In_{pure}}^{-1}$ ordered by the probability of belonging to the outer area (S_{Out}) and will result in a set that contains all the voxels that have a probability above t of belonging to the outer area. Due to the inversion, and assuming that the foreground object is limited by the image boundaries, its Euler characteristic will now be equal to 2 (spherical topology).

Finally, in step 4, set $F_{Out_{pure}} = Col(F_{Out_{pv}}, K_{Out_{pure}}, S_{Out})$ with the constraint set $K_{Out_{pure}} = (S_{Out} > (1 - t))^-$. This will continue the collapsing of the previous structure $F_{Out_{pv}}$ ordered by the probability of belonging to the outer area (S_{Out}) and will result in a set that contains all the voxels that have a probability above $(1 - t)$ of belonging to the outer area while maintaining its previous Euler characteristic of 2.

In short, this sequence of operations will produce a $F_{In_{pure}}$ and a $F_{In_{pv}}$ set with a Euler characteristic of 1 and a $F_{Out_{pure}}$ and a $F_{Out_{pv}}$ set with a Euler characteristic of 2. Please refer to Fig. 2 for a visual representation of the results of the pipeline.

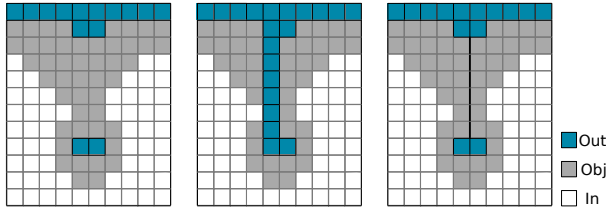


Fig. 3. Left: Disconnected $K_{Out_{pure}}$, Centre: Voxel based skeletonisation of $S_{Out} \cup S_{Obj}$, constrained by K_{Out} . Right: The same collapse procedure in the Khalimsky space. Notice the change of topology of the S_{Obj} structure when using a normal skeleton.

Apart from the minimality of the skeletons and the existence of lumps, the cubic complex has several other benefits over voxel based topology correction as the same sequence of steps will result in incorrect topologies when the thickness of the structure of interest is close to the image resolution. The problem is represented visually in Fig. 3.

2.3 Multi-stage Laplace Equation and Thickness Measurement

In order to calculate thickness, a unique association between two points is required. We use the Laplace equation, a second order partial differential equation (PDE), solved between two enclosed boundaries Ω and Ω' . The classic equation takes the form

$$\nabla^2 \phi = 0 \tag{1}$$

with the value at ϕ_Ω and $\phi_{\Omega'}$ set up as boundary conditions. For this work, instead of a single Laplacian field for all the pure voxels as in [5,6,16,13,1], a multiple Laplacian field is solved. The idea behind this approach came from the fact that methods that integrate through PV using ray casting [1] have problems estimating surface normals correctly. The use of a topologically correct extended Laplacian field implicitly solves this problem as the direction of integration becomes known.

The first stage, represented in Fig. 4 as a light blue colour, is solved between the $F_{In_{pure}}$ and the $F_{In_{pv}}$ edges, where the outer (0, 1, 2)-face of the respective complexes are fixed and set to 0 and 100. The second stage, represented in Fig. 4 as a white colour, is solved between the $F_{In_{pv}}$ and the $F_{Out_{pv}}$ edges, where the outer (0, 1, 2)-face of the respective complexes are set to 100 and 200. Finally, in the third stage, represented in Fig. 4 as a light grey colour, is solved between the $F_{Out_{pv}}$ and the $F_{Out_{pure}}$ edges, where the outer (0, 1, 2)-face of the respective complexes are set to 200 and 300. By solving the Laplace equation between these four equipotential lines, a smooth transition field arises. Note that the pattern of streamlines and isolines is independent of the choice of boundary condition voltages, as long as the four voltages, here set to 0, 100, 200 and 300, are different, equally spaced and increasing. Considering that the original image might have

an anisotropic voxel grid, and assuming that the value of the Laplacian in each element is the value at the centre of mass of the element, the finite difference approximation can be seen as:

$$\phi_{i+1}(x, y, z) = \frac{s_y^2 s_z^2 D(x \pm s_x, y, z) + s_x^2 s_z^2 D(x, y \pm s_y, z) + s_x^2 s_y^2 D(x, y, z \pm s_z)}{2(s_y^2 s_z^2 + s_x^2 s_z^2 + s_x^2 s_y^2)} \tag{2}$$

with $D(x \pm s_x, y, z) = \phi_i(x + s_x, y, z) + \phi_i(x - s_x, y, z)$ and similarly for the y and z directions. Here s_x, s_y and s_z equal to the distance between the centre of mass of neighbouring m-faces in the x, y and z respectively. In this case, s_x, s_y and s_z will be equal to half the original voxel size in the x, y and z directions on the \mathbb{Z}^3 image. From the resultant Laplacian field, the normals to the direction of the Laplacian isolines, denoted by T , are calculated using finite differences.

In order to measure thickness, the length of the streamlines between the inner and outer surface has to be measured by integrating the vector field T . This can be achieved by the use of two first order linear PDEs. Instead of the basic form proposed by Yezzi [16], where the speed of the advancing front is assumed to be 1, we use a more generalised form of the PDE. Here,

$$\nabla L \cdot T = f \tag{3}$$

for an unknown function L and assuming that T and f are known. In our case, and differently from [6,16,13,11], the value of f will be spatially varying and equal to the probability of belonging to the object, $f = S_{Obj}$. This value will act as *time cost* and will make the value of L equivalent to the time of arrival in a level-set framework. Even though collisions of the advancing front might exist, they are not a problem due to the upwind nature of the integration and the existence of the vector field T . Let $L_{0(x,y,z)}$ be a function that measures the time of arrival (arc length of the streamline according to the time cost f) from the boundary of set F_{In_pure} to a point in the object and $L_{1(x,y,z)}$ be the time of arrival from the boundary of set F_{Out_pure} to the point in the object. Thus, similarly to [16],

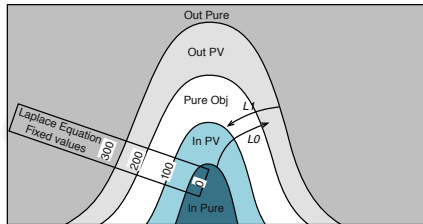


Fig. 4. Multi-Stage Laplace equation: The lines between each class are set to a fixed potential and the Laplace equation is solved. The distance L_0 and L_1 is integrated from opposite sides of the object, following the perpendicular to Laplacian field isolines.

$$L0_{(x,y,z)} = \frac{1}{(a_y a_z |T_x| + a_x a_z |T_y| + a_x a_y |T_z|)} \left[f_{(x,y,z)} a_x a_y a_z + a_y a_z |T_x| * L0_{(x \mp a_x, y, z)} + a_x a_z |T_y| * L0_{(x, y \mp a_y, z)} + a_x a_y |T_z| * L0_{(x, y, z \mp a_z)} \right] \tag{4}$$

and equivalently for $L1$ by replacing the \mp with \pm . Here, a_x , a_y and a_z are the size of the m -face in the x, y, z dimension respectively, and $f_{(x,y,z)} = S_{Obj}(x, y, z)$. The notation $x \pm a_x$ and $x \mp a_x$ and similarly for y, z, a_y and a_z are defined as

$$x \pm a_x = \begin{cases} x + a_x & |Tx| > 0 \\ x - a_x & |Tx| < 0 \end{cases} \quad x \mp a_x = \begin{cases} x - a_x & |Tx| > 0 \\ x + a_x & |Tx| < 0 \end{cases} \tag{5}$$

As some elements are infinitesimally thin and this would cause numerical instability, instead of the real structure size, a will be defined as the size of an influence zone. The influence area size will be equal to αs , with $\alpha \in]0, 1]$, in the infinitesimally thin direction of an m -face and equal to $(2 - \alpha)s$ in the other directions. For example, a 2-face in the xy plane will have $a_x = (2 - \alpha)s_x$, $a_y = (2 - \alpha)s_y$ and $a_z = \alpha s_z$. If $\alpha = 1$, the above equation will reduce to the classic anisotropic form and the behaviour will be equivalent to considering each element as a normal \mathbb{Z}^3 voxel, while α close to 0 will be the equivalent to having an infinitesimally thin element for every m -face with $m < 3$. Please refer to Fig. 5 for a visual explanation of the influence areas in a 2D complex.

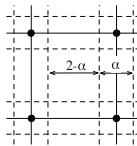


Fig. 5. A 2D complex represented by the full lines and the points. The influence area represented in dashed lines and determined by the value of α .

The final value of thickness is then defined as $Thick = L0 + L1$. In order to un-bias any further statistical analysis, the value of thickness should only be calculated either at the mid-harmonic point within the object, where $L0 = L1$, or at one of the mid-isopotential lines. The ribbon containing the thickness measurements will thus have spherical topology (Euler characteristic of 2).

3 Experiments and Results

The experimental section of this paper is divided into two subsections. First, the performance of the proposed algorithm is assessed against a digital phantom with known ground truth thickness and the results compared to two state-of-the-art methods. The proposed method is then applied to brain MRI data in order

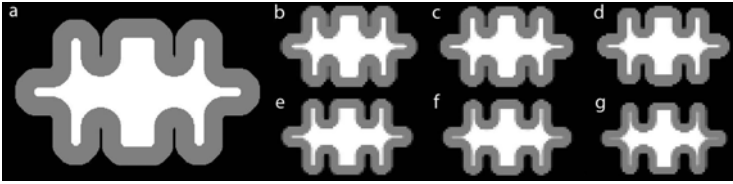


Fig. 6. a) A slice of the 3D simulated high resolution phantom with $4.8mm$ thickness. b) The same slice in the down-sampled version of the phantom with thickness 4.8 , c) 4.6 , d) 4.4 , e) 4.2 , f) 4.0 and g) $3.6mm$, equivalent to a thinning procedure of $0.4mm$ per year, sampled at year 0, 0.5, 1, 1.5, 2 and 3.

to assess group separation in terms of cortical thickness between Alzheimer's disease diagnosed patients and controls.

Phantom validation: In order to evaluate if the proposed method can accurately retrieve the underlying thickness of an object, 6 folded 3D digital phantoms with spherical topology and known ground truth thickness were created (Fig. 6), resulting in six $300 \times 250 \times 250$ isotropic images with 3 structures - In, Out and Object. The thickness of the object is changing with time from 4.8 to $3.6mm$ in order to simulate a thickness lost in an object over several time points. These high resolution phantoms were then down-sampled by 5 in order to simulate PV effect and the thickness of the down-sampled structures was then measured. We compare the proposed method with the classic Jones et al. [6] method that works on a binary version of the segmentation. Here, two threshold values (0.5 and 0.9) were used to binarise the data. The method proposed by Acosta et al. [1] was also used for this comparison as it also integrates through partial volume using a ray casting technique.

Results show that the method developed by Jones et al. is very sensitive to threshold selection, mainly because there is no topological information when extracting the boundaries of the object of interest. In real world scenarios, the selection of a threshold becomes more critical as the segmentation is less well behaved. Thus, methods that require threshold selection should not be used without optimisation of the threshold parameter. Furthermore, this optimisation requires a ground truth, which is not available on clinical data, and has to be done per dataset, hampering the utilisation of this technique. The method proposed by Acosta et al. is not dependant on a threshold value as it integrates through PV using a ray casting technique. However, the choice of stopping criteria for the ray casting is problematic in convoluted areas, greatly increasing the standard deviation of the estimated thickness. When applied to real data, this increase in standard deviation will reduce the statistical power for group analysis. Using a one-sample t-test, none of the methods estimate a thickness that statistically differs from the ground truth. However, the proposed method achieves better accuracy and precision than the other two methods. Furthermore, the proposed method is the only one that guarantees spherical topology

of the estimated thickness. The pattern of difference between ground truth and estimated thickness is represented in Fig. 7(c,d). This difference is, in overall, very close to zero for KCTE but very variable for Acosta *et al.* [11].

Brain MRI analysis: To further investigate the ability of the proposed method to extract topologically correct measurements from real world segmentations, the thickness of the cortical layer was calculated on the AIBL dataset. The purpose of this study was to evaluate group separation between the cortical thickness of controls and Alzheimer’s Disease (AD) diagnosed patients. From the full database, a subset of 54 AD diagnosed patients and 54 age- and gender-matched controls with T1-weighted 3T MRI volumetric images acquired using a 3D MPRAGE sequence (typically $1.20 \times 1.00 \times 1.00\text{mm}$) was selected. Images were segmented using an Expectation-Maximization based probabilistic framework [anonymous] and thickness was calculated with $t = 0.05$ equivalent to 5% probability. The segmentations were then parcellated into different areas using the AAL atlas and two-tailed unequal-variance two-group t-tests were calculated between AD patients and controls over each AAL region.

Fig. 8 (top left) shows the statistical difference between AD patients and controls in the different areas of the brain projected onto the AAL template, using the proposed method. It shows an overall reduction of the average cortical thickness in AD patients and also an expected symmetrical pattern of cortical loss, with highly statistically significant differences in thickness on the temporal, occipital and mid-frontal areas, clinically known to be affected in AD. More importantly, Fig. 8 (bottom) shows that the surface with the estimated thickness is homotopic to a sphere for KTE (Euler characteristic of 2) while both Acosta *et al.* and Jones *et al.* obtain surfaces with inconsistent topologies.

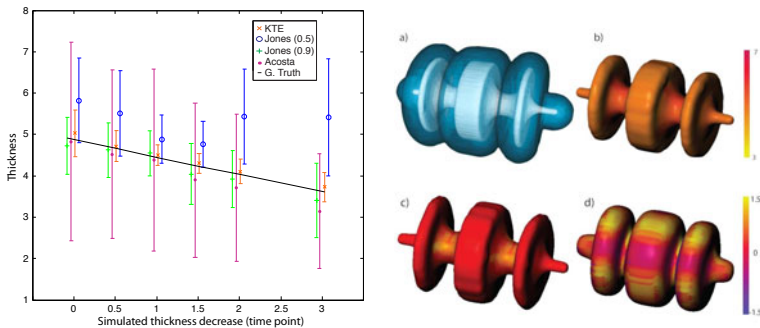


Fig. 7. Left: Mean and standard deviation of the estimated thickness at time-points 1 to 3 when compared to the ground truth in black. Right: a) The 3D phantom reconstruction, b) the surface with the estimated thickness for time-point 0 with KTE, and c) the difference between the ground truth thickness and the estimated one at time-point 0 for KTE and d) for Acosta *et al.*

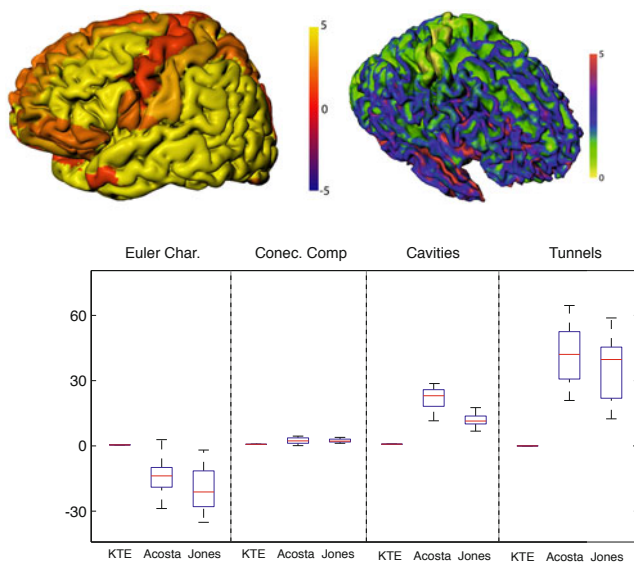


Fig. 8. Top left: Log of the p-values of statistical difference between AD patients and controls, where positive and negative represents thinning and thickening of the are for AD patients when compared to controls, using KTE; Top right: An example of an unsmoothed topologically correct thickness surface from the AIBL database; Bottom: The Euler characteristic and the number of connected components, cavities and tunnels for the estimated thickness surface

4 Conclusions

In this paper we present a new method to extract measurements of thickness from objects that are homotopic to a sphere. The main contribution of this work lies in three points: Khalimsky Cubic Complex based homotopic operations are used to correct the topology of the segmentation; a multi-stage Laplacian is then used to differentiate between pure from PV voxels; finally, the proposed method achieves sub-voxel accuracy without the use of *ad hoc* ray casting techniques by integrating through PV voxels following the Laplacian streamline field.

Experiments on digital phantoms with known ground truth thickness shows that the proposed method is more accurate and precise than two state of the art methods in retrieving a thickness measurements, while enforcing spherical topology. Quantitative analysis on brain data showed that the proposed algorithm is able to retrieve highly significant group differences between the cortical thickness of AD patients and controls, while maintaining spherical topology.

Acknowledgments

This study was supported by a scholarship from the Fundação para a Ciência e a Tecnologia, Portugal (Scholarship number SFRH/BD/43894/2008). Matthew J.

Clarkson is supported by the UCLH/UCL Comprehensive Biomedical Research Centre grant 168. The authors thank Drs. Hugues Talbot and Michel Couprie for their support.

References

1. Acosta, O., Bourgeat, P., Zuluaga, M.A., Fripp, J., Salvado, O., Ourselin, S.: Alzheimer's Disease Neuroimaging Initiative: Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. *Medical Image Analysis* 13(5), 730–743 (2009)
2. Aganj, I., Sapiro, G., Parikshak, N., Madsen, S.K., Thompson, P.M.: Measurement of cortical thickness from MRI by minimum line integrals on soft-classified tissue. *Human Brain Mapping* 30(10), 1097–1193 (2009)
3. Cointepas, Y., Bloch, I., Garnero, L.: A cellular model for multi-objects multi-dimensional homotopic deformations. *Pattern Recognition* 34(9), 1785–1798 (2001)
4. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. *P. Natl. Acad. Sci. Usa* 97(20), 11050–11055 (2000)
5. Hutton, C., Vita, E.D., Ashburner, J., Deichmann, R., Turner, R.: Voxel-based cortical thickness measurements in MRI. *NeuroImage* 40(4), 1701–1710 (2008)
6. Jones, S.E., Buchbinder, B.R., Aharon, I.: Three-dimensional mapping of cortical thickness using Laplace's equation. *Human Brain Mapping* 11(1), 12–32 (2000)
7. Khalimsky, E., Kopperman, R., Meyer, P.: Computer graphics and connected topologies on finite ordered sets. *Topology and its Applications* 36(1) (July 1990)
8. Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I., Evans, A.C.: Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage* 27(1), 210–221 (2005)
9. Kovalevsky, V.: Finite topology as applied to image-analysis. *Computer Vision Graphics And Image Processing Vision Graph* 46(2), 141–161 (1989)
10. Lehmann, M., Crutch, S.J., Ridgway, G.R., Ridha, B.H., Barnes, J., Rossor, M.N., Fox, N.C.: Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. *Neurobiology of Aging* (2009)
11. Lohmann, G., Preul, C., Hund-Georgiadis, M.: Morphology-based cortical thickness estimation. In: Taylor, C.J., Noble, J.A. (eds.) *IPMI 2003*. LNCS, vol. 2732, pp. 89–100. Springer, Heidelberg (2003)
12. Passat, N., Couprie, M., Bertrand, G.: Minimal simple pairs in the 3D cubic grid. *Journal of Mathematical Imaging and Vision* 32(3), 239–249 (2008)
13. Rocha, K.R., Yezzi Jr, A.J., Prince, J.L.: A hybrid Eulerian-Lagrangian approach for thickness of annular tissues. *IEEE TMI* 3765 (2005)
14. Rosas, H.D., Salat, D.H., Lee, S.Y., Fischl, B., Greve, D.N., Hevelone, N., Hersch, S.M.: Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain* 131(pt. 4), 1057–1068 (2008)
15. Scott, M.L.J., Bromiley, P.A., Thacker, N., Hutchinson, C.E., Jackson, A.: A fast, model-independent method for cerebral cortical thickness estimation using MRI. *Medical Image Analysis* 13(2), 269–285 (2009)
16. Yezzi, A.J., Prince, J.L.: An Eulerian PDE approach for computing tissue thickness. *IEEE Transactions on Medical Imaging* 22(10), 1332–1339 (2003)

A Convex Max-Flow Segmentation of LV Using Subject-Specific Distributions on Cardiac MRI

Mohammad Saleh Nambakhsh^{1,3}, Jing Yuan⁴, Ismail Ben Ayed²,
Kumaradevan Punithakumar², Aashish Goela^{5,6}, Ali Islam^{5,6},
Terry Peters^{1,3}, and Shuo Li^{1,2}

¹ Biomedical Engineering Program, University of Western Ontario, London, Canada

² GE Healthcare, London, ON, Canada

³ Imaging Research Laboratories, Robarts Research Institute, London, ON, Canada

⁴ Computer Science department, University of Western Ontario, London, Canada

⁵ Department of Medical Imaging, University of Western Ontario, London, Canada

⁶ St. Joseph's Health Care, London, ON, Canada

Abstract. This work studies the convex relaxation approach to the left ventricle (LV) segmentation which gives rise to a challenging multi-region separation with the geometrical constraint. For each region, we consider the global Bhattacharyya metric prior to evaluate a gray-scale and a radial distance distribution matching. In this regard, the studied problem amounts to finding three regions that most closely match their respective input distribution model. It was previously addressed by curve evolution, which leads to sub-optimal and computationally intensive algorithms, or by graph cuts, which result in heavy metrication errors (grid bias). The proposed convex relaxation approach solves the LV segmentation through a sequence of convex sub-problems. Each sub-problem leads to a novel bound of the Bhattacharyya measure and yields the convex formulation which paves the way to build up the efficient and reliable solver. In this respect, we propose a novel flow configuration that accounts for labeling-function variations, in comparison to the existing flow-maximization configurations. We show it leads to a new convex max-flow formulation which is dual to the obtained convex relaxed sub-problem and does give the exact and global optimum to the original non-convex sub-problem. In addition, we present such flow perspective gives a new and simple way to encode the geometrical constraint of optimal regions. A comprehensive experimental evaluation on sufficient patient subjects demonstrates that our approach yields improvements in optimality and accuracy over related recent methods.

1 Introduction

Left ventricle (LV) segmentation of 2D cardiac magnetic resonance (CMR) is one of the fundamental steps to diagnose coronary heart disease characterized by heart wall motion abnormalities, ventricular enlargement, aneurysms, scars (it occurs only on the myocardium region), strain, EF and etc.

Manual segmentation of CMR images is extremely tedious and time consuming¹. As such, automatic or semi-automatic algorithms are highly desired. Despite so many studies in the literature allocated to the task, challenges inherent to CMR images such as including papillary muscles in blood cavity within endocardial region, and presence of noise and artifacts prevented the algorithms to sufficiently address the problem for routine clinical use [11].

The problem of myocardial segmentation is commonly formulated as an optimization of a cost functional of the image features, which typically are solved using active contours or discrete graph-cut methods. Active contours have been the prevalent choice in medical image analysis as they are able to introduce a wide range of photometric and geometric constraints [2,3,12,14]. Generally, these constraints use pixel-wise penalties learned from image models in the training set that suffer from inability to distinguish between connected cardiac regions with almost the same intensity profile [2]. Also, such choice suffers from local optima, slow convergence, high sensitivities to initialization, and numerical complexity as well.

Training-based algorithms have several drawbacks including the difficulty in capturing the inter- and intra-subject variations and pathological characteristics. They further need a large amount of training dataset [2]. The recent active curve studies in [22,23] advanced it by building subject-specific models of a manually segmented frame in the current cardiac sequence.

In the discrete setting, although graph cut guarantees convergence to a global optima in nearly real time [6], it is unable to support global functionals that arise in our problem. Nevertheless, the recent works [19,14] use relaxations via bounds or approximations of the energy functional. Although these works led to substantial improvements over active contours in terms of speed, optimality, and accuracy, they still suffer from metrication errors [16].

Alternatively, continuous convex relaxation approaches share the advantages of both active curves and graph cuts, which recently have attracted a significant research attention in image segmentation problems. Some current studies based on convex optimization have shown its potential in solving the classical piecewise constant Mumford-Shah segmentation model [15], and its multiphase (multiregional) variant [13].

In this study, segmentation of the LV endo- and epi-cardial boundaries in a CMR sequence is formulated in an iterative convex max-flow relaxation approach with two original discrete cost functions. We solve a sequence of sub-problems which results in an algorithm robust to initial conditions. Unlike active contours, it also does not require a large number of iterative updates of the labeling and the time-consuming kernel density estimates (KDEs). The proposed formulation avoids a complex training process and the need for tremendous amount of training data. It leads to a segmentation faster than level-set methods and improvements in optimality and accuracy of the results. We further prove that the novel convex max-flow formulation obtains accurate and global solutions to the original, nonconvex sub-problems. Moreover, the proposed method is

¹ Almost over 200 2D images per subject.

shape-invariant and handles geometric variations of the LV intrinsically while it readily includes the papillary muscles into the endocardial or cavity region.

2 Convex Max-Flow Based Bhattacharyya Distribution Matching

In this work, we make the multi-region segmentation of LV based on the principle of Bhattacharyya distribution matching. More specifically, our task is to find the optimal regions of Blood Cavity Ω_c and Myocardium Ω_m , which best match the given distributions \mathcal{M}_c and \mathcal{M}_m based on the Bhattacharyya measure.

2.1 Bhattacharyya Distribution Matching

Let $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathcal{Z} \subset \mathbb{R}^n$ be an image function which maps domain Ω to the space \mathcal{Z} of a photometric variable such as a color vector. Our studies in this work consist of the sub-problem to find the region $\mathcal{R} \subset \Omega$ whose distribution most closely matches the given reference distribution \mathcal{M} by Bhattacharyya measure:

$$\mathcal{B}(\mathcal{P}, \mathcal{M}) = \sum_{z \in \mathcal{Z}} \sqrt{\mathcal{P}(z)\mathcal{M}(z)}, \tag{1}$$

where $\mathcal{P}(z)$ gives the nonparametric estimate of the distribution within \mathcal{R} :

$$\mathcal{P}(z) = \frac{\int_{\mathcal{R}} \mathcal{K}_z(x) dx}{|\mathcal{R}|}, \quad \forall z \in \mathcal{Z} \tag{2}$$

with $|\mathcal{R}|$ the area of region \mathcal{R} , i.e. $|\mathcal{R}| = \int_{\mathcal{R}} dx$; $\mathcal{K}_z(\cdot)$ is a kernel function, typically Gaussian:

$$\mathcal{K}_z(x) = \frac{1}{(2\pi\sigma^2)^{(n/2)}} \exp\left(-\frac{\|z - I(x)\|^2}{2\sigma^2}\right). \tag{3}$$

Studies such as [2] showed the Bhattacharyya distribution matching outperforms other methods in terms of robustness and accuracy.

In this study, our object is to find the optimal regions of Blood Cavity Ω_c and Myocardium Ω_m , which match the given distributions \mathcal{M}_c and \mathcal{M}_m respectively, subject to the minimum total perimeter of the two regions Ω_c and Ω_m , i.e. tight region boundaries are preferred. Beside this, we naturally adopt the geometrical constraint $\mathcal{R}_c \subset \mathcal{R}_m$ in our segmentation approach.

The problem therefore reduces to the following optimization problem:

$$\min_{\mathcal{R}_c, \mathcal{R}_m} -\mathcal{D}_c(\mathcal{R}_c) - \mathcal{D}_m(\mathcal{R}_m) + \lambda\left(|\partial\mathcal{R}_c| + |\partial\mathcal{R}_m|\right) \tag{4}$$

subject to the geometrical prior

$$\mathcal{R}_c \subset \mathcal{R}_m,$$

where $|\partial\mathcal{R}_{c,m}|$ measures the perimeter of \mathcal{R}_c and \mathcal{R}_m respectively and first and second terms in (4) are defined in the following:

$$\mathcal{D}_c(\mathcal{R}_c) = \alpha_c \mathcal{B}(\mathcal{P}_c^I, \mathcal{M}_c^I) + (1 - \alpha_c) \mathcal{B}(\mathcal{P}_c^D, \mathcal{M}_c^D) \quad (5)$$

$$\mathcal{D}_m(\mathcal{R}_m) = \alpha_m \mathcal{B}(\mathcal{P}_m^I, \mathcal{M}_m^I) + (1 - \alpha_m) \mathcal{B}(\mathcal{P}_m^D, \mathcal{M}_m^D) \quad (6)$$

where first term in both (5) and (6) is the Bhattacharyya of intensity distribution and the second term is the Bhattacharyya of radial distance distribution (distance from centroid of the regions). Note that in (6) only the myocardial strip is included in the intensity distribution and the cavity region is excluded from the intensity distributions. α_c, α_m are weights of the cavity and the myocardial regions which weights the intensity and the radial distance constraints.

2.2 Alternating Optimization Approach

For (4), we propose an effective alternating-optimization approach which splits (4) into two subproblems and tackles them sequentially at each iteration, see Alg. 1

Algorithm 1. Alternating Optimization Approach

For each k -iteration,

- Fix \mathcal{R}_m^k and optimize (4) over \mathcal{R}_c by

$$\mathcal{R}_c^{k+1} := \arg \min_{\mathcal{R}_c} -\mathcal{D}_c(\mathcal{R}_c) + \lambda |\partial\mathcal{R}_c|, \quad \text{s.t. } \mathcal{R}_c \subset \mathcal{R}_m^k. \quad (7)$$

- Fix \mathcal{R}_c^{k+1} and optimize (4) over \mathcal{R}_m by

$$\mathcal{R}_m^{k+1} := \arg \min_{\mathcal{R}_m} -\mathcal{D}_m(\mathcal{R}_m) + \lambda |\partial\mathcal{R}_m|, \quad \text{s.t. } \mathcal{R}_c^{k+1} \supset \mathcal{R}_m. \quad (8)$$

In view of the two subproblems (7) and (8), they essentially share the same type of optimization formulation which solves

$$\min_{\mathcal{R}} -\mathcal{D}(\mathcal{R}) + \lambda |\partial\mathcal{R}| \quad (9)$$

subject to the respective geometrical constraint in (7) and (8).

In the following, we therefore focus on the optimization problem (9), which corresponds to two subproblems (7) and (8) with incorporated geometrical priors. We further show it can be optimized globally and effectively by a novel iterative convex max-flow approach. It easily adapts the geometrical constraint by using the sufficiently large flow capacities (data cost) as standard max-flow scheme [7]. To this end, we only consider the optimization of (9) without any geometrical constraint which is assumed to be presented in flow settings.

2.3 Global Minimization of Energy Upper Bound

Let $u : \Omega \rightarrow \{0, 1\}$ be the labeling function of $\mathcal{R} \subset \Omega$, i.e. $u(x) = 1$ when $x \in \mathcal{R}$ and $u(x) = 0$ otherwise. The distribution in (2) can be rewritten, in terms of $u(x)$, as:

$$\mathcal{P}(z) = \frac{\int_{\Omega} \mathcal{K}_z(x)u(x) dx}{\int_{\Omega} u(x) dx}, \quad \forall z \in \mathcal{Z} \tag{10}$$

then the Bhattacharyya measure of (11) is:

$$\mathcal{B}(u, \mathcal{M}) = \sum_{z \in \mathcal{Z}} \left(\frac{\int_{\Omega} \mathcal{T}_{M,z}(x)u(x) dx}{\int_{\Omega} u(x) dx} \right)^{1/2} \tag{11}$$

where $\mathcal{T}_{M,z}(x) = \mathcal{K}_z(x)\mathcal{M}(z)$.

The perimeter of a region \mathcal{R} can be evaluated by the weighted boundary of the region or indicator function $u(x) \in \{0, 1\}$ [9][10]:

$$|\partial\mathcal{R}| = \int_{\Omega} C(x) |\nabla u(x)| dx, \quad C(x) \geq 0. \tag{12}$$

In this paper, we focus on the case where $C(x) = \lambda$ is a positive constant. The results can be easily extended to the more general $C(x)$.

In view of (11) and (12), (9) can be reformulated by:

$$\min_{u(x) \in \{0,1\}} \left\{ E(u) := - \sum_{z \in \mathcal{Z}} \left(\frac{\int_{\Omega} \mathcal{T}_{M,z}(x)u(x) dx}{\int_{\Omega} u(x) dx} \right)^{1/2} + \int_{\Omega} C(x) |\nabla u(x)| dx \right\}. \tag{13}$$

Direct computation of (13) is challenging, both theoretically and numerically, due to the constraint of a pointwise binary variable and the nonlinearity of its energy function [2]. To this end, we propose a novel iterative convex relaxation solution to (13), which globally minimizes a sequence of upper bounds of $E(u)$, denoted $E^{i+1}(u; u^i)$, $i = 1 \dots$:

$$u^{i+1} = \arg \min_{u \in \{0,1\}} E^{i+1}(u; u^i), \quad i = 1 \dots \tag{14}$$

where

$$E(u) \leq E^{i+1}(u; u^i), \quad i = 1 \dots \tag{15}$$

and

$$\|E(u) - E^{i+1}(u; u^i)\| \rightarrow 0, \text{ when } \|u - u^i\| \rightarrow 0 \tag{16}$$

² Unlike the classical continuous min-cut problem in [15][20], (13) is nonconvex even when $u \in \{0, 1\}$ is relaxed by $u \in [0, 1]$.

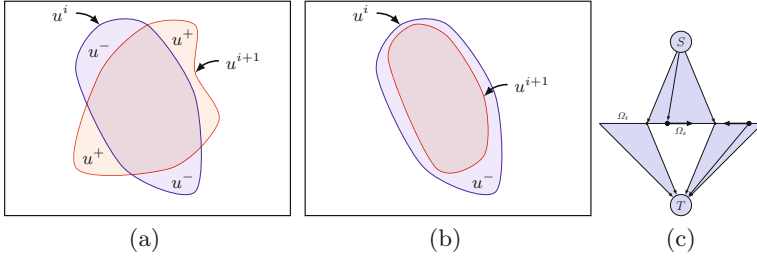


Fig. 1. Iterative bound optimization: (a) Illustrates the proposed bound; (b) Illustrates the bound in [11]. $u^i(x)$ denotes the labeling at iteration i ; (c) The proposed spatially continuous max-flow configuration.

One can easily show that the sequence of solutions u^i to the problems in (14) under constraints (15) and (16) yields a monotonically decreasing sequence of cost functions E and, therefore, a minimum of E at convergence, which can be summarized as the following propositions:

Proposition 1. *Sequence $E(u^i)$ is monotonically decreasing, i.e. $E(u^{i+1}) \leq E(u^i)$.*

Proposition 2. *Sequence $E(u^i)$ is convergent.*

Due to limited space Proofs of Prop. 1 and Prop. 2 is presented on request.

Estimation of Energy Upper Bounds. In view of Prop. 1 and Prop. 2, we derive estimation of the energy upper bound $E^{i+1}(u; u^i)$, which is *convex with respect to labeling function $u(x)$* . We will further show that $E^{i+1}(u; u^i)$ can be minimized *globally and exactly* over $u(x) \in \{0, 1\}$ in the following part.

Let $u^i(x) \in \{0, 1\}$ denote the binary labeling function, obtained at the previous i -th step, as the starting point of the $(i + 1)$ -th iteration. At this stage, we are seeking an optimal labeling function $u(x) \in \{0, 1\}$ which minimizes the energy upper bound. Let us express the difference between two functions $u(x)$ and the given $u^i(x)$ as a function of two new variables $u^+(x) \in \{0, 1\}$ and $u^-(x) \in \{0, 1\}$ (refer to Fig. 1a):

1. $u^+(x)$ indicates the area where $u^i(x)$ is 0 and $u(x)$ becomes 1, i.e. area increase:

$$u^+(x) := \begin{cases} u(x), & \text{where } u^i(x) = 0 \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

2. $u^-(x)$ indicates the area where $u^i(x)$ is 1 and $u(x)$ becomes 0, i.e. area decrease:

$$u^-(x) := \begin{cases} 1 - u(x), & \text{where } u^i(x) = 1 \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

The following describes the upper bound we propose.

Proposition 3. *Given a labeling $u^i(x) \in \{0, 1\}$, for any labeling function $u(x) \in \{0, 1\}$, we have the following upper bound which depends on $u^+(x)$ and $u^-(x)$:*

$$E(u) \leq -\mathcal{D}(u^i) + F^{i+1}(u; u^i) \tag{19}$$

where

$$F^{i+1}(u; u^i) = \int_{\Omega} C_v^i u^- dx + \int_{\Omega} C_w^i u^+ dx + \int_{\Omega} C |\nabla u| dx \tag{20}$$

and

$$C_v^i(x) = \sum_{z \in \mathcal{Z}} \frac{\mathcal{D}_{z,i} \mathcal{T}_{M,z}(x)}{\int_{\Omega} \mathcal{T}_{M,z}(x) u^i(x) dx}, \quad C_w^i(x) = \sum_{z \in \mathcal{Z}} \frac{\mathcal{D}_{z,i}}{2 \int_{\Omega} u^i(x) dx}, \tag{21}$$

$$\mathcal{D}_{z,i} = \left(\frac{\int_{\Omega} \mathcal{T}_{M,z}(x) u^i(x) dx}{\int_{\Omega} u^i(x) dx} \right)^{1/2}. \tag{22}$$

Due to limited space a detailed proof of Prop. 3 is given on request.

Prop. 3 leads us to the following conclusions:

- **Energy upper bound:** In view of (19), we have the energy upper bound

$$E^{i+1}(u; u^i) = -\mathcal{D}(u^i) + F^{i+1}(u; u^i)$$

where $\mathcal{D}(u^i)$ is a constant corresponding to the given u^i .

- **Continuous min-cut model:** Hence we can minimize $E^{i+1}(u; u^i)$ over $u(x) \in \{0, 1\}$ by

$$\min_{u(x) \in \{0,1\}} F^{i+1}(u; u^i) \tag{23}$$

where $F^{i+1}(u; u^i)$ is given in (20). We call (23) the *continuous min-cut model*.

It then proposes the following iterative minimization procedure:

1. Start with an arbitrary initial labeling $u^1(x)$ ($i = 1$);
2. For the i -th outer iteration, set the flow configuration based on labeling function $u^i(x)$. Compute $C_v^i(x)$ and $C_w^i(x)$ by (21), and solve (23) to obtain $u^{i+1}(x) \in \{0, 1\}$. In the following section, we propose a fast convex max-flow approach to computing (23) globally and exactly;
3. Let $i = i + 1$ and repeat the above two steps until convergence.

3 Continuous Max-Flow Approach

The new continuous min-cut model in (23) depends on labeling-function variations (u^+ and u^-), not on a single labeling function as with previous models [8,15,21,20]. In this section, we propose and solve the convex relaxation of (23), called convex relaxed min-cut model, through its dual model, so-called convex max-flow model which is constructed upon a novel flow configuration (see Fig. 1c). We show the proposed convex relaxed min-cut model solves (23) exactly and globally by means of its convex max-flow model. The convex max-flow formulation naturally leads to a fast algorithm to (23) [21,20].

3.1 Convex Relaxed Min-Cut Model

Given the current labeling function $u^i(x) \in \{0, 1\}$, we define:

$$\Omega = \Omega_s \cup \Omega_t, \quad \Omega_s \cap \Omega_t = \emptyset \tag{24}$$

where Ω_s is indicated by $u^i = 1$ and Ω_t by $u^i = 0$.

From the definitions of u^+ (17) and u^- (18), we reformulate (23) by:

$$\min_{u(x) \in \{0,1\}} \int_{\Omega_t} C_v^i(1-u) dx + \int_{\Omega_s} C_w^i u dx + \int_{\Omega} C(x) |\nabla u| dx \tag{25}$$

and propose its convex relaxation, i.e. *convex relaxed min-cut model*, by:

$$\min_{u(x) \in [0,1]} \int_{\Omega_t} C_v^i(1-u) dx + \int_{\Omega_s} C_w^i u dx + \int_{\Omega} C(x) |\nabla u| dx. \tag{26}$$

3.2 Convex Max-Flow Model

We propose the novel configuration of flows (see Fig. 1c for a 1D example) which accounts for the previous partition defined by Ω_s and Ω_t , which is in contrast the existing continuous flow setting [21,20]. Given Ω_s and Ω_t , we define three types of flows: the source flow $p_s(x)$ directed from s to $\forall x \in \Omega_s$, the sink flow $p_t(x)$ directed from $\forall x \in \Omega_t$ to t , and the spatial flow $p(x)$ given $\forall x \in \Omega$.

$p_s(x)$, $p_t(x)$ and $p(x)$ are constrained by the following flow capacities:

$$p_s(x) \leq C_v^n(x), \quad p_t(x) \leq C_w^n(x), \quad |p(x)| \leq C(x), \quad \text{a.e } x \in \Omega; \tag{27}$$

and flow conservation conditions:

$$-p_s(x) + \text{div } p(x) = 0, \quad p_t(x) + \text{div } p(x) = 0, \quad \text{a.e } x \in \Omega_t. \tag{28}$$

The continuous max-flow model can, therefore, be formulated by maximizing the total flow from the source s :

$$\max_{p_s, p_t, p} \int_{\Omega_s} p_s dx \tag{29}$$

subject to the flow constraints (27) and (28).

By means of convex analysis, we can prove the following result:

Proposition 4. *For the convex models (26) and (29), we have the equivalence:*

$$(29) \iff (26).$$

Due to limited space, its proof is given on request.

In addition, we can further show that the binary-valued optimization problem in (25) can be solved globally and exactly by its convex relaxation (26). Clearly, (26) is convex. Let $u^*(x)$ be its global optimum. We define the γ -upper level set $u_\gamma^*(x) \forall x \in \Omega$, for any $\gamma \in [0, 1]$:

$$u_\gamma^*(x) = \begin{cases} 1, & \text{when } u^*(x) > \gamma \\ 0, & \text{when } u^*(x) \leq \gamma \end{cases}, \tag{30}$$

Therefore, we have

Proposition 5. *When $u^*(x)$ gives one global optimum of (26), its thresholding $u_\gamma^*(x) \in \{0, 1\}$ by (30), for any $\gamma \in [0, 1]$, solves the binary-valued optimization problem in (25) globally and exactly.*

The proof of Prop. 5 follows the ideas of [21,20] (proposed on the request).

3.3 Convex Max-Flow Based Algorithm

We propose a fast algorithm based on the convex max-flow model (29). In this regard, we define the following augmented Lagrangian function:

$$\begin{aligned}
 L_c(p_s, p_t, p, u) &= \int_{\Omega_s} p_s \, dx + \int_{\Omega_s} (\operatorname{div} p - p_s)u \, dx + \int_{\Omega_t} (\operatorname{div} p + p_t)u \, dx \\
 &\quad - \frac{c}{2} \|\operatorname{div} p - p_s\|_{\Omega_s}^2 - \frac{c}{2} \|\operatorname{div} p + p_t\|_{\Omega_t}^2
 \end{aligned} \tag{31}$$

where $c > 0$. Finally, we derive the continuous max-flow algorithm (Alg.) from the classical augmented Lagrangian method [17,18], whose convergence properties can be found in [5].

Algorithm 2. Continuous Max-Flow Based Algorithm

- $i = 1$: Initialize flows p_s^1, p_t^1, p^1 and labeling $u^1 \in \{0, 1\}$. Then start $i + 1$ -th iteration;
- Fix p^i and u and maximize in a closed form $L_c(p_s, p_t, p, u)$ over p_s and p_t :

$$(p_s^{i+1}, p_t^{i+1}) := \arg \max_{p_s, p_t} L_c(p_s, p_t, p^i, u);$$

- Fix p_s^{i+1}, p_t^{i+1} and u and maximize iteratively $L_c(p_s, p_t, p, u)$ over p via the simple projection-descent step:

$$p^{i+1} := \arg \max_p L_c(p_s^{i+1}, p_t^{i+1}, p, u);$$

- Update labeling function u by

$$u(x) = \begin{cases} u(x) + c(\operatorname{div} p^{i+1}(x) - p_s^{i+1}(x)), & \forall x \in \Omega_s \\ u(x) + c(\operatorname{div} p^{i+1}(x) + p_t^{i+1}(x)), & \forall x \in \Omega_t \end{cases} .$$

Repeat the above two steps until convergence.

4 Experiments

To evaluate the proposed method(CMFM), 20 subjects with 120 short axis cardiac cine CMRI with pixel spacing of 0.7812×0.7812 and slice thickness of 10mm including apical, mid-cavity, and basal slices were used. The manual

segmentation of the first frame of each slice was chosen to learn the model distributions. For evaluation purpose, automatic segmented cavity and myocardium were compared to the corresponding manually segmented regions. Also, for comparison the level-set segmentation (*LSM*) in [2] and max-flow(graph-cut) method (*MFM*) in [4] were run over the same datasets.

After tuning of the algorithm in terms of the different parameters, the following parameters were used for all the dataset identically. $\sigma = 10^{-1}$ for the intensity distributions of the myocardium, $\sigma = 10$ for the intensity distributions of the cavity, and $\sigma = 2$ for the distance distributions in (3); $\alpha_c = 0.95$ in (5) and $\alpha_m = 0.6$ in (6). We also used $\lambda = 0.3$ in (27) and $c = 0.05$ in (31) for the convex max-flow optimizer.

4.1 Illustration of Segmented Images and Quantitative Evaluations

Fig.2 shows the results of one subject over mid-cavity, apical, and basal slices. The yellow and red curves correspond to the endo- and epicardium boundaries, respectively. The third row shows the apical frames where it is challenging to segment the cavity as it is a small region with motion artifacts. Examples illustrate that the proposed method perfectly excludes the papillary muscles from the myocardium. The results also demonstrate that the proposed method handles wide variations of the cavity in terms of scale and shape.

We further assessed the similarities between the ground truth and the segmentations via contour-based measure such as the Root Mean Squared Error (*RMSE*), and region-based measure such as Dice Metric (*DM*). Let V_a and V_m be automated and manually segmented volumes. *DM* is given by $DM = 2 \frac{V_a \cap V_m}{V_a + V_m}$. *DM* measures the similarity between the automatically detected and ground-truth regions. Higher *DM* shows better performance while it is always between 0 and 1. *RMSE* evaluates automatic epi- and endocardium boundaries by assigning 240 points along the boundary and computing the perpendicular distances between manual contour and them [4]. The lower *RMSE* shows the results are more close to the ground truth.

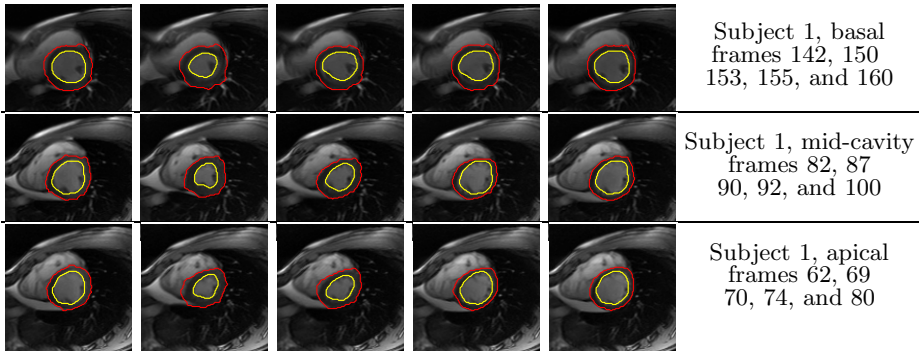


Fig. 2. Results for one subject at different slice levels: Basal (1st row); Mid-cavity(2nd row); Apical(3th row)

We examined quantitatively and qualitatively the *reliability* of the algorithm by evaluating the *reliability function*, i.e., the complementary cumulative distribution function (*ccdf*) of the obtained Dice metrics (Fig.3a and b).

Furthermore, one of the important quantitative clinical measures is LV volume which is the volume within the cavity. Fig.3c shows comparison of automated cavity volume and the manual one for 20 subjects over 20 phases.

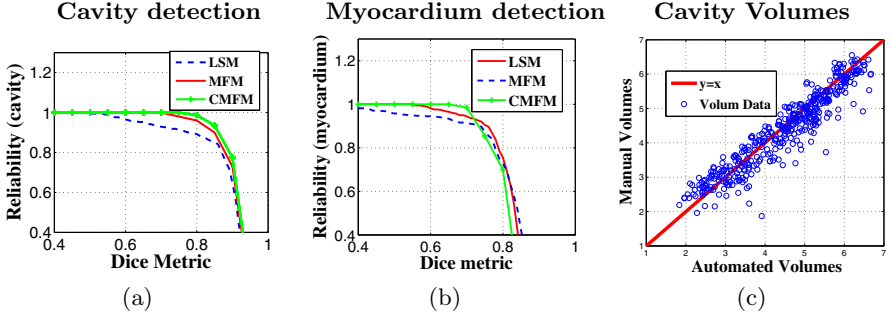


Fig. 3. Comparisons of (a) Cavity and (b) Myocardium Dice Metric reliability ($\mathcal{R}(d) = Pr(DM > d)$ d : horizontal axes) for 2280 images acquired from 20 subjects by the proposed method (CMFM), MFM [4] and LSM [2] algorithms; (c) Comparisons of manual and automatic segmented volumes of the cavity for the same database

Table 1 reports the *DM* statistics, reliability of *DM* higher than (0.80), average *RMSE* and average time is consumed to process a frame over 20 subjects for the proposed method (CMFM), MFM [4] and LSM [2].

Table 1. *DM*, Reliability of *DM*, the average *RMSE* (in pixels), and average time load to process a frame over 20 subjects (2280 images) for the proposed method (CMFM), MFM [4] and LSM [2]; Statistics of the *DM* expressed as mean \pm standard deviation; Algorithms have been run over 2 GHz CPU

Methods	Dice Metric		Reliability ($d = 0.8$)		<i>RMSE</i>		<i>time per</i> frame (sec)
	Cavity	Myocardium	Cavity	Myocardium	Cavity	Myocardium	
CMFM	0.92 ± 0.07	0.80 ± 0.10	1	0.70	1.35	2.05	1.79
MFM	0.91 ± 0.04	0.81 ± 0.06	0.97	0.79	1.60	1.99	0.14
LSM [2]	0.88 ± 0.09	0.81 ± 0.10	0.89	0.75	2.46	1.89	15.96

5 Conclusion

This study investigated fast detection of the *LV* endo- and epicardial boundaries in a cardiac magnetic resonance (*MR*) sequence. The solution is obtained by the proposed convex relaxation optimization of two original discrete cost functions

constructed of a global geometric and a global photometric constraint based on the Bhattacharyya similarity. Quantitative evaluations demonstrated that the results correlate well with independent manual segmentations drawn by an expert. The proposed method outperformed a recent active contour method in [2], and graph-cut method in [4]. The proposed approach removes the need of a complex training, while handles intrinsically geometric variations of the *LV* without biasing the solution towards a finite set of shapes. It also includes papillary muscles in the endocardial region.

References

1. Ben Ayed, I., Chen, H.M., Punithakumar, K., Ross, I., Li, S.: Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the Bhattacharyya measure. In: CVPR 2010 (2010)
2. Ben Ayed, I., Li, S., Ross, I.: Embedding overlap priors in variational left ventricle tracking. *IEEE Trans. Med. Imaging* 28(12), 1902–1913 (2009)
3. Ben Ayed, I., Li, S., Ross, I., Islam, A.: Myocardium tracking via matching distributions. *Int. J. of Comput. Assist. Radiol. and Surg.* 4(1), 37–44 (2009)
4. Ben Ayed, I., Punithakumar, K., Li, S., Islam, A., Chong, J.: Left ventricle segmentation via graph cut distribution matching. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 901–909. Springer, Heidelberg (2009)
5. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific (1999)
6. Boykov, Y., Funka Lea, G.: Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vision* 70(2), 109–131 (2006)
7. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 1124–1137 (2004), <http://dx.doi.org/10.1109/TPAMI.2004.60>
8. Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J., Osher, S.: Fast global minimization of the active contour/snake model. *J. Math. Imaging Vis.* 28(2), 151–167 (2007)
9. Chan, T., Shen, J.H.: *Image Processing And Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia (2005)
10. Giusti, E.: *Minimal surfaces and functions of bounded variation*. Australian National University, Canberra (1977)
11. Jolly, M.-P.: Automatic recovery of the left ventricular blood pool in cardiac cine MR images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 110–118. Springer, Heidelberg (2008)
12. Kaus, M.R., von Berg, J., Weese, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* 8(3), 245–254 (2004), <http://www.sciencedirect.com/science/article/B6W6Y-4D09D3J-1/2/d7268f23efbbbfa83da4665d311dee58>
13. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. Tech. report, HCI, IWR, Uni. Heidelberg (2008)
14. Liu, H., Chen, Y., Ho, H.P., Shi, P.: Geodesic active contours with adaptive neighboring influence. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 741–748. Springer, Heidelberg (2005)

15. Nikolova, M., Esedoglu, S., Tony, F.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* 66(5), 1632–1648 (2006), <http://link.aip.org/link/?SMM/66/1632/1>
16. Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach for computing minimal partitions. In: *CVPR 2009* (2009)
17. Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. *J. Optimiz. Theory App.* 12, 555–562 (1973)
18. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.* 1(2), 97–116 (1976)
19. Rother, C., Minka, T., Blake, A., Kolmogorov, V.: Cosegmentation of image pairs by histogram matching - incorporating a global constraint into MRFs. In: *CVPR 2006* (2006)
20. Yuan, J., Bae, E., Tai, X.C., Boycov, Y.: A study on continuous max-flow and min-cut approaches. Part I: Binary labeling. Tech report CAM-10-61, UCLA (2010)
21. Yuan, J., Bae, E., Tai, X.C.: A study on continuous max-flow and min-cut approaches. In: *CVPR 2010* (2010)
22. Zhu, Y., Papademetris, X., Sinusas, A.J., Duncan, J.S.: Segmentation of the left ventricle from cardiac MR images using a subject-specific dynamical model. *IEEE Trans. Med. Imaging* 29(4), 669–687 (2010)

Entangled Decision Forests and Their Application for Semantic Segmentation of CT Images

Albert Montillo^{1,2}, Jamie Shotton², John Winn², Juan Eugenio Iglesias^{2,3},
Dimitri Metaxas⁴, and Antonio Criminisi²

¹ GE Global Research Center, Niskayuna, NY, USA
montillo@ge.com

² Microsoft Research, Cambridge, UK
{jamie.shotton, jwinn, antcrim}@microsoft.com

³ University of California, Los Angeles, USA
jeiglesias@ucla.edu

⁴ Rutgers University, Piscataway, NJ USA
dnm@rutgers.edu

Abstract. This work addresses the challenging problem of simultaneously segmenting multiple anatomical structures in highly varied CT scans. We propose the entangled decision forest (EDF) as a new discriminative classifier which augments the state of the art decision forest, resulting in higher prediction accuracy and shortened decision time. Our main contribution is two-fold. First, we propose *entangling* the binary tests applied at each tree node in the forest, such that the test result can depend on the result of tests applied earlier in the same tree and at image points offset from the voxel to be classified. This is demonstrated to improve accuracy and capture long-range semantic context. Second, during training, we propose injecting randomness in a guided way, in which node feature types and parameters are randomly drawn from a learned (non-uniform) distribution. This further improves classification accuracy. We assess our probabilistic anatomy segmentation technique using a labeled database of CT image volumes of 250 different patients from various scan protocols and scanner vendors. In each volume, 12 anatomical structures have been manually segmented. The database comprises highly varied body shapes and sizes, a wide array of pathologies, scan resolutions, and diverse contrast agents. Quantitative comparisons with state of the art algorithms demonstrate both superior test accuracy and computational efficiency.

Keywords: Entanglement, auto-context, decision forests, CT, segmentation.

1 Introduction

This paper addresses the challenging problem of automatically parsing a 3D Computed Tomography (CT) scan into its basic components. Specifically, we wish to recognize and segment organs and anatomical structures as varied as the aorta, pelvis, and the lungs, simultaneously and fully automatically. This task is cast as a voxel classification problem and is addressed via novel modifications to the popular decision forest classifier [1,2].

The decision forest is experiencing rapid adoption in a wide array of information processing applications [3-8]. It can be used for clustering, regression, and as in this paper, for classification. The classifier has many attractive qualities that make it well suited for practical problems and close to an ideal universal learner [9]. It scales well computationally to large training sets, handles multi-class classification in a natural way, and the knowledge it has learned can be inspected and interpreted. In a typical image classification task [5], each pixel is classified separately. To improve segmentation results, constraints in the form of local consistency or semantic (e.g. anatomical) context are applied, but this requires either a separate random field [10] or multi-pass processing [5,11].

In this paper, we extend the decision forest classifier to directly enforce local consistency and semantic context without applying additional methods or passes. We show how this extension also speeds training and improves test accuracy. The two main contributions are as follows. First, to construct a tree node, n , at level, L , in the forest, we design new *entanglement* features which exploits the uncertain partial semantic information learned (or at test time, inferred) by the previous $L-1$ levels of the forest about the classification of voxels in a neighborhood. Since the nodes in the resulting classifier share information with each other, we call the new classifier an *entangled* decision forest. Second, during training we randomly sample feature types and parameters from a learned, non-uniform *proposal distribution* rather than from the uniform distribution used (implicitly) in previous decision forest research [1,2,5,6,7,14,20]. The random draws select, with greater probability, the feature types and parameters that tend to be relevant for classification, allowing higher accuracy for the same number of features tested. We show how these two contributions allow faster training and prediction, more accurate prediction, and how the combination of these contributions yields best performance.

In [5], a separate decision forest is grown in each successive round of classifier training. The forest grown in each round uses semantic information learned during a previous round encoded in the form of a bag of textons that characterize decision paths down the tree of the previous round's forest. Similarly, in [11], a separate probabilistic boosting tree (PBT) [12] is constructed in each successive round. The classifier grown in each round uses semantic information from a previous round encoded as the most likely class label in a spatially offset region. These works inspired our development of EDF to use the semantic information learned *in previous levels, in a single round* of classifier construction. This yields a simpler, more efficient classifier than sequences of forests or PBTs and enables higher accuracy, faster training and prediction, and requires less memory.

Our problem is defined as follows. We define demonstrate the utility of our solution to segment 12 anatomical structures in large field of view CT. We are given density and ground truth labels at each voxel in a set of training volumes. Our goal is to infer the probability of each organ label for each voxel of unseen test scans. The task is challenging due to the extremely large variations of both healthy structures and pathology in the abdominal-thoracic region. Variations include organ location and shape, contrast agent presence/absence, scanner field of view, and image resolution.

The most closely related work for high-speed CT segmentation is non-rigid marginal space learning (MSL) [13] which uses a boosted classifier to predict the parameters of an active shape model. MSL can segment the liver boundary in 10

seconds; in contrast, our method requires only 12 seconds to segment 12 organs simultaneously. The active shape model of MSL offers resilience to noise; our method also offers flexibility to handle organs only partially visible within the scanner field of view or which have topological changes due to pathology (cysts, tumors) as demonstrated in the results section.

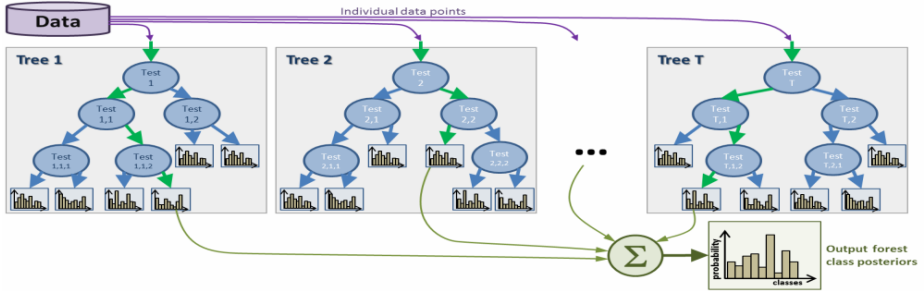


Fig. 1. Decision forest overview. During training, multiple trees are grown, using all training data for each tree. During testing, to classify a voxel it is initialized at the root node of each tree, and recursively sent left or right (red arrows) according to binary tests stored at each node. The voxel is classified using the average of the T posterior label distributions, with one coming from the leaf reached in each of the T trees.

The EDF is a new discriminative classifier which improves the accuracy and speed of the state of the art decision forest for image analysis. Our methodology may be used to improve results for other applications that rely upon the decision forest classifier, including MS lesion segmentation [8], brain segmentation [6], myocardium delineation [7], and beyond these medical applications for broad applicability in the field of computer vision, such as for object recognition [10].

2 Methods

2.1 Decision Forest Background

We begin with a brief review of randomized decision forests [1,2]. A decision forest is an ensemble of T decision trees. During *training*, the data (Fig. 1), consists of the set of data points from all training images, $S = \{v_i, l_i\}_1^N$. Each data point, s_i , consists of the voxel position, v_i , and its label, l_i . Tree t_i , receives the full set S and its root node selects a test to split S into two subsets to maximize information gain. A test consists of a feature (e.g. an image feature) and a feature response threshold. The left and right child nodes receive their respective subsets of S and the process is repeated at each child node to grow the next level of the tree. Growth stops when one or more stopping criteria, such as minimal information gain or a maximum tree depth occur. Each tree is unique because each tree node selects a random subset of the features and thresholds to try. During *testing*, the data (Fig. 1) consists of the voxel positions in a test image. The voxels are routed to one leaf in each tree by applying the test (selected during training) which is stored in each node. The test is applied to the voxel in the

test image. The test result guides the voxel to the left or right child node, and this is repeated until a leaf node is reached. An empirical distribution over classes learned from the training data is stored at each leaf. The voxel is classified by averaging the class distributions from the set of leaves it reached. The following section describes the features we use to define the node tests of our decision forest.

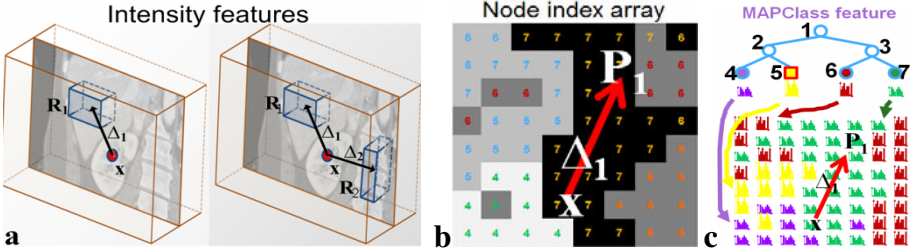


Fig. 2. Intensity and MAPClass features. (a) Intensity features measure image information from regions offset from the voxel to classify at \mathbf{x} . (b) MAPClass feature retrieves the label that the classifier currently predicts at location \mathbf{P}_1 offset from \mathbf{x} . Implementation-wise, we maintain a node index array which associates with each voxel the current tree node ID (represented by the number in each voxel). (c, top) This allows us to determine the current label posterior in the tree for the voxel at location \mathbf{P}_1 . (c, bottom) Conceptually, the tree induces a vector image of class posteriors which we used when developing the MAPClass and TopNClasses features.

2.2 Context Rich, Long-Range Visual Features

It has been shown [22] that to classify a voxel at a given location anatomical context from regions up to 200mm away are often very helpful. Therefore we do not use traditional features such as Haar wavelets [15] whose range is too short. Instead we construct two types of long-range, context-rich features. The first capture “appearance context”, the later capture “semantic context”. This will be explained next.

Appearance features. We construct intensity features that are spatially defined by (1) their position, \mathbf{x} , centered on the voxel to be labeled (Fig. 2a), and (2) one or two rectangular probe regions, \mathbf{R}_1 and \mathbf{R}_2 , offset from \mathbf{x} by displacements Δ_1 and Δ_2 which can be up to 200mm in each dimension (x, y, z). We construct two categories of intensity features. The first category consists of the mean CT intensity at a probed region, \mathbf{R}_1 (Fig 2a, left), while the second consists of the difference in the mean intensity at probed regions, \mathbf{R}_1 and \mathbf{R}_2 (Fig 2a, right). These are defined as follows:

$$f_{Intensity}(\mathbf{x}; \Delta_1, \mathbf{R}_1) = \bar{I}(\mathbf{R}_1(\mathbf{x} + \Delta_1)) \quad (1)$$

$$f_{IntensityDiff}(\mathbf{x}; \Delta_1, \mathbf{R}_1, \Delta_2, \mathbf{R}_2) = \bar{I}(\mathbf{R}_1(\mathbf{x} + \Delta_1)) - \bar{I}(\mathbf{R}_2(\mathbf{x} + \Delta_2)) \quad (2)$$

During training, the features to try at each node are parameterized by dimensions of \mathbf{R}_1 and \mathbf{R}_2 , offsets Δ_1 and Δ_2 and an intensity threshold α . These parameters are

chosen randomly to define the intensity test: $f(\cdot) > \alpha$. Once training has finished, the max information gain node test along with its optimal features are frozen and stored within the node for later use during testing.

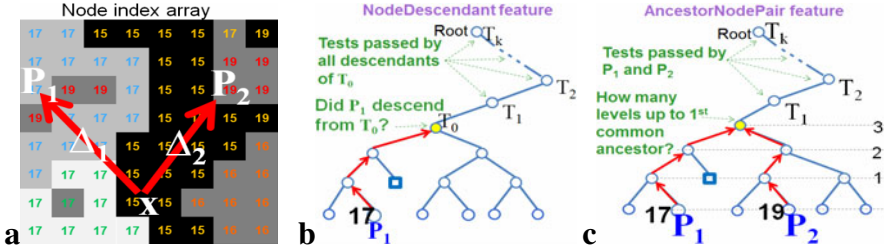


Fig. 3. Additional entanglement features. (a) Node index array associates voxels with intensity and tree node indices (same format as Fig. 2b but for a deeper tree level). (b) NodeDescendant feature tests whether probe voxel at \mathbf{P}_1 descends from a node (T_0 in this case). (c) AncestorNodePair feature tests whether the nodes of voxels \mathbf{P}_1 and \mathbf{P}_2 have a common ancestor $< \tau$ levels away.

Semantic context entanglement features. We now describe the first contribution of our paper. The basic idea is that during testing on novel images, we exploit the confident voxel label predictions (peaked distributions) that can be found using just early levels of the forest to aid the labelling of nearby voxels. This provides semantic context similar to auto-context [5,11], but does so within the same forest. We define four types of long range entanglement features to help train the node currently being grown using knowledge learned in already trained split nodes of the forest. Two features (*MAPClass* and *TopNClasses*) are based on the posterior class distribution of the nodes corresponding to probed voxels, and two (*NodeDescendant* and *AncestorNodePair*) are based on the location of the nodes within the trees.

We construct *MAPClass entanglement features* which use the maximum a posteriori label of a neighboring voxel at \mathbf{P}_1 in order to reduce uncertainty about the label at \mathbf{x} (Fig 2b). When such semantic context is helpful to classify the voxel at \mathbf{x} the feature yields high information gain and may become the winning feature for the node during tree growth. *MAPClass* tests whether the MAP class in the posterior of a probed voxel $\mathbf{P}_1 = \mathbf{x} + \Delta_1$ is equal to a particular class, C :

$$f_{MAPClass}(\mathbf{x}; \Delta_1, \mathbf{P}_1, C) = \begin{cases} \arg \max_c p(c; n_{p_1}) = C & 1 \\ \text{otherwise} & 0 \end{cases} \quad (3)$$

where $p(c; n_{p_1})$ is the posterior class distribution of the node of \mathbf{P}_1 . This posterior can be retrieved from the tree because we (1) train and test voxels in breadth first fashion and (2) maintain an association between voxels and the tree node ID at which they reside while moving down the tree. This association is a node index array (Fig 2b).

We define *TopNClasses entanglement features* denoted TopNClasses , where $N \in \{2, 3, 4\}$ that generalize the MAPClass feature. A TopNClass feature tests whether a particular class C is in the top N classes of the posterior class distribution of the probe voxel at $\mathbf{P}_1 = \mathbf{x} + \Delta_1$. The feature is defined as:

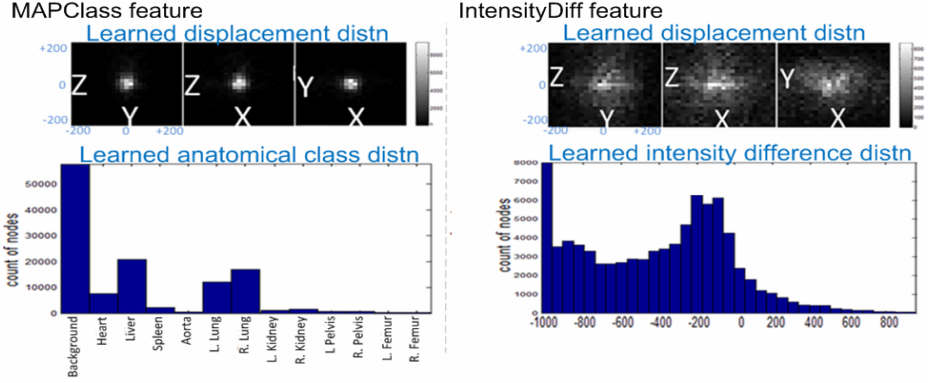


Fig. 4. Learned parameter distributions are clearly non-uniform. (left) Learned displacement and anatomical class distributions for MAPClass feature. (right) Displacement and intensity difference distributions for IntensityDiff feature.

$$f_{\text{TopNClasses}}(\mathbf{x}; \Delta_1, \mathbf{P}_1, N, C) = \begin{cases} C \in \text{top } N \text{ classes of } p(c; n_{\mathbf{p}_1}) & 1 \\ \text{otherwise} & 0 \end{cases} \quad (4)$$

We define *NodeDescendant entanglement features* which test whether a region near voxel \mathbf{x} has a particular appearance. The neighboring region is centered at voxel \mathbf{P}_1 (Fig. 3a,b). The test is whether the node currently corresponding to \mathbf{P}_1 descends from a particular tree node, T_0 . If it does, then we know \mathbf{P}_1 has satisfied the appearance test ($T_1 \dots T_k$) above T_0 in the tree in a particular way to arrive at T_0 .

We also define *AncestorNodePair entanglement features*. This type of feature tests whether two regions near voxel \mathbf{x} have passed similar appearance and semantic tests. The neighboring regions are centered at voxels \mathbf{P}_1 and \mathbf{P}_2 (Fig. 3a). The test is whether the nodes currently corresponding to \mathbf{P}_1 and \mathbf{P}_2 have their first common ancestor $< \tau$ tree levels above the current level (Fig. 3c). The threshold controls the required degree of similarity: the lower τ , the greater the required appearance and context similarity needed to pass the test, because the lower τ , the greater the number of tests, ($T_1 \dots T_k$), above the common ancestor.

2.3 Feature Selection Is Guided by Learned Proposal Distributions

This section describes the second contribution of our paper. We match the distribution of feature types and their parameters proposed at each tree node during training to the ones that tend to be most useful for training. The decision forest still chooses the winning feature, but each node chooses from features sets that are likely to be useful based on prior experience. The basic idea is to help the classifier explore more of the

sweet spot of feature space and hopefully find superior features. Since our features contain several parameters, the joint feature space is too large to search exhaustively. However, only a small subset of feature space tends to be relevant. Therefore, rather than drawing feature parameters from a uniform distribution over parameter space, we draw from a learned distribution. Specifically, we train an initial decision forest, F_{temp} , on our training data and record the distribution of accepted (winning) feature parameters and feature types across all its tree nodes. F_{temp} is then discarded, and we then use parameter distributions as the proposal distributions in a subsequent training of a final decision forest. While this requires additional training, it imposes no time penalty for prediction.

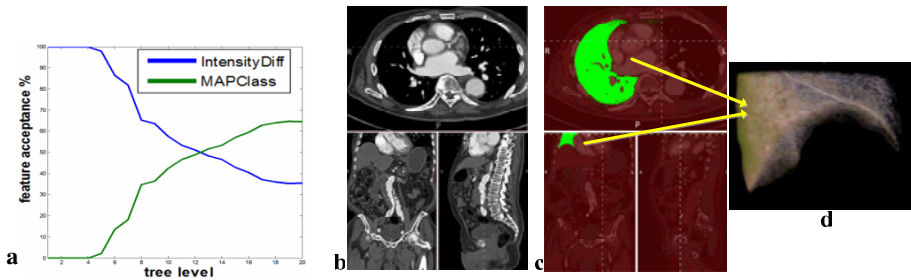


Fig. 5. (a) Example of learned feature type distribution by level. (b-d) *See results section*: (b) right lung is only partially visible in the scan. (c) EDF correctly assigns high posterior probability and (d) segments visible portion (3D rendering).

We learn *parameter proposal distributions*. The learned distribution of displacements tends to be Gaussian distributed and centered on the voxel to be labeled (Fig. 4 top row). Acceptance distributions of the remaining parameters also have non-uniform distributions (Fig. 4 bottom row). We draw feature parameters from these distributions during training. This tends to provide the forest with more useful features to choose from at each node and can improve final classifier accuracy.

We also learn *feature type proposal distributions*. Similarly, the distribution of feature types for each tree level is learned. Drawing feature types from this distribution can also improve classifier accuracy. Fig. 5a shows how the ratio of feature types varies with tree depth if we construct a forest using just MAPClass and IntensityDiff features. Early in tree growth appearance features dominate, while entanglement features dominate deeper levels. As more information gets inferred from intensity, entanglement features exploit semantic context and neighborhood consistency.

3 Results

3.1 Experimental Setup

We evaluate our EDF model on a database which consists of 250 large field of view CT scans in which each voxel has an intensity and is assigned one of 12 labels from a

set of very diverse anatomical structures {*heart, liver, spleen, aorta, left/right lung, left/right femur, left/right pelvis, left/right kidney*} or the *background* class label. This database was chosen because it was designed to include wide variations in patient health and scan protocol. We randomly selected 200 volumes for training and 50 for testing.

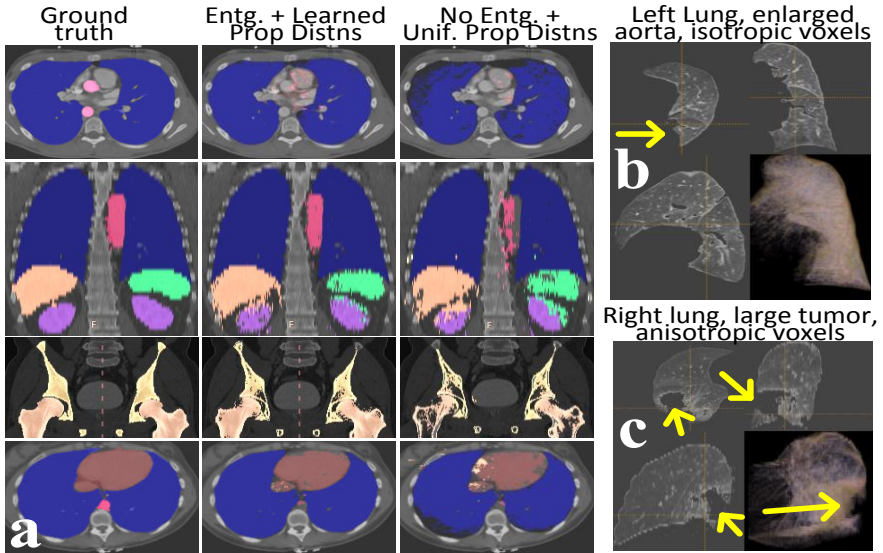


Fig. 6. Qualitative segmentation results. (a) Entanglement and learned proposal distributions (column 2) provide marked improvement compared to not using them (column 3). Four different subjects shown with axial slices (rows 1,4), and coronal (rows 2,3). (b) 2x2 panel showing intensities of voxels comprising an EDF segmented left lung distorted by enlarged aorta; volume rendering in lower right quadrant (c) EDF accurately segments despite severe anomaly and voxel anisotropy.

3.2 Qualitative Results

The EDF achieves a visually accurate segmentation of organs throughout the 50 test volumes. Example segmentations are shown in Fig. 6a where the first column is the ground truth organ segmentation, and the second column is the EDF segmentation result. We see good agreement for the lungs (blue) shown in rows 1, 2 and 4, for the liver (orange), spleen (green), and kidneys (purple) shown in row 2, for the femur bones (tan) in row 3, and for the heart (dark brown) in row 4. Column 3 shows the result using our decision forest but without entanglement features and without the learned proposal distributions. Node entanglement noticeably improve the lungs in row 1, the aorta (red), kidneys, and spleen in row 2, the femurs in row 3 and the lungs and heart in row 4.

The algorithm handles many complexities commonly found in the clinic. Fig. 6b shows how our algorithm correctly segmented the lung (physician verified) despite the fact that the patient had a severely enlarged aorta which caused a distortion

(see yellow arrow). Fig. 6c shows how EDF accurately segments despite a large tumor (arrows) and severe anisotropy in the voxel dimensions. Fig. 5b shows a case in which only a portion of the patient’s lungs were in the scanner’s field of view. EDF correctly assigns high posterior probability to lung pixels (see right lung in Fig. 5c) and properly segments the portion in the scanner.

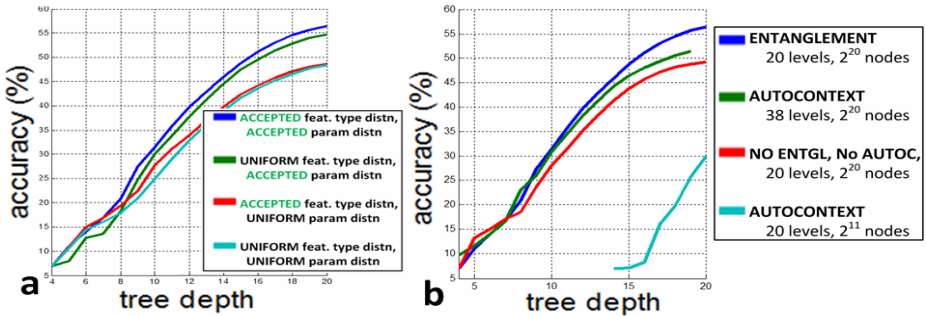


Fig. 7. Quantitative impact of each contribution. (a) Learning both proposal distributions increases accuracy. (b). Entanglement (blue) provides greater accuracy and prediction speed than auto-context (green). Note: green curve should be plotted at depths 20-38, but for comparison we plot it at depths 1-19.

3.3 Quantitative Results Including the Impact of Each of Our Contributions

Accuracy measure. For a quantitative analysis we measured the EDF segmentation accuracy across all 50 test scans using the average class Jaccard similarity coefficient [16]. The metric is the ratio of the intersection size (ground truth and predicted labels) divided by the size of their union. While the EDF achieves >97% average voxel accuracy throughout the volumes in our database, we use the Jaccard metric in this section, because we feel it is a more honest and reliable metric for segmentation accuracy and is not unduly influenced by the background class.

We measured *the impact of the learned proposal distributions*. Specifically, to understand the impact of using the acceptance distribution as proposal distributions (section 2.3), we trained the decision forest in four different ways: (1) using uniform feature type and uniform feature parameter distributions for baseline performance (light blue curve, Fig. 7a), (2) using learned (i.e. accepted) feature type distribution with uniform feature parameter distributions (red curve), (3) using uniform feature type distributions with learned feature parameter distributions (green curve), (4) using learned feature type and learned parameters distributions (dark blue curve). Learning only the feature type distribution yields a negligible improvement to baseline (red vs light blue). Learning feature parameter distribution boosts accuracy significantly (green vs red). Learning both distributions yields the best performance without penalty at lower depths (dark blue vs green) and boosts accuracy over baseline by 8% (dark blue vs light blue).

We compared our entanglement method to auto-context [5, 11], a state of the art approach which has yielded some of the best accuracy and speed for multi-structure segmentation. Specifically, we define the same auto-context features as [11] for our

decision forest. Auto-context requires multiple complete decision forests to be constructed. The auto-context feature defines semantic context to help classify a voxel at location \mathbf{x} by examining the class predicted for a probe voxel by a previous decision forest. For our comparison we conducted four experiments. *First*, we trained our decision forest 20 levels deep without entanglement and without auto-context for a baseline performance (red curve, Fig. 7b). *Second*, we trained a two-round, auto-context decision forest (ADF) using 20 total levels (light blue curve). Here we constructed a sequence of two decision forests with the same total number of levels as the baseline classifier, in order to achieve the same prediction *time*. Specifically, we used the output from the first 10 levels of the baseline as the input to the second round, 10 level forest. The second round forest uses the prediction from the first round to form auto-context features and also uses our intensity based features. *Third*, we trained another ADF, but this time with an equal *modeling capacity* to the baseline, (i.e. we trained the same number of tree nodes, requiring roughly the same amount of memory and training time). For this test, we used the final output from the first 19 levels of the baseline classifier as the input to train a second round, 19 level forest, for a total of 38 levels in the ADF. In this way, the ADF consists of $2 * 2^{19} = 2^{20}$ maximum possible nodes. *Fourth*, we trained the proposed EDF method as a single, 20 level deep forest using entanglement (dark blue curve). When the ADF is constrained to give its prediction in the same time as the baseline classifier, it yields much lower accuracy (light blue vs red). When the ADF is allowed more time for prediction using 38 levels, it beats the baseline (green versus red). However, we find *considerably better accuracy* using the EDF method (dark blue curve vs green). In addition to beating the performance of ADF, it reduces the prediction time by 47% since the EDF requires 18 fewer levels (20 vs 38).

In separate tests, we varied the test:train ratio. We found only minor degradation in accuracy. Using 50 images for test and 195 for training, accuracy = 56%; using 75 test and 170 train, accuracy = 56%; using 100 test and 145 train, accuracy = 54%.

For *efficiency considerations*, we find our parallel tree implementation of EDF segments novel volumes in just 12 seconds per volume (a typical volume is 512x512x424) using a standard Intel Xeon 2.4GHz computer (8 core) with 16GB RAM running Win7 x64. A very good, coarse labeling (at 8x downsampling) can be achieved in <1 second. Training on the 200 volumes, which need only be done once, requires about 8 hours.

4 Discussion

To the best of our knowledge, EDF segments volumetric CT at a speed equal to or better than state of the art methods. For example nonrigid marginal space learning (MSL) [13] can segment the outer liver surface in 10 seconds; EDF simultaneously segments 12 organs, including the liver, in 12 seconds.

Our existing implementation of the EDF could be used to automatically measure organ properties (e.g. volume, mean density). It could also be used to initialize interactive segmentation methods [17, 18] or to identify biologically homologous structures that guide non-rigid registration [19].

The EDF is a reusable algorithm. Applying it to segment abdominal-thoracic organs requires no specialization for a particular organ, nor any image alignment; we only assume the patient is supine. Applying it to CT merely requires normalized intensities (i.e. Hounsfield units). This suggests that EDF could be used to segment other organs, or to segment other modalities. Our formulations of node entanglement and the learning of proposal distribution are generic. These methods amplify the value of many hand-crafted, image-based features that have been defined in the literature for specific classification problems. EDF could be directly used to improve the results of other applications [5,6,8,7,] or combined with complementary methods [21] to improve CT image segmentation using decision forests.

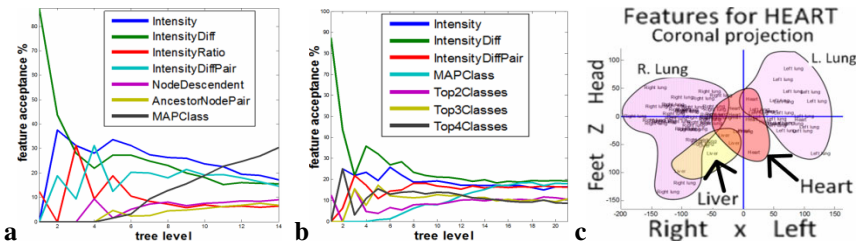


Fig. 8. EDF reveals how and what it learns. (a, b) relative importance of feature types at each level of forest growth. (c) Location and organ class of the top 50 features used to identify heart voxels. The hand-drawn regions here group these locations for different MAPClass classes C .

Theoretically speaking, compared to black-box learning methods (e.g. neural networks), one can query the EDF to understand what it has learned. For example in our EDF experiments, we queried the EDF to reveals what features it is using to learn at each level of growth. Our tests show that NodeDescendent entanglement (purple) achieves peak utilization before AncestorNodePair entanglement (tan) shown in Fig. 8a, while MAPClass (black) enjoys an ever increasing utilization rate with increasing depth. When we compared MAPClass to TopNClasses (Fig 8b) we found that Top4Classes (black) peaks, then Top3Classes (tan), and finally MAPClass peaks (light blue).

The EDF can also reveal the anatomical context that it has learned for each structure. By rendering a *scatter plot* of the top contributing features for a target structure, we can visualize the contextual information learned for that structure. For example, Fig. 8c shows how the MAPClass feature learns to segment a heart voxel, located at the blue cross-hair intersection. To find the top contributing features, we express information gain (5) as a sum of the information gain from each class:

$$G(F, A, B) = \sum_c \left(-p(F_c) \log p(F_c) - \left[-\sum_{F_{c_1}}^{A_{c_1}} p(A_{c_1}) \log p(A_{c_1}) - \sum_{F_{c_2}}^{B_{c_2}} p(B_{c_2}) \log p(B_{c_2}) \right] \right) \quad (5)$$

where F is the set of voxels being split into partitions A and B , and c is the index over classes. This enables us to rank learned node features based on how much they contributed to identifying the voxels of a given class by increasing the information gain for that class. Fig 8c shows a projection of the 3D scatter plot onto a coronal

plane. The semantic context that favors classifying a voxel as heart includes other heart pixels nearby (red region), lungs to the right and left (purple regions), and liver below the right lung (yellow region). All of this is learned by the EDF automatically.

5 Conclusions

This paper has proposed the entangled decision forest (EDF) as a new discriminative classifier which achieves higher prediction accuracy and shortened decision time. Our first contribution is to entangle the tests applied at each tree node with other nodes in the forest. This propagates knowledge from one part of the forest to another which speeds learning, improves classifier generalization and captures long range-semantic context. Our second contribution is to inject randomness in a guided way through the random selection of feature types and parameters drawn from learned distributions. Our contributions are an intrinsic improvement to the underlying classifier methodology and augment features defined in the literature.

We demonstrated EDF effectiveness on the very challenging task of simultaneously segmenting 12 organs in large field of view CT scans. The EDF achieves accurate voxel-level segmentation in 12 seconds per volume. The method handles large population variation and protocol variations. We suggest the method may be useful in other body regions and modalities.

References

1. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. *Neural Computation* 9(7), 1545–1588 (1997)
2. Breiman, L.: Random Forests. *Machine Learning* 45(1), 5–32 (2001)
3. Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Petrich, W., Hamprecht, F.A.: A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 10, 213 (2009)
4. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of SBFSEM volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) *DAGM 2008*. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
5. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: *Proc. of CVPR*, pp. 1–8 (2008)
6. Yi, Z., Criminisi, A., Shotton, J., Blake, A.: Discriminative, semantic segmentation of brain tissue in MR images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 558–565. Springer, Heidelberg (2009)
7. Lempitsky, V.S., Verhoek, M., Noble, J.A., Blake, A.: Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In: *Functional Imaging and Modeling of the Heart*, pp. 447–456 (2009)
8. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 111–118. Springer, Heidelberg (2010)
9. Hastie, T., Tibshirani, R., Friedman, J.H.: *The Elements of Statistical Learning*. Springer, Heidelberg (2009)

10. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comp. Vision* 81(1), 2–23 (2009)
11. Tu, Z., Bai, X.: Auto-context and Its Application to High-Level Vision Tasks and 3D Brain Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(10), 1744–1757 (2010)
12. Tu, Z.: Probabilistic boosting tree: Learning discriminative models for classification, recognition, and clustering. In: *Proc. of ICCV*, pp. 1589–1596 (2005)
13. Zheng, Y., Georgescu, B., Comaniciu, D.: Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) *IPMI 2009. LNCS*, vol. 5636, pp. 411–422. Springer, Heidelberg (2009)
14. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Machine Learning* 36(1), 3–42 (2006)
15. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. *Int. J. Comp. Vision* 57(2), 137–154 (2004)
16. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge. *Int. J. Comp. Vision* 88(2), 303–338 (2010)
17. Rother, C., Kolmogorov, V., Blake, A.: GrabCut -Interactive Foreground Extraction using Iterated Graph Cuts. In: *SIGGRAPH*, vol. 23(3), pp. 309–314 (2004)
18. Criminisi, A., Sharp, T., Blake, A.: GeoS: Geodesic Image Segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part I. LNCS*, vol. 5302, pp. 99–112. Springer, Heidelberg (2008)
19. Konukoglu, E., Criminisi, A., Pathak, S., Robertson, D., White, S., Siddiqui, K.: Robust Linear Registration of CT Images using Random Regression Forests. In: *SPIE Medical Imaging*, vol. 7962, p. 79621X (2011)
20. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in CT volumes. In: *Proc. of MICCAI-PMMIA* (2009)
21. Iglesias, J., Konukoglu, E., Montillo, A., Tu, Z., Criminisi, A.: Combining Generative & Discriminative Models for Semantic Segmentation of CT Scans via Active Learning. In: *Proc. of Info. Proc. In: Medical Imaging* (2011)
22. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Scans, In: *MICCAI-MCV Workshop* (2010)

Venous Tree Separation in the Liver: Graph Partitioning Using a Non-ising Model

Thomas O'Donnell¹, Jens N. Kaftan^{1,2,4}, Andreas Schuh^{1,3}, Christian Tietjen⁴,
Grzegorz Soza⁴, and Til Aach²

¹ Siemens Corporate Research, 755 College Rd East, Princeton, NJ 08540, USA

² Institute of Imaging and Computer Vision, RWTH Aachen University,
52056 Aachen, Germany

³ Section of Biomedical Image Analysis, Department of Radiology,
University of Pennsylvania, Philadelphia, PA 19104, USA

⁴ Siemens Healthcare Sector, Computed Tomography, 91301 Forchheim, Germany
`tom.odonnell@siemens.com`

Abstract. Entangled tree-like vascular systems are commonly found in the body (e.g., in the peripheries and lungs). Separation of these systems in medical images may be formulated as a graph partitioning problem given an imperfect segmentation and specification of the tree roots. In this work, we show that the ubiquitous Ising-model approaches (e.g., Graph Cuts, Random Walker) are not appropriate for tackling this problem and propose a novel method based on recursive minimal paths for doing so. To motivate our method, we focus on the intertwined portal and hepatic venous systems in the liver. Separation of these systems is critical for liver intervention planning, in particular when resection is involved. We apply our method to 34 clinical datasets, each containing well over a hundred vessel branches, demonstrating its effectiveness.

Keywords: Vessel, Tree, Separation, Liver, Ising, Minimal Path.

1 Introduction

The venous vasculature of the liver is composed of the portal venous system and the hepatic venous system. The portal system receives venous blood from the gastrointestinal track and distributes it to the hepatic lobules. The lobules serve to produce bile, break down insulin, and cleanse the blood, among other functions. The blood is then drained from the lobules into the hepatic system and exits the liver through the vena cava [7].

The identification of the individual portal and hepatic systems is critical for liver intervention planning. In particular, resection of any type requires that there is a sufficiently large portion of the liver remaining and that it is completely supplied by *all* vascular systems including the venous systems [5]. In addition, those regions that are not adequately supplied must be removed. Planning therefore must integrate knowledge of the individual 3D venous systems to identify liver regions that could safely be disconnected from blood supply [13].

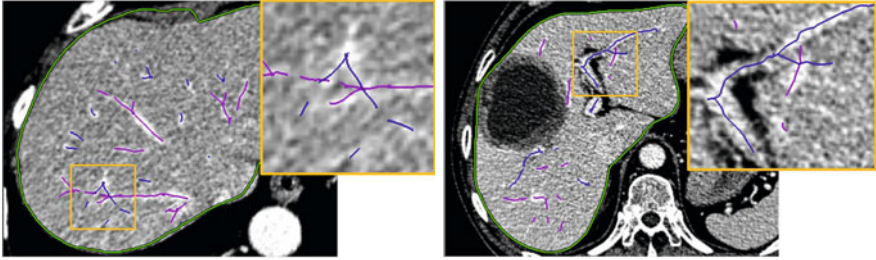


Fig. 1. Examples of overlapping liver vessel trees. The hepatic vessels are shown in purple, and the portal vessels in blue. Note that the centerlines from adjacent slices are shown as well.

The same requirements hold, of course, for living donor liver transplantation [10]. And, in the field of minimally invasive surgery, where less complete visualization during the procedure is generally permitted, this awareness of the vessel topology thus takes on even greater significance [4]. With this in mind, the liver may be thought of as being made up of segments (“lobes”) that form functionally independent units whose definition is fundamentally tied to the individual venous systems. Indeed, the identification and separation of the portal and hepatic trees form the foundation of the commonly employed Couinaud scheme for lobe partitioning [3].

In this paper we present a technique for the fully-automatic separation of the liver venous vasculature into portal and hepatic systems from a CTA volume given an *automatically generated* segmentation mask of the venous anatomy and the specification of the portal and hepatic roots. While this may appear to be a straightforward task, it is in fact quite challenging due to two reasons. First, the segmentation mask often contains errors. Specifically, there may both exist loops within the mask’s individual portal and hepatic systems as well as improper connections between them. By nature, the portal and hepatic systems are intertwined. Due to partial volume effects and errors in the segmentation mask, there are frequently several links between the systems. Disconnecting these links becomes necessary in order to understand where one system ends and the other begins. However, given the large quantity of liver vessel branches and the connections between them, the number of possible combinations of cuts to separate the mask’s two systems is huge (see Fig. 1). Second, standard methods for determining which cuts are appropriate (e.g., MinCut/MaxFlow, Random Walker) are not straightforwardly applicable for this problem domain as we shall see.

First, however, we must transform the automatically generated segmentation mask into a form that may be more easily analyzed. In reality, the portal and hepatic venous systems consist of separate groups of trees or sets of trees, while the underlying topology of the automatic segmentation mask we are presented with is almost always a graph (rather than a tree) caused by spurious links due to segmentation errors. We extract this underlying graph as follows: To the

segmentation mask of the venous anatomy [8], we apply a skeletonization. We then recover a centerline graph from this skeletonization and refine it by automatically removing extraneous short branches [9], and fusing branches which are very close to each other. The branches may then be described as nodes of a graph and we assign weights to branch connections (“edges”) based on intensity and connection angle, among other features. Our goal is then to partition this graph. (Note again that the vessel *branches* are nodes and the *connections* between them are edges - not the other way around). In this partitioning, “flow” must be respected. By this we mean that there is a directed path from *one and only one* root (portal or hepatic) to any branch. The direction of that path is the flow. It can be only implicitly determined if all connections between the roots have been severed. We must choose the best connections to sever such that there is no flow “upstream” to the roots.

Graph-based segmentation methods have become quite popular in the vision literature in recent years. However, the most well-known ones, (MinCut/MaxFlow [2], Random Walker [6]), are based on the Ising model (i.e., they assume node interactions can be based solely on nearest neighbors). This assumption does not apply to our domain as can be seen in Fig. 2a. There are two branches (nodes) represented by dotted lines and one connection (junction) represented by a dot. In algorithms which follow the Ising model, it is possible, without looking at the context, to assign a cost to this edge (e.g., a weak cost motivating the edge to be cut, or a strong one so that it might be kept). This cost might be based on, for example, the acuteness of angle, or local intensity differences. An appropriate choice for such edge weight, however, might be also dependent on further, non-local features such as a flow feature, which we cannot know ahead of time. In this context, the flow is only defined when the systems are separated. If there is supposed to be a flow into this edge as in Fig 2b then the edge weight should be strong as it is part of a normal bifurcation. If, however, two flows converge on the edge, it should be considered weak as in Fig. 2c. It is impossible, therefore, to assign the edge strength without consulting the flow context. Moreover, the flow context is based on the roots which are non-local to the vast majority of nodes. Hence, the Ising model does not apply.

Given that standard graph partitioning algorithms are not applicable, we introduce a novel recursive minimal path technique to perform this separation: We first run Dijkstra’s algorithm from all portal (hepatic) roots to find the cheapest paths to any hepatic (portal) roots. If no path exists both structures are already separated. Otherwise, there is at least one path between the portal and hepatic systems that must be cut. We examine the candidate edges on the cheapest paths and sort them in terms of weakness. We remove the weakest candidate and recursively apply the algorithm until the systems are separated. Upon return we re-establish the previously removed edge, and select instead the second weakest candidate and recursively apply the algorithm again. When we exit the algorithm we will have collected multiple lists of cuts (hypotheses), each of which partitions the graph. We then evaluate which cut is best with respect to the overall anatomical separation and topology.

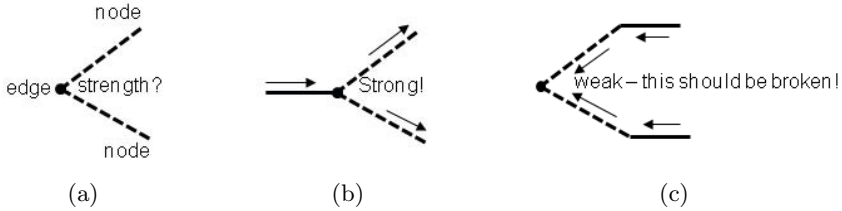


Fig. 2. In the Ising model, it is assumed the edge connecting two nodes (dashed lines representing vessel branches) can be assigned a strength based solely on its nearest neighbors. However, in our domain, the edge strength is dependent on the flow which is *non-local*. (a) The strength of this edge is not well-defined without further knowledge. (b) If the flow is in the direction indicated by the arrows, then the edge forms a natural bifurcation. The edge should be kept and therefore be strong. (c) If the flow is in the opposite direction, the edge should be broken, therefore be weak. However, the flow emanates from the roots and the roots are non-local. Therefore, techniques such as MinCut/MaxFlow and Random Walker which employ an Ising model are not appropriate.

Our contributions are the introduction of a new method for separating intertwined, connected, and crossing tree systems using a non-Ising partitioning scheme. We exploit the fact that there are typically only a small number of points at which the trees touch (on the order of three to four). This facilitates employing a recursive minimal path technique that would be prohibitively expensive in other situations. Because we can employ this approach, the concept of flow may be incorporated into the search. Moreover, the result is a set of multiple separation hypotheses from which we may choose the best (compare this to, for example, Mincut/Maxflow which results in a single solution). We can then evaluate these solutions and determine which one is most appropriate using additional global characteristics (e.g., overall separation, distribution of mass, global morphology, etc.).

2 Related Work

There have been previous attempts to separate portal and hepatic venous systems; however, this action usually takes place in the context of the vessel segmentation itself. Typical for such approaches is the one of Bauer et al., who preprocess the volume into a set of tube-like elements and then, starting from the designated roots, incorporate these into tree structures as they are encountered [1]. During this process, they make use of the orientation information, which is of a local nature, specific to the relationship between the tree under construction and the tube-like element under consideration. Our approach, by contrast, examines a potential break point between trees in a global, and hence more stable and robust fashion. We can examine an entire path, from root to root, to see for example, if a rapid change of angle between two branches is local

noise, or based on a trend involving many branches (by adjusting M in Fig. 3). Hence, even subtle changes in orientation may be identified.

Similarly to [1], Selle et al. construct the trees by adding branches one by one [13]. They consult a graph-like construction of the venous system but do not employ minimal paths. Rather, they use the graph to enforce the constraint that the radii must decrease as they move towards the leaves. In reality, however, radius changes are often too minute to distinguish them on a local level. Soler et al. take a similar approach but include angle information [14]. Details are, however, omitted. In a different domain, Park et al. follow a reverse strategy [11]. Starting from the terminal ends of segmented pulmonary vessels, they backtrack to their origin to classify the branches into veins and arteries. This method employs radius and orientation information via local spheres but there is no recursive algorithm for determining the selections nor a means for evaluating the result in a global fashion. In the same domain, Saha et al., [12] propose a morphological separation using a fuzzy distance transform to iteratively separate the vascular systems starting at large scales and progressing towards smaller scales. This approach may miss cuts which do not distinguish themselves in larger scales.

3 Our Method: Recursive Minimal Paths

Given a segmentation of the venous system, we perform a skeletonization and recursively step through it to form a centerline network. This network is refined to remove small branches and fuse nearby ones. We describe the network as a graph $G = (V, E)$, where $V = \{v_1, \dots, v_n\}$ represents the set of n branches (“nodes”). Each branch v_i consists of an ordered series of n_i sites in voxel space, $v_i = \{\mathbf{v}_i^{(1)}, \dots, \mathbf{v}_i^{(n_i)}\}$. We assume two disjoint sets of branches that serve as roots: $V_{r_p} = \{v_{r_{p_1}}, \dots, v_{r_{p_n}}\}$, a set of p_n portal roots, and $V_{r_h} = \{v_{r_{h_1}}, \dots, v_{r_{h_n}}\}$, a set of h_n hepatic roots, such that $V_{r_p} \in V$, $V_{r_h} \in V$, and $V_{r_p} \cap V_{r_h} = \emptyset$. We seek to partition the branches, V , into two disjoint sets, P and H , corresponding to the portal and hepatic systems such that $V_{r_p} \in P$ and $V_{r_h} \in H$. The set of edges, $E = \{e_{ij}\}$, represents the connections between branches such that there exists an edge, e_{ij} , between two branches v_i and v_j if the two branches share a common position, the so-called junction, \mathbf{j}_{ij} . Edges are made up of parts of each of the branches and joined at the junction (see Fig. 3a). We define them in this way so that it is easy to extract features such as “angle between branches” that cannot be defined by a single point (the junction). That is, $\exists e_{ij}$ for all pairs (v_i, v_j) with

$$e_{ij}^i = \begin{cases} \{\mathbf{v}_i^{(1)}, \dots, \mathbf{v}_i^{(m)}\} & \text{if } \mathbf{j}_{ij} = \mathbf{v}_i^{(1)} = \mathbf{v}_j^{(1)} \vee \mathbf{j}_{ij} = \mathbf{v}_i^{(1)} = \mathbf{v}_j^{(n_j)} \\ \{\mathbf{v}_i^{(n_i)}, \dots, \mathbf{v}_i^{(n_i-m)}\} & \text{if } \mathbf{j}_{ij} = \mathbf{v}_i^{(n_i)} = \mathbf{v}_j^{(1)} \vee \mathbf{j}_{ij} = \mathbf{v}_i^{(n_i)} = \mathbf{v}_j^{(n_j)} \end{cases},$$

$$e_{ij}^j = \begin{cases} \{\mathbf{v}_j^{(1)}, \dots, \mathbf{v}_j^{(m)}\} & \text{if } \mathbf{j}_{ij} = \mathbf{v}_i^{(1)} = \mathbf{v}_j^{(1)} \vee \mathbf{j}_{ij} = \mathbf{v}_i^{(n_i)} = \mathbf{v}_j^{(1)} \\ \{\mathbf{v}_j^{(n_j)}, \dots, \mathbf{v}_j^{(n_j-m)}\} & \text{if } \mathbf{j}_{ij} = \mathbf{v}_i^{(1)} = \mathbf{v}_j^{(n_j)} \vee \mathbf{j}_{ij} = \mathbf{v}_i^{(n_i)} = \mathbf{v}_j^{(n_j)} \end{cases},$$

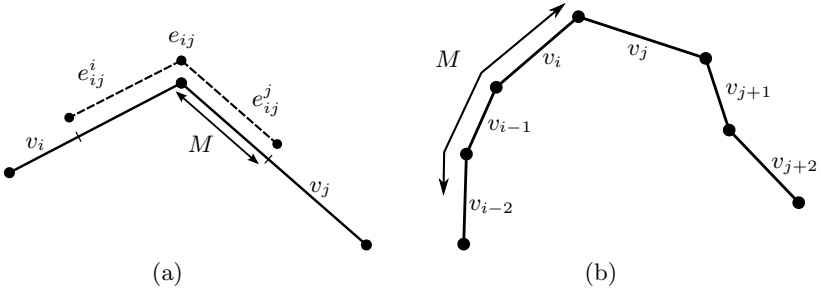


Fig. 3. (a) The edge e_{ij} , is composed of two ordered halves e_{ij}^i and e_{ij}^j of equal length, M created from branches v_i and v_j , respectively. (b) Once a path is determined, M can be expanded to give a more robust estimate of the angle at the edge.

where $m = \min(n_i, n_j, M)$. Thus an edge, e_{ij} , is composed of two ordered halves e_{ij}^i and e_{ij}^j of equal length. The length is limited by a globally defined margin, M . We compute several features on each for the nodes.

From these features we compute the edge costs based on intensity, radius, distance to roots, and angle information and combine these costs into a final weighted cost. We compute the intensity feature on each edge by first averaging the intensities over its extent:

$$I(e_{ij}) = \frac{1}{2m} \left(\sum_{k=1}^m I(e_{ij}^i(k)) + \sum_{k=1}^m I(e_{ij}^j(k)) \right), \quad (1)$$

where $e_{ij}^i(k)$ is the k^{th} position in the i^{th} edge half and $I(e_{ij}^i(k))$ is the intensity at that position. To compute the actual cost $C_I(e_{ij})$, we determine the Gaussian distribution, L_I , of the intensity feature with, μ_I and σ_I being the mean and standard deviation of the intensity feature from all edges in that dataset, respectively,

$$L_I(e_{ij}) = \frac{1}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{(I(e_{ij}) - \mu_I)^2}{2\sigma_I^2}}. \quad (2)$$

We want the intensity cost, $C_I(e_{ij})$, to be low if the intensity of an edge is below average (and thus likely to represent inappropriate connections between branches - the ones we'd like to break) and high if it is higher than average, thus

$$C_I(e_{ij}) = \begin{cases} \frac{\sqrt{2\pi\sigma_I^2}}{2} \cdot L_I(e_{ij}) & \text{if } I(e_{ij}) - \mu_I \leq 0 \\ 1 - \frac{\sqrt{2\pi\sigma_I^2}}{2} \cdot L_I(e_{ij}) & \text{otherwise} \end{cases}. \quad (3)$$

The radius cost $C_R(e_{ij})$ is computed similarly to $C_I(e_{ij})$ except that we substitute $R(\cdot)$, the Euclidean distance to the nearest non-vessel voxel, for $I(\cdot)$ in (1). Thus edges representing vessels with a small average radius are assigned a low cost (and are thus more likely to be broken) and edges with a larger average radius, a higher one. The utilized angle feature equals

$$\Phi(e_{ij}) = \frac{1}{2} \left(\left\langle \text{LSQ}(w(e_{ij}^i)), \text{LSQ}(w(e_{ij}^j)) \right\rangle + 1 \right), \quad (4)$$

where $\langle \cdot \rangle$ represents the dot product and $\text{LSQ}(\cdot)$ the normalized Least Squares fit to edge halves, e_{ij}^i and e_{ij}^j , converted into world coordinates via $w(\cdot)$ (using $M = 5$, cf. Fig. 3a). Since the dot product ranges from $[-1, 1]$, we normalize the values to $[0, 1]$. Again, we compute $C_\Phi(e_{ij})$ similarly to the intensity cost by creating a Gaussian distribution, $L_\Phi(e_{ij})$ (cf. (2)), and assigning low costs to low values (acute angles) as previously described.

The cost $C_D(e_{ij})$ represents the penalty for being close to the roots. To compute it we determine the distance to the closest root (portal or hepatic) for all edges, set up a Gaussian distribution based on the average values and standard deviation for all edges in the dataset, and again create a cost function. This cost function is low if the edge is far from a root (and thus more likely to be incorrect).

Next, we form a weighted sum of these costs: $C_I(e_{ij})$, $C_R(e_{ij})$, $C_\Phi(e_{ij})$, and $C_D(e_{ij})$ to get a final cost, $C_{\text{total}}(e_{ij})$, for each edge. Given this graph with the edge costs computed above, we are now ready to separate V into P and H using the following recursive method: First, we determine if there exists any path from V_{r_p} to V_{r_h} . If no (further) path exists both structures are separated. Otherwise, we determine the existing paths $\Lambda = \{\lambda_1, \dots, \lambda_{n_\lambda}\}$ from all portal roots V_{r_p} to any hepatic root, V_{r_h} . A path λ_q may be described as:

$$\lambda_q = \{e_{ab}e_{bc} \dots e_{gh}e_{hi}e_{ij}e_{jk}e_{kl} \dots e_{xy}e_{yz}\}, \quad (5)$$

where $a \in V_{r_p}$ and $z \in V_{r_h}$. We then select the minimal cost path that contains the cheapest edge among all edges of all paths in Λ . Once we have moved from

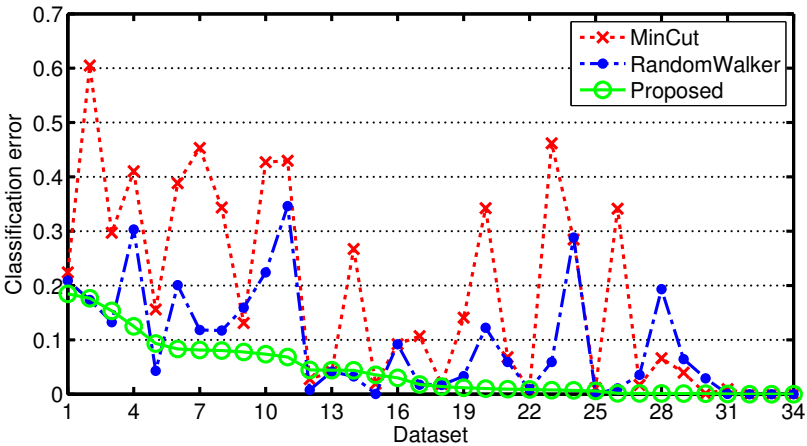


Fig. 4. Performance of the different separation algorithms. One can observe that the proposed algorithm (green, solid line) results in a lower classification error for almost every dataset with a classification error being always below 20%.

the abstraction level of edges and nodes to complete paths, we are now able to recompute the total edge costs along the selected path using $M = 20$ (see Fig. 3b) for the angle feature, $C_\Phi(e_{ij})$, allowing a more robust estimate of this feature¹. We then remove the cheapest edge of the selected path from the graph, store this edge on a list, k , and call the algorithm recursively. Upon return (the stack is popped) we try the second cheapest edge of the selected path, again calling the algorithm recursively.

The result is a set of cuts (hypotheses) $K = \{k_1, \dots, k_{n_k}\}$ where k_i is a list of edges e_{ij} that, when removed, partitions V into P and H . To determine the best k_i , we evaluate each hypotheses with respect to the overall anatomical separation and topology. In particular, both trees should be of approximately the same size, with the size of a tree $T = \{P, H\}$ being defined as

$$A_{k_i}(T) = \sum_{v_j \in T} \|v_j\|, \quad (6)$$

with $\|v_j\|$ being the length of branch v_j . At the same time, both trees shall be well separated in space maximizing

$$S_{k_i} = \|\mu_p - \mu_h\|_2 / \|\sigma_p^2 + \sigma_h^2\|_2, \quad (7)$$

where μ_p, μ_h are the mean positions and σ_p^2, σ_h^2 are the variances of the portal and hepatic voxel positions, respectively.

The complexity of the algorithm depends on the size of the graph, the number of connections between the roots, and the $B = 2$ attempts made for each level of recursion to find the best edge to cut. Since Dijkstra is of complexity $O(E + V \cdot \log V)$, our algorithm has a complexity of $O(\# \text{ of connections} \cdot B \cdot (E + V \cdot \log V))$. Note that the complexity only depends on the size (number of branches and connections between them) of the segmented vasculature and is hence not necessarily dependent on the actual image size or resolution.

4 Results

We evaluated our algorithm on 34 CTA liver datasets of varying quality, extent of disease, and protocols with slice thicknesses between 1 – 2 mm. All scans were routinely acquired on Siemens equipment at different clinical sites. Note that only vessel separation and not vessel segmentation is the focus of this paper. Evaluation was performed as follows: For each image volume, an automatic segmentation was run [8] resulting in a centerline graph that was edited manually if necessary to insure that all branches were reachable by at least one of a set of specified roots. This graph, as is, was input to the different separations

¹ We were unable to use such a large M previously because the angle computation could only be made on the sites that made up the directly involved branches. Once we have an entire path, the sampling for this computation can be much larger including multiple branches along the path, and therefore is more stable.

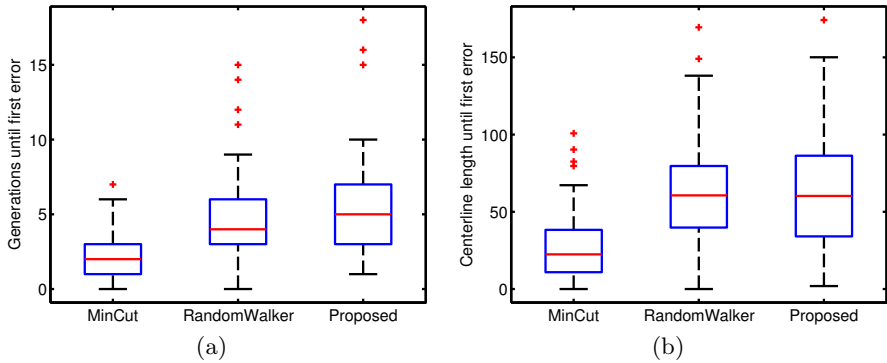


Fig. 5. Detailed analysis of the classification error shown as box plots. The distance from the root to the first misclassified was measured in terms of number of generations (edges along the path), as well as centerline length (in mm). On each box, the central mark denotes the median, the edges of the box the 25th and 75th percentile while the whiskers extend to the most extreme (considered) data points.

algorithms including our proposed algorithm. Ground truth was generated by manually labeling each branch of the graph as portal or hepatic. Validation involved comparing the manually specified labels to automatically computed ones. Classification error was on a branch-basis, i.e., number of branches correctly classified as portal or hepatic (Fig. 4). Also, we determined how far from the roots the first error appeared (Fig. 5).

For all the algorithms, the edge costs were computed by uniformly weighting (all weights equal one), $C_I(e_{ij})$, $C_R(e_{ij})$, $C_\Phi(e_{ij})$, $C_D(e_{ij})$ to obtain a final cost, $C_{\text{total}}(e_{ij})$. For our proposed algorithm, to determine the cut with the best separation, the topology metric was given twice the weight of the separation metric. $B = 2$ attempts were made for each level of recursion. For MinCut, the n-links used the weights as above and the t-links were assigned as weight 1 to both source and sink. For the Random Walker (RW) algorithm, again the same edge weights were employed.

One can observe that the proposed method results in a smaller classification error for almost every dataset. By exploiting the concept of flow in the separation, we were able to achieve an average classification error as low as 4.4% with a standard deviation of 5.3%, whereas the second best method, namely RW, resulted in an average classification error of 9.2% with a corresponding standard deviation of 9.9%. Run time was less than 15 seconds for all cases on a Dell Precision M4400.

Note that MinCut suffers also in this context often from the well-known “small-cut” problem. This can be particularly observed when examining Fig. 5. It shows that if a sub-tree is spuriously classified, the error appears significantly closer to the roots for MinCut than for both other methods. Finally, Fig. 6 shows exemplarily the separation result of each of the compared methods, with the misclassified branches highlighted in red.

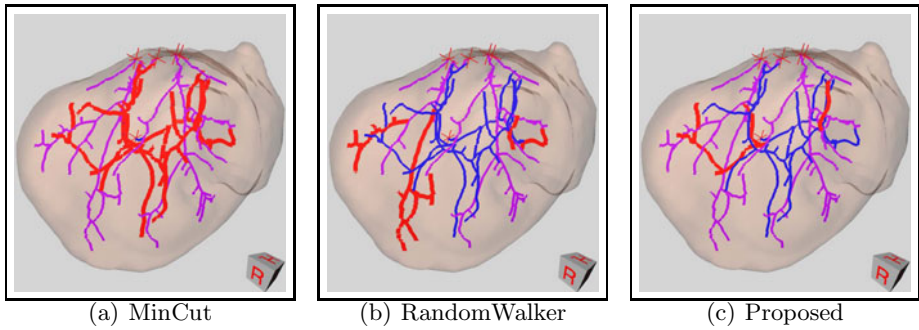


Fig. 6. Separation results showing the portal and hepatic trees in blue and purple, respectively. Incorrect classified branches are highlighted in red (best viewed in color).

5 Conclusions

Graph partitioning algorithms that employ the Ising model have become extremely commonplace in the medical imaging community. One goal of our work is to point out that these models are not always appropriate and to provide an alternative means to solve the partitioning in these situations.

Standard graph partitioning algorithms are typically limited to local features since algorithms following the Ising model consider only nearest neighbors in their energy formulations. The flow, however, depends on the root positions relative to an edge. The introduction of a recursive separation technique based on minimal paths allows the sound integration of both local and global features. It significantly reduces the classification error compared to conventional techniques. Moreover, user interaction can be easily integrated into our framework by either proposing the n -best solutions to the user in case of ambiguities or by presenting specific paths between differently labeled roots for manual separation. As the proposed algorithm does not incorporate any domain knowledge, we anticipate its applicability for other vessel separation objectives such as pulmonary vessel separation and for other imaging modalities such as MRI. Hence, we plan to investigate its performance for those applications in more detail in the future.

References

1. Bauer, C., Pock, T., Sorantin, E., Bischof, H., Beichel, R.: Segmentation of Interwoven 3D Tubular Tree Structures Utilizing Shape Priors and Graph Cuts. *Medical Image Analysis* 14(2), 172–184 (2009)
2. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(9), 1124–1137 (2004)
3. Couinaud, C.: *Le Foie - Etudes anatomiques et chirurgicales*. Masson (1957)
4. Fishman, E.K., Kuszyk, B.S., Heath, D.G., Gao, L.: Surgical Planning for Liver Resection. *Computer* 29(1), 64–72 (1996)

5. Glombitza, G., Lamadé, W., Demiris, A.M., Göpfert, M.R., Mayer, A., Bahner, M.L., Meinzer, H.P., Richter, G., Lehnert, T., Herfarth, C.: Virtual Planning of Liver Resections: Image Processing, Visualization and Volumetric Evaluation. *International Journal of Medical Informatics* 53(2-3), 225–237 (1999)
6. Grady, L.: Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(11), 1768–1783 (2006)
7. Gray, H.: *Anatomy of the Human Body*. Lea & Febiger, 1918, Philadelphia (2000)
8. Kaftan, J.N., Tek, H., Aach, T.: A Two-Stage Approach for Fully Automatic Segmentation of Venous Vascular Structures in Liver CT Images. In: *SPIE Medical Imaging*, vol. 7259, pp. 725911–1–12 (2009)
9. Kiraly, A.P., Helferty, J.P., Hoffman, E.A., McLennan, G., Higgins, W.E.: Three-Dimensional Path Planning for Virtual Bronchoscopy. *IEEE Transactions on Medical Imaging* 23(11), 1365–1379 (2004)
10. Low, G., Wiebe, E., Walji, A.H., Bigam, D.L.: Imaging Evaluation of Potential Donors in Living-donor Liver Transplantation. *Clinical Radiology* 63(2), 136–145 (2008)
11. Park, S., Bajaj, C., Gladish, G.: Artery-Vein Separation of Human Vasculature from 3D Thoracic CT Angio Scans. In: *Computational Modeling of Objects Presented in Images: Fundamentals, Methods, and Applications (CompIMAGE)*, pp. 23–30 (2006)
12. Saha, P.K., Gao, Z., Alford, S.K., Sonka, M., Hoffman, E.A.: Topomorphologic Separation of Fused Isointensity Objects via Multiscale Opening: Separating Arteries and Veins in 3-D Pulmonary CT. *IEEE Transactions on Medical Imaging* 29(3), 840–851 (2010)
13. Selle, D., Preim, B., Schenk, A., Peitgen, H.-O.: Analysis of Vasculature for Liver Surgical Planning. *IEEE Transactions on Medical Imaging* 21(11), 1344–1357 (2002)
14. Soler, L., Delingette, H., Malandain, G., Montagnat, J., Ayache, N., Koehl, C., Dourthe, O., Malassagne, B., Smith, M., Mutter, D., Marescaux, J.: Fully Automatic Anatomical, Pathological, and Functional Segmentation from CT Scans for Hepatic Surgery. *Computer Aided Surgery* 6(3), 131–142 (2001)

Efficient Algorithms for Segmenting Globally Optimal and Smooth Multi-surfaces*

Lei Xu¹, Branislav Stojkovic¹, Yongding Zhu¹, Qi Song²,
Xiaodong Wu², Milan Sonka², and Jinhui Xu¹

¹ Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY 14260, USA
{lxu,bs65,yzhu3,jinhui}@buffalo.edu

² Department of Electrical and Computer Engineering
University of Iowa
Iowa City, IA 52242, USA
{qi-song,xiaodong-wu,milan-sonka}@uiowa.edu

Abstract. Despite extensive studies in the past, the problem of segmenting globally optimal single and multiple surfaces in 3D volumetric images remains challenging in medical imaging. The problem becomes even harder in highly noisy and edge-weak images. In this paper we present a novel and highly efficient graph-theoretical iterative method with bi-criteria of global optimality and smoothness for both single and multiple surfaces. Our approach is based on a volumetric graph representation of the 3D image that incorporates curvature information. To evaluate the convergence and performance of our method, we test it on a set of 14 3D OCT images. Our experiments suggest that the proposed method yields optimal (or almost optimal) solutions in 3 to 5 iterations. To the best of our knowledge, this is the first algorithm that utilizes curvature in objective function to ensure the smoothness of the generated surfaces while striving for achieving global optimality. Comparing to the best existing approaches, our method has a much improved running time, yields almost the same global optimality but with much better smoothness, which makes it especially suitable for segmenting highly noisy images.

1 Introduction

In this paper, we consider the problem of segmenting single and multiple surfaces in 3D volumetric images. Efficient detection of globally optimal surface in volumetric images is one of the most challenging problems in imaging processing. The problem arises not only in the segmentation problem of biomedical

* The research of the first three and the last authors was supported in part by NSF through a CAREER Award CCF-0546509 and a grant IIS-0713489. The research of the other three authors was supported in part by the NSF grants CCF-0844765 and the NIH grants K25 CA123112 and R01 EB004640.

images (e.g., CT, MRI, Ultrasound, Microscopy, Optical Coherence Tomography (OCT)), but also in many other fundamental optimization problems such as surface reconstruction [13], data mining [7], ore mining [4], metric labeling [3], and radiation treatment planning [3]. Most known segmentation methods used today are energy-based approaches, which can be classified into two categories, variational segmentation and combinatorial segmentation. Variational methods include snake [11], region competition, geodesic active contours [2], active appearance models, and other methods based on level-sets [14]. In most cases, variational segmentation uses variational optimization techniques that find only a local minima of the corresponding energy function. Some versions of active contours compute a globally optimal solution, but only in 2-D. The combinatorial segmentation methods are mainly based on graph algorithms, such as minimum spanning tree [5], shortest paths and their 3D extensions [15], and graph cuts [1]. Each of these approaches has respective strengths and weakness, depending on the optimality of the cost function. In recent years, Wu et al. proposed a new detection method for optimal surfaces [13][16]. Their optimality is controlled by a cost function using only the voxel weights and some geometric constraints on the connectivity of the surface. Their approach can compute a globally optimal surface in polynomial time. However, since their cost function only considers voxel weights, their surface often contains spikes and jaggedness in noisy regions.

To provide better solution, we study in this paper a new variant of the segmentation problem, called *Optimal Smooth Surfaces (OSS)*. In the OSS problem, the objective function is a linear combination of voxel weights and their smoothness, measured by the curvature of each voxel on the segmented surfaces. Thus our problem aims to simultaneously achieve global optimality of the surfaces and local smoothness. To ensure the local smoothness, OSS uses (mean) curvature as a penalty to avoid spikes (as spikes often have very large curvature). The additional requirement on the smoothness dramatically increases the hardness of the problem. There are evidences suggesting that the OSS problem is NP-hard. To our best knowledge, there is no efficient volumetric image segmentation algorithm which guarantees both global optimality and smoothness.

For the OSS problem, we present a novel graph-theoretical approach. Our approach first uses shortest paths (or edge-disjoint shortest paths) to compute a set of local segmentations, and then stitches them to form a global surface (or surfaces). The obtained surfaces are then iteratively improved through local computation, with each iteration strictly reducing the objective value. We tested our approach using a set of OCT images (e.g., 14 data sets). Our experiments show that within a small constant (3-5) number of iterations, our approach converges rather quickly to optimal or almost optimal. Compared to existing best approaches ([10]) using OCT images, our approach avoids to perform the rather costly minimum cut computation, and hence significantly improves the running time. Our segmentation has a total weight which is almost equal to that of the weight-only approach, but with much improved smoothness. In all our experiments, our approach almost eliminates all spikes, seemingly suggests that our approach could handle highly noisy images.

Since all computations in our approach are based on local information. Our approach is particularly suitable for parallel implementation on GPUs, which could potentially allow us to segment 3 or 4D volumetric images in real time.

2 Problem Description

In general, the multi-surfaces segmentation problem can be viewed as an optimization problem for finding a set of surfaces with the minimum cost—such cost being edge and/or region based—so that the obtained surface set is feasible. In this section, we define what is meant by a feasible surface set and the cost of a set of surfaces. We first briefly review what is curvature and mean curvature of a surface.

2.1 Curvature and Mean Curvature

Curvature captures the intrinsic geometric properties of surfaces and has been widely used for measuring smoothness [12,9]. For 3D surfaces, there are two types of commonly used curvatures, mean curvature and Gaussian curvature. Mean curvature is defined as the mean of two 2D principal curvatures measured from two orthogonal directions which have the maximum and minimum bending rates. The Gaussian curvature is defined as the product of the two principal curvatures. In the OSS problem, we use the mean curvature for two main reasons. (i) Gaussian curvature is not indicative as a shape descriptor in cases where the minimal curvature is zero (e.g., cylinder). (ii) The mean curvature is relatively easier to approximate by two orthogonal directions. Note that when the surface is unknown in advance, it is highly challenging to compute the exact curvature. Thus most of existing techniques compute only approximation curvature [6]. In our proposed approaches, we use the curvatures along the x and y directions to approximate the mean curvature, which gives us the freedom to de-couple the curvature computation in the two directions. We call the curvatures in the two directions as x-curvature and y-curvature respectively.

Each x and y curvature is the curvature of a 2D curve. For a 2D curve given explicitly as a function of $y = f(x)$, its curvature is $\kappa = \frac{|y''|}{(1+y'^2)^{3/2}}$. For a more general curve with parametrization $s(t) = (x(t), y(t))$, its curvature at $s(t)$ is
$$\kappa(t) = \frac{x(t)y''(t) - y(t)x''(t)}{(x'(t)^2 + y'(t)^2)^{3/2}}.$$

2.2 Single Surface Detection

The single surface detection (OSS-1) problem can be defined as follows: Given a 3D image I (see Fig. 1) of size $n_x \times n_y \times n_z$ and with each voxel $I(x, y, z)$ associated with a non-negative weight $w(x, y, z)$ representing the inverse of the probability that $I(x, y, z)$ is a boundary voxel, find a terrain surface S (i.e., each column of I intersects S at one voxel) with the minimum cost, where the cost of S is defined as $c(S) = \sum_{I(x,y,z) \in S} \alpha w(x, y, z) + (1 - \alpha)|\kappa(x, y, z)|$,

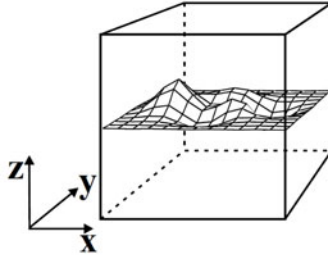


Fig. 1. Orientation of surface

where $\kappa(x, y, z)$ is the mean curvature of S at $I(x, y, z)$, $\alpha \in [0, 1]$ is a weighting constant, and the voxel weight is computed from some standard edge-detection technique. Computing the weight of the OCT data sets examined in this paper was described in [10] and [8]. An optimal surface is the one with the minimum total sum of voxel costs among all feasible surfaces that can be defined in the 3D image. The feasibility of a surface is constrained by the smoothness parameter κ_{\max} which is the maximum curvature allowed of the surface, such that for each voxel $I(x, y, z)$ on the surface, $|\kappa(x, y, z)| \leq \kappa_{\max}$. In the OSS problem, $\kappa(x, y, z) \geq 0$ since we choose $\kappa = \frac{|y''|}{(1+y'^2)^{3/2}}$ to compute it.

2.3 Multiple Surface Detection

The multiple surface detection (OSS-M) problem is an extension of OSS-1 problem discussed in the above section. In simultaneously detecting k ($k \geq 2$) distinct but interrelated surfaces, the optimality is not only determined by the inherent costs and smoothness properties of the individual surfaces, but also confined by their interrelations. Thus additional constraints are added to model the desired relations between the surfaces. Let each surface of interest S_i denoted by a function $f_i(x, y)$ mapping (x, y) pairs to z -values. For each pair of surfaces S_i and S_j , a constraint will be added to require that $|f_i(x, y) - f_j(x, y)| > \delta$ for all (x, y) , where δ is surface interaction parameter. While in general a pair of surfaces may be allowed to cross, having crossing surfaces does not make sense for the 3D images in many practical problems where the surface are expected not to intersect or overlap, and thus we will also assume $\delta > 0$. Then a set of surfaces are considered feasible if each individual surface in the set satisfies the given smoothness constraints for the surface and each pair of surfaces does not cross. The multiple surface detection problem is to find a set of feasible surfaces with minimum cost.

3 Algorithms

To solve the OSS-1 and OSS-M problem, we propose a graph-theoretical based iterative approach. In this approach, our main idea is to reduce the problem to

a sequence of shortest path problem in 2D spaces. To illustrate the idea more clearly, we present the algorithm for the OSS-1 and OSS-M problem respectively in the following sections.

3.1 Algorithm for Single Surface Detection

Instead of solving the problem in one shot, our strategy to detect the surface with global optimum (including smoothness) in the OSS-1 problem is iterative local improvement. More specifically, the algorithm for solving the OSS-1 problem contains two main steps. First, a feasible surface S with less cost is constructed as the initial solution. Second, the surface S is iteratively improved to global optimum by local adjustment. We call the first step as *initialization* and second step as *stitching*.

Initialization of OSS-1. In the initialization step, our goal is to construct a feasible surface S with less cost. To achieve this, our approach uses the following two observations: (a) The curvature can be computed independently in the x and y directions (e.g., the x-curvature of $I(x, y, z)$ depends only on voxels in the neighboring columns $Col(x - 1, y)$, $Col(x, y)$ and $Col(x + 1, y)$); (b) The curvature in one direction can be computed by using only a constant number (e.g., 3) of neighboring voxels. With these observations, we show (1) how to find $S(1, :, :)$ in the first x-slice $I(1, :, :)$ and (2) how to expand it to the rest x-slices.

In (1), $S(1, :, :)$ is actually the curve which is the intersection between the first x-slice $I(1, :, :)$ and S . To find this curve, our idea is to construct a weighted directed graph $G_1 = (V_1, E_1)$ of $I(1, :, :)$ and reduce it to a shortest path problem in G_1 which could be solved optimally. G_1 is constructed according to $I(1, :, :)$ as follows. Every vertex $V(u, c, b) \in V_1$ is a 3-tuple of voxels $u = I(x_u, y_u, z_u)$, $c = I(x_c, y_c, z_c)$ and $b = I(x_b, y_b, z_b)$ in $I(1, :, :)$, where the three voxels satisfy the following three conditions,

$$\begin{cases} x_u = x_c = x_b = 1, & \text{(c1)} \\ y_u = y_c + 1 \text{ and } y_b = y_c - 1, & \text{(c2)} \\ \kappa[V(u, c, b)] \leq \kappa_{\max} & \text{(c3)} \end{cases}$$

where (c1) is obvious, (c2) means that the three voxels should be from three consecutive neighboring columns $Col(x, y - 1)$, $Col(x, y)$ and $Col(x, y + 1)$ respectively, and (c3) is for guaranteeing the smoothness. More precisely, let $\kappa[V(u, c, b)]$ be the curvature of voxel c when voxel u and voxel b as its y-direction neighbors. We have

$$\kappa[V(u, c, b)] = \frac{1}{2}(\kappa_x[V(u, c, b)] + \kappa_y[V(u, c, b)]) \quad (1)$$

where $\kappa_x[V(u, c, b)]$ and $\kappa_y[V(u, c, b)]$ denote the x-direction and y-direction curvature of voxel c respectively. Since we do not consider x-direction curvature in this case, $\kappa[V(u, c, b)] = \frac{1}{2}\kappa_y[V(u, c, b)]$. The cost assigned to $V(u, c, b)$ is

$$cost[V(u, c, b)] = \alpha w(x_c, y_c, z_c) + (1 - \alpha)\kappa[V(u, c, b)] \quad (2)$$

An edge is added for a pair of vertices $V(u, c, b)$ and $V(u', c', b')$ if $c = u'$ and $b = c'$ which indicate that the pair of vertices share two common voxels. After adding two dummy vertices in G_1 (one as source and the other as sink), the problem in the first x-slice $I(1, :, :)$ can be reduced to a shortest path problem in G_1 . Since G_1 is a DAG (directed acyclic graph), it could be solved optimally by topological sort algorithm in linear time $O(n_y n_z \kappa_{\max}^3)$. The cost of the resulting path includes the weight and the y-curvature of each voxel on the path. Thus $S(1, :, :)$ is formed by the voxels appeared on the path.

In (2), we generalize the above idea and expand $S(1, :, :)$ on the rest x-slice one by one to form a feasible surface S . There are two differences which could be illustrated by example of generating $S(i, :, :)$ from partial surface $S(1 : i - 1, :, :)$. First, in addition to the above three conditions (c1),(c2) and (c3), every vertex $V(u, c, b) \in V_i$ in $G_i = (V_i, E_i)$ has to satisfy one more condition (c4) which restricts the y-coordinate y_c of voxel c by

$$\kappa[I(i - 1, y_c, z_l)] \leq \kappa_{\max}, \quad (c4)$$

where $I(i - 1, y_c, z_l)$ is the voxel of S on column $Col(i - 1, y_c)$ and $\kappa[I(i - 1, y_c, z_l)]$ is the curvature of that voxel. By adding (c4), the smoothness property of voxel $I(i - 1, y_c, z_l)$ is preserved because its three neighboring voxels are fixed and the smoothness is not violated before we compute $S(i, :, :)$. The second difference is on the cost assigned to the vertex $V(u, c, b)$. More specifically, the equation (2) is replaced by the following one,

$$\begin{aligned} cost[V(u, c, b)] = & \alpha(2w(x_c, y_c, z_c) + w(i - 1, y_c, z_l)) \\ & + (1 - \alpha)(\kappa[V(u, c, b)] + \Delta\kappa[I(i - 1, y_c, z_l)]) \end{aligned} \quad (3)$$

where $w(i - 1, y_c, z_l)$ is the cost of voxel $I(i - 1, y_c, z_l)$ on surface S and $\Delta\kappa[I(i - 1, y_c, z_l)]$ is the change of curvature of voxel $I(i - 1, y_c, z_l)$ since it does not have voxel c as its x-direction neighbor previously. In addition to the weight of voxel $I(u, c, b)$ itself, we need to include the weight adjustment for $I(i - 1, y_c, z_l)$ caused by voxel $I(u, c, b)$ and thus the coefficient 2 before $w(x_c, y_c, z_c)$ follows. The new cost formula (3) reflects the overall change of cost for chosen voxel c on frame $S(i, :, :)$. The size of G_i is reduced from $O(n_y n_z \kappa_{\max}^3)$ to $O(n_y \kappa_{\max}^4)$. Finally, $S(i, :, :)$ is formed by the voxels appeared on the shortest path of G_i .

Stitching of OSS-1. To form an optimal surface from S constructed in the above section, our idea is to iteratively improve S by local adjustment called stitching. In this section, we first interpret what is stitching and then show how to iteratively apply it to obtain an optimal surface.

Firstly, it is easy to see that before S achieves optimum, there must exist an x-frame $S(i, :, :)$ on which the voxels are not entirely located at optimal positions. Otherwise S must be optimal. Thus in order to improve S , one way is to adjust $S(i, :, :)$ locally when we fix the rest of S other than $S(i, :, :)$. To illustrate the idea of stitching more clearly, let $S_l(1 : i - 1, :, :)$ and $S_r(i + 1 : n_x, :, :)$ be two partial surfaces. To stitch $S_l(1 : i - 1, :, :)$ and $S_r(i + 1 : n_x, :, :)$ in slice $I(i, :, :)$, our main idea is to fix the selected voxels in slices $I(i - 1, :, :)$ and $I(i + 1, :, :)$, and

convert the stitching problem into a 2D shortest path problem in graph $G_s(i)$ which is similar to the graph G_i discussed in the above section. However, there are two differences from the initialization step. Firstly, in addition to the above four conditions (c1)-(c4), every vertex $V(u, c, b) \in V_s(i)$ in $G_s(i) = (V_s(i), E_s(i))$ has to satisfy one more condition (c5) which restricts the y-coordinate y_c of voxel c to preserve the smoothness property of S . More precisely,

$$\kappa[I(i + 1, y_c, z_r)] \leq \kappa_{\max}, \quad (c5)$$

where $I(i+1, y_c, z_r)$ is the voxel of S on column $Col(i+1, y_c)$ and $\kappa[I(i+1, y_c, z_r)]$ is the curvature of that voxel. Secondly, different from the initialization step, the cost function in this case includes the weight and the mean curvature of voxels in slice $I(i, :, :)$, as well as the adjustment of the x-curvatures for those selected voxels in slices $I(i - 1, :, :)$ and $I(i + 1, :, :)$. We replace equation (3) by

$$cost[V(u, c, b)] = \alpha costw[V(u, c, b)] + (1 - \alpha) costk[V(u, c, b)], \quad (4)$$

where

$$costw[V(u, c, b)] = w(x_c, y_c, z_c) + w(i - 1, y_c, z_i) + w(i + 1, y_c, z_r) + 2(w(x_c, y_c, z_c) - w(x'_c, y'_c, z'_c)) \quad (5)$$

and

$$costk[V(u, c, b)] = \kappa[V(u, c, b)] + \Delta\kappa[I(i - 1, y_c, z_i)] + \Delta\kappa[I(i + 1, y_c, z_r)], \quad (6)$$

and $w(x'_c, y'_c, z'_c)$ is the weight cost of voxel on column $Col(i, y_c)$ of unadjusted $S(i, :, :)$. We set coefficient 2 before $(w(x_c, y_c, z_c) - w(x'_c, y'_c, z'_c))$ in equation (5) to reflect the weight adjustment for both selected voxels in slices $I(i - 1, :, :)$ and $I(i + 1, :, :)$. In equation (6), we include the curvature of voxels in slice $I(i, :, :)$ as well as adjustment of the x-curvatures for both selected voxels in slices $I(i - 1, :, :)$ and $I(i + 1, :, :)$. Thus an locally improved $S(i, :, :)$ can be obtained by shortest path algorithm on graph $G_s(i)$ within $O(n_y \kappa_{\max}^3)$ time.

Since stitching on one x-frame could locally improve S , our strategy is repeatedly apply the stitching algorithm on the obtained surface with each time shifting the to-be-stitched slices by on slice. For instance, if the current iteration of stitching is on slices $I(1, :, :), I(3, :, :), \dots, I(2i - 1, :, :), \dots$ (see Fig. 2 (a)-(d)), then

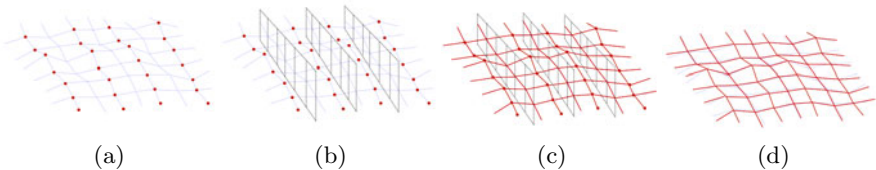


Fig. 2. One iteration of stitching. (a) Before stitching; (b) Fixed the red voxels; (c) Stitching; (d) The new surface.

in the next iteration, the stitching is on slices $I(2, :, :), I(4, :, :), \dots, I(2i, :, :), \dots$. It can be shown (through mathematical proof and experiments) that the cost of the surface will monotonically decrease in each iteration and an optimal solution can be finally obtained. In all our experiments with the OCT images, our approach converges to the optimal (or almost optimal) within a small constant (3-5) number of iterations.

3.2 Algorithm for Multiple Surface Detection

To solve the OSS-M problem, we extend the algorithm of OSS-1 problem discussed in the above section. The algorithm for the OSS-M problem also contains two main steps, initialization and stitching. Thus we only focus on the extension of algorithm to avoid repetition.

Initialization of OSS-M. In the initialization step, our goal is to construct a set of feasible surfaces S_1 to S_k with less cost, where k is the number of surfaces in the 3D image. To achieve this, we show (1) how to find $S_1(1, :, :)$ to $S_k(1, :, :)$ in the first x-slice $I(1, :, :)$ and (2) how to expand $S_1(1, :, :)$ to $S_k(1, :, :)$ to the rest x-slices. In (1), we first construct G'_1 similar with G_1 in the above section. In order to find $S_1(1, :, :)$ to $S_k(1, :, :)$, we need to compute k shortest edge-disjoint paths. It could result in a wrong solution if we try to solve the problem by executing k times the following two steps: compute a shortest path and remove its arcs from the graph G'_1 . To see this, let us consider the example in Fig. 3a. In this example, we intend to determine two edge-disjoint shortest paths from 1 to 4. The resulting network is the one shown in Fig. 3b, where the two shortest path from 1 to 4 is $\{1, (1, 2), 2, (2, 4), 4\}$ and $\{1, (1, 3), 3, (3, 4), 4\}$. Note that the shortest path in Fig. 3a $\{1, (1, 2), 2, (2, 3), 3, (3, 4), 4\}$ is not one of the two shortest disjoint paths. To overcome this difficulty, one way is using the Suurballe's algorithm. In this algorithm, the k shortest arc-disjoint paths problem is considered to be equivalent to a minimum cost flow problem in which a feasible flow of value k is sought from an initial vertex to a terminal vertex such that the total cost is minimized. The flow in each arc can not exceed the unity. Obviously, if such a

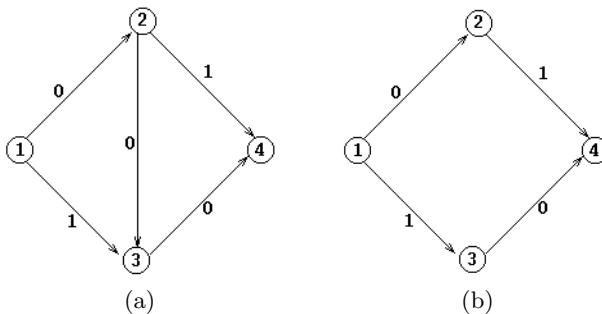


Fig. 3. Counter-example of applying shortest path algorithms k times

flow does not exist, the same can be concluded about the k disjoint paths. In order to obtain the k paths, we have to construct a new graph G''_1 , whose set of arcs is the set of arcs for which the flow is one. To determine the shortest disjoint path, the second disjoint shortest, the third disjoint shortest, and so up to the k -th shortest disjoint path, we have to execute k times the following two steps (1) Compute a shortest path in the graph G''_1 and (2) Remove its arcs. However, it is difficult to directly apply this algorithm to our problem because (1) the surface interaction parameter δ is not easily encoded as a constraint and (2) the running time of minimum cost flow algorithm is $O(E \log(E + V \log V))$. Since our goal is to find a set of feasible solutions, our strategy is to compute the shortest path one by one, and remove δ vertices above and below for each vertex on the shortest path. To avoid the effect caused by the deletion of vertices in the previous step of computing shortest path, we relax κ_{\max} when we compute the second to k shortest paths. Finally, we sort the voxels on each column by their z -coordinates. If any voxel of $S_1(1, :, :)$ to $S_k(1, :, :)$ violates the smoothness requirement, we replace it by its neighbor within the same column and thus make the result feasible. In such a way, since k is a small constant number (3 or 4), the running time is still linear to the size of graph G''_1 . Thus in the next round, we expand $S_1(1, :, :)$ to $S_k(1, :, :)$ on x -frame $I(2, :, :)$ one by one. By repeating the same procedure $(n_x - 1)$ times, we expand $S_1(1, :, :)$ to $S_k(1, :, :)$ to a set of feasible surfaces.

Stitching of OSS-M. To obtain a set of surfaces with global optimum, we follow the same idea of algorithm for OSS-1 problem. For instance, if the current iteration of stitching is on slices $I(1, :, :), I(3, :, :), \dots, I(2i - 1, :, :), \dots$, then in the next iteration, the stitching is on slices $I(2, :, :), I(4, :, :), \dots, I(2i, :, :), \dots$. For one iteration, the set of k surfaces is stitched one by one similar to OSS-1's. However, to avoid intersection of surfaces, we add one more restriction (c6) of the y -coordinate y_c of voxel c on the to-be-stitched x -frame of surface S ,

$$y_l < y_c < y_u, \quad (\text{c6})$$

where y_l (y_u) is the y -coordinate of the voxels on the same column of y_c and belonged to the surface below (above) to S .

4 Experiments

The proposed algorithms were examined on 14 datasets of 3D OCT images ($200 \times 200 \times 256$). To prove the effectiveness of our algorithms, we mainly focused on segmenting surface 1, 6 and 7 since they were computed by the same weight function [8]. The average of the two tracings from two human experts were used as the reference standard. The unsigned surface positioning errors were computed for each surface as a distance between the computed results and the reference standard. In our experiments, we apply our algorithm to the down-sampled 3D images ($100 \times 100 \times 128$). The results were up-sampled to the size of original 3D images before we compare them to the reference standard. Our

algorithm is implemented by C++ and LEDA (Library of Efficient Data types and Algorithms)-5.2. The experiments were conducted on a Linux workstation (2.4 GHz, 4GB memory).

5 Results

Mean \pm SD in μm		
Surface 1	Surface 6	Surface 7
3.65 ± 2.70	2.35 ± 1.56	7.9 ± 4.89

To test the performance of our algorithms, we mainly focus on the results of OSS-M for the case of $k = 3$ (i.e., detecting surfaces 1,6, and 7), since OSS-1very well detects surface 6. The computed unsigned position errors are summarized in the table to the right where $\alpha = 0.75$ and $\kappa_{\max} = 5$. Compared to previous result reported in the literature [10] in which all 7 surfaces were detected simultaneously, our segmentation results are more accurate for surface 1 and 6 (i.e., 4.0 ± 1.2 and 3.5 ± 2.0 in [10]), and comparable for surface 7 (i.e., 7.8 ± 2.5 in [10]). An example result with $\alpha = 0.75$ and $\kappa_{\max} = 5$ shown in Fig. 4(a) is compared to the reference standard in Fig. 4(b). Our result (shown in green line) is consistent with the reference standard (shown in red line).

To analyze the convergence of OSS-M, we measure the cost deduction in each iteration. Fig. 5(a) shows that for each iteration, the deduction of the surfaces' cost is positive before it goes down to 0. This means that (1) the cost of all surfaces is monotonically decreased, (2) the proposed method yields optimal solutions in at most 5 iterations, and (3) the largest cost deduction occurs in the first two iterations.

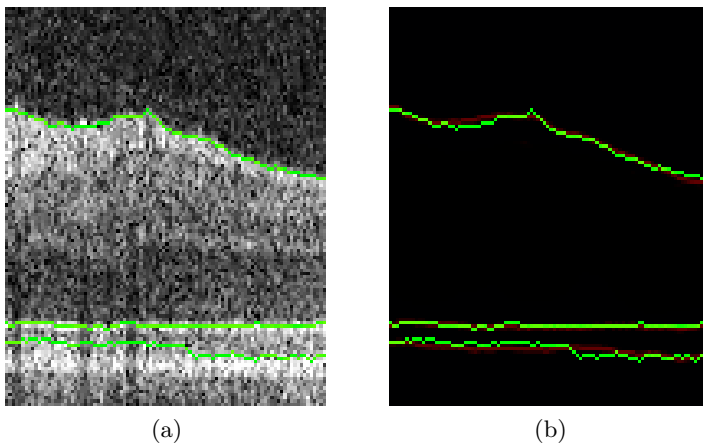


Fig. 4. (a) Example result shown on a 2-D scan from one of the 3D images; (b) Red: reference standard; Green: our result

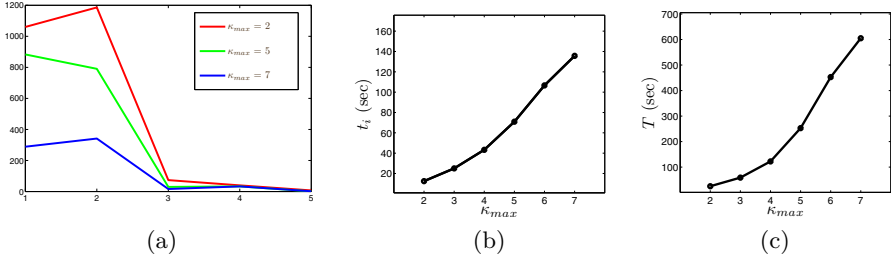


Fig. 5. (a) Cost deduction for all surfaces in each iteration; (b) Average running time for an iteration; (c) Total running time

To evaluate the running time of our algorithm, we measure the average running time (in seconds) t_i for a stitching iteration (obtained from dividing the total time by the number of iterations; see Fig. 5 (b)) and the total running time T (see Fig. 5 (c)) for different κ_{max} values. It is interesting to see that even though the searching space increases super linearly with respect to κ_{max} , the running time is roughly a linear function of κ_{max} .

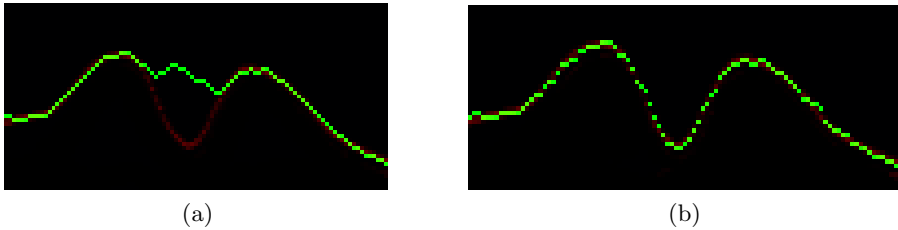


Fig. 6. (a) $\kappa_{max} = 2$; (b) $\kappa_{max} = 5$

To examine the influence of κ_{max} on the segmentations, we intentionally choose some small κ_{max} (e.g., $\kappa_{max} = 2$). Fig. 6 shows that for too small κ_{max} value, the algorithm may generate wrong result and could easily correct it when increasing κ_{max} . Our experiments show that $\kappa_{max} = 5$ is sufficient for all tested images. This also means that the average running time of our algorithm is around 4 minutes.

To visualize the improved smoothness by our algorithm, we measure the total curvature and total weight of a particular surface for different α values. Fig. 7(a) shows the surface for $\alpha = 1$, meaning that the cost function uses only weight. Fig. 7(b) is the same surface with $\alpha = 0.75$. From Fig. 7, we can see that our algorithm could improve the smoothness of the surface and remove all spikes by incorporating curvature into the cost function. The total curvature improves about 5% while the total weight increases only 0.2%.

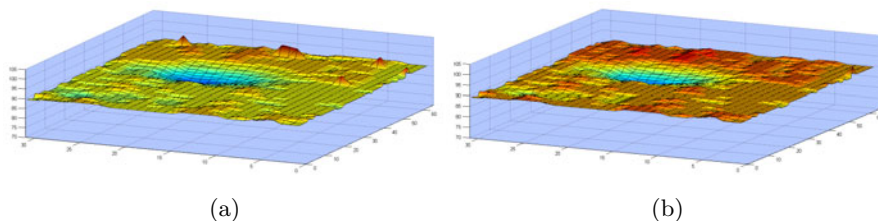


Fig. 7. (a) $\alpha = 1$ (weight only); (b) $\alpha = 0.75$ ($0.75 \times \text{weight} + 0.25 \times \text{curvature}$)

6 Discussion and Conclusion

We have presented efficient algorithms for segmenting globally optimal and smooth multi-surfaces. These are the first algorithms that utilize curvature in the objective function to ensure smoothness. Comparing to the best existing approaches using graph cut algorithms [10], our method can detect more smooth surfaces with much improved running time. The proposed method is also suitable for parallel implementation on GPUs, which could potentially allow us to segment highly noisy volumetric images in real time.

References

1. Boykov, Y., Kolmogorov, V.: Computing geodesics and minimal surfaces via graph cuts. In: ICCV, pp. 26–33 (2003)
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. *International Journal of Computer Vision* 22(1), 61–79 (1997)
3. Chekuri, C., Khanna, S., Naor, J., Zosin, L.: A linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM J. Discrete Math.* 18(3), 608–625 (2004)
4. Denby, S.: Genetic algorithms: A new approach to pit optimization. In: Proc. 24th Int. Symp. on the Application of Computers and Operations Research in the Mineral Industry, pp. 126–133 (1993)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. *International Journal of Computer Vision* 59(2), 167–181 (2004)
6. Flynn, P.J., Jain, A.K.: On reliable curvature estimation. In: CVPR 1989: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 110–116 (1989)
7. Fukuda, T., Morimoto, Y., Morishita, S., Tokuyama, T.: Data mining with optimized two-dimensional association rules. *ACM Trans. Database Syst.* 26(2), 179–213 (2001)
8. Garvin, M.K., Abramoff, M.D., Kardon, R., Russell, S.R., Wu, X., Sonka, M.: Intraretinal layer segmentation of macular optical coherence tomography images using optimal 3-d graph search. *IEEE Trans. Med. Imaging* 27(10), 1495–1505 (2008)
9. Ge, C.L.W.Y., Vining, D.J.: Segmentation in virtual colonoscopy using a geometric deformable model. *Computerized Medical Imaging and Graphics* 30(1), 17–30 (2006)

10. Haeker, M., Wu, X., Abràmoff, M.D., Kardon, R., Sonka, M.: Incorporation of regional information in optimal 3-D graph search with application for intraretinal layer segmentation of optical coherence tomography images. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 607–618. Springer, Heidelberg (2007)
11. Kass, M., Witkin, A.P., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
12. Leventon, M.E., Faugeras, O., Eric, W., Grimson, W.E.L., Wells III, W.M.: Level set based segmentation with intensity and curvature priors (2000)
13. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(1), 119–134 (2006)
14. Osher, S., Paragios, N.: *Geometric Level Set Methods in Imaging, Vision, and Graphics*. Springer-Verlag New York, Inc., Secaucus (2003)
15. Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.): *EMMCVPR 2005*. LNCS, vol. 3757. Springer, Heidelberg (2005)
16. Wu, X., Chen, D.Z.: Optimal net surface problems with applications. In: Widmayer, P., Triguero, F., Morales, R., Hennessy, M., Eidenbenz, S., Conejo, R. (eds.) *ICALP 2002*. LNCS, vol. 2380, pp. 1029–1042. Springer, Heidelberg (2002)

Graph Cuts with Invariant Object-Interaction Priors: Application to Intervertebral Disc Segmentation

Ismail Ben Ayed^{1,2}, Kumaradevan Punithakumar^{1,2}, Gregory Garvin²,
Walter Romano², and Shuo Li^{1,2}

¹ GE Healthcare, London, ON, Canada

² The University of Western Ontario, ON, Canada

Abstract. This study investigates novel *object-interaction* priors for graph cut image segmentation with application to intervertebral disc delineation in magnetic resonance (MR) lumbar spine images. The algorithm optimizes an original cost function which constrains the solution with learned prior knowledge about the geometric interactions between different objects in the image. Based on a global measure of similarity between distributions, the proposed priors are intrinsically invariant with respect to translation and rotation. We further introduce a scale variable from which we derive an original *fixed-point equation (FPE)*, thereby achieving scale-invariance with only few fast computations. The proposed priors relax the need of costly pose estimation (or registration) procedures and large training sets (we use a single subject for training), and can tolerate shape deformations, unlike template-based priors. Our formulation leads to an *NP-hard* problem which does not afford a form directly amenable to graph cut optimization. We proceeded to a relaxation of the problem via an *auxiliary function*, thereby obtaining a nearly real-time solution with few graph cuts. Quantitative evaluations over 60 intervertebral discs acquired from 10 subjects demonstrated that the proposed algorithm yields a high correlation with independent manual segmentations by an expert. We further demonstrate experimentally the invariance of the proposed geometric attributes. This supports the fact that a single subject is sufficient for training our algorithm, and confirms the relevance of the proposed priors to disc segmentation.

1 Introduction

Accurate segmentation of lumbar spine discs in magnetic resonance (MR) images is useful in quantifying intervertebral disc degeneration (IDD) and assisting surgical spine procedures [10]. Quantitative disc measurements often resort to time-consuming, manual segmentations [12]. Related works generally focused on automating vertebrae segmentation [7,8,4], and only few studies investigated disc segmentation [10,5,15] or detection [6,14]. The problem is acknowledged difficult because of the similarities in intensity and shape profiles between the discs and their surrounding regions (refer to the example in Fig. 2 b). Based on standard

techniques such as fuzzy clustering [10], watersheds [5], and edge detection [15], existing methods require intensive user inputs (e.g. user landmarks for each disc [10]), a large training set, pose registration, and a heavy computational load. The detection algorithms in [6,14] used probabilistic graphical models to embed prior information on the geometric interaction between pairs of discs. Enforcing the distance between the discs to fall within a known range, these algorithms led to promising detection results. Unfortunately, they yield only disc centroids, not segmentations. Furthermore, their interaction priors are not scale-invariant.

In the scope of image segmentation, embedding priors on the geometric interactions between different objects has been generally avoided, mainly because it leads to challenging optimization problems. Existing geometric priors commonly bias each single target region towards a known set of *template* shapes or atlases, independently of other related regions [10,13,17]. Although very useful in some cases, template-based priors require costly pose estimation (or registration), as well as a large set of training examples. Furthermore, they are sensitive to shape deformations. An unknown pathological case outside the set of learned templates, for instance a degenerated disc with shape irregularities, may not be recovered.

A recent notable study by Toshev et al. [16] demonstrated that a *global* shape description can yield very competitive image segmentation results. The shape model of an object is the histogram of distances and orientations corresponding to all pairs of points on the training shape, and the segmentation is sought following the optimization of the L_1 similarity between histograms, an *NP-hard* problem which the authors solve via semidefinite programming relaxation [16]. Unlike template-based priors, this global shape description is translation invariant, requires a single training example, and can tolerate shape deformations. Unfortunately, the description is not scale or rotation invariant and, therefore, requires heavy computations to handle scale variations. It is worth noting that the contribution in [16] follows on several recent segmentation studies which have shown the usefulness of intensity (or color) priors based on global measures of similarity between distributions [2,11,1]. Although helpful, such intensity priors are not sufficient to obtain satisfying segmentations when the target regions and their surrounding structures have almost the same intensity profiles (refer to the example in Fig. 2 b).

Inspired by the global shape and intensity descriptions in [16,2,11,1], we propose novel *object-interaction* priors and their application to intervertebral disc delineation in MR lumbar spine images. The algorithm optimizes an original cost function which constrains the solution with learned prior knowledge about the geometric interactions between different objects in the image. Based on the Bhattacharyya measure and the distributions of the geometric relationships between pairs of points within different objects, the proposed priors are intrinsically invariant with respect to translation and rotation. We further introduce a scale variable from which we derive an original *fixed-point equation (FPE)*, thereby achieving scale-invariance with only few fast computations. The proposed priors relax the need of costly pose estimation (or registration) procedures and large

training sets (we used a single subject for training), and can tolerate shape deformations, unlike template-based priors. Our formulation leads to an *NP-hard* problem which does not afford a form directly amenable to efficient graph cut optimization. We proceeded to a relaxation of the problem via an *auxiliary function*, thereby obtaining a nearly real-time solution with few graph cuts. Quantitative evaluations over 60 intervertebral discs acquired from 10 subjects demonstrated that the proposed algorithm yields a high correlation with independent manual segmentations by an expert. We further demonstrate experimentally that the proposed geometric attributes do not vary significantly from one subject to another (refer to the illustration in Fig. 4). This experiment supports the fact that a single subject is sufficient for training, and confirms the relevance of the proposed priors to disc segmentation.

2 Formulation

Consider a MR spine image function $I(p) = I_p : \mathcal{P} \subset \mathbb{R}^2 \rightarrow \mathcal{I} \subset \mathbb{R}$, with \mathcal{P} the image domain and \mathcal{I} the set of intensity variables. Given a simple user input, which consists of an elliptic approximation of the boundary of only one disc, the purpose is to delineate all subsequent discs in the image (cf. the examples in Fig. 3). For each disc, the solution is efficiently obtained following the minimization of an original *discrete* energy containing three distribution similarity measures. The first measure is an intensity prior which embeds information about image data within the target disc. The last two measures are geometric priors which embed information about the interactions between neighboring discs.

2.1 The Energy

General definitions and notations: To introduce the energy, we first consider the following general definitions for any labeling (or segmentation) $\mathcal{L}(p) = \mathcal{L}_p : \mathcal{P} \rightarrow \{0, 1\}$, any function $J(p) = J_p : \mathcal{P} \rightarrow \mathcal{J}$, and any set of variables $\mathcal{J} \subset \mathbb{R}$.

- $\mathbf{R}_1^{\mathcal{L}}$ and $\mathbf{R}_0^{\mathcal{L}}$ are the complementary regions defined by $\mathbf{R}_1^{\mathcal{L}} = \{p \in \mathcal{P} / \mathcal{L}_p = 1\}$ and $\mathbf{R}_0^{\mathcal{L}} = \{p \in \mathcal{P} / \mathcal{L}_p = 0\} = \mathcal{P} \setminus \mathbf{R}_1^{\mathcal{L}}$.
- $\mathbf{P}_{\mathcal{L}, \mathcal{J}}$ is the kernel density estimate (KDE) of the distribution of function $J : \mathcal{P} \rightarrow \mathcal{J}$ within region $\mathbf{R}_1^{\mathcal{L}}$:

$$\forall j \in \mathcal{J}, \quad \mathbf{P}_{\mathcal{L}, \mathcal{J}}(j) = \frac{\sum_{p \in \mathbf{R}_1^{\mathcal{L}}} K(j - J_p)}{\mathbf{A}(\mathbf{R}_1^{\mathcal{L}})} \text{ with } K(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(y)^2}{2\sigma^2}} \quad (1)$$

$\mathbf{A}(\mathbf{R})$ denotes the number of pixels within region \mathbf{R} , and K is the Gaussian kernel (σ is the width of the kernel).

- $\mathcal{B}_{\mathcal{J}}(f, g)$ is the *Bhattacharyya* coefficient measuring the amount of overlap (similarity) between two distributions f and g : $\mathcal{B}_{\mathcal{J}}(f, g) = \sum_{j \in \mathcal{J}} \sqrt{f(j)g(j)}$. Note that the values of $\mathcal{B}_{\mathcal{J}}$ are always in $[0, 1]$, where 0 indicates that there is no overlap, and 1 indicates a perfect match between the distributions.

The intensity prior: Let \mathcal{M}_I denotes a model distribution of intensity learned from image data I within the user-provided elliptic region (cf. the examples in Fig. 3). The minimization of this prior identifies each of the subsequent discs as a region whose intensity distribution most closely matches \mathcal{M}_I :

$$\mathbf{B}_I(\mathcal{L}) = -\mathcal{B}_I(\mathbf{P}_{\mathcal{L},I}, \mathcal{M}_I) = -\sum_{i \in \mathcal{I}} \sqrt{\mathbf{P}_{\mathcal{L},I}(i) \mathcal{M}_I(i)} \quad (2)$$

Although helpful, this prior is not sufficient to obtain satisfying segmentations because the discs and some surrounding structures have almost the same intensity profiles (Fig. 2 b depicts an example of segmentation with this prior).

The object-interaction priors: Assume an elliptic approximation of the boundary of a previously segmented disc is given. Let $O \in \mathcal{P}$ be the center of the ellipse and \mathbf{u} a unit vector pointing along the minor axis (an illustration is given in Fig. 1). For each point $p \in \mathcal{P}$, let $d\mathbf{p}$ the vector pointing from O towards p . Consider the following geometric functions:

$$\begin{cases} C(p) = C_p = \frac{\langle d\mathbf{p}, \mathbf{u} \rangle}{\|d\mathbf{p}\|} = \cos(\alpha_p) : \mathcal{P} \rightarrow \mathcal{C} \subset [-1, 1] \\ D(p) = D_p = \|p - O\| : \mathcal{P} \rightarrow \mathcal{D} \subset \mathbb{R} \end{cases} \quad (3)$$

C measures the cosine of the angle between vectors $d\mathbf{p}$ and \mathbf{u} , whereas D evaluates the distance between p and O . To constrain the segmentation with prior geometric information, we propose to optimize the following two constraints.

1. The angle-distribution prior: We assume that the distribution of angle function C within the target disc follows a model \mathcal{M}_C which can be learned from a different training subject. To find a disc region whose angle-function distribution most closely matches \mathcal{M}_C , we propose to minimize:

$$\mathbf{B}_C(\mathcal{L}) = -\mathcal{B}_C(\mathbf{P}_{\mathcal{L},C}, \mathcal{M}_C) = -\sum_{c \in \mathcal{C}} \sqrt{\mathbf{P}_{\mathcal{L},C}(c) \mathcal{M}_C(c)} \quad (4)$$

This geometric prior is invariant to translation, rotation, and scale of a pair of discs. We examined such invariance experimentally: using manual segmentations, we plotted in the first line of Fig. 4 the angle-function distributions corresponding to 10 different subjects and 3 different disc pairs. The high similarity between these 10 distributions supports the fact that a single subject is sufficient for training, and confirms the invariance of the prior.

2. The distance-distribution prior: We assume that the distribution of distance function D within the target disc follows a model \mathcal{M}_D learned from the training subject. Our purpose is to find a region whose distance distribution most closely matches \mathcal{M}_D by minimizing

$$\mathbf{B}_D(\mathcal{L}) = -\mathcal{B}_D(\mathbf{P}_{\mathcal{L},D}, \mathcal{M}_D) = -\sum_{d \in \mathcal{D}} \sqrt{\mathbf{P}_{\mathcal{L},D}(d) \mathcal{M}_D(d)} \quad (5)$$

This geometric prior is invariant to translation and rotation, but not to scale. Using manual segmentations, we plotted in the second line of Fig. 4 the distance distributions corresponding to 10 different subjects and 3 different disc pairs. The distributions have similar shapes, but different supports. The shifts between the distributions are due to inter-subject variations in scale.

Introducing a scale variable: To achieve scale invariance of the distance prior, we relax the assumption that the distribution of D within the target disc follows exactly the learned model \mathcal{M}_D . We rather assume that it belongs to the following set of distributions parametrized with a scale variable s :

$$\{\mathcal{M}_D(\cdot, s) : \mathcal{D} \times \mathbb{R} \rightarrow [0, 1] / \mathcal{M}_D(d, s) = \mathcal{M}_D(d + s), s \in \mathbb{R}\} \quad (6)$$

In this case, we rewrite the distance-distribution prior as follows

$$\mathbf{B}_D(\mathcal{L}, s) = -\mathcal{B}_D(\mathbf{P}_{\mathcal{L}, D}, \mathcal{M}_D(\cdot, s)) = - \sum_{d \in \mathcal{D}} \sqrt{\mathbf{P}_{\mathcal{L}, D}(d) \mathcal{M}_D(d + s)} \quad (7)$$

where s is an additional variable which has to be optimized with the labeling. Based on global rather pixel-wise information, the proposed geometric priors

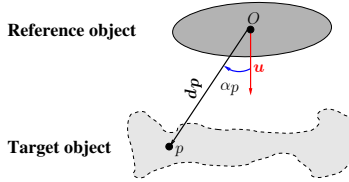


Fig. 1. The geometric relationships between pairs of points within different objects: the upper object (reference) is segmented with the proposed method at a previous stage. The lower object is the target sought at the current stage of the segmentation process.

relaxes (1) extensive learning/modeling of geometric characteristics (we use a single subject for training) and (2) complex optimization with respect to several translation and rotation parameters.

We propose to minimize an energy containing the intensity and object-interaction priors as well as a regularization term for smooth boundaries:

$$\{\mathcal{L}^{opt}, s^{opt}\} = \arg \min_{\mathcal{L}, s} \mathcal{F}(\mathcal{L}, s) = \underbrace{\mathbf{B}_I(\mathcal{L})}_{Image \text{ prior}} + \underbrace{\beta(\mathbf{B}_C(\mathcal{L}) + \mathbf{B}_D(\mathcal{L}, s))}_{Object-interaction \text{ priors}} + \underbrace{\lambda \mathbf{S}(\mathcal{L})}_{Smoothness} \quad (8)$$

$\mathbf{S}(\mathcal{L})$ ensures label consistency of neighboring pixels (\mathcal{N} is a neighborhood system): $\mathbf{S}(\mathcal{L}) = \sum_{\{p, q\} \in \mathcal{N}} \frac{\delta_{\mathcal{L}(p) \neq \mathcal{L}(q)}}{\|p - q\|}$ with $\delta_{x \neq y} = 1$ if $x \neq y$ and $\delta_{x \neq y} = 0$ if $x = y$. γ and λ are positive constants balancing the contribution of each term.

2.2 Optimization

Energy (8) depends on two type of variables (labeling \mathcal{L} and scale variable s). Therefore, we proceed to an iterative two-step optimization strategy. The first step consists of fixing s and optimizing $\mathcal{F}(\mathcal{L}, s)$ with respect to the labeling via auxiliary-function relaxation and graph cuts. The second step consists of finding the optimal scale variable via fixed-point-equation updates, given the labeling provided by the first step. The algorithm iterates these two steps until convergence. Each step decreases $\mathcal{F}(\mathcal{L}, s)$ with respect to a variable. Thus, the algorithm is guaranteed to converge.

Step 1–Graph Cut optimization via auxiliary-function relaxation: The *global* terms \mathbf{B}_I , \mathbf{B}_C and \mathbf{B}_D in (8) are not directly amenable to max-flow optimization because they do not reference pixel or pixel-neighborhood penalties. They evaluates a global similarity measure between distributions and, therefore, the ensuing optimization problem is challenging and *NP-hard*. To obtain a solution efficiently, we proceed to an *auxiliary-function relaxation* of the problem in (8). Rather than optimizing directly cost function \mathcal{F} , the relaxation optimizes iteratively a sequence of upper bounds of $\mathcal{F}(\mathcal{L}, s)$, denoted $\mathcal{A}(\mathcal{L}, \mathcal{L}^n, s)$, $n = 1 \dots$:

$$\mathcal{L}^{n+1} = \arg \min_{\mathcal{L} \in \{0,1\}} \mathcal{A}(\mathcal{L}, \mathcal{L}^n, s), \quad n = 1 \dots \quad (9)$$

under the following constraints:

$$\begin{cases} \mathcal{F}(\mathcal{L}, s) \leq \mathcal{A}(\mathcal{L}, \mathcal{L}^n, s), & n = 1 \dots \\ \mathcal{F}(\mathcal{L}, s) = \mathcal{A}(\mathcal{L}, \mathcal{L}, s) \end{cases} \quad (10)$$

\mathcal{A} is called auxiliary function of \mathcal{F} . Such relaxations are commonly used in the Nonnegative Matrix Factorization (NMF) literature for challenging problems [9]. Using the constraints in (10), and by definition of minimum in (9), one can show that the sequence of solutions in (9) yields a monotonically decreasing sequence of \mathcal{F} :

$$\mathcal{F}(\mathcal{L}^{(n)}, s) = \mathcal{A}(\mathcal{L}^{(n)}, \mathcal{L}^{(n)}, s) \geq \mathcal{A}(\mathcal{L}^{(n+1)}, \mathcal{L}^{(n)}, s) \geq \mathcal{F}(\mathcal{L}^{(n+1)}, s) \quad (11)$$

Furthermore, $\mathcal{F}(\mathcal{L}^{(n)}, s)$ is lower bounded because the Bhattacharyya measures are upper bounded by one. Therefore, $\mathcal{F}(\mathcal{L}^{(n)}, s)$ converges to a minimum of \mathcal{F} , and the solution is obtained by the optimal labeling at convergence.

Auxiliary function of the proposed energy: To introduce an auxiliary function of the proposed energy, let us first consider the following proposition:

Proposition 1: *Given a fixed labeling \mathcal{L}^n , for any labeling \mathcal{L} verifying $\mathbf{R}_1^{\mathcal{L}} \subset \mathbf{R}_1^{\mathcal{L}^n}$ and $\forall \alpha \in [0, 1]$, we have the following upper bound of the proposed energy:*

$$\begin{aligned} \mathcal{F}(\mathcal{L}, s) \leq \mathcal{A}(\mathcal{L}, \mathcal{L}^n, s, \alpha) &= \mathcal{A}_I(\mathcal{L}, \mathcal{L}^n, 0, \alpha) + \beta(\mathcal{A}_C(\mathcal{L}, \mathcal{L}^n, 0, \alpha) + \mathcal{A}_D(\mathcal{L}, \mathcal{L}^n, s, \alpha)) \\ &+ \lambda \mathbf{S}(\mathcal{L}) \end{aligned} \quad (12)$$

where $\mathcal{A}_J(\mathcal{L}, \mathcal{L}^n, s, \alpha)$ has the following general form for any function $J \in \{I : \mathcal{P} \rightarrow \mathcal{I}, C : \mathcal{P} \rightarrow \mathcal{C}, D : \mathcal{P} \rightarrow \mathcal{D}\}$:

$$\mathcal{A}_J(\mathcal{L}, \mathcal{L}^n, s, \alpha) = \sum_{p \in \mathbf{R}_0^{\mathcal{L}}} m_{p,J,s}(\mathbf{0}) + (1 - \alpha) \sum_{p \in \mathbf{R}_1^{\mathcal{L}}} m_{p,J,s}(\mathbf{1}), \quad (13)$$

with $m_{p,J,s}(\mathbf{0})$ and $m_{p,J,s}(\mathbf{1})$ given for each p in \mathcal{P} by:

$$\begin{cases} m_{p,J,s}(\mathbf{0}) = \frac{\mathcal{L}_p^n}{\mathbf{A}(\mathbf{R}_1^{\mathcal{L}^n})} \left(\mathbf{B}_J(\mathcal{L}^n, s) + \sum_{j \in \mathcal{J}} K(j - J_p) \sqrt{\frac{\mathcal{M}_J(j+s)}{\mathbf{P}_{\mathcal{L}^n, J}(j)}} \right) \\ m_{p,J,s}(\mathbf{1}) = \frac{\mathbf{B}_J(\mathcal{L}^n, s)}{\mathbf{A}(\mathbf{R}_1^{\mathcal{L}^n})} \quad \text{with} \quad \mathbf{B}_J(\mathcal{L}, s) = - \sum_{j \in \mathcal{J}} \sqrt{\mathbf{P}_{\mathcal{L}, J}(j) \mathcal{M}_J(j+s)} \end{cases} \quad (14)$$

To prove proposition 1, we apply the same principle steps of the proof we detailed recently in [1] to each of the Bhattacharyya constraints in cost function (8). We omit the details here due to space limit.

One can further verify that, for $\alpha = 0$, $\mathcal{A}(\mathcal{L}, \mathcal{L}, s, \alpha) = \mathcal{F}(\mathcal{L}, s)$, i.e., the bound in (12) is an auxiliary function of the proposed energy. This instructs us to consider the following procedure to optimize \mathcal{F} over the labeling:

- $n = 0$; Initialize the labeling $\mathcal{L}^0 = \mathcal{L}_0$; Initialize $\alpha: \alpha = \alpha_0$ with $0 < \alpha_0 < 1$
- Repeat the following two steps until convergence
 1. $\mathcal{L}^{(n+1)} = \arg \min_{\mathcal{L}: \mathbf{R}_1^{\mathcal{L}} \subset \mathbf{R}_1^{\mathcal{L}^n}} \mathcal{A}(\mathcal{L}, \mathcal{L}^n, s, \alpha)$
 2. Decrease $\alpha: \alpha = \alpha^\rho$ with $\rho > 1$

Graph Cuts: Now notice that the auxiliary function $\mathcal{A}(\mathcal{L}, \mathcal{L}^n, s, \alpha)$ in step 1 of the above procedure has the form of the sum of *unary* and *pairwise (sub-modular)* penalties, unlike the initial energy \mathcal{F} . In combinatorial optimization, a global optimum of such form can be computed efficiently in low-order polynomial time via a graph cut [3]. The graph-cut (or max-flow) algorithm of Boykov and Kolmogorov is well established in the computer vision literature. Therefore, we omit the details of this algorithm here, and refer the reader to [3].

Step 2–Fixed-point-equation (FPE) updates of the scale variable: With the labeling fixed, this step optimizes \mathcal{F} with respect to s . Following the variable change $d \leftarrow d - s$, the derivative of \mathcal{F} with respect to s reads:

$$\frac{\partial \mathcal{F}}{\partial s} = -\beta \frac{\partial \sum_{d \in \mathcal{D}} \sqrt{\mathbf{P}_{\mathcal{L}, D}(d) \mathcal{M}_D(d+s)}}{\partial s} = -\beta \sum_{d \in \mathcal{D}} \frac{\partial \mathbf{P}_{\mathcal{L}, D}(d-s)}{\partial s} \sqrt{\frac{\mathcal{M}_D(d)}{2\mathbf{P}_{\mathcal{L}, D}(d-s)}} \quad (15)$$

Using the KDE expression in (11), we also have:

$$\frac{\partial \mathbf{P}_{\mathcal{L}, D}(d-s)}{\partial s} = \frac{\sum_{p \in \mathbf{R}_1^{\mathcal{L}}} \frac{\partial K(d-s-D_p)}{\partial s}}{\mathbf{A}(\mathbf{R}_1^{\mathcal{L}})} = \frac{\sum_{p \in \mathbf{R}_1^{\mathcal{L}}} (d-s-D_p) K(d-s-D_p)}{\sigma^2 \mathbf{A}(\mathbf{R}_1^{\mathcal{L}})} \quad (16)$$

Embedding this derivative in (15), setting the obtained expression equal to zero, and after some algebraic manipulations, the necessary condition for a minimum of \mathcal{F} with respect to s can be expressed as the following fixed-point equation:

$$s - g(s) = 0 \quad \text{where} \quad g(s) = \frac{\sum_{d \in \mathcal{D}} \sum_{p \in \mathbf{R}_1^{\mathcal{L}}} (d - D_p) K(d - D_p - s) \sqrt{\frac{\mathcal{M}_D(d)}{\mathbf{P}_{\mathcal{L}, D}(d-s)}}}{\sum_{d \in \mathcal{D}} \sum_{p \in \mathbf{R}_1^{\mathcal{L}}} K(d - D_p - s) \sqrt{\frac{\mathcal{M}_D(d)}{\mathbf{P}_{\mathcal{L}, D}(d-s)}}} \quad (17)$$

Therefore, the solution of (17) can be obtained by the following *fixed-point-equation updates*:

$$s^{n+1} = g(s^n), \quad n = 1, 2, \dots \quad (18)$$

Let s^{opt} be the limit of sequence s^n at convergence. We have:

$$s^{opt} = \lim_{n \rightarrow +\infty} s^{n+1} = \lim_{n \rightarrow +\infty} g(s^n) = g(\lim_{n \rightarrow +\infty} s^n) = g(s^{opt}) \quad (19)$$

Consequently, s_{opt} is a solution of the necessary condition obtained in (17).

2.3 Experiments

The evaluation was carried out over 10 midsagittal T2-weighted MR lumbar spine images¹. We proceeded to a leave-one-out approach: for each testing case, we used a single different subject for training. A total of 60 lumbar discs were automatically delineated, and the results were compared to independent manual segmentations by an expert. Although we focus on lumbar spine discs here (6 discs per subject) as in [6], the formulation can be readily extended to the whole spine without additional user effort.

In the following, we first describe a typical example which illustrates explicitly the effect of the proposed object-interaction priors, the effect of the fixed-point equation we derived, and the robustness of the algorithm with respect to pose variations. Then, we describe a quantitative and comparative performance analysis using several accuracy measures. Finally, we give a representative sample of the results, and examine experimentally the invariance of the proposed geometric attributes, i.e., the angle and distance distributions, over ground-truth segmentations of 10 subjects and several disc pairs.

A Typical example: The continuous green curve in Fig. 2 (b) depicts the segmentation boundary obtained *without* the object-interaction priors. This solution, which corresponds to $\beta = 0$ in problem (8), included erroneously some neighboring structures in the final disc region (the ground truth is depicted with the discontinuous yellow curve). On the contrary, *with* the proposed priors, the algorithm yielded a solution very close to the ground truth (Fig. 2 a). The blue curve depicts an elliptical approximation of a previously segmented disc. Fig. 2 (c) depicts the disc pair in the sole training subject, and illustrates how the algorithm handles successfully the differences in pose (translation, rotation, and scale) between the training and testing pairs. The second line in Fig. 2 illustrates the effect of the fixed-point computations corresponding to the same example. (d) shows the fast convergence of such computations: the optimal scale variable s^{opt} is typically obtained within less than 30 iterations. The discontinuous line in Fig. 2 (e) depicts the distance model \mathcal{M}_D learned from the training pair in (c), whereas the continuous one depicts the distance distribution corresponding to the ground truth in the test image in (a). The shift between these two distributions is due to the difference in scale between the training and testing subjects.

¹ 2D T2-weighted MR spine images are commonly used in clinical practice thanks to their short acquisition time and ability to depict disc degeneration [10].

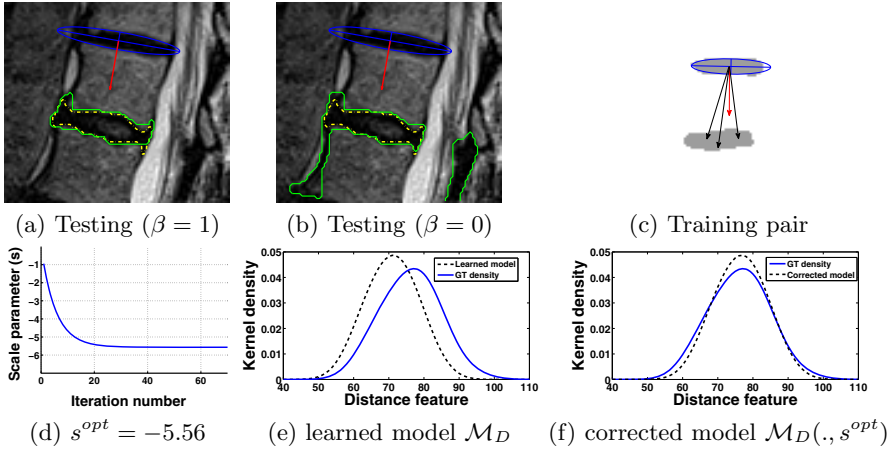


Fig. 2. A typical example. The green curves in (a) and (b) depict respectively the segmentation obtained with and without the object-interaction priors. The yellow discontinuous curve in (a)-(b) depicts the ground truth. (c) shows the disc pair of the training subject. (d) shows the fast convergence of the proposed fixed-point-equation computations: s^{opt} is obtained within less than 30 iterations (1.38 sec). (e) and (f) show respectively the learned model (\mathcal{M}_D) and the model corrected with s^{opt} ($\mathcal{M}_D(\cdot, s^{opt})$), both displayed with the ground-truth distance distribution (GT density). The number of graph cut iterations is 9 (3.3 sec). $\lambda = 2.5 \times 10^{-4}$. The distributions were estimated using 192 bins and a kernel width $\sigma = 15$. *The proposed formulation handles intrinsically translation and rotation variations, compute efficiently the optimal scale, and does not enforce a systematic bias towards the shape of the training discs.*

Fig. 2 (f) shows that $\mathcal{M}_D(\cdot, s^{opt})$, i.e., the model corrected with the optimal scale variable s^{opt} , befits much better the ground-truth distribution. The overall computation time is 4.68 sec (The graph cuts took 3.3 sec and the fixed-point iterations 1.38 sec). It is worth noting that the algorithm did not bias the solution towards the shape of the training discs in (c), unlike template-based priors.

Visual inspection: Fig. 3 depicts a representative sample of the results for the lumbar spines of three subjects. The green curves depict the segmentations obtained with the proposed algorithm, whereas the red curves correspond to the manual ground truth. The initial simple user input is depicted by the blue ellipses in the first line of the figure. The proposed formulation deals successfully with the variations in shape and pose of the discs, although neither a heavy training nor a costly pose optimization are required.

Quantitative performance evaluations: We assessed the similarities between the ground truth and the segmentations obtained with the proposed method over 10 subjects. We used three measures: the Root Mean Squared Error (*RMSE*), the Dice metric (*DM*), and the correlation coefficient (*r*). *DM* is commonly used to measure the similarity (overlap) between the automatically detected and ground-truth regions [2]: $DM = \frac{2\mathbf{A}_{am}}{\mathbf{A}_a + \mathbf{A}_m}$, with \mathbf{A}_a , \mathbf{A}_m , and \mathbf{A}_{am} corresponding

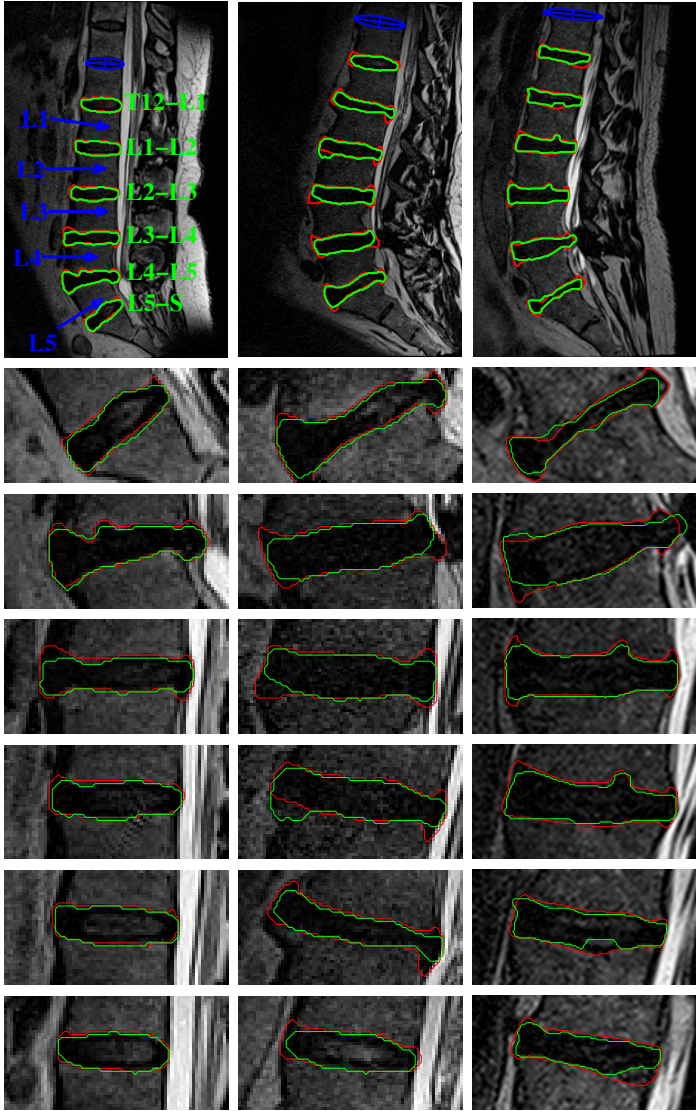


Fig. 3. A representative sample of the results for 3 subjects. Each column depicts the results for one subject: the first line shows the whole image, whereas the rest of the lines show the segmentation results corresponding individual discs from $L5 - S$ to $T12 - L1$ (refer to the standard annotation in the top left image). The green curves depict the segmentations obtained with the proposed algorithm. The red curves correspond to the manual ground truth. The initial simple user input is depicted by the blue ellipses in the first line of the figure. $\lambda = 2.5 \times 10^{-4}$. $\beta = 0.25$.

Table 1. Quantitative performance evaluations over 10 subjects. The parameters were unchanged for all the subjects: $\lambda = 2.5 \times 10^{-4}$ and $\beta = 0.25$.

<i>RMSE mean (in mm)</i>	<i>DM mean</i>	<i>Correlation coefficient (r)</i>
2.73	0.88	0.98

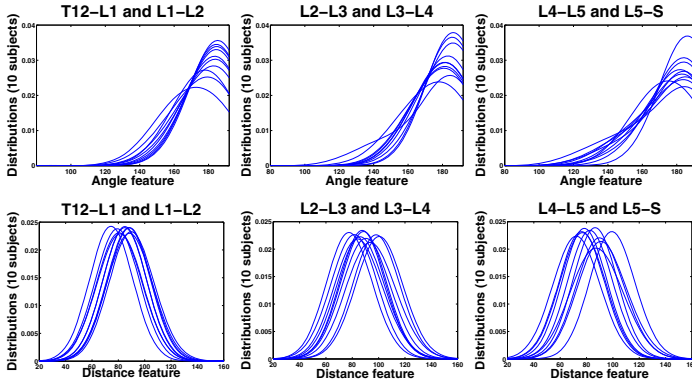


Fig. 4. Invariance of the angle and distance distributions over 10 subjects and 3 different disc pairs (refer to the annotation in the top left image in Fig. 3). The distributions were estimated using 192 bins and a kernel width $\sigma = 15$.

respectively to the areas of the segmented region, the hand-labeled region, and the intersection between them². *RMSE* evaluates the perpendicular distances from manual to automatic boundaries: $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \|u_i - v_i\|^2}$, with u_i a point on the automatically detected boundary and v_i the corresponding point on the manually traced boundary (we set $N = 240$)³. Table 1 reports the *DM* mean, the *RMSE* mean and the correlation coefficient between manual and automatic region areas.

Invariance of the geometric distributions: Using ground truth segmentations, we plotted in Fig. 4 the angle and distance distributions corresponding to 10 different subjects and 3 disc pairs. The figures demonstrate that the angle distributions are very similar (refer to first line in Fig. 4). The distance distributions have similar shapes, but different supports. These shifts, which are due to inter-subject variations in scale, can be handled efficiently with the proposed fixed-point-equation computations. This experiment supports the fact that a single subject is sufficient for training, and confirms the relevance of the proposed priors to disc segmentation.

² The higher the *DM*, the better the performance. *DM* is always in $[0, 1]$. $DM > 0.80$ indicates an excellent agreement between manual and automatic segmentations.

³ The lower *RMSE*, the better the conformity of the results to the ground truth.

References

1. Ben Ayed, I., Chen, H.M., Punithakumar, K., Ross, I., Li, S.: Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the bhattacharyya measure. In: CVPR, pp. 3288–3295 (2010)
2. Ben Ayed, I., Li, S., Ross, I.: Embedding overlap priors in variational left ventricle tracking. *IEEE Trans. on Medical Imaging* 28(12), 1902–1913 (2009)
3. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 26(9), 1124–1137 (2004)
4. Carballido-Gamio, J., Belongie, S.J., Majumdar, S.: Normalized cuts in 3-d for spinal mri segmentation. *IEEE Trans. on Medical Imaging* 23(1), 36–43 (2004)
5. Chevrefils, C., Cheriet, F., Aubin, C.E., Grimard, G.: Texture analysis for automatic segmentation of intervertebral disks of scoliotic spines from mr images. *IEEE Trans. on Information Technology in Biomedicine* 13, 608–620 (2009)
6. Corso, J.J., Alomari, R.S., Chaudhary, V.: Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 202–210. Springer, Heidelberg (2008)
7. Huang, S.H., Chu, Y.H., Lai, S.H., Novak, C.L.: Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE Trans. on Medical Imaging* 28(10), 1595–1605 (2009)
8. Klinder, T., Ostermann, J., Ehm, M., Franz, A., Kneser, R., Lorenz, C.: Automated model-based vertebra detection, identification, and segmentation in ct images. *Medical Image Analysis* 13(3), 471–482 (2009)
9. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS, pp. 556–562 (2000)
10. Michopoulou, S.K., Costaridou, L., Panagiotopoulos, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A.: Atlas-based segmentation of degenerated lumbar intervertebral discs from mr images of the spine. *IEEE Trans. on Biomedical Engineering* 56(9), 2225–2231 (2009)
11. Mukherjee, L., Singh, V., Dyer, C.R.: Half-integrality based algorithms for cosegmentation of images. In: CVPR. pp. 2028–2035 (2009)
12. Niemelainen, R., Videman, T., Dhillon, S., Battie, M.: Quantitative measurement of intervertebral disc signal using mri. *Clin. Radiol.* 63, 252–255 (2008)
13. Rousson, M., Paragios, N.: Prior knowledge, level set representations & visual grouping. *International Journal of Computer Vision* 76(3), 231–243 (2008)
14. Schmidt, S., Kappes, J.H., Bergholdt, M., Pekar, V., Dries, S.P.M., Bystrov, D., Schnörr, C.: Spine detection and labeling using a parts-based graphical model. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
15. Shi, R., Sun, D., Qiu, Z.L., Weiss, K.: An efficient method for segmentation of mri spine images. In: IEEE ICME, pp. 713–717 (2007)
16. Toshev, A., Taskar, B., Daniilidis, K.: Object detection via boundary structure segmentation. In: CVPR, pp. 950–957 (2010)
17. Vu, N., Manjunath, B.S.: Shape prior segmentation of multiple objects with graph cuts. In: CVPR (2008)

CoRPORATE: Cortical Reconstruction by Pruning Outliers with Reeb Analysis and Topology-Preserving Evolution

Yonggang Shi¹, Rongjie Lai², and Arthur W. Toga^{1,*}

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

² Dept. of Mathematics, University of Southern California, Los Angeles, CA, USA
yshi@loni.ucla.edu

Abstract. In this paper we propose a novel system for the accurate reconstruction of cortical surfaces from magnetic resonance images. At the core of our system is a novel framework for outlier detection and pruning by integrating intrinsic Reeb analysis of Laplace-Beltrami eigenfunctions with topology-preserving evolution for localized filtering of outliers, which avoids unnecessary smoothing and shrinkage of cortical regions with high curvature. In our experiments, we compare our method with FreeSurfer and illustrate that our results can better capture cortical geometry in deep sulcal regions. To demonstrate the robustness of our method, we apply it to over 1300 scans from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). We show that cross-sectional group differences and longitudinal changes can be detected successfully with our method.

1 Introduction

The reconstruction of the cortical surface from structural magnetic resonance (MR) images is a critical and challenging problem in neuro-imaging research and clinical studies [26,24]. Compared with the volumetric representation, a surface representation of the cortex provides a more natural platform for characterizing and mapping its macro anatomy, i.e., the convolution pattern. This can potentially help more precisely align the functional organization of the brain across population [3]. Such advantages, however, depend heavily on the accuracy of the surface in representing the sulcal and gyral folding pattern of the cortex. In this paper, we propose a new and fully automated method of cortical reconstruction based on Reeb analysis and localized evolution for outlier detection and pruning, which avoids unnecessary smoothing and shrinkage, and leads to more faithful representation of cortical anatomy, especially in the deep sulcal regions.

Cortical reconstruction is a well-studied problem in neuroimage analysis and many algorithms have been developed in the last two decades [9,2,16,4,6]. Different techniques used in important stages of the reconstruction process, such

* This work was supported by grants 5P41RR013642 and 5R01MH080892-03 from NIH and W81XWH-10-1-0882 from DoD.

as skull stripping and topological correction, have resulted in various levels of robustness and accuracy across systems. Even with these differences, one common idea adopted in the most popular systems in reconstructing the gray matter (GM) surface is to first extract the white matter (WM) surface and then evolve it toward the boundary of the GM and cerebrospinal fluid (CSF). To prevent leakage into non-cortical regions, smoothness regularization was incorporated *globally* on the surface to balance the evolution force derived from the image data. This regularization-based method, however, has two problems. First, it only helps but does not guarantee the occurrence of leakage can be prevented. Second, shrinkage can arise, especially in deep sulcal regions, since the regularization is enforced everywhere on the surface even if it is not necessary. On the other hand, localized outlier detection and removal with boundary deformation was proposed recently for the reconstruction of sub-cortical structures [20]. The outlier detection in this approach uses projection onto subspaces spanned by Laplace-Beltrami (LB) eigen-functions, thus computationally it is prohibitively expensive for the significantly more complicated cortical surface.

To overcome the limitations of global regularization in cortical reconstruction, we propose in this paper a new framework for outlier detection and pruning that integrates intrinsic Reeb analysis on the mesh representation of a cortical surface with topology-preserving evolution in its mask representation in the image domain. By analyzing the Reeb graph of LB eigen-functions of the mesh representation at saddle points, we develop an intrinsic approach for outlier detection. Compared with the outlier detection method in [20], the method we develop here is computationally tractable for large meshes since only a small number of eigen-functions are used. The location of the outliers are fed back to the mask representation in the image domain and a topology-preserving evolution algorithm is developed to remove the outliers while keeping other parts intact. Building upon this core technique, we develop a novel system for cortical reconstruction that generates accurate surface representations of cortical anatomy from MR images. We compare our method with state-of-the-art results from FreeSurfer [2] and show that the localized outlier removal process in our method leads to more accurate representation of cortical geometry in deep sulcal regions. We also demonstrate the robustness of our method by applying it to cross-sectional and longitudinal analyses of the large scale data from ADNI [11].

The rest of the paper is organized as follows. In section 2, we describe the pre-processing steps for cortical reconstruction and the topology-preserving evolution algorithm for generating the initial mask and mesh representations of the cortical surface that could have outliers. Using the mesh representation, the intrinsic Reeb analysis for outlier detection is proposed in section 3. A topology-preserving algorithm is then developed to remove the outliers in the mask representation, whose boundary is then used to generate a smooth mesh representation of the cortical geometry. Experimental results are presented in section 4, where we compare our method with FreeSurfer and apply it to large scale data set from ADNI. Finally, conclusions are made in section 5.

2 Surface Reconstruction with Topology-Preserving Evolution

In this section, we first describe the preprocessing steps that generate the tissue map and skeletal map of deep sulcal regions from a T1-weighted MR image. Using these maps, a topology-preserving evolution algorithm is then developed to extract the WM and GM boundary without smoothness regularization.

2.1 Preprocessing

To reconstruct cortical models from a T1-weighted MR image, we build a workflow of preprocessing steps based upon publicly available tools that represent successful experiences in previous research. The first step uses a meta algorithm [8] for skull stripping and automatically generates an image that only contains the brain as shown in Fig. 1(a). After that, inhomogeneity correction is

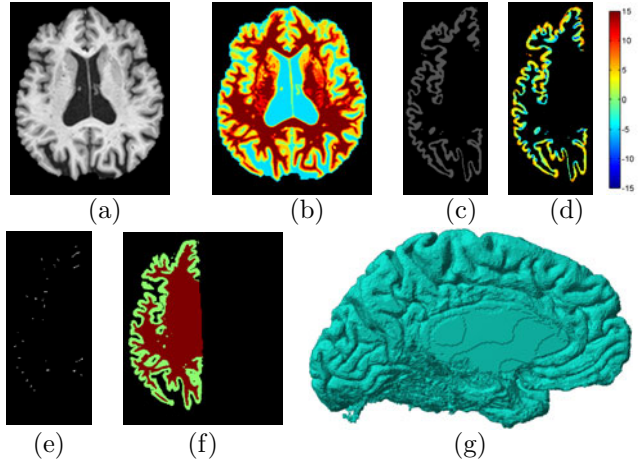


Fig. 1. An illustration of the preprocessing steps. (a) Skull-stripped image. (b) Tissue map. (c) GM band of the left hemisphere. (d) Flux map in the GM band. (e) Skeletal map in the GM band. (f) Evolution speeds. (g) GM surface with outliers.

applied [23]. The tissue classification algorithm in BrainSuite [16] is used to generate the tissue map T_{map} as shown in Fig. 1(b), which encodes partial volume effects by assigning each voxel a number in the range $[0, 3]$. Using the maximum likelihood principle, the WM region is defined as: $\mathcal{WM} = \{2.5 < T_{map} \leq 3\}$, the GM region is composed of two sub-regions: $\mathcal{GM}_{wm} = \{2 \leq T_{map} \leq 2.5\}$ including voxels with both gray and white matter and $\mathcal{GM}_{csf} = \{1.5 < T_{map} < 2\}$ including voxels with both gray matter and CSF, and the CSF region is defined as: $\mathcal{CSF} = \{T_{map} \leq 1.5\}$. To separate the cerebral cortex from the cerebellum, we register the skull-stripped image to the LPBA40 atlas [17] with a nonlinear registration algorithm [1] and remove the cerebellum based on labels in the atlas space. We also fill in the putamina, caudates, and ventricles with WM labels. For more accurate measurement of GM thickness in the parahippocampal regions, we remove the hippocampus and amygdala with the automated segmentation algorithm in FSL [13] and fill those regions with the CSF label. We divide the brain into left and right hemispheres and reconstruct cortical surfaces for each hemisphere. The WM is separated using labels in the LPBA40 atlas. The GM

and CSF are divided according to their distances to the WM of the left and right hemispheres. After that, our method processes each hemisphere separately.

To accurately model the GM surface buried in deep sulcal regions, we compute the Hamilton-Jacobi skeleton [22] in the GM band of each hemisphere as shown in Fig. 1(c). For voxels in \mathcal{GM}_{wm} , we use the distance transform of the \mathcal{WM} region to compute the flux. For voxels in \mathcal{GM}_{csf} , the flux is computed with the distance transform of the \mathcal{GM}_{wm} region. The flux map in the GM band is shown in Fig. 1(d). Starting from voxels on the GM/WM boundary, a homotopy-preserving thinning algorithm is applied to the GM band [22]. After removing voxels on the GM/CSF boundary with high flux values, we obtain the skeletal map that represents the GM boundary in buried sulcal regions, which we denote as \mathcal{S}_{gm} and plot as the bright regions in Fig. 1(e).

2.2 Topology-Preserving Evolution

Based on the tissue map and skeletal map of the GM band, we design topology-preserving evolution algorithms to reconstruct the WM and GM surface with genus-zero topology. Our evolution algorithm is derived from the fast evolution algorithm in [18] that realizes surface evolution by simply updating two lists of boundary voxels. We augment this algorithm with topological and well-composedness constraints such that the boundary surface will be a genus-zero manifold [7]. This not only ensures the correct topology, but also allows spectral analysis to be performed on this surface.

Given a volumetric mask composed of an interior region A_o and background region A_c , we define two sets of boundary voxels:

$$\begin{aligned} L_{in} &= \{\mathbf{p} | \mathbf{p} \in A_o \text{ s.t. } \exists \mathbf{q} \in N_6(\mathbf{p}) \cap A_c\}, \\ L_{out} &= \{\mathbf{p} | \mathbf{p} \in A_c \text{ s.t. } \exists \mathbf{q} \in N_6(\mathbf{p}) \cap A_o\}. \end{aligned} \quad (1)$$

where $N_6(\cdot)$ denotes six-connected neighbors of a voxel, and the six-connectedness is necessary for the well-composedness of the object and background region [7]. We also define an indication function for fast access of the regional information about the voxels:

$$\phi(\mathbf{p}) = \begin{cases} 3 & \text{if } \mathbf{p} \in A_c \setminus L_{out}; \\ 1 & \text{if } \mathbf{p} \in L_{out}; \\ -1 & \text{if } \mathbf{p} \in L_{in}; \\ -3 & \text{if } \mathbf{p} \in A_o \setminus L_{in}. \end{cases} \quad (2)$$

Starting from an initial mask, we run the fast evolution algorithm in Table 1, which we adapt from [18], in two stages to find the WM and GM boundary of each hemisphere. At the first stage, we extract the WM boundary using the following evolution speed:

$$F_{wm} = \begin{cases} 1 & \text{if } \mathbf{p} \in \mathcal{WM}; \\ -1 & \text{otherwise.} \end{cases} \quad (3)$$

In the second stage, we continue the evolution process according to the following speed to extract the GM boundary:

$$F_{gm} = \begin{cases} 1 & \text{if } \mathbf{p} \in (\mathcal{GM}_{wm} \cup \mathcal{GM}_{csf} \cup \mathcal{WM}) \setminus \mathcal{S}_{gm}; \\ -1 & \text{if } \mathbf{p} \in \mathcal{S}_{gm}; \\ -1 & \text{otherwise.} \end{cases} \quad (4)$$

As an illustration, the evolution speeds for the left hemisphere of an image are illustrated in Fig. 1(f), where the red region represents $F_{wm} = 1$ and $F_{gm} = 1$, and the green region represents $F_{wm} = -1$ and $F_{gm} = 1$. In each iteration, this algorithm checks whether the addition or removal of a voxel \mathbf{p} from the two lists will change the well-composedness of the mask. This can be achieved by testing whether two singular configurations occur in the 26-connected neighborhood of each voxel [7]. To ensure there is no topological change during the deformation process, only simple points in the two lists are updated [5].

At the end of each stage of the evolution algorithm, the indication function ϕ provides a mask representation for the WM and GM surface. Because well-composed and genus-zero topology is ensured, we can find the mesh representation of the WM and GM surface by extracting the faces intersected by cuboids of the object and background region in the mask representation [20]. By dividing each cuboid face into two triangles, we obtain a triangular mesh representation $(\mathcal{V}, \mathcal{T})$ for each surface, where \mathcal{V} is the set of vertices and \mathcal{T} is the set of triangles. Because each triangle is on the intersection of two cuboids in the object and background region, we can define two maps $\mathcal{O}_N : \mathcal{T} \rightarrow A_o$ and $\mathcal{C}_N : \mathcal{T} \rightarrow A_c$ that map each triangle in the mesh representation back to corresponding cuboids in the object and background region of the mask representation in the image domain. This allows us to move freely between mesh-based feature analysis and topology-preserving evolution in the image domain. As an illustration, the GM surface extracted from the image in Fig. 1(a) is shown in Fig. 1(g). We can see clearly there are outliers on the surface due to artifacts in the preprocessing steps such as skull stripping and tissue classification. To remove these outliers while keeping other smooth parts intact, we will design intrinsic feature analysis and localized evolution algorithms in the next section.

Table 1. Fast Algorithm for Topology-preserving Evolution

- | |
|---|
| <ul style="list-style-type: none"> – Step 1: Initialize arrays ϕ and the lists L_{out} and L_{in}. – Step 2: Compute the evolution speed F. – Step 3: Boundary deformation by updating L_{in}, L_{out} and ϕ. <ul style="list-style-type: none"> • Outward evolution. For each Well-Composed and simple point $\mathbf{x} \in L_{out}$ with $F(\mathbf{x}) > 0$, switch \mathbf{x} to L_{in} and set $\phi(\mathbf{x}) = -1$. $\forall \mathbf{y} \in N_6(\mathbf{x})$ with $\phi(\mathbf{y}) = 3$, add \mathbf{y} to L_{out} and set $\phi(\mathbf{y}) = 1$. • Eliminate redundant points in L_{in}. For each point $\mathbf{x} \in L_{in}$, if $\forall \mathbf{y} \in N_6(\mathbf{x}), \phi(\mathbf{y}) < 0$, delete \mathbf{x} from L_{in}, and set $\phi(\mathbf{x}) = -3$. • Inward evolution. For each Well-Composed and simple point $\mathbf{x} \in L_{in}$ with $F(\mathbf{x}) < 0$, switch \mathbf{x} to L_{out} and set $\phi(\mathbf{x}) = 1$. $\forall \mathbf{y} \in N_6(\mathbf{x})$ with $\phi(\mathbf{y}) = -3$, add \mathbf{y} to L_{in} and set $\hat{\phi}(\mathbf{y}) = -1$. • Eliminate redundant points in L_{out}. For each point $\mathbf{x} \in L_{out}$, if $\forall \mathbf{y} \in N(\mathbf{x}), \phi(\mathbf{y}) > 0$, delete \mathbf{x} from L_{out} and set $\phi(\mathbf{x}) = 3$. – Step 4: Stop if no element is switched; otherwise, go back to Step 2. |
|---|

3 Localized Outlier Pruning with Reeb Analysis

Starting from an initial surface with outliers, which can be either a WM or GM surface, we develop a localized filtering algorithm in this section to prune these outliers and generate a smooth surface with the correct topology. We first describe the background of Laplace-Beltrami eigen-functions on manifolds, which have received increasing interests in medical image analysis [15,14,12,19]. An automated outlier detection algorithm will then be presented based on the Reeb graph of eigen-functions. After that, evolution speeds will be designed to remove outliers without affecting other regions. Finally, an energy minimization problem is solved to generate the smooth surface representation of the cortex.

3.1 Laplace-Beltrami Eigen-Functions

Given a manifold \mathcal{M} , the eigen-functions of its Laplace-Beltrami (LB) operator $\Delta_{\mathcal{M}}$ is defined as

$$\Delta_{\mathcal{M}}f = -\lambda f, \tag{5}$$

where λ is the eigenvalue and f is the corresponding eigen-function. The spectrum of $\Delta_{\mathcal{M}}$ is discrete and we can order the eigenvalues according to their magnitude as $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots$. For λ_i , we denote the corresponding eigen-function as f_i . For a triangulated surface $\mathcal{M} = (\mathcal{V}, \mathcal{T})$, such as the WM and GM surface, we can compute its LB spectrum by solving a generalized matrix eigenvalue problem:

$$Qf = \lambda Uf \tag{6}$$

where the two matrices Q and U are formed using the finite element method, and f is a vector representing the value of the eigen-function on the vertex set \mathcal{V} . More specifically, the matrices are defined as:

$$Q_{ij} = \begin{cases} \frac{1}{2} \sum_{\mathcal{V}_j \in N(\mathcal{V}_i)} \sum_{\mathcal{T}_l \in N(\mathcal{V}_i, \mathcal{V}_j)} \cot \theta_l^{i,j}, & \text{if } i = j; \\ -\frac{1}{2} \sum_{\mathcal{T}_l \in N(\mathcal{V}_i, \mathcal{V}_j)} \cot \theta_l^{i,j}, & \text{if } \mathcal{V}_j \in N(\mathcal{V}_i); \\ 0, & \text{otherwise.} \end{cases}$$

$$U_{ij} = \begin{cases} \frac{1}{12} \sum_{\mathcal{V}_j \in N(\mathcal{V}_i)} \sum_{\mathcal{T}_l \in N(\mathcal{V}_i, \mathcal{V}_j)} A_l, & \text{if } i = j; \\ \frac{1}{12} \sum_{\mathcal{T}_l \in N(\mathcal{V}_i, \mathcal{V}_j)} A_l, & \text{if } \mathcal{V}_j \in N(\mathcal{V}_i); \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where $N(\mathcal{V}_i)$ is the set of vertices in the 1-ring neighborhood of the i -th vertex \mathcal{V}_i , $N(\mathcal{V}_i, \mathcal{V}_k)$ is the set of triangles sharing the edge $(\mathcal{V}_i, \mathcal{V}_k)$, $\theta_l^{i,j}$ is the angle in the triangle \mathcal{T}_l opposite to the edge $(\mathcal{V}_i, \mathcal{V}_j)$, and A_l is the area of the l -th triangle \mathcal{T}_l . Once the matrices are constructed, we can use a sparse solver such as MATLAB to compute the eigen-functions.

3.2 Outlier Detection with Reeb Analysis

The Reeb graph of a function f on a manifold \mathcal{M} is defined as follows.

Definition 1. Let $f : \mathcal{M} \rightarrow \mathbb{R}$. The Reeb graph $R(f)$ of f is the quotient space with its topology defined through the equivalent relation $x \simeq y$ if $f(x) = f(y)$ for $\forall x, y \in \mathcal{M}$.

For general surfaces, we can assume its LB eigen-functions are Morse functions [25], which means the critical points of f are non-degenerate and the Reeb graph $R(f)$ encodes the topology of \mathcal{M} . This means that Reeb graphs of eigen-functions on a genus-zero surface have tree structures. For smooth cortical surfaces, we have developed intrinsic characterizations using the Reeb graphs of LB eigen-functions, which typically have a chain structure [21]. For a surface with outliers, the Reeb graphs will have branches. We assume the outliers are smaller than the main body of the cortical surface, so the Reeb graphs of the noisy surface can be considered as a main axis with small branches. Because Reeb graphs split at saddle points, our idea is to detect outliers by locating and analyzing the branches of Reeb graphs at the saddle points of eigen-functions.

Let $N_-(\mathcal{V}_i) = \{\mathcal{V}_j \in N(\mathcal{V}_i) | f(\mathcal{V}_j) < f(\mathcal{V}_i)\}$ and $N_+(\mathcal{V}_i) = \{\mathcal{V}_j \in N(\mathcal{V}_i) | f(\mathcal{V}_j) > f(\mathcal{V}_i)\}$ denote the lower and upper neighbors in the 1-ring neighborhood of a vertex \mathcal{V}_i . Let $\#$ denote the number of connected components in a set. If $\#N_-(\mathcal{V}_i) + \#N_+(\mathcal{V}_i) \geq 4$, we denote \mathcal{V}_i as a saddle point. We sample level contours of the function f on the surface at the values:

$$f_-(\mathcal{V}_i) = \frac{f(\mathcal{V}_i) + \max_{\mathcal{V}_j \in N_-(\mathcal{V}_i)} f(\mathcal{V}_j)}{2}$$

$$f_+(\mathcal{V}_i) = \frac{f(\mathcal{V}_i) + \min_{\mathcal{V}_j \in N_+(\mathcal{V}_i)} f(\mathcal{V}_j)}{2}.$$

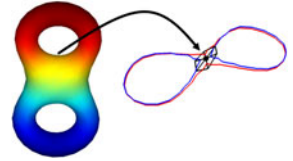


Fig. 2. The level contours in the neighborhood of a saddle point

Let C_- and C_+ denote the level contours of f on \mathcal{M} at the value $f_-(\mathcal{V}_i)$ and $f_+(\mathcal{V}_i)$ that intersect the 1-ring neighborhood of \mathcal{V}_i , respectively. As an illustration, we plot the level contours at a saddle point of the first LB eigen-function of a double tori in Fig. 2. We can see that the level sets of f change topology in the neighborhood of the saddle point, thus $\#C_- > \#C_+$ or $\#C_- < \#C_+$. Without loss of generality, we consider the case $\#C_+ > \#C_-$. The increase of components in C_+ means the addition of a branch in the Reeb graph. Since we assume outliers are much smaller than the main body of the cortical surface, we choose the shortest level contour in C_+ as the possible branch created by outliers and denote it as C_+^* . For a genus-zero surface, this contour C_+^* partitions the surface into two components and we denote the component where all vertices have a function value higher than $f_+(\mathcal{V}_i)$ as $R(C_+^*)$. We consider this component as an outlier if it satisfies two conditions:

$$\int_{C_+^*} ds < \alpha \quad \text{and} \quad \int_{R(C_+^*)} dA / \int_{C_+^*} ds > \beta \quad (8)$$

where the parameters α and β are thresholds selected to identify sharp and small outliers. For the case $\#C_- > \#C_+$, we can detect an outlier similarly. As an example, we plot in Fig. 3(a) and (b) the outliers detected via the Reeb analysis

of eigen-functions on the surface in Fig. 1(g). The outliers detected by the first eigen-function f_1 are plotted in yellow and the additional outliers detected with f_2, f_3, \dots, f_{30} are plotted in red. This shows more outliers can be detected with higher order eigen-functions.

3.3 Localized Filtering for Outlier Pruning

Let R denote the set of triangles in the outliers detected via Reeb analysis on the mesh representation of a WM or GM surface. By using the two maps $\mathcal{O}_N : \mathcal{T} \rightarrow A_o$ and $\mathcal{C}_N : \mathcal{T} \rightarrow A_c$, we can map the outliers to the mask domain and define:

$$\begin{aligned} L_{in}^* &= \hat{L}_{in} \setminus \cup_{\mathcal{T}_i \in R} \mathcal{O}_N(\mathcal{T}_i) \\ L_{out}^* &= \hat{L}_{out} \setminus \cup_{\mathcal{T}_i \in R} \mathcal{C}_N(\mathcal{T}_i) \end{aligned} \quad (9)$$

where \hat{L}_{in} and \hat{L}_{out} are the two voxel lists on the boundary of the mask representation of the WM or GM surface. To filter out the outliers locally, we define the following evolution speed to iteratively deform the mask representation of the surface and prune the outliers:

$$F(\mathbf{p}) = \begin{cases} -1 & \text{if } \mathbf{p} \in L_{in} \setminus L_{in}^*, \text{ and } \overline{H}(\mathbf{p}) \geq 1/2; \\ 1 & \text{if } \mathbf{p} \in L_{out} \setminus L_{out}^* \text{ and } \overline{H}(\mathbf{p}) < 1/2; \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

where the function $\overline{H} = H(\phi) \otimes G(0; \sigma)$, which is the convolution of the Heaviside function and a Gaussian kernel $G(0; \sigma)$ with zero mean and standard deviation σ [18]. The evolution process converges when no element switch happens.

After the outlier pruning, we extract the triangular mesh representation of the mask boundary and solve an energy minimization problem to generate the final surface

$$E = \|x - \hat{x}\|^2 + \gamma \|\Delta \hat{x}\|^2 \quad (11)$$

where x is the vector of coordinates of all vertices, Δ is the discrete Laplacian matrix of the mesh, and γ is a regularization parameter. Note that we have chosen the Laplacian instead of the gradient operator in the regularization term

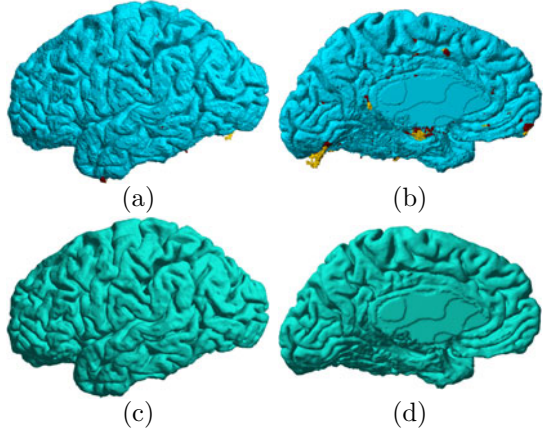


Fig. 3. (a)(b) Outlier detection results. Yellow: outliers detected with f_1 . Red: additional outliers detected with f_2, \dots, f_{30} . (c)(d) The smooth surface with outliers pruned.

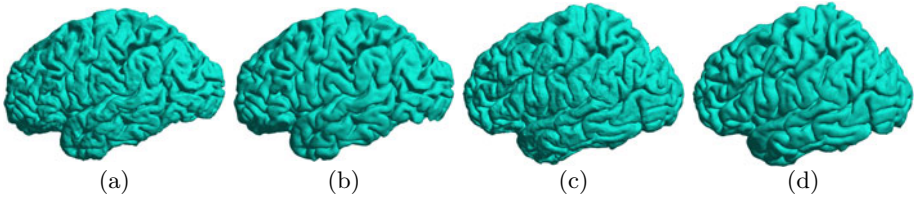


Fig. 4. (a)(c) Left hemispherical GM surfaces extracted by our method. (b)(d) Results from FreeSurfer.

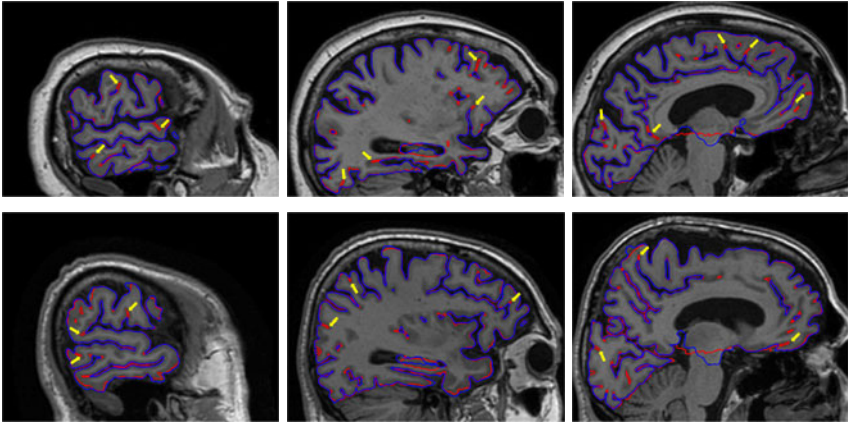


Fig. 5. The intersection of cortical surfaces with three sagittal slices. Red Contour: our method. Blue contour: FreeSurfer. Top row: data set one. Bottom row: data set two. Yellow arrows highlight deep sulcal regions better captured by our method.

to avoid the shrinkage effect from the gradient operator. The solution of this quadratic problem gives us the coordinates of vertices on the smoothed surface:

$$\hat{x} = (I + \gamma \Delta' \Delta)^{-1} x. \quad (12)$$

As an illustration, the smoothed GM surface after outlier pruning is shown in Fig. 3(c) and (d).

4 Experiments

In this section, we present experimental results to demonstrate the proposed cortical reconstruction algorithm. For all experiments, we use the first 30 eigenfunctions for Reeb analysis and the same set of parameters are used, which are $\alpha = 50$ and $\beta = 2$ for outlier detection, $\sigma = 3$ for the Gaussian kernel in outlier removal, and $\gamma = 10$ for mesh smoothing. The computational time of our method to reconstruct the WM and GM surfaces from an MRI image is approximately three hours on a single CPU. In the first experiment, we perform comparisons with FreeSurfer to demonstrate that the localized outlier pruning process can provide more accurate representation of cortical geometry in deep sulcal regions.

In the second experiment, we apply our method to the large scale data set from ADNI to demonstrate its robustness.

4.1 Comparison with FreeSurfer

To demonstrate that the localized outlier pruning in our method helps capture GM boundary in deep sulcal regions, we compare our method with FreeSurfer [2]. We applied both methods to two MR images and the extracted left hemispherical GM surfaces are shown in Fig. 4. We can see the surfaces from both methods represent almost identical convolution patterns. For a more clear comparison, we have plotted the intersection of the surfaces with sagittal slices in Fig. 5. Overall we can see the red and blue contours overlap in most places. In deeper sulcal regions, however, we can clearly see that the red contours from our method do a better job in reaching deeper parts of the sulcal region as highlighted by the yellow arrows. This shows the localized pruning process in our method leads to more accurate representation of the GM boundary in deep sulcal regions and holds the potential of generating more accurate representations of cortical anatomy.

4.2 Large Scale Validation on ADNI Data

We validate the robustness of our method by applying it to the large scale data from ADNI [11]. We used both the baseline scans of 221 normal controls (NC), 396 patients with mild cognitive impairment (MCI), 193 patients with Alzheimer’s disease (AD) and the follow-up scans at Month 24 of 166 NC, 249 patients with MCI, 109 patients with AD. We first extracted the GM surface of the left hemisphere from the ICBM single subject atlas [10] and project the labels in the MRI volume to the GM surface as shown in Fig. 6 (a) and (b). We extracted the WM and GM surfaces of the left hemisphere from all baseline and Month 24 data of ADNI. All GM surfaces are mapped to the atlas surface using the intrinsic embedding we developed in [21] to generate the gyral parcellation. For each point on a GM surface, its GM thickness is computed as the distance to the corresponding WM surface.

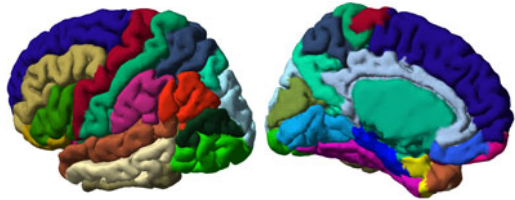


Fig. 6. Gyral labels on the atlas surface

Using the average GM thickness of each gyrus as the predictor variable, we performed statistical analysis to test for group differences. The first experiment compared the group differences of GM thickness at baseline. The second experiment investigated the longitudinal changes of GM thickness in each gyrus. We list the p-values of the 10 gyri with the most significant changes in Table 2 and 3. From the cross-sectional results we can see that our method successfully detects significant group differences in gyri considered affected by AD. From the longitudinal results we can see that our method shows significant changes in

Table 2. Cross-sectional results at baseline

	Para-hippo	Ento-rhinal	Sup. Temp.	Mid. Temp.	Inf. Temp.	Fusi-form	Inf. Pariet.	Inf. Occip.	Sup. Front.	Mid. Front.
NC vs AD	1.2e-19	2.3e-19	3.5e-21	1.5e-14	8.2e-11	9.5e-11	1.7e-7	1.8e-7	7.5e-7	2.1e-6
NC vs MCI	2.2e-6	2.7e-7	4.4e-7	1.6e-5	2.4e-2	4.5e-6	1.3e-2	3.3e-1	1.5e-4	5.4e-3
MCI vs AD	4.5e-9	1.8e-7	1.0e-8	1.2e-5	1.4e-6	1.5e-3	8.2e-4	4.3e-5	5.6e-2	2.3e-3

Table 3. Longitudinal results

	Para-hippo	Ento-rhinal	Sup. Temp.	Mid. Temp.	Inf. Temp.	Fusi-form	Inf. Pariet.	Inf. Occip.	Sup. Front.	Mid. Front.
NC	5.1e-1	1.6e-1	3.0e-5	5.6e-1	2.3e-1	9.1e-1	2.4e-1	2.4e-1	4.7e-3	4.2e-2
MCI	4.2e-8	1.6e-6	1.9e-9	3.2e-4	9.6e-3	2.3e-2	8.1e-3	3.3e-1	4.9e-3	1.7e-3
AD	4.4e-2	1.8e-1	2.0e-10	8.9e-5	3.8e-3	6.7e-3	3.7e-2	5.0e-1	1.2e-3	9.8e-3

those gyri for the patient groups. These results demonstrate the robustness of our method in large scale studies.

5 Conclusions

In this paper we developed a novel system for cortical reconstruction based on the idea of localized outlier pruning. Preliminary results demonstrate that our method can obtain more accurate representation of cortical geometry in deep sulcal regions than state-of-the-art results. The robustness of our method is validated via applications to large scale data from the ADNI. For future work, we will optimize the computational efficiency of the proposed method and perform more through comparisons with existing tools for cortical reconstruction.

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. *Med. Image. Anal.* 12(1), 26–42 (2008)
2. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis i: segmentation and surface reconstruction. *NeuroImage* 9, 179–194 (1999)
3. Dubois, J., Benders, M., Borradori-Tolsa, C., et al.: Primary cortical folding in the human newborn: an early marker of later functional development. *Brain* 131(8), 2028–2041 (2008)
4. Han, X., Pham, D.L., Tosun, D., et al.: CRUISE: Cortical reconstruction using implicit surface evolution. *NeuroImage* 23, 997–1012 (2004)
5. Han, X., Xu, C., Prince, J.: A topology preserving level set method for geometric deformable models. *IEEE Trans. Pattern Anal. Machine Intell.* 25(6), 755–768 (2003)
6. Kim, J.S., Singh, V., Lee, J.K., et al.: Automated 3-d extraction and evaluation of the inner and outer cortical surfaces using a laplacian map and partial volume effect classification. *NeuroImage* 27, 210–221 (2005)
7. Latecki, L.J.: 3D well-composed pictures. *Graph. Models Image Process.* 59(3), 164–172 (1997)

8. Leung, K., Parker, D., Cunha, A., et al.: IRMA: An image registration meta-algorithm evaluating alternative algorithms with multiple metrics. In: Proc. Int. Conf. on Scientific and Statistical Data Base Management, pp. 612–617 (2008)
9. Mangin, J.F., Frouin, V., Bloch, I., et al.: From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision* 5(4), 297–318 (1995)
10. Mazziotta, J.C., Toga, A.W., Evans, A.C., et al.: A probabilistic atlas and reference system for the human brain: international consortium for brain mapping. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 356, 1293–1322 (2001)
11. Mueller, S., Weiner, M., Thal, L., et al.: The Alzheimer’s disease neuroimaging initiative. *Clin. North Am.* 15, 869–877 (2005)
12. Niethammer, M., Reuter, M., Wolter, F.-E., Bouix, S., Peinecke, N., Koo, M.-S., Shenton, M.E.: Global medical shape analysis using the laplace-beltrami spectrum. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 850–857. Springer, Heidelberg (2007)
13. Patenaude, B.: Bayesian Statistical Models of Shape and Appearance for Subcortical Brain Segmentation. Ph.D. thesis, University of Oxford (2007)
14. Qiu, A., Bitouk, D., Miller, M.I.: Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace-Beltrami operator. *IEEE Trans. Med. Imag.* 25(10), 1296–1306 (2006)
15. Reuter, M.: Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions. *Int’l Journal of Computer Vision* 89(2-3), 287–308 (2010)
16. Shattuck, D., Leahy, R.: BrainSuite: An automated cortical surface identification tool. *Med. Image. Anal.* 8(2), 129–142 (2002)
17. Shattuck, D., Mirza, M., Adisetiyo, V., et al.: Construction of a 3d probabilistic atlas of human brain structures. *NeuroImage* 39(3), 1064–1080 (2008)
18. Shi, Y., Karl, W.C.: A real-time algorithm for the approximation of level-set-based curve evolution. *IEEE Trans. Image Processing* 17(5), 645–657 (2008)
19. Shi, Y., Lai, R., Krishna, S., et al.: Anisotropic Laplace-Beltrami eigenmaps: Bridging Reeb graphs and skeletons. In: Proc. MMBIA, pp. 1–7 (2008)
20. Shi, Y., Lai, R., Morra, J., et al.: Robust surface reconstruction via laplace-beltrami eigen-projection and boundary deformation. *IEEE Trans. Med. Imag.* 29(12), 2009–2022 (2010)
21. Shi, Y., Sun, B., Lai, R., Dinov, I., Toga, A.W.: Automated sulci identification via intrinsic modeling of cortical anatomy. In: Jiang, T., Navab, N., Plum, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 49–56. Springer, Heidelberg (2010)
22. Siddiqi, K., Bouix, S., Tannebaum, A., Zuker, S.: Hamilton-Jacobi skeletons. *Int’l Journal of Computer Vision* 48(3), 215–231 (2002)
23. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imag.* 17(1), 87–97 (1998)
24. Thompson, P.M., Hayashi, K.M., Sowell, E.R., et al.: Mapping cortical change in alzheimer’s disease, brain development, and schizophrenia. *NeuroImage* 23, S2–S18 (2004)
25. Uhlenbeck, K.: Generic properties of eigenfunctions. *Amer. J. of Math.* 98(4), 1059–1078 (1976)
26. Van Essen, D.C., Drury, H.A., Joshi, S., Miller, M.I.: Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proc. Natl. Acad. Sci. USA* 95, 788–795 (1998)

Globally Optimal Tumor Segmentation in PET-CT Images: A Graph-Based Co-segmentation Method

Dongfeng Han¹, John Bayouth¹, Qi Song², Aakant Taurani²,
Milan Sonka¹, John Buatti¹, and Xiaodong Wu^{1,2}

¹ Department of Radiation Oncology

² Department of Electrical and Computer Engineering
The University of Iowa, Iowa City, IA, USA

handongfeng@gmail.com,

{john-bayouth,qi-song,aakant-aurani,milan-sonka,
john-buatti,xiaodong-wu}@uiowa.edu

Abstract. Tumor segmentation in PET and CT images is notoriously challenging due to the low spatial resolution in PET and low contrast in CT images. In this paper, we have proposed a general framework to use both PET and CT images simultaneously for tumor segmentation. Our method utilizes the strength of each imaging modality: the superior contrast of PET and the superior spatial resolution of CT. We formulate this problem as a Markov Random Field (MRF) based segmentation of the image pair with a regularized term that penalizes the segmentation difference between PET and CT. Our method simulates the clinical practice of delineating tumor simultaneously using both PET and CT, and is able to concurrently segment tumor from both modalities, achieving globally optimal solutions in low-order polynomial time by a single maximum flow computation. The method was evaluated on clinically relevant tumor segmentation problems. The results showed that our method can effectively make use of both PET and CT image information, yielding segmentation accuracy of 0.85 in Dice similarity coefficient and the average median hausdorff distance (HD) of 6.4 mm, which is 10 % (resp., 16 %) improvement compared to the graph cuts method solely using the PET (resp., CT) images.

1 Introduction

Image-guided radiation therapy utilizes on-board daily image-based target and critical structure position verification that enables precise delivery of a high radiation dose to an image-defined target volume while maintaining a low dose to the surrounding critical structures. Accurate target definition is essential to obtain the full benefit of using image-guided radiotherapy for head-and-neck cancer (HNC) or lung cancer. Standard treatment paradigms utilize manual target

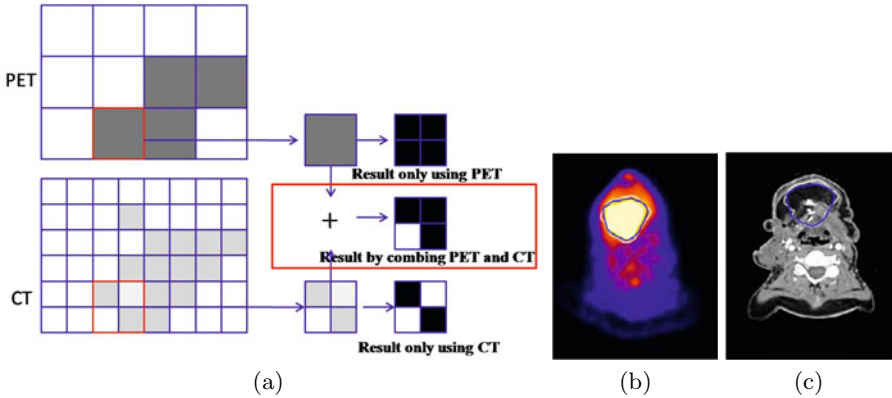


Fig. 1. The PET-CT co-segmentation. (a) The idea of the proposed method. For two co-registered PET and CT images, the superior contrast of PET and the superior spatial resolution of CT can be helpful to obtain a more accurate segmentation. Simultaneously using these information is superior than only using one modality image. (b) and (c) are the registered PET and CT images. The white contour is the ground truth and the blue contour is our co-segmentation result. The tumor is difficult to identify in a CT image. Using both PET and CT images simultaneously, the tumor can be segmented with a high accuracy.

volume definitions on computed tomography (CT) scans. This is potentially inaccurate and is subject to both high intra- and inter-observer variability because the tumors are similar in density to the surrounding soft tissues. It is difficult to obtain a consistent tumor definition in an objective way by manual contouring.

To improve visualization of the tumor volume, functional imaging using F-fluorodeoxyglucose (FDG) positron emission tomography (PET) has been combined with treatment planning CT to assist radiation oncologists in target delineation. PET-CT has become a standard method for tumor staging and has increasingly been used by radiation oncologists in tumor target delineation. During an FDG-PET scan, a tracer dose of the radioactive material, FDG, is injected intravenously. The FDG uptake is increased in cells with high metabolic rate, hence diseased areas (such as tumor, inflammation) in FDG-PET appear as high-uptake hot spots.

Although the high contrast between tumor and normal tissue on PET images can reduce the inter- and intra-observed variability in tumor localization, observed variability in tumor delineation with the qualitative use of FDG-PET is still high and often inconsistent with CT-based anatomically defined tumor contours. Accurate segmentation of an FDG-PET image for tumor delineation is a complex and unresolved problem. Attempts to make definitions more consistent through objective/automated segmentation methods (i.e., standardized uptake value threshold) have been fraught with difficulty. Factors such as the image reconstruction algorithm, the partial volume effect for objects smaller than twice the full width at half maximum, and the display windowing, image

registration between anatomically based multi-modality (CT/MR) datasets, and the threshold approach utilized, all influence image segmentation. Objective and robust tools are needed to utilize the novel information that molecular imaging techniques provide. This would be tremendously valuable in radiation therapy to capitalize on advances in the precision of radiation planning and delivery, as we have reached an era where our precision in delivering a radiation dose distribution that conforms three-dimensionally to the shape of our target is far greater than the precision with which the target can be defined. Currently, dual-modality PET-CT imaging is widely used in clinical therapy and increasingly directly in the treatment planning process albeit in a largely subjective and inconsistent fashion. Currently available automated tumor delineation methods have relied solely on a threshold of an absolute PET intensity value [3]. The CT image data is ignored and often contradicts the threshold based target definition. The complexity of human anatomy and neoplastic growth produce sophisticated features in images, which cannot fully be captured in single modality images. Because the information presented in PET and CT is distinct but complementary, the dual-modality PET-CT images provides great potentials to the segmentation fields. This complementary character is well known in clinical practice where the diagnostic accuracy of the combined PET-CT has proven superior to either PET or CT alone.

We propose an efficient graph-based method to utilize the strength of each system for target/tumor delineation: the superior contrast of PET and the superior spatial resolution of CT. Recently, graph-based optimization receives a lot of attention in computer vision area [2,5,11,10,4]. The basic idea is to transform the target problem as an energy minimization problem, which can be efficiently solved by certain discrete optimization tools. In this paper, a general framework has been proposed to use both PET and CT images simultaneously for the dual-modality co-segmentation, which is formulated as a Markov Random Field (MRF) optimization problem. We introduce a new energy term into the objective energy function penalizing the segmentation difference between PET and CT images to achieve tumor segmentation in PET and CT simultaneously. Our new algorithm guarantees to achieve global optimal solutions with respect to the co-segmentation energy function in a low-order polynomial time by computing a single maximum flow in the constructed graph. This has clinical implications for improving both radiation therapy target definition and response assessment for clinical oncology. In Fig 1, we show an example of the PET-CT co-segmentation.

1.1 Related Work

Image segmentation from co-registered PET-CT has been attracted increasing attentions in recently years. Yu *et al.* [14] demonstrated the effectiveness of automated region-based segmentation of head-and-neck tumor from PET-CT by validating the results with manual tracings of radiation oncologists. The usefulness of combining the information of both PET and CT images for segmentation was shown by Baardwijk *et al.* [1]. Yu *et al.* [13] further demonstrated the effectiveness of using co-registered segmentation to identify useful textural features

for distinguishing tumor from normal tissue in the head and neck regions. The FDG-PET images coregistered with CT images was also proved to be superior to CT images in reducing inter-observer variability by Riegel *et al.* [8]. Potesil *et al.* [7] proposed a method utilizing the initial robust hot spot detection and segmentation performed in PET to provide the basic tumor appearance and shape model for classifying voxels in CT. Unfortunately, all those previous methods lack the capability of fully making use of the strengths of both PET and CT in the sense of concurrently segmenting the tumor from both modalities. Xia *et al.* [12] presented a systematic solution for the co-segmentation of brain PET-CT images with the MAP-MRF model, which was achieved by solving a maximum a posteriori (MAP) problem using the expectation-maximization (EM) algorithm with simulated annealing. Their method suffered long execution times due to the use of simulated annealing.

The limitations of existing approaches is the main impetus for designing a new efficient method for simultaneously segmenting the tumor from both PET and CT with high accuracy.

2 Problem Formulation

We formulate the task of co-segmenting as a binary labeling of Markov Random Field (MRF) on a graph corresponding to the input PET and CT images. The method attempts to simultaneously minimize the total MRF energy for both PET and CT images as well as penalizing the segmentation difference between two images.

In our co-segmentation problem, a discrete random variable f_u is introduced for each voxel u in either the input PET or CT image. Denote by \mathbf{f}_P (resp., \mathbf{f}_C) the set of variables corresponding to the voxels in the input PET (resp., CT) image. In the preprocessing phase, we apply the image registration algorithm to register the PET image to the CT image and upsample the PET image to make sure that both have the same size. We thus assume that there is a one-to-one correspondence between \mathbf{f}_P and \mathbf{f}_C and denote u' the voxel in the CT image corresponding the voxel u in the PET image. Each label f in \mathbf{f}_P or \mathbf{f}_C takes a label value from the label set $\mathcal{L} = \{0, 1\}$ indicating that the voxel is in the foreground ($f = 1$) or in the background ($f = 0$). Then, every possible assignment of the random variables in \mathbf{f}_P (resp., \mathbf{f}_C) define the tumor region in the PET (resp., CT) image. We certainly can compute an optimal tumor segmentation in the PET or CT image separately to minimize the corresponding MRF energy by applying Boykov and Funka-Lea's graph cuts method [2]. However, that method does not make use of the information from the other modality.

We propose to introduce a third set of co-segmentation binary variables \mathbf{f}_{P-C} each associated with a pair of corresponding voxels (u, u') in the PET and CT images to utilize the strength of both systems for segmentation. \mathbf{f}_{P-C} is used to incorporate the energy penalty of the segmentation difference between the PET and CT. That is, if the labelings of the pair of corresponding voxels (u, u') , f_u and $f_{u'}$, are the same, then no penalty is enforced; otherwise, we penalize the

disagreement. Furthermore, the PET and CT may weight in differently for the final co-segmentation result. Thus, a different penalty is enforced based on the agreement of the co-segmentation label $f_{(u,u')}$ with f_u from $f_{u'}$.

Hence, the problem of PET-CT co-segmentation is to minimize the following energy function (Eq. (II)).

$$\mathcal{E}_{PET-CT} = E_P(\mathbf{f}_P) + E_C(\mathbf{f}_C) + E_{P-C}(\mathbf{f}_{P-C}), \quad (1)$$

where $E_P(\mathbf{f}_P)$ and $E_C(\mathbf{f}_C)$ are the MRF energy functions for the PET and CT, respectively, and the energy term $E_{P-C}(\mathbf{f}_{P-C})$ is used to penalize the segmentation difference between the PET and CT. The co-segmentation energy term $E_{P-C}(\mathbf{f}_{P-C})$ makes use of the high contrast of PET and the high spatial resolution of CT to link the PET segmentation and the CT segmentation as a co-segmentation process.

The MRF Segmentation Energy for the PET. Let \mathcal{N}_P denote the neighborhood system used in the input PET image \mathcal{I}_P (e.g., the 4-neighbor setting). The MRF energy term $E_P(\mathbf{f}_P)$ consists of a data term and a smoothness term defined, as follows.

$$E(\mathbf{f}_P) = \sum_{u \in \mathcal{I}_P} d_u(f_u) + \sum_{(u,v) \in \mathcal{N}_P} w_{u,v}(f_u, f_v). \quad (2)$$

The data term $d_u(f_u)$ is the likelihood that imposes individual penalty for assigning a label f_u (i.e., the foreground (tumor) or the background) to the voxel u . The smoothness term $w(f_u, f_v)$, representing the interaction potential [2], measures the cost of assigning different labels to two neighboring voxels u and v in \mathcal{I}_P .

$$w_{u,v}(f_u, f_v) = \begin{cases} \epsilon(u, v), & \text{if } f_u \neq f_v \\ 0, & \text{if } f_u = f_v \end{cases} \quad (3)$$

where $\epsilon(u, v)$ is the smoothness value computed from the neighbouring voxels u and v . Note that we can obtain an optimal segmentation using only the PET image \mathcal{I}_P with respect to the energy function $E(\mathbf{f}_P)$ by applying the graph cuts method [2].

The MRF Segmentation Energy for the CT. Let \mathcal{N}_C denote the neighborhood system used in the input CT image \mathcal{I}_C . The MRF energy term $E_C(\mathbf{f}_C)$ has the same form as the energy term $E_P(\mathbf{f}_P)$ for the CT.

$$E(\mathbf{f}_C) = \sum_{u' \in \mathcal{I}_C} d_{u'}(f_{u'}) + \sum_{(u',v') \in \mathcal{N}_C} w_{u',v'}(f_{u'}, f_{v'}). \quad (4)$$

Again, an optimal segmentation using only the CT image \mathcal{I}_C with respect to the energy function $E(\mathbf{f}_C)$ can be obtained by applying the graph cuts method [2].

The Co-Segmentation Energy Term. $E_{P-C}(\mathbf{f}_{P-C})$ penalizes the segmentation difference between the PET \mathcal{I}_P and the CT \mathcal{I}_C . Each variable $f_{(u,u')} \in \mathbf{f}_{P-C}$ associated with a pair of corresponding voxels (u, u') in \mathcal{I}_P and \mathcal{I}_C takes a label from the label set $\mathcal{L} = \{0, 1\}$. If $f_{(u,u')} = 1$, then no matter f_u agrees with $f_{u'}$ or not, the voxels u and u' are classified as foreground (tumor); otherwise, the voxels u and u' are classified as background. We use the function $\gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'})$ to penalize the disagreement of f_u and $f_{u'}$ and resolve the disagreement based on the salient features of the PET and CT. The generalized Potts model is used in our method, as follows.

$$\gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'}) = \begin{cases} 0, & \text{if } f_u = f_{(u,u')} = f_{u'}, \\ \varphi_1(u, u'), & \text{if } f_u \neq f_{u'}, f_{(u,u')} = f_{u'}, \\ \varphi_2(u, u'), & \text{if } f_u \neq f_{u'}, f_{(u,u')} = f_u, \end{cases} \quad (5)$$

where $\varphi_1(u, u')$ and $\varphi_2(u, u')$ are the values to penalize the segmentation difference between u and u' .

Thus the co-segmentation energy term is defined as follows.

$$E_{P-C}(\mathbf{f}_{P-C}) = \sum_{u \in \mathcal{I}_P, u' \in \mathcal{I}_C} \gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'}). \quad (6)$$

3 The Graph Optimization Method

This section presents our graph-based algorithm for the PET-CT co-segmentation problem, which achieves globally optimal solution with respect to the energy function \mathcal{E}_{PET-CT} defined in Eq. (1). The algorithm runs in a low-order polynomial time by computing a minimum-cost s - t cut in the constructed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$.

Each variable $f_u \in \mathbf{f}_P$ (resp., $f_{u'} \in \mathbf{f}_C$) corresponds to exactly one node in \mathcal{V}_P (resp., \mathcal{V}_C). To simplify the notation, in graph \mathcal{G} , we also use u (resp., u') to denote the corresponding node of voxel u (resp., u') in \mathcal{I}_P (resp., \mathcal{I}_C). In addition, a graph node $z_{u,u'} \in \mathcal{V}_{P-C}$ is introduced for every variable $f_{(u,u')} \in \mathbf{f}_{P-C}$. Since our goal is to formulate the co-segmentation problem as a minimum-cost s - t cut problem, two terminal nodes, a source s and a sink t , are added. Hence, the graph node set $\mathcal{V} = \mathcal{V}_P \cup \mathcal{V}_C \cup \mathcal{V}_{P-C} \cup \{s, t\}$. As in Boykov and Funka-Lea's graph cuts method [2], we introduce t -edges and n -edges for the sub-node sets \mathcal{V}_P and \mathcal{V}_C separately, and additional d -edges between the corresponding nodes of \mathcal{V}_{P-C} and \mathcal{V}_P , and of \mathcal{V}_{P-C} and \mathcal{V}_C .

(1) **t-Edges** are used to incorporate the data term of the MRF segmentation energy (i.e., the data terms in Eqs. (2) and (4)). For each node u in \mathcal{V}_P , we put an edge from s to u with the edge cost of $d_u(f_u = 1)$ and an edge from u to t with the edge cost $d_u(f_u = 0)$. As in the graph cuts method [2], two subsets \mathcal{F} and \mathcal{B} of seed voxels are manually identified with \mathcal{F} of the foreground seeds and \mathcal{B} of the background seeds. For any voxel in \mathcal{F} (\mathcal{B}), we add hard constrains

to make sure it is foreground (background) in the final result. Instead of being used for the hard constraints, \mathcal{F} and \mathcal{B} are also used to compute the Gaussian Mixture Model (GMM) to represent the intensity distribution of each voxel in \mathcal{I}_P . Each of the background GMM and the foreground GMM is a full-covariance Gaussian mixture with K components ($K = 15$ in our study). The negative log likelihoods of a voxel for being background or foreground obtained by this method are used in $d_u(f_u)$. Denote by I_u the intensity value of voxel u in \mathcal{I}_P . The values of $d_u(f_u)$ are computed as in the following table.

$d_u(\cdot)$	$u \in \mathcal{F}$	$u \in \mathcal{B}$	$u \notin \mathcal{F} \cup \mathcal{B}$
$f_u = 0$	$+\infty$	0	$-\lambda_1 \log Pr(I_u \text{'background'})$
$f_u = 1$	0	$+\infty$	$-\lambda_2 \log Pr(I_u \text{'foreground'})$

In the same way, we introduce t -edges for each node $u' \in \mathcal{V}_C$.

(2) **n-Edges** are used to enforce the smoothness term of the MRF segmentation energy (i.e., the smoothness terms in Eqs. (2) and (4)). For each voxel pair (u, v) in the PET image \mathcal{I}_P , if $(u, v) \in \mathcal{N}_P$ (recall that \mathcal{N}_P is the neighboring setting of \mathcal{I}_P), we add two n -edges in \mathcal{G} , one from the node $u \in \mathcal{V}_P$ to the node $v \in \mathcal{V}_P$ and the other in the opposite direction from v to u . The cost of each edge is $\epsilon_{u,v}$, with

$$\epsilon(u, v) = \begin{cases} \lambda_3 e^{(-\theta_1 \|I_u - I_v\|)}, & (u, v) \in \mathcal{N}_P \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

The n -edges are introduced for the sub-node set \mathcal{V}_C in the same way with the neighboring setting \mathcal{N}_C .

(3) **d-Edges** are used to penalize the segmentation difference between the PET and CT images. Recall that for each corresponding pair of nodes (u, u') with $u \in \mathcal{V}_P$ and $u' \in \mathcal{V}_C$, we introduce a node $z_{u,u'} \in \mathcal{V}_{P-C}$. We now put two edges between u and $z_{u,u'}$, one from u to $z_{u,u'}$ and the other from $z_{u,u'}$ to u , each with a cost of $\varphi_1(u, u')$. Those d -edges are used to penalize the case that $f_u \neq f_{u'}$, but $f_{(u,u')} = f_{u'}$. Similarly, two edges between $z_{u,u'}$ and u' are added with each of cost $\varphi_2(u, u')$, penalizing the case that $f_u \neq f_{u'}$, but $f_{(u,u')} = f_u$. Both $\varphi_1(u, u')$ and $\varphi_2(u, u')$ are computed, as follows.

$$\begin{cases} \varphi_1(u, u') = \lambda_4 e^{(-\theta_2 \|I_u - I_{u'}\|)}, \\ \varphi_2(u, u') = \lambda_5 e^{(-\theta_3 \|I_u - I_{u'}\|)} \end{cases} \quad (8)$$

We thus finish the construction of the graph \mathcal{G} from the input PET and CT images \mathcal{I}_P and \mathcal{I}_C . Fig. 2 shows an example construction of the graph. For the 3D images, we use the 6-neighboring system for both \mathcal{N}_P and \mathcal{N}_C (this can be easily extended to any other form of the neighborhood system for the images).

We next show that the minimum-cost s - t cut $\mathcal{C}^* = (S^*, \bar{S}^*)$ defines a segmentation minimizing the objective energy function Eq. (11). Note that for a corresponding pair of voxels (u, u') in the PET and CT, if $f_u = f_{u'}$, then any d -edge between u and $z_{u,u'}$, and between $z_{u,u'}$ and u' cannot be in \mathcal{C}^* . Thus,

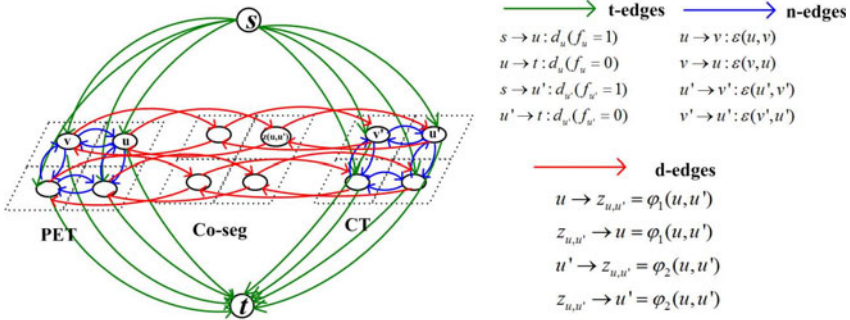


Fig. 2. Graph construction. Three types of edges are used for the graph construction. For each node $u \in \mathcal{V}_P \cup \mathcal{V}_C$, the edge cost of $s \rightarrow u$ is $d_u(f_u = 1)$, the edge cost of $u \rightarrow t$ is $d_u(f_u = 0)$. For two node $(u, v) \in \mathcal{N}_P$ ($(u, v) \in \mathcal{N}_C$), the edge cost of $u \rightarrow v$ is $\epsilon(u, v)$. For node $u \in \mathcal{V}_P$ and its correspondence $u' \in \mathcal{V}_C$, the edge cost for $u \rightarrow z_{u,u'}$ and $z_{u,u'} \rightarrow u$ are $\varphi_1(u, u')$. The edge cost for $u' \rightarrow z_{u,u'}$ and $z_{u,u'} \rightarrow u'$ are $\varphi_2(u, u')$. Refer to text for details.

we only consider an s - t cut not including that kind of d -edges, namely, an *admissible cut*. It is easy to see that any finite admissible s - t cut $\mathcal{C} = (S, \bar{S})$ defines a feasible configuration for $\mathbf{f}_P \cup \mathbf{f}_C \cup \mathbf{f}_{P-C}$ (i.e., a co-segmentation). We can also show that the total cost $w(\mathcal{C})$ of the cut \mathcal{C} equals the total energy \mathcal{E}_{PET-CT} of the segmentation. Divide the s - t cut \mathcal{C} into three disjoint sub-cuts: $((S \cap \mathcal{V}_P) \cup \{s\}, (\bar{S} \cap \mathcal{V}_P) \cup \{t\})$, $((S \cap \mathcal{V}_C) \cup \{s\}, (\bar{S} \cap \mathcal{V}_C) \cup \{t\})$, and those d -edges in \mathcal{C} . Using a similar argument in Boykov and Funka-Lea’s graph cuts method [2], we can prove that the total cost of the sub-cut $((S \cap \mathcal{V}_P) \cup \{s\}, (\bar{S} \cap \mathcal{V}_P) \cup \{t\})$ equals the MRF segmentation energy $E(\mathbf{f}_P)$ for the PET, and that the total cost of the sub-cut $((S \cap \mathcal{V}_C) \cup \{s\}, (\bar{S} \cap \mathcal{V}_C) \cup \{t\})$ equals the MRF segmentation energy $E(\mathbf{f}_C)$ for the CT.

We now calculate the total cost of those d -edges in \mathcal{C} , denoted \mathcal{C}_d . For any edge $(a, b) \in \mathcal{C}_d$ (i.e., $a \in S$ and $b \in \bar{S}$), four cases need to be considered: *Case 1)* $a \in \mathcal{V}_P$ and $b \in \mathcal{V}_{P-C}$; *Case 2)* $a \in \mathcal{V}_{P-C}$ and $b \in \mathcal{V}_C$; *Case 3)* $a \in \mathcal{V}_C$ and $b \in \mathcal{V}_{P-C}$; and *Case 4)* $a \in \mathcal{V}_{P-C}$ and $b \in \mathcal{V}_P$. Assume that the corresponding voxel pair is (u, u') .

For Case 1): $a \in \mathcal{V}_P$ and $b \in \mathcal{V}_{P-C}$. The cost of the edge (a, b) with $w(a, b) = w(u, z_{u,u'}) = \varphi_1(u, u')$. In this case, $f_u = 1$ and $f_{(u,u')} = f_{u'} = 0$. Based on Eq. (5), $\gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'}) = \varphi_1(u, u')$. That is, $w(a, b) = \gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'})$. With a similar argument, we can prove that $w(a, b) = \gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'})$ for all the other three cases.

Putting all those four cases together, we have that the total cost of \mathcal{C}_d with $w(\mathcal{C}_d) = \sum_{(a,b) \in \mathcal{C}_d} w(a, b) = \sum_{u \in \mathcal{V}_P, u' \in \mathcal{V}_C} \gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'})$ being aware of the fact that $\gamma_{u,u'}(f_u, f_{(u,u')}, f_{u'}) = 0$ if $f_u = f_{(u,u')} = f_{u'}$.

Hence, the total cost of an finite admissible s - t cut equals to the total energy of the its corresponding segmentation. Therefore, the optimal co-segmentation solution can be found by computing a minimum s - t cut in \mathcal{G} [2].

4 Experimental Methods

Implementation and Parameters Setting. Our algorithm was implemented with ISO C++ on a standard PC with a 2.46GHz Intel R CoreTM2 Duo processor and 4 GB memory, running 64-bit Windows system. The max flow library [2] was utilized as the optimization tool. For the PET image, we use the Elastix [6] tools to register it to CT image. After the registration, the PET and CT have the same resolution and each voxel $u \in \mathcal{V}_P$ and its correspondence $u' \in \mathcal{V}_C$ have the same location in PET and CT, respectively. In our experiments, we used the following parameter setting. The data term coefficients, $\lambda_1 = \lambda_2 = 1$. For the smoothness term coefficients, we set $\lambda_3 = 50$ and $\theta_1 = \theta_2 = \theta_3 = 1$. The coefficients in the co-segmentation term were set $\lambda_4 = 50$ and $\lambda_5 = 1$. Note that by choosing λ_4 and λ_5 with fixed constants, we do not fully explore the power of our model in Eq. (5).

Data Collection. We used FDG-PET-CT images of a group of 16 patients with head-and-neck cancer (HNC) for the present study. The reconstructed matrix size for each CT slice was 512×512 , with a voxel size of $0.97 \times 0.97 \times 2$ mm. For the PET images, the reconstructed matrix size was 168×168 , with a voxel size of $3.39 \times 3.39 \times 2$ mm. The PET images were co-registered to the CT images and interpolated to the same voxel size so that there was a one-to-one correspondence between PET and CT. The radiation oncologist with expertise in HNC manually segmented the primary tumors from the 16 PET-CT images as the ground truth.

Evaluation Strategy. The segmentation performance was evaluated using the Dice similarity coefficient (DSC) and median hausdorff distance (HD). A DSC of 0 means that the segmentations have no overlap at all, while 1 indicates a perfect matching. A small median HD value indicates an accurate segmentation and a large value gives an indication that the segmentation method is not accurate. All DSC and median HD values were computed in 3-D. We compared three methods on 16 PET-CT datasets: (1) the proposed PET-CT co-segmentation method (Co-Seg); (2) the graph cuts method only using PET (PET-Seg) [2]; and (3) the graph cuts method only using CT (CT-Seg) [2]. We manually identified a set of foreground and a set of background seeds for each dataset and used for all the three methods.

5 Results and Discussion

The results are summarized in Figure 3 showing the DSCs and median HDs for the three methods. Our method obtained tumor segmentation accuracy characterized by $DSC = 0.86 \pm 0.051$ on average, which was higher than that of the PET-Seg method (0.78 ± 0.045) and the CT-Seg method (0.51 ± 0.12). The median Hausdorff distance mostly ranged between 5-7 mm. The average value was 6.4 ± 0.56 mm, which was much smaller than those of PET-Seg (7.6 ± 0.97 mm) and CT-Seg (11.6 ± 2.73 mm). We noted that the dataset No. 3 has a relatively low accuracy, which is due to the fact that the ground truth is not good enough.

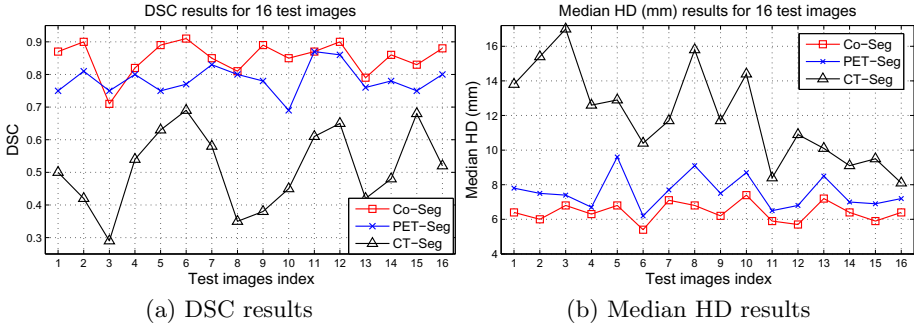


Fig. 3. DSC and median HD results for different methods

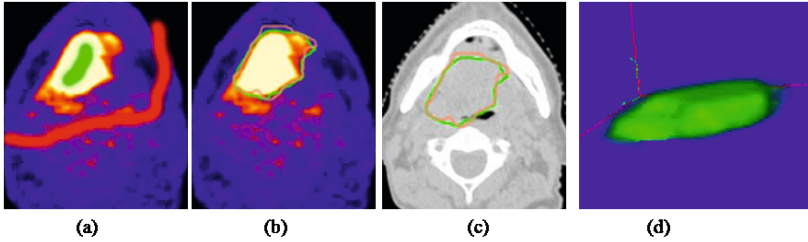


Fig. 4. A tumor segmentation example. (a) 2-D slice of 3-D PET image with seeds highlighted. The read line and the green line were background and foreground seeds. (b) The results in PET image. Manual segmentation were showed with orange and our results were showed with green. (c) The results in CT image. (d) 3-D representation of the tumor (green).

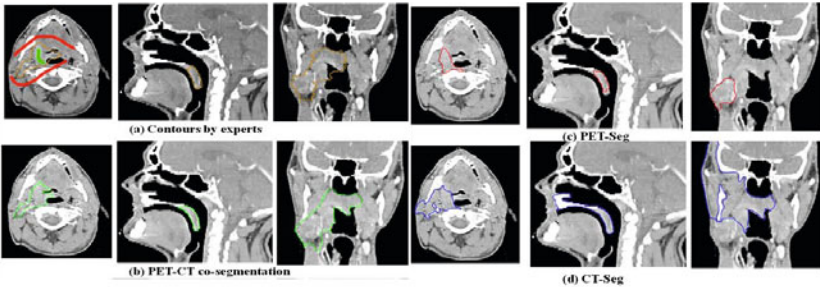


Fig. 5. Segmentation result comparisons. (a) The manual segmentation results. The foreground seeds are shown in green and the background seeds are shown in red. (b) The results by our co-segmentation method (Co-Seg). (c) The results by the graph cuts method solely using the PET image (PET-Seg). (d) The results by the graph cuts method solely using the CT image(CT-Seg).

Fig. 4 showed a typical tumor segmentation result by our method. The visual comparison with the independent standard showed our method matched well to the ground truth. Fig. 5 illustrates the segmentation results on one head-and-neck test dataset for the comparison purpose of the three methods. The results demonstrate the high accuracy of our Co-Seg method comparing to the independent stand. In most of cases, the results using only PET had smaller segmentation volumes than the co-segmentation results. Due to the low resolution in PET, the voxels along the tumor boundary suffered the partial volume effect thus the tumor volume can not be fully identified using solely PET datasets. On the other hand, though with high resolution, the low contrast of the CT datasets led to the biggest segmentation errors among the three methods. Our co-segmentation approach can successfully incorporate both the information from PET and CT. Another issue we need to discuss is the role of co-registration as a pre-processing step. Currently, we assume there is one-to-one correspondence between PET and CT after registration. However, the registration errors will affect the results. To overcome this problem, introducing a high order term may reduce such errors and we will investigate it in our future work.

The running time of our method includes two parts: (1) PET-CT registration; and (2) PET-CT co-segmentation. For each image with size of $512 \times 512 \times 100$, the execution time for the registration was about 60s. In the PET-CT co-segmentation process, we used the ROI to reduce the computing time. Commonly, the size of ROI was about $200 \times 200 \times 60$. The time for co-segmentation was about 20s. We note that our current code is not fully optimized and there is much room for improvement.

6 Conclusion and Future Work

In this paper, we have proposed a novel method to make use of the strength of both the PET and CT systems for tumor segmentation. The tumor in the PET and CT images is concurrently segmented with the awareness of the information provided by the other modality. Our method was tested on segmenting 16 patients' datasets with challenging neck-and-head tumor. The results showed high accuracy on real clinical data comparing to those obtained by the method solely using PET or CT image data. In the future, one interesting work is to extend our idea of using dual-modality PET-CT images to [9], in which the authors tried to use surface context to improve tumor segmentation only in CT images. Additional, we plan to further improve the accuracy by studying adaptive parameter selection, and investigate the use of atlas-based segmentation method to reduce the human interaction.

Acknowledgements

This research was supported in part by the NSF grants CCF-0830402 and CCF-0844765, and the NIH grants R01 EB004640 and K25 CA123112.

References

1. Baardwijk, A., Bosmans, G., Boersma, L., et al.: PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *International Journal of Radiation Oncology, Biology and Physics* 68(3) (2007)
2. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision* 70(2), 109–131 (2006)
3. Ford, E., Kinahan, P., Hanlon, L., et al.: Tumor delineation using PET in head and neck cancers: threshold contouring and lesion volumes. *Medical Physics* 33, 4280 (2006)
4. Han, D., Bayouth, J., Bhatia, S., Sonka, M., Wu, X.: Motion Artifact Reduction in 4D Helical CT: Graph-Based Structure Alignment. *Medical Computer Vision*. In: *Recognition Techniques and Applications in Medical Imaging (MICCAI MCV)*, pp. 63–73 (2011)
5. Han, D., Wu, X., Sonka, M.: Optimal multiple surfaces searching for video/image resizing—a graph-theoretic approach. In: *IEEE 12th International Conference on Computer Vision, ICCV 2009*, pp. 1026–1033 (2009)
6. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging* 29(1), 196–205 (2010)
7. Potesil, V., Huang, X., Zhou, X.: Automated tumor delineation using joint PET/CT information. In: *Proc. SPIE International Symposium on Medical Imaging: Computer-Aided Diagnosis*, vol. 6514 (2007)
8. Riegel, A., Berson, A., Destian, S., et al.: Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *International Journal of Radiation Oncology, Biology and Physics* 65(3) (2006)
9. Song, Q., Chen, M., Bai, J., Sonka, M., Wu, X.: Surface–region context in optimal multi-object graph-based segmentation: Robust delineation of pulmonary tumors. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011. LNCS*, vol. 6801, pp. 61–72. Springer, Heidelberg (2011), (this volume)
10. Song, Q., Wu, X., Liu, Y., Haeker, M., Sonka, M.: Simultaneous searching of globally optimal interacting surfaces with shape priors. In: *CVPR* (2010)
11. Song, Q., Wu, X., Liu, Y., Smith, M., Buatti, J., Sonka, M.: Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5762, pp. 827–835. Springer, Heidelberg (2009)
12. Xia, Y., Wen, L., Eberl, S., Fulham, M., Feng, D.: Segmentation of dual modality brain PET/CT images using the MAP-MRF model. In: *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pp. 107–110. IEEE, Los Alamitos (2008)
13. Yu, H., Caldwell, C., Mah, K., Mozeg, D.: Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Transactions on Medical Imaging* 28(3), 374–383 (2009)
14. Yu, H., Caldwell, C., Mah, K., et al.: Automated radiation targeting in head-and-neck cancer using region-based texture analysis of PET and CT images. *International Journal of Radiation Oncology, Biology and Physics* 75(2), 618–625 (2009)

Approximations of the Diffeomorphic Metric and Their Applications in Shape Learning

Xianfeng Yang¹, Alvina Goh², and Anqi Qiu^{1,3}

¹ Division of Bioengineering, National University of Singapore, Singapore

² Department of Mathematics, National University of Singapore, Singapore

³ Clinical Imaging Research Center, National University of Singapore, Singapore

Abstract. In neuroimaging studies based on anatomical shapes, it is well-known that the dimensionality of the shape information is much higher than the number of subjects available. A major challenge in shape analysis is to develop a dimensionality reduction approach that is able to efficiently characterize anatomical variations in a low-dimensional space. For this, there is a need to characterize shape variations among individuals for N given subjects. Therefore, one would need to calculate $\binom{N}{2}$ mappings between any two shapes and obtain their distance matrix. In this paper, we propose a method that reduces the computational burden to N mappings. This is made possible by making use of the first- and second-order approximations of the metric distance between two brain structural shapes in a diffeomorphic metric space. We directly derive these approximations based on the so-called conservation law of momentum, i.e., the diffeomorphic transformation acting on anatomical shapes along the geodesic is completely determined by its velocity at the origin of a fixed template. This allows for estimating morphological variation of two shapes through the first- and second-order approximations of the initial velocity in the tangent space of the diffeomorphisms at the template. We also introduce an alternative representation of these approximations through the initial momentum, i.e., a linear transformation of the initial velocity, and provide a simple computational algorithm for the matrix of the diffeomorphic metric. We employ this algorithm to compute the distance matrix of hippocampal shapes among an aging population used in a dimensionality reduction analysis, namely, ISOMAP. Our results demonstrate that the first- and second-order approximations are sufficient to characterize shape variations when compared to the diffeomorphic metric constructed through $\binom{N}{2}$ mappings in ISOMAP analysis.

Keywords: diffeomorphic metric, exponential map, ISOMAP, hippocampal shape.

1 Introduction

Brain shape analysis characterizes morphological variation of different subjects relative to the coordinates of a template shape. Compared to brain volumetric analysis, shape analysis can better distinguish patients from normal controls in

a variety of neurodevelopment disorders, psychiatric and neurodegenerative diseases [119]. Thus, there has been great emphasis on developing new approaches for shape analysis of brain anatomical models. However, structural shapes live in a high dimensional space whose dimensionality is much higher than the number of subjects available. A major challenge in shape analysis is to develop a dimensionality reduction approach that efficiently characterizes anatomical variations in a low-dimensional space for facilitating the visualization and classification of shapes, and detecting the association between shapes and clinical measures. ISOMAP is one of several widely used dimensionality reduction methods where a shape distance matrix is used for computing a quasi-isometric, low-dimensional embedding of a set of high-dimensional shapes [7,15]. It estimates the intrinsic geometry of a shape manifold based on a rough estimate of each shape's neighbors on the manifold. However, the computation of the matrix whose elements quantify geodesic distances between any two in the set of N shapes requires $\binom{N}{2}$ shape mappings, which is very computationally intensive. In this paper, we derive approximations of the geodesic distance between two shapes in a diffeomorphic metric space to reduce the computational burden of the distance matrix from $\binom{N}{2}$ to N mappings.

Our proposed approximations of the geodesic distance between two shapes are consistently designed under the framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) for anatomical surface models [14]. Even though there are many available registration algorithms for aligning brain anatomies [22,5,20,3], high-dimensional diffeomorphic mapping approaches are widely used due to their mapping accuracy [10,14,2]. In particular, LDDMM places anatomical shapes in a metric space and provides diffeomorphic transformations, i.e., one-to-one, smooth (forward and inverse), acting on the ambient space [8,24,28,14]. Thus, under LDDMM, connected sets remain connected, disjoint sets remain disjoint, and the smoothness of neuroanatomical features such as curves and surfaces is preserved. As these transformations are generated by the flow of diffeomorphisms, they are not additive and provide guaranteed bijections between anatomies. The length of the optimal flow trajectory connecting two anatomies in the metric space gives a geodesic metric distance that can be used to quantify the closeness in shapes between different instances. [15] computed the matrix of the diffeomorphic metric distances through $\binom{N}{2}$ LDDMM mappings and applied multi-dimensional scaling to embed hippocampal shapes in a finite Euclidean space for distinguishing demented subjects from normal controls. [7] reduced this matrix computation whose element was estimated by a sum of piece-wise diffeomorphic metric distances among neighborhood shapes. In this way, [7] assumed that neighborhood shapes are similar so that their diffeomorphic metrics can be approximated from a small deformation model rather than the large deformation model as given in [15]. ISOMAP was then employed to embed the whole brain shapes in a finite Euclidean space that characterizes the brain morphology in the healthy aging and Alzheimer's disease. Nevertheless, it did not evaluate shape information that could be lost in the small deformation model.

In this paper, we present the first- and second-order approximations of the diffeomorphic metric in the large deformation setting so that we are able to construct the matrix of the diffeomorphic metric among N subjects using N LD-DMM mappings given a fixed template. In our earlier work [26], we made use of the first-order approximation, showed how a shape can be approximated as a linear combination of its neighbors, and proposed the Locally Linear Diffeomorphic Metric Embedding algorithm. These approximations are directly derived based on the so-called conservation law of momentum, i.e., the diffeomorphic transformation acting on anatomical shapes along the geodesic is completely determined by its velocity at the origin of a fixed template [16]. We directly estimate morphological variation of two shapes through the first- and second-order approximations of the initial velocity in the tangent space of the diffeomorphisms at the template, which facilitates the metric distance computation as a inner product of these approximations in the diffeomorphic metric space. We also introduce an alternative representation of these approximations through the initial momentum, i.e., a linear transformation of the initial velocity, and provide a simple computation algorithm for the matrix of the diffeomorphic metric. We discuss the performance of these approximations in shape learning using ISOMAP by comparing to the distance matrix directly computed from $\binom{N}{2}$ mappings.

2 Methods

We represent brain anatomical shapes using a surface model S throughout the paper. In order to capture the variations between any two shapes, there is a need to consider them in a metric space. We first review briefly, the construction of the metric space of shapes and the group of diffeomorphic transformations relating two shapes, followed by the definition of the metric distance measuring the similarity between the two shapes in terms of velocity fields of the diffeomorphisms. We then derive the first-order and second-order approximation of the metric distance for the efficient computation and provide the alternative computation based on the initial momentum, which is a linear transformation of the initial velocity field of a diffeomorphic flow, for the surface model.

2.1 Review: Diffeomorphic Metric

Given a template surface S_{temp} , the metric space \mathcal{S} is constructed as an orbit of S_{temp} under the group of diffeomorphic transformations \mathcal{G} , i.e., $\mathcal{S} = \mathcal{G} \cdot S_{\text{temp}}$. The diffeomorphic transformations (one-to-one, smooth forward and inverse transformation) are introduced as transformations of the coordinates on the background space $\Omega \subset \mathbb{R}^3$, i.e., $\mathcal{G} : \Omega \rightarrow \Omega$. Notice that the mappings under this construction are diffeomorphisms living in the infinite dimensional setting and it is significantly less clear how to generate such mappings, when compared to other settings where there is a finite number of parameters, e.g., the finite dimensional rotation groups.

One approach, proposed by Grenander and Miller [9] and adopted in this paper, is to construct diffeomorphisms $\phi_t \in \mathcal{G}$ as a flow of ordinary differential equations (ODEs), where $\phi_t, t \in [0, 1]$ obeys the following equation:

$$\dot{\phi}_t = v_t(\phi_t), \quad \phi_0 = \text{Id}, \quad t \in [0, 1], \tag{1}$$

where Id denotes the identity map and v_t are the associated velocity vector fields. The vector fields v_t are constrained to be sufficiently smooth, so that Eq. (1) is integrable and generates diffeomorphic transformations over finite time. The smoothness is ensured by forcing v_t to lie in a smooth Hilbert space $(V, \|\cdot\|_V)$ with s -derivatives having finite integral square and zero boundary [4, 23]. In our case, we model V as a reproducing kernel Hilbert space with a linear operator L and its kernel $k_V = L^{-1}$ associated with the norm square $\|u\|_V^2 = \langle Lu, u \rangle_2 = \langle k_V^{-1}u, u \rangle_2, \forall u \in V$, where $\langle \cdot, \cdot \rangle_2$ denotes the \mathbf{L}^2 inner product. The group of diffeomorphisms $\mathcal{G}(V)$ is the solutions of Eq. (1) with the vector fields satisfying $\int_0^1 \|v_t\|_V dt < \infty$.

Now, given the template surface $S_{\text{temp}} \in \mathcal{S}$ and a target surface $S_{\text{targ}} \in \mathcal{S}$, the geodesic $\phi_t, t \in [0, 1]$ which lies in the manifold of diffeomorphisms and connects S_{temp} and S_{targ} , is defined as

$$\phi_0 = \text{Id}, \quad \phi_1 \cdot S_{\text{temp}} = S_{\text{targ}}.$$

A metric $\rho(S_{\text{temp}}, S_{\text{targ}})$ between S_{temp} and S_{targ} is then defined as the Riemannian length of ϕ_t , computed as the integral of the norm of the vector field $\|v_t\|_V$ associated with ϕ_t . Based on the fact that energy-minimizing curves coincide with constant-speed length-minimizing curves, we can compute the length of the geodesic (the metric between S_{temp} and S_{targ}) through the following variational problem as

$$\begin{aligned} \rho(S_{\text{temp}}, S_{\text{targ}})^2 &= \inf_{\substack{v_t: \dot{\phi}_t = v_t(\phi_t), \\ \phi_0 = \text{Id}}} \int_0^1 \|v_t\|_V^2 dt \\ &= \inf_{\substack{v_t: \dot{\phi}_t = v_t(\phi_t), \\ \phi_0 = \text{Id}}} \int_0^1 \langle Lv_t, v_t \rangle_2 dt \quad \text{such that } \phi_1 \cdot S_{\text{temp}} = S_{\text{targ}}. \end{aligned} \tag{2}$$

One can prove that the velocity field v_t defined in Eq. (2) satisfies the following property at all times [16]:

Property 1. For all $u \in V$,

$$\langle Lv_t, u \rangle_2 = \langle Lv_0, (D\phi_t)^{-1}u(\phi_t) \rangle_2,$$

where $D\phi_t$ is the Jacobian matrix of ϕ_t .

As a direct consequence of Property 1, given the initial velocity vector field v_0 , one can generate a unique time-dependent diffeomorphic transformation. We thus rewrite the metric distance between S_{temp} and S_{targ} in the form of

$$\rho(S_{\text{temp}}, S_{\text{targ}})^2 = \inf_{v_0: \phi_t = v_t(\phi_t)} \langle Lv_0, v_0 \rangle_2, \tag{3}$$

where v_t satisfies Property \square . We denote such a diffeomorphic transformation that brings S_{temp} to S_{targ} as $\phi_t^{v_0}$, $\phi_1^{v_0} \cdot S_{\text{temp}} = S_{\text{targ}}$. $\phi_t^{v_0}$ is also the exponential map on $\mathcal{G}(V)$ for its right invariant Riemannian structure, i.e., $\phi_t^{v_0} = \text{Exp}_{\text{Id}}(tv_0)$. Next, we show that by approximating the exponential map to the first (or second) order leads to the first-order (or second-order) approximation of the metric distance between any two shapes.

2.2 The First- and Second-Order Approximations of the Diffeomorphic Metric

We now define the first- and second-order approximations of the diffeomorphic metric between two brain anatomical shapes through an exponential map. We denote the brain anatomical shapes of the i^{th} and j^{th} subjects as S_i and S_j and assume the diffeomorphic transformations that bring the template S_{temp} to S_i and S_j to be $\phi_t^{v_0^i}$ and $\phi_t^{v_0^j}$, respectively. Based on the composition law on $\mathcal{G}(V)$, S_i is related to S_j as $S_j = (\phi_1^{v_0^j} \circ (\phi_1^{v_0^i})^{-1}) \cdot S_i$. The geodesic curve ϕ_t , connecting S_i and S_j , at $t = 1$ is then given as

$$\phi_1 = \text{Exp}_{\text{Id}}(v_0^j) \circ \text{Exp}_{\text{Id}}(-v_0^i). \tag{4}$$

If we denote $\mu = \text{Log}_{\text{Id}}(\text{Exp}_{\text{Id}}(v_0^j) \circ \text{Exp}_{\text{Id}}(-v_0^i))$, from the weakened version of the Baker-Campbell-Hausdorff formula for the infinite-dimensional diffeomorphic group [25], μ can be expanded at identity as

$$\mu = v_0^j - v_0^i + \frac{1}{2} [v_0^j, -v_0^i] + \dots, \tag{5}$$

where $[\cdot, \cdot]$ represents the Lie bracket of vector fields. ϕ_t in Eq. (4) can be written as

$$\phi_1 = \text{Exp}_{\text{Id}}(\mu) = \text{Exp}_{\text{Id}}((v_0^j - v_0^i) + \frac{1}{2} [v_0^j, -v_0^i] + \dots).$$

The first-order approximation of ϕ_1 in Eq. (4) can be written as

$$\phi_1 = \text{Exp}_{\text{Id}}(\mu) \approx \text{Exp}_{\text{Id}}(v_0^j - v_0^i).$$

Therefore, the metric between S_i and S_j can be approximated to the first-order as

$$\rho(S_i, S_j)^2 \approx \rho^{1^{\text{st}}}(S_i, S_j)^2 = \langle L(v_0^j - v_0^i), v_0^j - v_0^i \rangle_2. \tag{6}$$

The second-order approximation of ϕ_1 can be similarly defined as

$$\phi_1 = \text{Exp}_{\text{Id}}(\mu) \approx \text{Exp}_{\text{Id}}((v_0^j - v_0^i) + \frac{1}{2} [v_0^j, -v_0^i]).$$

This computation is not straightforward because of the involvement of the Lie bracket of vector fields. Following the derivation given by Younes [27], the Lie bracket of vector fields is computed as the derivatives of the vector fields being taken at $s, t = 0$, i.e.,

$$\begin{aligned} [v_0^j, -v_0^i] &= \frac{d}{dt} \frac{d}{ds} (\text{Id} + tv_0^j) \circ (\text{Id} - sv_0^i) \circ (\text{Id} + tv_0^j)^{-1} |_{s=0, t=0} \\ &= -\frac{d}{dt} D(\text{Id} + tv_0^j)v_0^i \circ (\text{Id} + tv_0^j)^{-1} |_{s=0, t=0} = Dv_0^i v_0^j - Dv_0^j v_0^i, \end{aligned}$$

where Dv is the Jacobian of v . Notice that in general, the Lie bracket, $Dv_0^i v_0^j - Dv_0^j v_0^i$, does not necessarily belong to the Hilbert space, V . However, we will assume that we consider larger spaces that would contain it. For example, one could be the union of V and a space of vector fields with one degree differentiable lower than $v \in V$. Under this assumption, the second-order approximation of ϕ_t can then be written in the form of

$$\phi_1 = \text{Exp}_{\text{Id}}(\mu) \approx \text{Exp}_{\text{Id}}((v_0^j - v_0^i) + \frac{1}{2}(Dv_0^i v_0^j - Dv_0^j v_0^i)).$$

This yields the second-order approximation of the metric between S_i and S_j as

$$\begin{aligned} \rho(S_i, S_j)^2 &\approx \rho^{2^{\text{nd}}}(S_i, S_j)^2 \tag{7} \\ &= \langle L(v_0^j - v_0^i + \frac{1}{2}(Dv_0^i v_0^j - Dv_0^j v_0^i)), v_0^j - v_0^i + \frac{1}{2}(Dv_0^i v_0^j - Dv_0^j v_0^i) \rangle_2. \end{aligned}$$

2.3 Alternative Computation through Initial Momentum

Now, we apply the property of the duality isometry in Hilbert spaces and derive the equivalent representations of ρ , $\rho^{1^{\text{st}}}$, and $\rho^{2^{\text{nd}}}$ in terms of the momentum m_t . m_t is defined as a linear transformation of v_t through kernel $k_V = L^{-1}$ associated with the reproducing kernel Hilbert space V . More precisely, k_V maps v_t to m_t , i.e., $k_V : v_t \rightarrow m_t = k_V^{-1}v_t$. Therefore, for each $u \in V$, $\langle m_t, u \rangle_2 = \langle k_V^{-1}v_t, u \rangle_2$, where $\langle \cdot, \cdot \rangle_2$ denotes the \mathbf{L}^2 inner product. By combining the definition of momentum with Property 1, Miller et al. [16] obtained the following law.

Conservation Law of Momentum. *For all $u \in V$,*

$$\langle m_t, u \rangle_2 = \langle m_0, (D\phi_t)^{-1}u(\phi_t) \rangle_2.$$

This uniquely specifies m_t as a linear form on V , given the initial momentum m_0 and the evolving diffeomorphism ϕ_t . We see that making a change of variables yields the following expression relating m_t to the initial momentum m_0 and the geodesic ϕ_t connecting S_{temp} and S_{targ} ,

$$m_t = |D\phi_t^{-1}|(D\phi_t^{-1})^\top m_0 \circ \phi_t^{-1}. \tag{8}$$

Substituting $v_t = k_V m_t$ into Eq. (3), the diffeomorphic metric can be rewritten as

$$\rho(S_{\text{temp}}, S_{\text{targ}})^2 = \inf_{\substack{m_0: \phi_t = k_V m_t(\phi_t), \\ \phi_0 = \text{Id}}} \langle m_0, k_V m_0 \rangle_2 \quad \text{s.t. } \phi_1 \cdot S_{\text{temp}} = S_{\text{targ}}, \tag{9}$$

where m_t satisfies Eq. (8) (16). According to the conservation law of momentum and the duality isometry between the momentum and velocity field, the first- and second-order approximations of the diffeomorphic metric in Eqs. (6)-(7) can be expressed as

$$\rho^{1st}(S_i, S_j)^2 = \langle (m_0^j - m_0^i), k_V(m_0^j - m_0^i) \rangle_2 \tag{10}$$

$$\rho^{2nd}(S_i, S_j)^2 = \langle (m_0^j - m_0^i) + \frac{1}{2}L[D(k_V m_0^i)(k_V m_0^j) - D(k_V m_0^j)(k_V m_0^i)], \tag{11}$$

$$k_V(m_0^j - m_0^i) + \frac{1}{2}[D(k_V m_0^i)(k_V m_0^j) - D(k_V m_0^j)(k_V m_0^i)] \rangle_2.$$

2.4 Numerical Implementation

In this section, we demonstrate that Eqs. (10) and (11) lead to a simple computation of shape reconstruction using initial momentum, when surface-based brain anatomies are represented as triangular meshes in a discrete setting.

Assume that surfaces S_{temp} and S_{targ} are represented as triangular meshes. Let $S_{temp} = \mathbf{x} = (x_l)_{l=1}^n$ and $S_{targ} = \mathbf{y} = (y_l)_{l=1}^m$, where the number of vertices on S_{temp} and S_{targ} are n and m , respectively. For simplicity, we denote the kernel of V as $k_V \text{Id}_{3 \times 3}$, where k_V is a scalar kernel and $\text{Id}_{3 \times 3}$ is an 3×3 identity matrix. We assume that the geodesic vector fields connecting the two are splines such that $v_t(\cdot) = \sum_{l=1}^n k_V(\phi_t(x_l), \cdot)\alpha_l(t)$, where $\alpha_l(t)$ is the momentum vector of the l^{th} point at time t . The momentum m_t is given as a sum of Dirac measures, $m_t = \sum_{l=1}^n \alpha_l(t) \otimes \delta_{\phi_t(x_l)}$. Therefore, for any $u \in V$, $\langle m_t, u \rangle_2 = \sum_{l=1}^n \langle \alpha_l(t), u(\phi_t(x_l)) \rangle_2$.

The first- and second-order approximations of the metric between S_i and S_j in Eqs. (10) and (11) are expressed in terms of the initial momentum vectors as

$$\rho^{1st}(S_i, S_j)^2 = \sum_{l=1}^n \sum_{q=1}^n (\alpha_l^j - \alpha_l^i)^\top \left[k_V(x_l, x_q) (\alpha_q^j - \alpha_q^i) \right], \tag{12}$$

$$\rho^{2nd}(S_i, S_j)^2 = \sum_{l=1}^n \sum_{q=1}^n \tag{13}$$

$$\left[(\alpha_l^j - \alpha_l^i) + \frac{1}{2} \sum_{k=1}^n k_V^{-1}(x_l, x_k) \left[\left(\sum_{p=1}^n \alpha_p^i \nabla_1 k_V(x_k, x_p) \right)^\top \right] \left(\sum_{p=1}^n k_V(x_k, x_p) \alpha_p^j \right) \right. \\ \left. - \left(\sum_{p=1}^n \alpha_p^j \nabla_1 k_V(x_k, x_p) \right)^\top \left(\sum_{p=1}^n k_V(x_k, x_p) \alpha_p^i \right) \right]^\top \\ \left[k_V(x_l, x_q) (\alpha_q^j - \alpha_q^i) + \frac{1}{2} \left[\left(\sum_{p=1}^n \alpha_p^i \nabla_1 k_V(x_q, x_p) \right)^\top \right] \left(\sum_{p=1}^n k_V(x_q, x_p) \alpha_p^j \right) \right. \\ \left. - \left(\sum_{p=1}^n \alpha_p^j \nabla_1 k_V(x_q, x_p) \right)^\top \left(\sum_{p=1}^n k_V(x_q, x_p) \alpha_p^i \right) \right],$$

where α_i^i and α_q^j are initial momenta associated with $\phi_t^{v_0^i}$ and $\phi_t^{v_0^j}$ in short notations, respectively. $\nabla_1 k_V(x, y)$ is the derivative of k_V with respect to x .

We now summarize the pairwise metric distance computation among N subjects, $S_i, i = 1, 2, \dots, N$, where S_i is represented by a discrete surface such that $S_i = \mathbf{y} = (y_l)_{l=1}^m$. Given template $S_{\text{temp}} = \mathbf{x} = (x_l)_{l=1}^n$, we find the diffeomorphic transformation $\phi_t^{\alpha^i(0)}$ associated with initial momentum $\alpha^i(0)$ to align S_{temp} to S_i using a large deformation diffeomorphic metric surface mapping algorithm (LDDMM-surface) [28]. The first- and second-order approximations of the metric distance between the i^{th} and j^{th} subjects can be computed using Eqs. (12) and (13), respectively. Notice that this reduces the number of the surface-based LDDMM mappings from $\binom{N}{2}$ to N . We demonstrate how to use a matrix with the ij element as a diffeomorphic metric between subjects i and j in shape learning in §3

3 Results

In this section, we demonstrate applications of the first- and second-order approximations of the diffeomorphic metric in shape learning of aging hippocampal surface-based shapes of 196 subjects. Before doing so, we briefly describe the hippocampal segmentation and surface construction and then evaluate the two approximations of the diffeomorphic metric by comparing them with the metric directly computed by the LDDMM-surface mapping (so called a full metric ρ throughout the paper) [28].

196 T1-weighted magnetic resonance (MR) images of healthy subjects (age range: 18 – 94 years, mean age=56.5 years, standard deviation: 21.4 years) were randomly selected from the public available dataset of Open Access Series of Imaging Studies (OASIS) [13]. We automatically delineated the hippocampus from the intensity-inhomogeneity corrected T1-weighted MR images using a Markov random field (MRF) model that incorporated the hippocampal anatomy as a prior. The MRF model was first applied to label each voxel in the image volume as gray matter, white matter, CSF, or subcortical structures [6]. Due to the lack of constraints on structural shapes, this process introduced irregularities and topological errors (e.g. holes) at the hippocampal boundary. This increases shape variation and thus reduces statistical power to detect group differences. To avoid this issue, we generated the hippocampal shapes of each individual subject with the properties of smoothness and correct topology by injecting a template shape into them using the LDDMM-image mapping algorithm [17]. The hippocampal template shape was created from 41 manually labeled hippocampi via a large deformation diffeomorphic template generation algorithm [18]. The surface representation of the hippocampal shape was created by composing the diffeomorphic map on the template surface.

To compute the distance matrix of the full diffeomorphic metric, we applied $\binom{196}{2}$ LDDMM-surface mappings to find the optimal initial momentum between any two hippocampal surface shapes [28]. For the first- and second-order approximations of this distance matrix, we chose an average surface as the template

Table 1. Computational cost of the distance matrix using the full metric and its first- and second-order approximations

Computational cost	Full metric ρ	1 st order $\rho^{1^{st}}$	2 nd order $\rho^{2^{nd}}$
Number of CPUs	40	40	40
Number of mappings	$\binom{196}{2} = 19110$	196	196
Time cost	6 days	1 hour	1 hour

[18] and applied 196 LDDMM-surface mappings to align the template to each individual hippocampal surfaces. Then the distance matrices were estimated using Eqs. (12) and (13), respectively. Table 1 lists the computational cost of the distance matrix using the full metric and its first- and second-order approximations. Fig. 1 shows the scatter plots of the full metrics obtained from $\binom{196}{2}$ LDDMM-surface mappings against their first- and second-order approximations. Visually, both the first- and second-order approximations are consistent with the full metrics. Pearson’s correlation analysis showed significant correlations of the full metric matrix with its first- and second-order approximations (1st-order: $r = 0.975, p < 0.0001$; 2nd-order: $r = 0.972, p < 0.0001$).

To understand aging processes of the hippocampus, we assumed that the hippocampal shape variations due to aging can be captured by a low-dimensional space such that the relationship of the hippocampal shape with age is easily identified. We thus employed ISOMAP [21] to construct an embedding in a low-dimensional Euclidean space that preserves the pair-wise geodesic distances of hippocampal shapes in the diffeomorphic metric shape space. Fig. 2 shows the 196 hippocampal shapes in the first two dimensions of ISOMAP embeddings when the full metric, its first- and second-order approximations were used in ISOMAP, respectively. Visually, these embeddings from the full metric and its first- and second-order approximations are very similar. We quantified the similarity of the first two dimensions between these embeddings as follows. The embeddings of the 196 subjects obtained based on the full metric, and its first- and second-order approximations are denoted as $Y^{full} = [y_1^{full}, y_2^{full}, \dots, y_{196}^{full}]$, $Y^{1^{st}} = [y_1^{1^{st}}, y_2^{1^{st}}, \dots, y_{196}^{1^{st}}]$, $Y^{2^{nd}} = [y_1^{2^{nd}}, y_2^{2^{nd}}, \dots, y_{196}^{2^{nd}}]$, respectively. We define

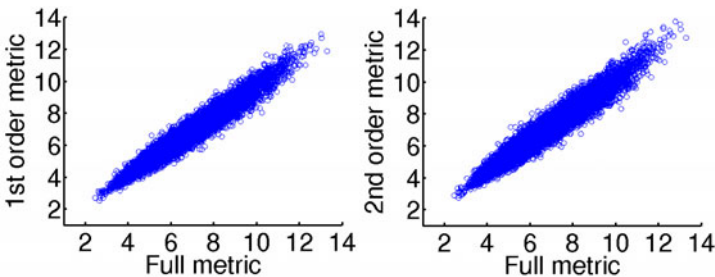


Fig. 1. Scatter plots of the full metrics against their first-order (left) and second-order (right) approximations

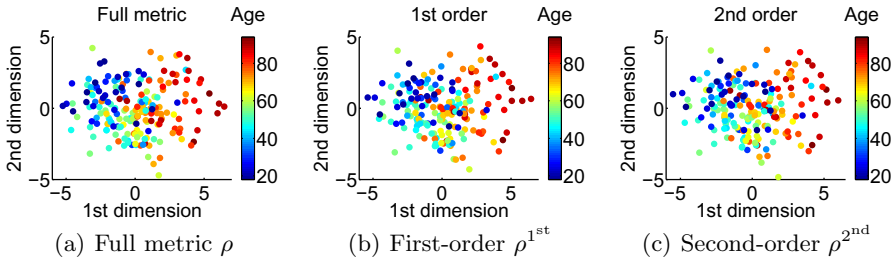


Fig. 2. Embedding obtained by ISOMAP. Figs. 2(a)-2(c) illustrate the representation of the 196 hippocampal shapes in the first two embedding dimensions computed from ISOMAP when the full metric, its first-order and second-order approximations are used, respectively.

Table 2. Correlation coefficients between age and each dimension of the ISOMAP embeddings obtained when the full metric, and its first- and second-order approximations were used, respectively

ISOMAP embedding	Full metric	1 st order $\rho^{1^{st}}$	2 nd order $\rho^{2^{nd}}$
1st dimension	0.65	0.6	0.59
2nd dimension	0.1	0.04	0.05

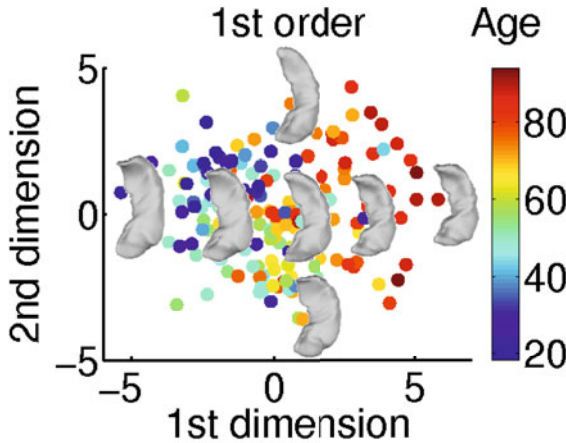


Fig. 3. Visualization of the hippocampal shapes in the ISOMAP embedding space generated based on the first-order approximation of the diffeomorphic metric. Color encodes the subject’s age information.

the difference between two embeddings Y^j, Y^k as $\epsilon(Y^j, Y^k) = \sum_{i=1}^{196} \frac{\|y_i^j - y_i^k\|^2}{196}$. We found that $\epsilon(Y^{full}, Y^{1^{st}})$ and $\epsilon(Y^{full}, Y^{2^{nd}})$ are respectively 0.21 and 0.29, which are close to 0.

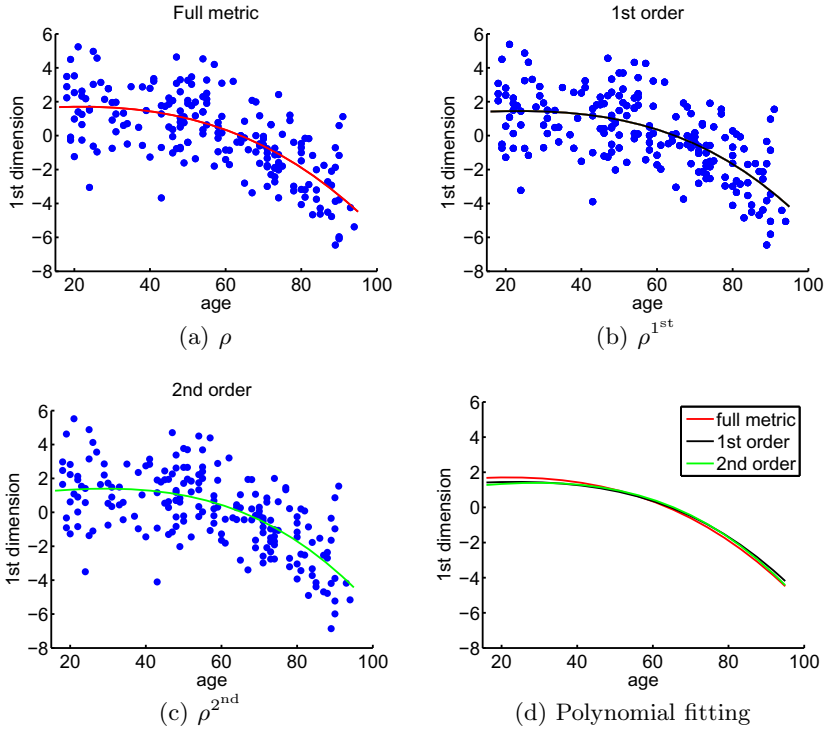


Fig. 4. The relation of age with the first dimension of the ISOMAP embedding. Figs. 4(a), 4(c) show the relationships obtained using the full metric, the first- and second-order approximations, respectively. The third-order polynomial fitting curve is shown in each panel. Fig. 4(d) illustrates the closeness of the polynomial fittings when the full metric (red), the first- (black) and second-order (green) approximations were used in ISOMAP.

Table 2 shows the results of Pearson’s correlation analysis between age and each dimension of the ISOMAP embedding. The first embedding dimension of the hippocampal shapes is the most correlated with age regardless whether the full metric or its approximations were used. Fig. 3 illustrates that the first dimension of the ISOMAP embedding encodes the hippocampal shape compression in the head-tail and lateral-medial directions as age increases. Such shape variations associated to the nonlinear aging process are exhibited regardless whether the full metric or its approximations were used in ISOMAP. We performed curve fitting using polynomial functions on the first dimension of LLDME embedding and ages of all 196 healthy subjects. We found that a polynomial function of degree 3 gives the best model fit for all three metrics. Fig. 4(d) illustrates the closeness of the nonlinear fittings when the full metric, and its approximations were used. This nonlinear relationship of the hippocampal morphometry with age has rarely been discovered through the hippocampal volume in cross-sectional studies even though [11] suggested two piece-wise linear functions to fit the

relationship between age and hippocampal volume with the acceleration of the hippocampal volume loss at age around 50s and 70s. However, our result is supported by [12], where the whole brain morphological feature was used to classify subjects aged above 50 into different aging groups. Thus, the brain morphology (e.g., volumes or shapes) may follow different trajectories at different age ranges. The hippocampal shape represented by the first ISOMAP dimension reveals such nonlinear age-related processes.

4 Conclusion

In this paper, we derive the first- and second-order approximations of the diffeomorphic metric in a shape space and demonstrate that both these approximations are close to the full metric. These approximations significantly speed up the computation of the pairwise distance matrix and thus facilitate a variety of shape learning methods, such as ISOMAP, where a distance matrix is needed. We show the low-dimensional embeddings of the hippocampal shapes obtained from ISOMAP based on the approximated distance matrix (the first-order and the second-order) are similar to that obtained based on the full metric. We notice that there are limitations on the proposed approach. First of all, the distance matrix computed based on the approximated diffeomorphic metric is dependent on the template used in LDDMM, while the distance matrix constructed based on the full metric is template free. In this paper, we did not discuss the potential influence of the template on the metric approximations. Instead, we used the average template and our results support that the metric approximation computed based on the average template is very much in line with the full metric. Second, we would expect that the second-order approximation of the diffeomorphic metric should be closer to the full metric than the first-order approximation. However, our results (Fig. 1) did not support this expectation. This is possibly due to the fact that the age-related variations of the hippocampal shapes can be sufficiently characterized by the first order of the diffeomorphic metric.

Acknowledgments. Work supported by grants A*STAR SERC 082-101-0025, A*STAR SICS-09/1/1/001, a center grant from the National Medical Research Council (NMRC/CG/NUHS/2010), the Young Investigator Award at National University of Singapore (NUSYIA FY10 P07), and National University of Singapore MOE AcRF Tier 1.

References

1. Apostolova, L.G., Dinov, I.D., Dutton, R.A., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M.: 3D comparison of hippocampal atrophy in amnesic mild cognitive impairment and alzheimer's disease. *Brain* 129, 2867–2873 (2006)
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113 (2007)
3. Avants, B., Gee, J.C.: Geodesic estimation for large deformation anatomical shape and intensity averaging. *NeuroImage* 23, 139–150 (2004)

4. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. *Quart. App. Math.* 56, 587–600 (1998)
5. van Essen, D.: A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage* 28, 635–662 (2005)
6. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355 (2002)
7. 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)
8. Glaunès, J., Qiu, A., Miller, M., Younes, L.: Large deformation diffeomorphic metric curve mapping. *IJCV* 80(3), 317–336 (2008)
9. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. *Quart. App. Math.* 56(4), 617–694 (1998)
10. Guimond, A., Roche, A., Ayache, N., Meunier, J.: Three-Dimensional Multimodal Brain Warping Using the Demons Algorithm and Adaptive Intensity Corrections. *IEEE TMI* 20(1), 58–69 (2001)
11. Jernigan, T.L., Archibald, S.L., Fennema-Notestine, C., Gamst, A.C., Stout, J.C., Bonner, J., Hesselink, J.R.: Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiol Aging* 22(4), 581–594 (2001)
12. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S.M., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage* 21(1), 46–57 (2004)
13. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cog. Neurosci.* 19, 1498–1507 (2007)
14. Miller, M.I., Qiu, A.: The emerging discipline of computational functional anatomy. *Neuroimage* 45, S16–S39 (2009)
15. Miller, M.I., Priebe, C.E., Qiu, A., Fischl, B., Kolasny, A., Brown, T., Park, Y., Ratnanather, J.T., Busa, E., Jovicich, J., Yu, P., Dickerson, B.C., Buckner, R.L.: Collaborative computational anatomy: An MRI morphometry study of the human brain via diffeomorphic metric mapping. *Human Brain Mapping* 30(7), 2132–2141 (2009)
16. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. *JMIV* 24, 209–228 (2006)
17. Qiu, A., Miller, M.I.: Multi-structure network shape analysis via normal surface momentum maps. *NeuroImage* 42(4), 1430–1438 (2008)
18. Qiu, A., Brown, T., Fischl, B., Ma, J., Miller, M.I.: Atlas generation for subcortical and ventricular structures with its applications in shape analysis. *IEEE TIP* 19(6), 1539–1547 (2010)
19. Qiu, A., Fennema-Notestine, C., Dale, A.M., Miller, M.I.: Regional shape abnormalities in mild cognitive impairment and Alzheimer’s disease. *Neuroimage* 45, 656–661 (2009)
20. Robbins, S., Evans, A.C., Collins, D.L., Whitesides, S.: Tuning and comparing spatial normalization methods. *Med. Image Anal.* 8(3), 311–323 (2004)
21. Tenenbaum, J.B., Silva, V.d., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290(5500), 2319–2323 (2000)
22. Thompson, P.M., Schwartz, C., Lin, R.T., Khan, A.A., Toga, A.W.: Three-dimensional statistical analysis of sulcal variability in the human brain. *J. Neurosci.* 16(13), 4261–4274 (1996)

23. Trouvé, A.: Diffeomorphism groups and pattern matching in image analysis. *IJCV* 28(3), 213–221 (1998)
24. Vaillant, M., Qiu, A., Glaunès, J., Miller, M.I.: Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. *NeuroImage* 34, 1149–1159 (2007)
25. Wojtyński, W.: One parameter subgroups and B-C-H formula. *Stud. Math.* 111, 163–185 (1994)
26. Yang, X., Goh, A., Qiu, A.: Locally Linear Diffeomorphic Metric Embedding (LLDME) for surface-based anatomical shape modeling. *NeuroImage* 56(1), 149–161 (2011)
27. Younes, L.: *Shapes and Diffeomorphism*. Springer, Heidelberg (2010)
28. Zhong, J., Qiu, A.: Multi-manifold diffeomorphic metric mapping for aligning cortical hemispheric surfaces. *Neuroimage* 49, 355–365 (2010)

Anisotropic Diffusion of Tensor Fields for Fold Shape Analysis on Surfaces

Maxime Boucher^{1,2,*}, Alan Evans², and Kaleem Siddiqi¹

¹ School of Computer Science and Centre for Intelligent Machines, McGill University

² McConnell Brain Imaging Center, Montreal Neurological Institute,
McGill University
boucher@bic.mni.mcgill.ca

Abstract. The folding pattern of the human cortical surface is organized in a coherent set of troughs and ridges, which mark important anatomical demarcations that are similar across subjects. Cortical surface shape is often analyzed in the literature using isotropic diffusion, a strategy that is questionable because many anatomical regions are known to follow the direction of folds. This paper introduces anisotropic diffusion kernels to follow neighboring fold directions on surfaces, extending recent literature on enhancing curve-like patterns in images. A second contribution is to map deformations that affect sulcal length, i.e., are parallel to neighboring folds, with other deformations that affect sulcal length, within the diffusion process. Using the proposed method, we demonstrate anisotropic shape differences of the cortical surface associated with aging in a database of 95 healthy subjects, such as a contraction of the cingulate sulcus, shorter gyri in the temporal lobe and a contraction in the frontal lobe.

Keywords: Surface Morphometry, Statistical Shape Analysis.

1 Introduction

The folding pattern of the human cortical surface results in a coherent set of troughs and ridges, which to some extent are similar across subjects. Folds can mark important anatomical boundaries and they vary in width and length across individuals. Differences in fold shape can be associated with aging, gender, disease and underlying white matter connectivity.

Several methods exist to analyze cortical surface fold shape. For example, one approach is to compute the average shape of a set of surfaces using an iterative registration process, following which the mechanical deformation field required to deform an individual surface onto the mean surface is used as a statistical shape model [4,9,11,12]. This approach faces a statistical estimation problem: the effect of a factor on the shape of the underlying cortical surface needs to be estimated in the presence of statistical noise. The presence of noise reduces the confidence that the estimated effect of a factor is different from

* Corresponding author.

zero and reduces the detection power of the statistical framework. One way to mitigate this problem is to locally average the shape measurements using surface diffusion kernels [4]. However, isotropic diffusion averages measurements across neighboring neuroanatomic regions, which is generally not desirable.

In this paper we propose to use anisotropic diffusion on surfaces to highlight patterns that occur on specific folds, an example of which is shown in Figure 1. We use the approach of surface-based morphometry for statistical analysis, where each individual cortical surface is viewed as a deformed version of a template surface. We use a simplicial mesh as a representation to find the correspondence between vertices on the template surface and each of the individual surfaces via a surface registration process. We then carry out anisotropic diffusion in tensor-based morphometry and statistical analysis on the deformation tensor of that mapping, which is the 2-dimensional version of the one used in [9].

There are two main contributions in this paper. The first is a method to anisotropically diffuse a scalar field on a surface, such that the diffusion speed is faster along neighboring fold orientations, as shown in Figure 1. This differs from other anisotropic diffusion schemes that are designed primarily to suppress diffusion across creases [5]. The second contribution is related to tensor diffusion on surfaces. The added complexity is that tensors have a directional component, which needs to be taken into consideration. We propose to keep tensors aligned with neighboring folds within the diffusion process. This requires us to rotate the tensor field within the diffusion scheme according to how the neighboring folds change in orientation. The method we develop is similar to the one proposed in [8] for enhancing curve like patterns in images.

The paper is organized as follows. We first describe a general setting to carry out tensor-based morphometry on surfaces, in Section 2.1. We then show how orientation estimates can be used to generate anisotropic filters on surfaces in Section 2.3. The method to rotate tensors according to neighboring fold orientation is developed in Section 2.4. Finally, in Section 3 we present experimental results on cross-sectional data that demonstrate that the anisotropic diffusion scheme is able to follow the orientation of neighboring folds, and that the method

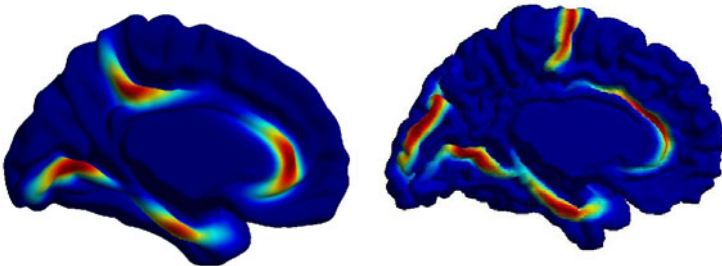


Fig. 1. We illustrate our anisotropic diffusion process on two examples, a template surface (left) and the mid-cortical surface of a subject (right). In both examples seed regions are placed and the color maps indicate how the diffusion follows the shape of folds.

offers increased sensitivity in detecting shape differences in test studies. Using the proposed coherence-enhancing filters, we analyze how the shape of cortical folds is correlated with aging, in a database of 95 healthy subjects. We are able to find significant shape differences of the cortical surface associated with aging, which are strongly localized along specific folds. Such results would not be possible using isotropic diffusion kernels.

2 Methods

We first describe a general setting to carry out morphometric analysis on surfaces in Sections 2.1 and 2.2. We then develop anisotropic filters on surfaces in Section 2.3. Finally Section 2.4 combines these methods for estimating fold orientation.

We begin by introducing some notation. In this paper, all operations are performed on a template surface $\bar{\mathcal{M}}$ with a Riemannian metric \mathbf{g} . The surface $\bar{\mathcal{M}}$ represents the template shape of one hemisphere of the cerebral cortex. The Riemannian metric used is therefore the one induced by the imbedding in \mathbb{R}^3 .

Bold face letters are used for vector fields and tensor fields, such as $\mathbf{u}, \mathbf{v} \in \mathcal{T}\bar{\mathcal{M}}$, where $\mathcal{T}\bar{\mathcal{M}}$ is the tangent bundle of $\bar{\mathcal{M}}$. We denote the inner product according to the Riemannian metric as $\mathbf{g}(\mathbf{u}, \mathbf{v})$. In order to keep the notation simple, we use a global coordinate frame formed of two tangent vector fields, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{T}\bar{\mathcal{M}}$. We use lower indices for vector fields and upper indices for covector fields, such that \mathbf{u} can be written in terms of the $[\mathbf{x}_1, \mathbf{x}_2]$ coordinate frame as

$$\mathbf{u} = u^1 \mathbf{x}_1 + u^2 \mathbf{x}_2. \tag{1}$$

Upper case letters are used for second order tensor fields (e.g.: \mathbf{T}), with lower case letters such as \mathbf{u}, \mathbf{v} for vector fields. The only exception to the upper/lower case convention for tensor fields is the Riemannian metric where \mathbf{g} is used.

Using the $[\mathbf{x}_1, \mathbf{x}_2]$ frame, we write \mathbf{g} the metric tensor expressed in this basis and g_{ij} , the i, j component of the Riemannian metric. This allows us to define a point-wise dot-product between the tangent vector fields \mathbf{u}, \mathbf{v} as

$$\mathbf{g}(\mathbf{u}, \mathbf{v}) = [u^1, u^2] \mathbf{g} \begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = \sum_{i,j \in \{1,2\}} g_{ij} u^i v^j. \tag{2}$$

We use \mathbf{g}^{-1} for the inverse of the metric tensor and g^{ij} the i, j component of this inverse. The L^2 norm of a second-order tensor field $\mathbf{T} \in \mathcal{T}\bar{\mathcal{M}} \times \mathcal{T}\bar{\mathcal{M}}$ is expressed as

$$\|\mathbf{T}\|_{\mathbf{g}}^2 = \int_{\bar{\mathcal{M}}} \|\mathbf{T}\|_{\mathbf{g}_p}^2 d\bar{\mathcal{M}} = \sum_{i,j,k,l \in \{1,2\}} \int_{\bar{\mathcal{M}}} g_{ik} g_{jl} T^{ij} T^{kl} d\bar{\mathcal{M}}, \tag{3}$$

where the notation $\|\mathbf{T}\|_{\mathbf{g}_p}^2$ represents the l^2 norm at a point $\mathbf{p} \in \bar{\mathcal{M}}$. Where it is convenient to do so, we will denote a multiplication between tensors, say

\mathbf{T}, \mathbf{S} , by writing the tensors next two each other and including a metric term if necessary. For example, a multiplication between two tensors is given by

$$\mathbf{TgS} = \sum_{j,k \in \{1,2\}} T^{ij} g_{jk} S^{kl}. \quad (4)$$

2.1 Morphometric Analysis on Surfaces

We now review the framework of tensor-based morphometry used for shape analysis [9]. The shape of each hemisphere of the human cerebral cortex is represented as a single surface embedded in Euclidean space \mathbb{R}^3 . Let $\mathcal{M}_s \subset \mathbb{R}^3, k = 1, \dots, N$ be surfaces representing the cerebral cortices of N individuals, one surface per hemisphere. We suppose that these surfaces are homeomorphic to a sphere.

The surfaces are mapped onto a common template $\bar{\mathcal{M}}$ through a process known as surface registration [7]. Surface registration produces a diffeomorphic map $\psi_s : \bar{\mathcal{M}} \rightarrow \mathcal{M}_s$ that matches the features of $\bar{\mathcal{M}}$ onto the corresponding features on \mathcal{M}_s . The amount of deformation required to map $\bar{\mathcal{M}}$ to \mathcal{M}_s is used to analyze the shape of individual surfaces. This amount is computable from the diffeomorphic map ψ_s . We obtain a tensor [9] representing the amount of deformation required to morph \mathcal{M}_s into $\bar{\mathcal{M}}$ as

$$\mathbf{T}_s = [\text{grad}(\psi_s)]^t [\text{grad}(\psi_s)]. \quad (5)$$

In this case, \mathbf{T}_s forms a 2x2 positive symmetric definite tensor field on the surface. A multivariate statistical test is then applied on the tensor field \mathbf{T}_s in order to see if there is any significant correlation with an exterior factor. The method for statistical analysis is described in [6,9] and we provide an example result in Fig. 5. Unless we refer to a specific subject, the $_s$ index will be dropped and we will use the notation \mathbf{T} for a tensor field on $\bar{\mathcal{M}}$ in the rest of this paper.

2.2 Diffusion of Tensor Fields

In this section, we present a method for diffusion of tensor fields on surfaces. The tensor laplacian, in terms of the vector fields $\mathbf{x}_1, \mathbf{x}_2$ is given as

$$\Delta \mathbf{T} = \sum_{i,j \in \{1,2\}} \nabla_{\mathbf{x}_i} g^{ij} \nabla_{\mathbf{x}_j} \mathbf{T}, \quad (6)$$

where ∇ is the covariant derivative. Equation 6 allows to perform diffusion on surfaces as

$$\frac{\partial \mathbf{T}}{\partial t} = \Delta \mathbf{T}. \quad (7)$$

2.3 Anisotropic Diffusion of Scalar Fields along Folds

The anisotropic filter for surface shape analysis we propose is designed to follow the orientation of neighboring crests and troughs, as shown in Figure 11. The

general anisotropic diffusion equation used to build these filters for a scalar field $f : \bar{\mathcal{M}} \rightarrow \mathbb{R}$ is given as [10]:

$$\frac{\partial f}{\partial t} = \sum_{i,j,k \in \{1,2\}} \frac{1}{\sqrt{|\mathbf{g}|}} \partial_{\mathbf{x}_i} \sqrt{|\mathbf{g}|} A^i_j g^{jk} \partial_{\mathbf{x}_k} f, \tag{8}$$

where $\mathbf{A} : \mathcal{T}\bar{\mathcal{M}} \rightarrow \mathcal{T}\bar{\mathcal{M}}$ is a symmetric positive definite operator which controls diffusion speed on $\bar{\mathcal{M}}$, and $|\mathbf{g}|$ is the determinant of \mathbf{g} .

The most direct method by which \mathbf{A} can be adapted to follow curve patterns on $\bar{\mathcal{M}}$ is to estimate the local orientation of the folding pattern. Let \mathbf{v}_θ be an estimate of the local fold orientation ($\|\mathbf{v}_\theta\|_{\mathbf{g}_p} = 1$) on $\bar{\mathcal{M}}$. Then \mathbf{A} can be chosen as

$$\mathbf{A} = \alpha \mathbf{I} + (1 - \alpha) ((\mathbf{v}_\theta \otimes \mathbf{v}_\theta) \mathbf{g}), \tag{9}$$

where $\alpha \in]0, 1]$ is the anisotropic factor, \mathbf{I} is the identity operator, \otimes is a tensor product. The challenge is to select \mathbf{v}_θ such that it is a good estimate of the local folding pattern orientation.

The principal curvature directions on a surface provide a first-hand estimate of the local orientation of its folds. The curvature of the surface is low in the direction parallel to the fold, and is high perpendicular to fold orientation. The principal curvatures are measured by the shape operator. Let \mathbf{n} be the unit surface normal. The shape operator \mathbf{C} of a surface is defined as the covariant derivative of the normal \mathbf{n} along \mathbf{u} as

$$\mathbf{C}\mathbf{u} = -\nabla_{\mathbf{u}}\mathbf{n}, \tag{10}$$

where ∇ is the covariant derivative in \mathbb{R}^3 . The covariant derivative of the normal is a tangent vector field on $\bar{\mathcal{M}}$. To focus on the magnitude of the principal curvatures, we use the absolute value of the curvature. Let, the following definition of \mathbf{S} be given as the non-negative symmetric tensor such that, if $\mathbf{v}_1, \mathbf{v}_2$ are the two eigenvectors of \mathbf{C} with eigenvalue λ_1, λ_2 . Then, \mathbf{S} is defined as

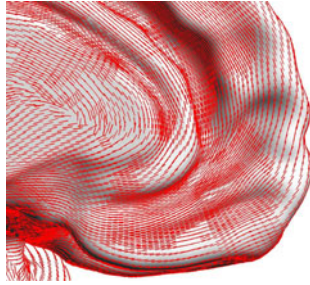
$$\mathbf{S} = |\lambda_1| \mathbf{v}_1 \otimes \mathbf{v}_1 + |\lambda_2| \mathbf{v}_2 \otimes \mathbf{v}_2. \tag{11}$$

In other words, the tensor \mathbf{S} is the tensor whose eigenvalues are $|\lambda_1|, |\lambda_2|$. This allows us to define an estimator of the local orientation \mathbf{v}_θ of neighboring folds as the eigenvector of \mathbf{S} with smallest eigenvalue, i.e.

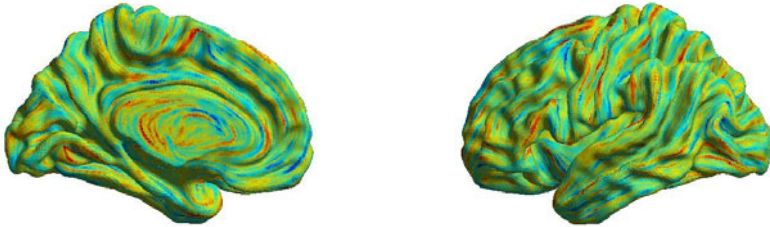
$$\mathbf{v}_\theta = \operatorname{argmin}_{\{\mathbf{v}_\theta \mid \|\mathbf{v}'_\theta\|_{\mathbf{g}_p} = 1\}} \mathbf{g}(\mathbf{v}'_\theta, \mathbf{S}\mathbf{g}\mathbf{v}'_\theta). \tag{12}$$

However, between creases where curvature is small, the principal directions of \mathbf{S} may not accurately reflect neighboring fold directions. Hence, we first apply a small amount of regularization to the tensor \mathbf{S} using tensor diffusion as expressed by Equation 7. Let, $\mathbf{S}(t)$ be the result of applying diffusion on \mathbf{S} for a time t according to Equation 7. We then apply Equation 12 to obtain the fold direction that is perpendicular to the direction of highest curvature.

Typically, we chose an anisotropic factor α of 0.1 for diffusion and a diffusion time of 100 to estimate local orientation. A few results of scalar field diffusion



(a) Generation of a vector field \mathbf{v}_θ using the normal curvature of neighboring folds. The vector field shown is parallel to the folds on the patch.



(b) Medial view of the Left Hemisphere (c) Exterior view of the Left Hemisphere

Fig. 2. A visualization of orientation fields on the template surface using line integral convolution. In a qualitative sense, the vector field shown follows the major folds on the cortical surface.

are shown in Figure 1 and Figure 2(a). It should be noted that the matrix field \mathbf{A} defined in this section will contain discontinuities, corresponding to the discontinuities of the vector field \mathbf{v}_θ . In order for the anisotropic diffusion process to be well defined, a very small amount of regularization is also applied to \mathbf{A} , again by using isotropic diffusion (Equation 7).

2.4 Anisotropic Diffusion of Tensor Fields on Surfaces

We now consider diffusion of tensor fields on surfaces. Recall that Equation 7 provides a diffusion process for tensors. The center column of Figure 3 illustrates that the basic diffusion equation provided in Equation 7 does not result in a diffusion scheme that follows local fold orientation. This is a problem for our application where we intend to study shape differences that affect fold shape. In particular, we wish to align the diffusion process using the orientation of neighboring folds.

To see how tensors can be realigned according to neighboring fold orientation, consider the following local rotation process. Let \mathbf{R}_θ be a rotation operator, such that it preserves the local orientation on surfaces and does not change the inner product between two vector fields, i.e. $\mathbf{g}(\mathbf{R}_\theta \mathbf{u}, \mathbf{R}_\theta \mathbf{v}) = \mathbf{g}(\mathbf{u}, \mathbf{v})$. Using matrix notation, a rotated frame $\mathbf{t}_1, \mathbf{t}_2$ can be written in term of the original frame as

$$\begin{bmatrix} \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix} = \mathbf{R}_\theta \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \mathbf{g}^{-\frac{1}{2}} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \mathbf{g}^{\frac{1}{2}} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}. \tag{13}$$

The technique that we propose in this paper is to compute an optimal rotation operator \mathbf{R}_θ such that it best preserves the tensor field \mathbf{S} . Then, the same rotation operator is applied to the deformation tensors computed in Equation 5.

We will now define a diffusion scheme for tensors in a rotated coordinate system. Consider a rotated differential operator as

$$\begin{aligned} \mathbf{R}_\theta^{-1} \nabla_{\mathbf{x}_k} (\mathbf{R}_\theta \mathbf{S} \mathbf{R}_\theta^t) \mathbf{R}_\theta^{t-1} &= \nabla_{\mathbf{x}_k} \mathbf{S} + v_k^\theta \left(\mathbf{g}^{-\frac{1}{2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{g}^{\frac{1}{2}} \mathbf{S} + \mathbf{S} \mathbf{g}^{\frac{1}{2}} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \mathbf{g}^{-\frac{1}{2}} \right) \\ &= \nabla_{\mathbf{x}_k} \mathbf{S} + v_k^\theta DS, \end{aligned} \tag{14}$$

where DS is defined as the term in parentheses and v_k^θ designates a parameter that needs to be estimated. Equation 14 allows us to define a rotated anisotropic diffusion scheme for tensors as

$$\frac{\partial \mathbf{S}}{\partial t} = \sum_{i,j,k \in \{1,2\}} \nabla_{\mathbf{x}_i} A^i_j g^{jk} (\nabla_{\mathbf{x}_k} \mathbf{S} + v_k^\theta DS). \tag{15}$$

Let $\mathbf{w} = [w_1, w_2]$ be an estimator for v_k^θ . We optimize the value of \mathbf{w} , by minimizing the norm of the differential of \mathbf{S} over the entire diffusion process as

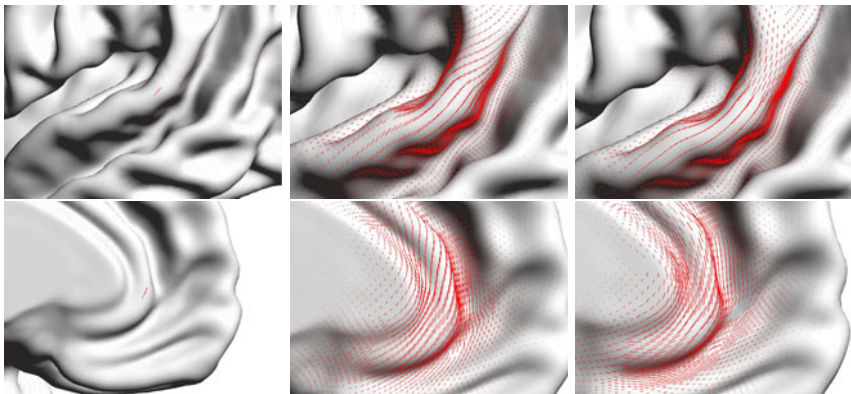


Fig. 3. An illustration of aligning tensors with neighboring folds while performing diffusion. An anisotropic second-order tensor (shown on the left) was diffused on a template surface (top: left superior temporal gyros, bottom: left rostral cingulate gyrus). The center needle map shows the direction of the highest eigenvalue of the diffused tensor. The diffusion results in a tensor field that does not maintain a constant orientation with nearby geometric folds. The right needle map shows how the method presented in Section 2.4 causes the diffused tensors to be aligned with nearby folds.

$$w_k = \operatorname{argmin}_{w'_k} \int_{[0,t]} \|\partial_{\mathbf{x}_k} \mathbf{S}(z) + DS(z)w'_k\|_{\mathbf{g}_p}^2 dz. \quad (16)$$

Using calculus of variations, we obtain that the solution \mathbf{w} of Equation 16 is

$$w_k \int_{[0,t]} \|DS(z)\|_{\mathbf{g}_p}^2 dz = - \int_{[0,t]} \operatorname{trace}(\partial_{\mathbf{x}_k} \mathbf{S}(z) \mathbf{g} DS(z) \mathbf{g}) dz, \quad (17)$$

where *trace* is the trace of the tensor. An illustration of performing diffusion while rotating tensors according to Equation 17 is provided in Figure 3. The left column shows a tensor field which is equal to zero everywhere, except at one location, where it equals $\mathbf{v}_\theta \otimes \mathbf{v}_\theta$. The result of diffusing this tensor field without computing an optimal rotation field is shown in the center column and the result of diffusion with the optimal rotation field is shown in the right column. The rotation causes the diffused tensor field to follow the orientation of neighboring folds.

3 Experimental Results

We now provide additional illustrations and promising applications of the proposed diffusion process. We begin by evaluating the method on the cortical surfaces of individual subjects. Here our goal is to select a few major folds on the cortical surface (such as the central sulcus, hippocampal gyrus and superior temporal sulcus), and to determine if the anisotropic diffusion process follows these folds. Each fold was first manually identified on the template surface $\bar{\mathcal{M}}$. Then a geodesic distance function was used to measure the distance to the fold on the template surface, as shown in Figure 4(a). The level sets of the geodesic distance function provide an approximation to the direction the diffusion should follow in order to remain parallel to the segmented fold.

We used surface registration to map the geodesic distance functions computed on the template surface onto 15 cortical surfaces (left hemispheres) of individual subjects. Let $\bar{\rho}$ be the geodesic distance function on the template surface and $\rho = \bar{\rho}(\psi^{-1})$ the mapped geodesic distance function on a specific subject (see Fig. 4(a)). Let $e^{t\Delta_\alpha}$ be the solution to the anisotropic diffusion given in Equation 8 with $\alpha = 0.1$. We first computed a diffused value for the geodesic distance function as $\mu = e^{t\Delta_\alpha} \rho$, where $e^{t\Delta_\alpha}$ is the result of applying diffusion for a time t according to Equation 8. Then, we determined the extent to which diffusion follows level sets of the geodesic distance function by measuring how it varies compared to μ as

$$E(\rho) = \sqrt{e^{t\Delta_\alpha} (\rho - \mu)^2}. \quad (18)$$

In the event where diffusion perfectly follows the level set curves of ρ , $E(\rho)$ is almost zero. We computed the diffusion until the kernel covered a region of about 4 cm on the cerebral cortex along the longest axis. We display the result for the three sulci in the second row of Figure 4. As a point of comparison, we used an

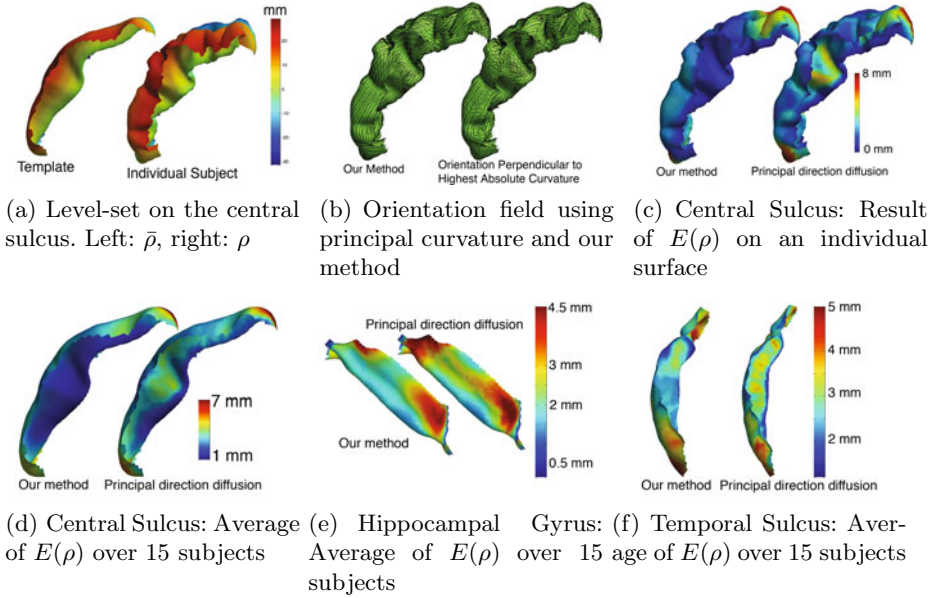


Fig. 4. Anisotropic diffusion is able to follow the level sets of a geodesic distance function to a segmented fold on individual cortical surfaces, as measured by $E(\rho)$. See the text for a discussion.

anisotropic diffusion scheme where the diffusion was constrained to the direction of the smaller principal curvature, as measured by $\mathbf{S}(0)$ (see the example on the left of Figure 4(b)). The result of this anisotropic kernel is shown in Figure 4 under the label “principal direction diffusion”.

In the second row of Figure 4, we show that the value of $E(\rho)$ is much smaller on average in our method as compared with principal direction diffusion. The orientation field produced by our method is more parallel to the major fold and is not influenced by the small changes in orientation along the sulcal wall, as opposed to that obtained using principal direction diffusion (see Fig. 4(b)). Also, most of the failure cases are due to the fact that the geodesic distance function is based only on the segmented fold, whereas our method captures neighboring fold orientation as well.

In a second experiment we used anisotropic diffusion to reveal significant shape differences in the left hemisphere within a population of 95 adults aged 19 to 85 years old, with a median age of 44 years, selected from the ICBM database. For illustration purposes, only the left hemisphere was used. We used multivariate tensor-based T -statistics on this population to compare anisotropic and isotropic diffusion [6].

To compute regression statistics, first a tensor field for each subjects was computed according to Equation 5. Then, each tensor field was diffused according to Equation 15 with parameters $\alpha = 0.1$ and for a time $t = 400$ (mm^2). Let $\hat{\mathbf{T}}_s$

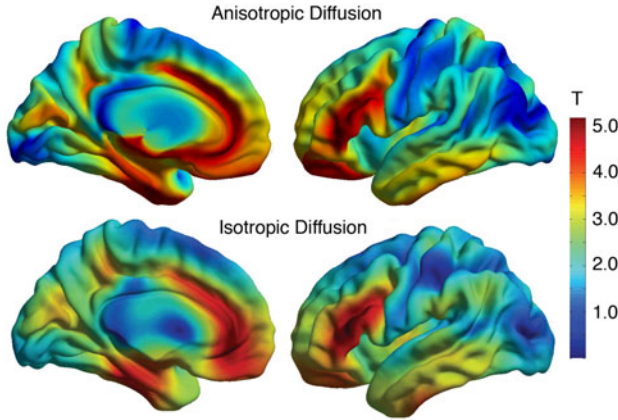


Fig. 5. Multivariate tensor statistics using isotropic and anisotropic diffusion. Anisotropic diffusion shows peak regions that are well localized around specific folds, whereas peaks for isotropic diffusion spread across multiple folds.

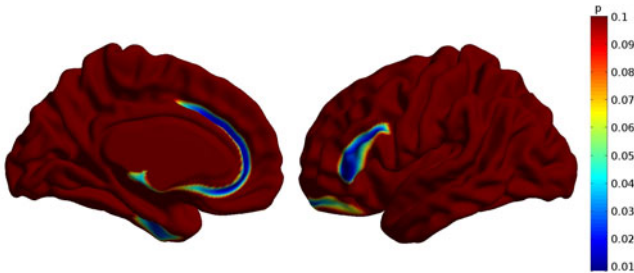


Fig. 6. Probability maps of multivariate tensor statistics using anisotropic diffusion. Statistically significant regions are well localized around specific folds.

be the resulting tensor field after diffusion. The following linear model was used to compute the regression with age and gender:

$$\text{Log}(\hat{\mathbf{T}}_s) = \hat{\mathbf{M}} + \hat{\mathbf{T}}_{age}age_s + \hat{\mathbf{T}}_{gender}gender_s + \mathbf{E}_s, \tag{19}$$

where Log is the tensor logarithm using the Log-Euclidean model of [11], age_s and $gender_s$ are the age and gender of subject s , and \mathbf{E}_s is the residual error of the linear model. A least-squares fit was used to find the regression tensor field corresponding to the mean $\hat{\mathbf{M}}$, age $\hat{\mathbf{T}}_{age}$ and gender $\hat{\mathbf{T}}_{gender}$. A multivariate T test for age correlation was then computed at each vertex using SurfStat [11].

We first provide a comparison between isotropic diffusion and anisotropic diffusion in Figure 5 in terms of the resulting T -test. Although a fair comparison is difficult because there are many diffusion parameters to be selected, Figure 5 demonstrates that results using anisotropic diffusion are well localized around specific folds. By this, we mean that regions where T is high do not spread over

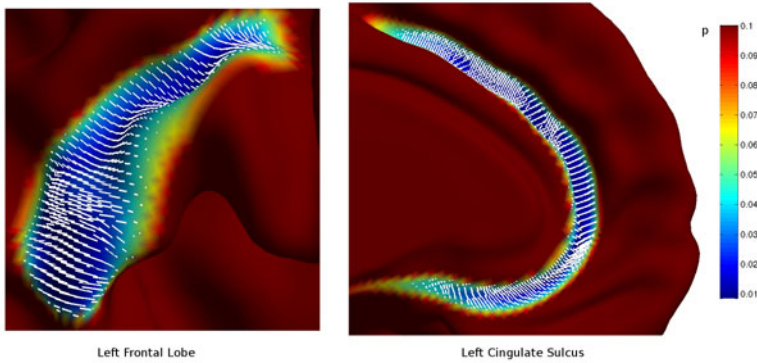


Fig. 7. A close up view of the maximum negative correlation with aging in the Cingulate sulcus and Frontal lobe. The white lines show the direction which indicates maximum negative correlation associated with aging.

multiple folds with anisotropic diffusion whereas regions where T is high with isotropic diffusion spread across several folds and also produce lower peak values for T .

We computed thresholds for significance to control for false positives using a permutation test with 10000 permutations. The threshold for significance ($p < 0.05$) using multivariate tensor-based morphometry was $|T| > 4.752$. We display significance results in Figure 6. These results indeed show that anisotropic diffusion is able to find regions of significant atrophy localized to a single fold.

We also tried to determine which direction showed the highest correlation with the subject's age and whether this correlation was positive or negative. We extracted from the tensor fields $\hat{\mathbf{T}}_{age}$, the orientation which showed the highest correlation with age:

$$\mathbf{v}_{max} = \underset{\{\mathbf{v}' \mid \|\mathbf{v}'\|_{\mathbf{g}_p} = 1\}}{\operatorname{argmax}} |\mathbf{g}(\mathbf{v}', \hat{\mathbf{T}}_{age} \mathbf{g} \mathbf{v}')|. \quad (20)$$

This orientation field of maximum correlation is shown in Figure 7 and corresponds to a negative correlation in both the cingulate sulcus and a region of the frontal lobe. The orientation field shows that the folds in question have a smaller width in the elderly populations. We tested the significance of the shape difference along this direction using the permutation test presented in 3 and found the result to be slightly more significant than presented in Figure 6.

4 Conclusion

This paper proposes a new method for morphometric studies of cortical surfaces. The proposed anisotropic tensor diffusion process on surfaces can be used to highlight deformations that affect sulcal folds by enhancing deformation maps along them. Using the proposed coherence-enhancing filters, we are able to find

anisotropic shape differences of the cortical surface associated with aging that are well localized along specific folds.

Acknowledgments

This work was supported by grants from FQRNT and NSERC.

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine* 56(2), 411–421 (2006)
2. Boucher, M., Evans, A., Siddiqi, K.: Oriented morphometry of folds on surfaces. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) *IPMI 2009*. LNCS, vol. 5636, pp. 614–625. Springer, Heidelberg (2009)
3. Boucher, M., Evans, A., Siddiqi, K.: A Texture Manifold for Curve-Based Morphometry of the Cerebral Cortex. In: *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pp. 174–183 (2011)
4. Chung, M., Worsley, K., Evans, A.: Tensor-based brain surface modeling and analysis. In: *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 467–473 (2003)
5. Desbrun, M., Meyer, M., Schröder, P., Barr, A.: Implicit fairing of irregular meshes using diffusion and curvature flow. In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 317–324 (1999)
6. Fillard, P., Arsigny, V., Ayache, N., Pennec, X.: A Riemannian Framework for the Processing of Tensor-Valued Images. In: Fogh Olsen, O., Florack, L.M.J., Kuijper, A. (eds.) *DSSCV 2005*. LNCS, vol. 3753, pp. 112–123. Springer, Heidelberg (2005)
7. Fischl, Sereno, Dale: Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9(2), 195–207 (1999)
8. Franken, E., Duits, R.: Crossing-preserving coherence-enhancing diffusion on invertible orientation scores. *IJCV* 85(3), 253–278 (2009)
9. Lepore, N., Brun, C.A., Chou, Y.-Y., Chiang, M.-C., Dutton, R.A., Hayashi, K.M., Luders, E., Lopez, O.L., Aizenstein, H., Toga, A.W., Becker, J.T., Thompson, P.M.: Generalized Tensor-Based Morphometry of HIV/AIDS Using Multivariate Statistics on Deformation Tensors. *IEEE Transactions on Medical Imaging* 27(1), 129–141 (2008)
10. Weickert, J.: *Anisotropic diffusion in image processing*, vol. 256 (1998)
11. Worsley, K.J., Taylor, J.E., Carbonell, F., Chung, M.K., Duerden, E., Bernhardt, B., Lyttelton, O., Boucher, M., Evans, A.C.: SurfStat: A Matlab toolbox for the statistical analysis of univariate and multivariate surface and volumetric data using linear mixed effects models and random field theory. *NeuroImage* 47, S102 (2009)

A Novel Longitudinal Atlas Construction Framework by Groupwise Registration of Subject Image Sequences

Shu Liao, Hongjun Jia, Guorong Wu, and Dinggang Shen

Department of Radiology,
Biomedical Research Imaging Center,
University of North Carolina at Chapel Hill
dgshen@med.unc.edu

Abstract. Longitudinal atlas construction is a challenging task in medical image analysis. Given a set of longitudinal images of different subjects, the task is how to construct the unbiased longitudinal atlas sequence reflecting the anatomical changes over time. In this paper, a novel longitudinal atlas construction framework is proposed. The main contributions of the proposed method lie in the following aspects: (1) Subject-specific longitudinal information is captured by establishing a robust growth model for each subject. (2) The trajectory constraints are enforced for both subject image sequences and the atlas sequence, and only one transformation is needed for each subject to map its image sequence to the atlas sequence while preserving the temporal correspondence. (3) The longitudinal atlases are estimated by groupwise registration and kernel regression, thus no explicit template is used and the atlases are constructed without introducing bias due to the selection of the explicit template. (4) The proposed method is general, where the number of longitudinal images of each subject and the time points at which the images are taken can be different. The proposed method is evaluated on a longitudinal database and compared with a state-of-the-art longitudinal atlas construction method. Experimental results show that the proposed method achieves more consistent spatial-temporal correspondence as well as higher registration accuracy than the compared method.

1 Introduction

Study of longitudinal changes of brain anatomical structures plays an important role in medical image analysis. Observing the anatomical shape variations over time in different subjects can provide important clues to study developmental trends across the life span [10]. For this purpose, images of the same subject at different time points are taken to observe subject-specific longitudinal changes.

Longitudinal atlas construction is an active research topic in longitudinal study in the past decade. It can be broadly classified into three categories: (1) Atlas construction by kernel regression [2,6] over the temporal domain; (2) Joint alignment of image sequences to a selected template space [3,7]; (3) Atlas construction by registration of cross-sectional images of different time points [4].

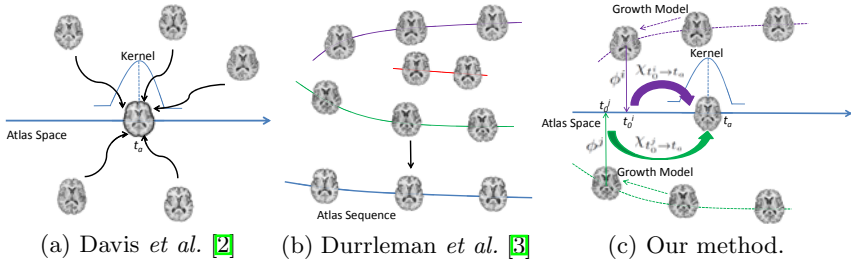


Fig. 1. (a) The method proposed by Davis *et al.* [2], where the atlas is built by kernel regression, and the contribution of each image to build the atlas is determined by the kernel. However, the subject-specific longitudinal information is not considered in this approach. (b) The approach proposed by Durrleman *et al.* [3], where the shape evolution model is constructed for each subject first, as indicated by solid lines, and then each subject's image sequence is registered to the atlas sequence. (c) The proposed method. In this method, for each subject, its corresponding growth model is first estimated based on 4D image registration method to establish the temporal correspondence within the image sequence. A single transformation ϕ_i is then estimated to map subject i 's image sequence to the atlas space. Each image can be warped to the atlas space of a certain time point by the composite deformation field formed by ϕ_i and its subject's growth model. It can be further warped to any time points in the atlas space by the evolution model χ of the atlas. Finally the atlas at each time point is built by performing kernel regression with respect to all the warped images.

Each stream of methods have their own advantages and disadvantages. The proposed method in this paper is mostly related to the methods in [2] and [3].

Davis *et al.* [2] proposed a kernel regression based framework for longitudinal atlas construction. This method extended the Nadaraya-Watson kernel regression method by formulating the regression problem in terms of the Fréchet mean. The longitudinal atlas at a particular time point is constructed by performing kernel regression on the Riemannian manifold represented by diffeomorphisms. The importance of each image during the atlas construction process is reflected by its weight assigned by the kernel. However, in [2], the subject-specific longitudinal information is not considered, thus it may lead to inconsistent temporal correspondence among images of the same subject taken at different time points. Figure 1 (a) illustrates the idea of this method.

Durrleman *et al.* proposed a joint spatial-temporal atlas construction framework in [3] to build the longitudinal atlas. In this method, the shape evolution model of each subject is first established by regression (i.e., indicated by the solid lines across the images of the same subject in Figure 1 (b)). It is not required in this method that each subject must have the same number of scans, or images must be scanned at the same time point. After building the shape evolution model of each individual subject, pairwise spatial-temporal registration is performed by aligning each subject's image sequence to the atlas sequence. The idea of this method is depicted in Figure 1 (b). However, this method requires an explicit template sequence to perform registration instead of estimating the

atlas with the groupwise registration scheme, which may lead to bias. Moreover, this approach was evaluated on the 2D human skull profiles only and has not been tested thoroughly on building real human brain atlas.

Therefore, we are motivated to propose a new longitudinal atlas construction method. The proposed method integrates both the subject-specific longitudinal information as well as the population shape variation information to estimate the atlas sequence. More precisely, the temporal correspondence among the image sequence of each subject is represented by its corresponding growth model, which is estimated based on the 4D image registration algorithm. Thus, images belonging to each subject can be warped to any single time point of the subject's image sequence by using the growth model, while enforcing the temporal trajectory constraints. On the other hand, the correspondence among the atlas sequence can also be represented by the evolution model in the atlas space. Specifically, the atlas sequence is estimated by performing groupwise registration from each subject's image sequence to the atlas space. Note that only a single transformation is needed for each subject to map its image sequence to the atlas space while preserving the temporal correspondence. Since the atlas sequence is estimated based on groupwise registration and kernel regression, no explicit template is assumed. Figure 1(c) illustrates the idea of the proposed method. The proposed method is evaluated on the longitudinal dataset in [8] and compared with the state-of-the-art atlas construction method proposed by Davis *et al.* [2]. It is observed that the proposed method achieves higher registration accuracy as well as better temporal correspondence compared with Davis's method [2].

2 Formulation and Properties of the Proposed Method

In this section, we describe the design details and advantages of the proposed longitudinal atlas construction framework. The whole framework can be summarized by Figure 2. The preprocessing step shown in Figure 2 includes histogram matching and rigid alignment of all the other time point images to the first time point image of each subject.

Suppose there are C different subjects, and each subject i ($i = 1, \dots, C$) has n_i longitudinal images taken at different time points. Let $S_{t_j^i}$ denote the j th time point image of subject i . The number of longitudinal images and the earliest time point of each subject are not necessarily the same. The task is to simultaneously estimate an atlas sequence which can reflect the longitudinal changes of the population.

Suppose there are N different time points t_1, \dots, t_N where we want to construct the atlas sequence, and let $T = \{t_1, \dots, t_N\}$ and the atlas at different time point t denoted as M_t , where $t \in T$. The whole framework to construct the longitudinal atlas can be formulated as the energy minimization problem as expressed in Equation 1:

$$E(M_t, \phi^i, \chi) = \sum_{t \in T} \sum_{i=1}^C \left\{ \frac{\sum_{\substack{t_0^i \leq t_j^i \leq t_{n_i-1}^i \\ t_0^i \leq t_j^i \leq t_{n_i-1}^i}} d(M_t, \chi_{t_0^i \rightarrow t}(I_{t_j^i}))^2 K_h(t - t_j^i)}{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} K_h(t - t_j^i)} + \Psi(\phi^i, \chi) \right\}, \tag{1}$$

where $I_{t_j^i}$ is the transformed image of $S_{t_j^i}$ to the time point t_0^i in the atlas space defined by Equation 2:

$$I_{t_j^i} = \phi^i \circ V_{(t_j^i \rightarrow t_0^i)}^i(S_{t_j^i}), \tag{2}$$

where $V_{(t_j^i \rightarrow t_0^i)}^i$ denotes the underlying growth model of subject i which can warp image $S_{t_j^i}$ of subject i to its first time point image $S_{t_0^i}$ while preserving the temporal smoothness. The growth model only needs to be estimated once. ϕ^i is the transformation which maps subject i 's space to the atlas space, and \circ denotes the operation of compositing deformations.

In Equation 1, χ is the underlying evolution model in the atlas space. $K_h(t)$ is the kernel satisfying $K_h(t) = \frac{1}{h}K(\frac{t}{h})$, where h is the bandwidth of the kernel and K is a function satisfying $\int K(t)dt = 1$. $d(\cdot)$ is the distance metric between images defined based on diffeomorphisms. $\Psi(\phi^i, \chi)$ is the overall regularization term defined by Equation 3:

$$\Psi(\phi^i, \chi) = \gamma_{\phi^i} Reg(\phi^i) + \gamma_{\chi} Reg(\chi), \tag{3}$$

where $Reg(\cdot)$ denotes the regularization function, and γ_{ϕ^i} and γ_{χ} are constants trading off the accuracies in image matching and the smoothness of the deformation field.

The physical meaning beneath Equation 1 is: all the longitudinal images of each subject are jointly considered as a sequence with trajectory constraints established by its growth model V^i . Each subject i 's image sequence can be

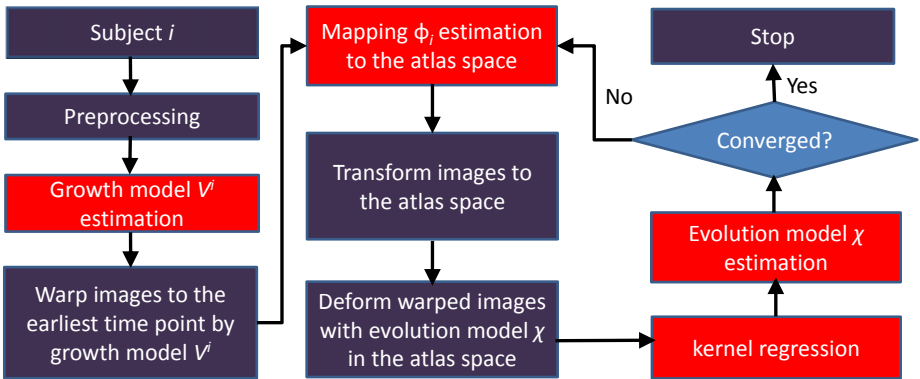


Fig. 2. Flow chart of the proposed method, where rectangles highlighted with red color denote the major components of the proposed approach

transformed to the atlas space by using a single transformation ϕ^i , and the transformed images can be further warped to any time point t in the atlas space by the evolution model χ . Finally the atlas sequence is estimated based on groupwise registration and kernel regression over all subject image sequences.

There are four main advantages of the proposed framework compared to conventional atlas construction approaches proposed in [2] and [3]. First, subject-specific longitudinal information is considered by building a growth model for each subject, which is different from the method proposed in [2], where subject-specific longitudinal information is not considered and thus possibly leads to inconsistent temporal correspondence. Second, the trajectory constraints are enforced by considering each subject's image sequence as a group to map to the atlas sequence, thus only one transformation ϕ^i is needed for each subject i to map its image sequence to the atlas sequence. Third, the atlas sequence is estimated based on groupwise registration and kernel regression over all subject image sequences, therefore the unbiased atlas sequence can be obtained, which is different from the method proposed in [3] where an explicit template is used. Finally, the proposed method is general, as the number of longitudinal images of each subject and the time points at which the images are taken can be different.

To minimize the energy function in Equation 1 with respect to M_t , ϕ^i and χ , the four step optimization strategy is adopted: (1) Growth model V^i estimation, (2) subject-specific transformation ϕ^i estimation, (3) atlas M_t construction by kernel regression, and (4) evolution model χ estimation in the atlas space. The proposed method is consisted of these four major components, as highlighted by red rectangles shown in Figure 2. Note that the growth model in the first step only needs to be computed once, and the last three steps need to be updated iteratively, which is also illustrated in Figure 2. In the following sections, details of each component are given on how to optimize Equation 1.

2.1 Growth Model Estimation for Each Subject

The first step of the proposed framework is to estimate the growth model of each subject after the preprocessing step as shown in Figure 2. The goal of this step is to recover the geometric changes of anatomical structures over time of each subject image sequence.

The growth model of each subject can be estimated based on 4D image registration methods (i.e., a method proposed in [9]). Therefore, after the 4D registration, temporal correspondences are established among the entire image sequence, which are represented by the respective deformation field from each time point image to the first time point image. This step can be summarized by Algorithm 1 below.

After estimating the growth model for each subject i by Algorithm 1, the longitudinal information of subject i contained in the image sequence can be propagated and aggregated to any time point based on the growth model. Without loss of generality, in this paper the images of each subject are all warped to its earliest time point.

Algorithm 1. Growth Model Estimation for Subject i

Input: The rigidly-aligned and histogram-matched image sequence $S_{t_0^i}, S_{t_1^i}, \dots, S_{t_{n_i-1}^i}$ of subject i , where n_i denotes the total number of images in the image sequence of subject i .

Output: Deformation fields $V_{(t_0^i \rightarrow t_0^i)}^i, V_{(t_1^i \rightarrow t_0^i)}^i, \dots, V_{(t_{n_i-1}^i \rightarrow t_0^i)}^i$ mapping from $S_{t_0^i}, S_{t_1^i}, \dots, S_{t_{n_i-1}^i}$ to $S_{t_0^i}$.

1. Construct the moving image sequence as $S_{t_0^i}, S_{t_1^i}, \dots, S_{t_{n_i-1}^i}$.
 2. Construct the reference image sequence by repeating the first time point image as $S_{t_0^i}, S_{t_0^i}, \dots, S_{t_0^i}$.
 3. Register the moving image sequence to the reference image sequence using the 4D registration method in [9]. Denote the resulting deformation field that warps $S_{t_j^i}$ ($j = 0, \dots, n_i - 1$) to $S_{t_0^i}$ as $V_{(t_j^i \rightarrow t_0^i)}^i$, where $V_{(t_0^i \rightarrow t_0^i)}^i$ is the identity deformation field.
 4. Return $V_{(t_0^i \rightarrow t_0^i)}^i, V_{(t_1^i \rightarrow t_0^i)}^i, \dots, V_{(t_{n_i-1}^i \rightarrow t_0^i)}^i$.
-

The growth model provides smooth and consistent temporal correspondence among image sequence of each subject. Another advantage of building the growth model for each subject is that if the longitudinal data of a particular subject is taken sparsely with large time gap (e.g., more than 3 years), the geometric changes of brain structures can be dramatic; In this case, the growth model can bridge the gap of the dramatic changes of anatomical structures by interpolating longitudinal images between two consecutive time points. The growth model estimation step for each subject only needs to be calculated once, which is also illustrated by Figure 2.

2.2 Transformation of Subject Image Sequence to the Atlas Space

After building the subject-specific growth model described in Section 2.1, the next step is to estimate the transformation from each subject space to the atlas space (i.e. ϕ^i in Equations 1 and 2) by fixing the rest of the variables such as M_t and χ in Equation 1.

To estimate ϕ^i , the image matching term $d(M_t, \chi_{t_0^i \rightarrow t}(I_{t_j^i}))$ in Equation 1 can be redefined as $d((\chi_{t_0^i \rightarrow t})^{-1}M_t, I_{t_j^i}^i)$ (i.e., each atlas is first warped to the earliest time point t_0^i of subject i in the atlas space, and the warped images of subject i at the same time point t_0^i are then matched to the warped atlases by

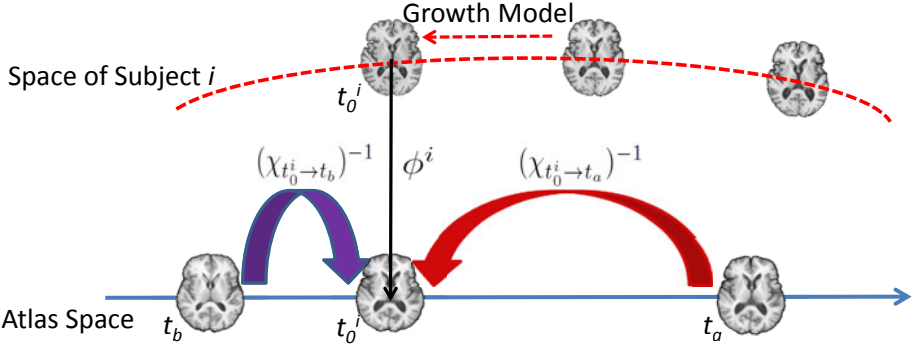


Fig. 3. Illustration of how to estimate the transformation ϕ^i from the subject space to the atlas space. First, each image of subject i is warped to the earliest time point t_0^i of subject i by the growth model. Then, for the current atlas sequence, each atlas at different time point is first warped to the subject i 's earliest time point t_0^i in the atlas space by applying the reversed evolution model $(\chi_{t_0^i \rightarrow t})^{-1}$. The transformation mapping from the subject space to the atlas space ϕ^i can be estimated by performing groupwise diffeomorphic registration from the warped images of subject i to the warped atlases both at t_0^i .

the transformation ϕ^i). This redefinition is valid only when the evolution model χ is a diffeomorphic transformation. Therefore, Equation 1 now becomes:

$$J(\phi^i) = \sum_{t \in T} \left\{ \frac{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} d((\chi_{t_0^i \rightarrow t})^{-1} M_t, I_{t_j^i})^2 K_h(t - t_j^i)}{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} K_h(t - t_j^i)} + \gamma_{\phi^i} \text{Reg}(\phi^i) \right\}, \tag{4}$$

where $I_{t_j^i} = \phi^i \circ V_{(t_j^i \rightarrow t_0^i)}^i(S_{t_j^i})$. Equation 4 reflects that atlases at different time points are inversely warped to the time point t_0^i by the reversed evolution model $(\chi_{t_0^i \rightarrow t})^{-1}$. Each image $S_{t_j^i}$ of subject i is aligned to the first time point t_0^i of subject i by its growth model $V_{(t_j^i \rightarrow t_0^i)}^i$. Thus ϕ^i can be estimated by groupwise registration between the warped images of subject i at the first time point t_0^i and all the warped atlases M_t also at time point t_0^i . This procedure can be summarized by Figure 3. In Figure 3, we aim to estimate the transformation from the subject space to the atlas space for subject i , where t_0^i denotes the earliest time point of subject i . We only need to estimate one transformation from the subject space to the atlas space at time t_0^i as images of the same subject can be warped to the earliest time point t_0^i based on the growth model built in Section 2.1.

After estimating ϕ^i for each subject, we can transform images in each subject's space to the common atlas space to construct and update the atlas sequence, which will be described in the next section.

2.3 Atlas Construction by Kernel Regression

After estimating the transformation ϕ^i from each subject i 's space to the atlas space, the next step is to construct and update the atlas sequence M_t by fixing the variables ϕ^i and χ in Equation 1.

By fixing ϕ^i and χ , the energy minimization problem in Equation 1 with respect to M_t now becomes:

$$J(M_t) = \sum_{i=1}^C \left\{ \frac{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} d(M_t, \chi_{t_0^i \rightarrow t}(I_{t_j^i}))^2 K_h(t - t_j^i)}{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} K_h(t - t_j^i)} \right\}. \tag{5}$$

The optimal solution M_t of Equation 5 can be obtained by Equation 6:

$$M_t = \arg \min_{M_{opt} \in \Lambda} \sum_{i=1}^C \left(\frac{\sum_{j=0}^{n_i-1} K_h(t - t_j^i) d(M_{opt}, \chi_{t_0^i \rightarrow t}(I_{t_j^i}))^2}{\sum_{j=0}^{n_i-1} K_h(t - t_j^i)} \right), \tag{6}$$

where Λ denotes the whole possible image space, C denotes the number of subjects, n_i denotes the number of longitudinal images belonging to the subject i , and t_j^i denotes the time point at which the j th longitudinal image of subject i is taken. $K_h(\cdot)$ is the kernel function, and the Gaussian kernel is adopted in this paper. $d(\cdot)$ is the distance metric between images defined based on diffeomorphisms.

Equation 6 actually denotes a kernel regression procedure to estimate M_t . In this paper, the greedy iterative algorithm proposed in 5 is adopted to estimate the optimal solution of Equation 6.

2.4 Evolution Model Estimation in the Atlas Space

After constructing the atlas sequence by kernel regression in Section 2.3, the last step of the proposed method is to estimate and update the evolution model χ in the atlas space by fixing variables ϕ^i and M_t in Equation 1. When ϕ^i and M_t are fixed, Equation 1 becomes:

$$J(\chi) = \sum_{t \in T} \sum_{i=1}^C \left\{ \frac{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} d(M_t, \chi_{t_0^i \rightarrow t}(I_{t_j^i}))^2 K_h(t - t_j^i)}{\sum_{t_0^i \leq t_j^i \leq t_{n_i-1}^i} K_h(t - t_j^i)} + \gamma_\chi Reg(\chi) \right\}. \tag{7}$$

Therefore, to estimate χ , first all images of each subject i are warped to the atlas space of the first time point of subject i by $\phi^i \circ V_{(t_j^i \rightarrow t_0^i)}^i(S_{t_j^i})$. Then, we can estimate $\chi_{t_0^i \rightarrow t}$ by registering all these warped images of subject i with M_t , thus the overall χ can be obtained by stitching all $\chi_{t_0^i \rightarrow t}$. Note that the kernel $K_h(\cdot)$ will be used to constrain the weight of these warped images.

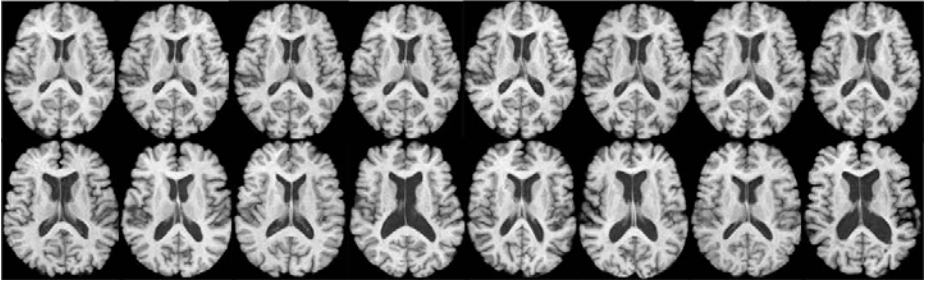


Fig. 4. First row shows images taken from the same subject from age 67 to age 74. Significant longitudinal changes can be observed. Second row shows images taken from different subjects, demonstrating the large structural variations across different subjects in the dataset.

3 Experimental Results

The proposed method is evaluated by building longitudinal atlases from the longitudinal dataset in [8]. Ten subjects are selected, with the ages ranging from 67 to 85 when the first time point images were taken. Each subject has around ten longitudinal images taken at different time points, and the period between each pair of consecutive time points is around one year. Each image is with resolution $256 \times 256 \times 124$. The first row in Figure 4 shows longitudinal images scanned from the same subject from ages 67 to 74, where significant longitudinal changes can be observed, especially at the ventricle region. Images shown in the second row of Figure 4 are taken from different subjects to demonstrate the large population shape variations in this dataset. The segmentation results of each image into three different types of tissues: white matter (WM), gray matter (GM) and ventricular CSF, are also available.

The proposed method is also compared with the state-of-the-art atlas construction algorithm proposed by Davis *et al.* [2]. Both methods were implemented based on the Insight Segmentation and Registration Toolkit (ITK) [4]. In this paper, the proposed method was evaluated both on the ability to capture the global shape variations among all the images and the ability to preserve the temporal correspondence among the longitudinal data of the same subject.

3.1 Experiments on Measuring Global Registration Accuracy

In this section, the proposed method is evaluated on measuring the global registration accuracy among all the aligned images together. Figure 5 shows the same cross section of the atlas constructed by the proposed method across different ages. It can be observed that the expansion of the lateral ventricle is captured.

¹ <http://www.itk.org/>

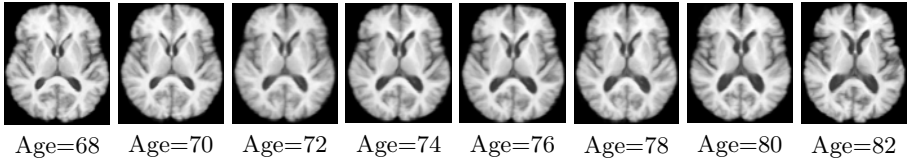


Fig. 5. The same cross-sectional images obtained from the atlas constructed by the proposed method at different ages. The expansion behavior of the lateral ventricle is captured.

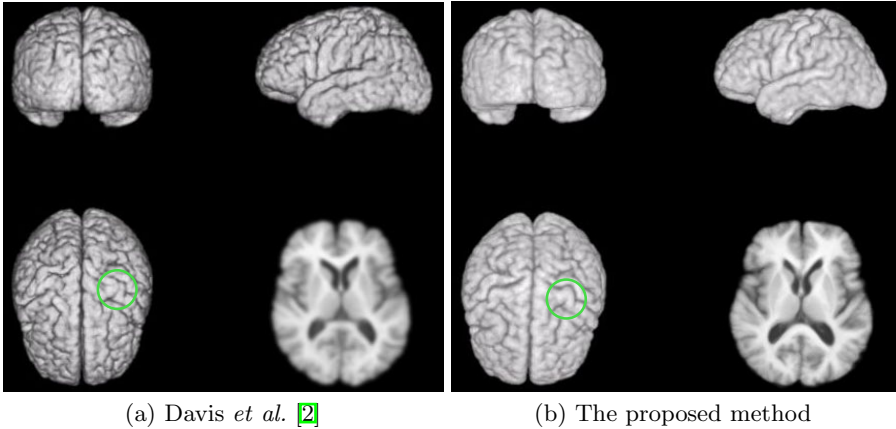


Fig. 6. 3D rendering of the atlas of age 74 constructed by: (a) the method proposed by Davis *et al.* [2] and (b) the proposed method. Significant differences are highlighted with the green circles.

To visually compare the atlas construction performance of the proposed method and the method proposed by Davis *et al.* [2], Figure 6 shows the 3D rendering of the atlas of age 74 constructed by both the proposed method and the method proposed by Davis *et al.* [2]. It can be observed that the atlas constructed by using the proposed method is sharper and preserves more anatomical details than the atlas constructed by the approach proposed by Davis *et al.* [2], where the regions with significant differences are highlighted by the green circles.

In this paper, we also quantitatively evaluate the proposed method by using the tissue overlap ratio [1]. It is defined as $P = \frac{\#(A \cap B)}{\#(A \cup B)}$, where A and B denote the regions of a specific tissue in the two images, and $\#(\cdot)$ denotes the number of voxels inside a region. In this paper, since there is no explicit template used for both Davis's method [2] and the proposed method, the segmentation result of the template image was obtained by majority voting from all aligned images by setting the tissue type of each voxel in the template image as the majority of tissue labels from all aligned images. The average values of P for WM, GM and ventricular CSF across different ages using Davis's method [2] and the proposed method are shown in Figures 7 (a) to (c). It can be observed that the proposed

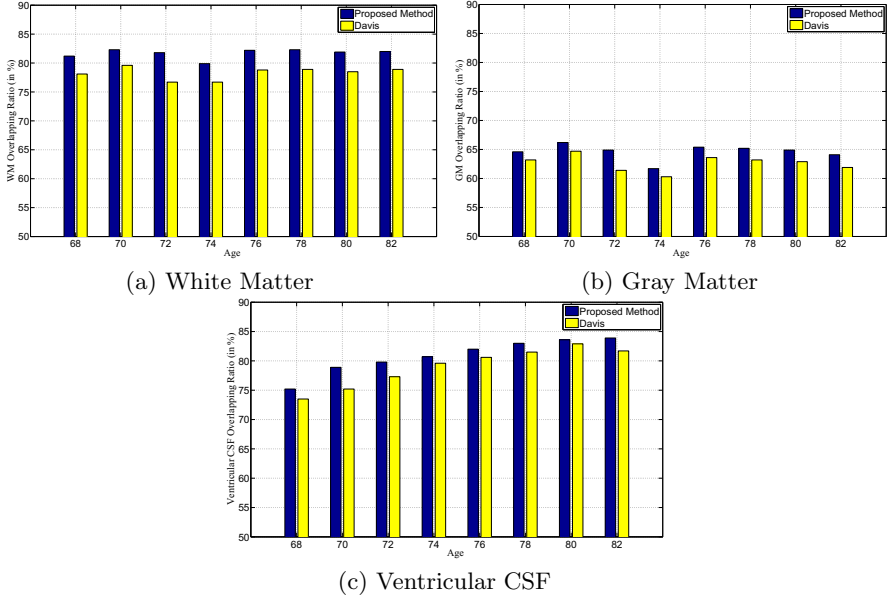


Fig. 7. Mean tissue overlap ratios for: (a) white matter, (b) gray matter, and (c) ventricular CSF across different ages by using Davis’s method [2] (yellow bars) and the proposed method (blue bars).

method consistently achieves higher tissue overlap ratio than Davis’s method [2]. More precisely, for each type of tissue and for each age, the tissue overlap ratio obtained by the proposed method is normally 2% to 3% higher than that obtained by Davis’s method [2], which is a significant improvement as the standard deviation of the tissue overlap ratios for each type of tissue at different age is no more than 0.1%.

3.2 Experiments on Measuring Individual Temporal Smoothness

Besides evaluating the global registration accuracy among all the images of the proposed method, we also measure the registration accuracy within each subject. We adopt the tissue overlap ratio measure similar to Section 3.1, but now each subject is considered as a separate group, and the overlap ratios for WM, GM and ventricular CSF are measured for each group.

The average tissue overlap ratios of WM, GM and ventricular CSF across the 10 groups (i.e., there are 10 subjects in total) are shown in Figures 8 (a) to (c) for different ages. It can be observed that the average subject-specific tissue overlap ratios are generally higher than those obtained from the whole population shown in Section 3.1, as the longitudinal changes within each subject are much smaller than the shape variations across different subjects. It is demonstrated that the proposed method still maintains higher tissue overlap ratios than Davis’s

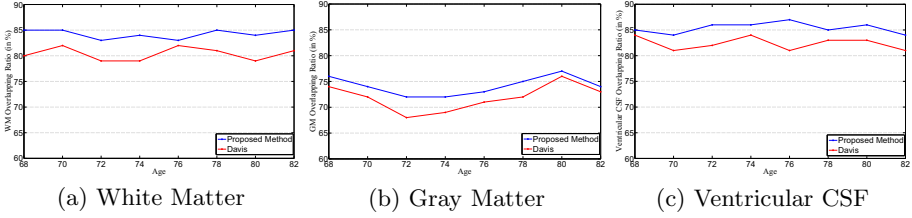


Fig. 8. Subject-specific mean tissue overlap ratios for: (a) white matter, (b) gray matter, and (c) ventricular CSF across different ages with Davis’s method [2] and the proposed method

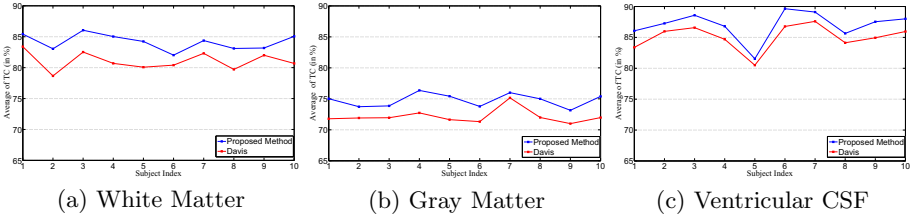


Fig. 9. Mean temporal consistency (TC) values for different subjects of: (a) white matter, (b) gray matter, and (c) ventricular CSF with Davis’s method [2] and the proposed method

method, which implies the more accurate registration results by the proposed method within each subject’s image sequence.

Moreover, to measure the temporal consistency across different ages of each subject, the temporal consistency (TC) factor of different types of tissues is calculated. The average TC factor is defined as: $TC = \frac{1}{|\Omega|} \sum_{i \in \Omega} (1 - L_i / (Y - 1))$, where Ω is the voxel set of region of interest to measure the temporal consistency, and $|\Omega|$ denotes the number of voxels in Ω . L_i denotes the number of tissue label changes of the corresponding voxel i across time, and Y denotes the number of longitudinal images of the subject. The average TC values for WM, GM and ventricular CSF of different approaches for each subject are shown in Figures 9 (a) to (c). It is observed that the proposed method consistently achieves higher TC values for different types of tissues for each subject compared to Davis’s method, which strongly implies the better temporal correspondence established by the proposed method. The improvement of the temporal consistency of the proposed method is significant as the average TC values for the proposed method are normally 2% to 3% higher than Davis’s method, while the standard deviations of the TC values are no more than 0.1%.

4 Conclusion

In this paper, a new framework to construct the longitudinal atlas is proposed. The proposed method takes both the subject-specific longitudinal changes and

population shape variations information into account when estimating the longitudinal atlas. The subject-specific longitudinal information is captured by establishing a growth model for each subject based on the 4D image registration algorithm. Then, transformations which map each subject's image sequence to the atlas space are estimated by performing diffeomorphic groupwise registration between the warped subject images and the warped atlases. Images of each subject are transformed to the atlas space by the estimated diffeomorphic transformations and warped to the time point of interest to build the atlas by the evolution model in the atlas space. The atlas is then estimated based on the kernel regression procedure. The proposed method is qualitatively and quantitatively compared with the state-of-the-art atlas building algorithm proposed by Davis *et al.* [2] on the longitudinal dataset. Experimental results show that the proposed method achieves higher registration accuracy as well as better temporal correspondence. Future work includes testing the proposed method to measure the longitudinal changes of deep brain structures such as hippocampus.

References

1. Crum, W.R., Rueckert, D., Jenkinson, M., Kennedy, D., Smith, S.M.: A framework for detailed objective comparison of non-rigid registration algorithms in neuroimaging. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 679–686. Springer, Heidelberg (2004)
2. Davis, B., Fletcher, E., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: ICCV, pp. 1–7 (2007)
3. Durrleman, S., Pennec, X., Trounev, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 297–304. Springer, Heidelberg (2009)
4. Gabriel, H., Yundi, S., Hongtu, Z., Mar, S., Martin, S., Marc, N.: Dti longitudinal atlas construction as an average of growth model. In: STIA (2010)
5. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
6. Khan, A., Beg, M.: Representation of time-varying shapes in the large deformation diffeomorphic framework. In: ISBI, pp. 1521–1524 (2008)
7. Perperidis, D., Mohiaddin, R.H., Rueckert, D.: Spatio-temporal free-form registration of cardiac mri sequences. *MedIA* 9, 441–456 (2005)
8. Resnick, S.M., Goldszal, A.F., Davatzikos, C., Golski, S., Kraut, M.A., Metter, E.J., Bryan, R.N., Zonderman, A.B.: One-year age changes in mri brain volumes in older adults. *Cerebral Cortex* 10, 464–472 (2000)
9. Shen, D., Davatzikos, C.: Measuring temporal morphological changes robustly in brain mr images via 4-dimensional template warping. *NeuroImage* 21, 1508–1517 (2004)
10. Yoon, U., Fonov, V.S., Perusse, D., Evans, A.C., Group, B.D.C.: The effect of template choice on morphometric analysis of pediatric brain data. *NeuroImage* 45, 769–777 (2009)

A Probabilistic Framework to Infer Brain Functional Connectivity from Anatomical Connections

Fani Deligianni¹, Gael Varoquaux^{2,3}, Bertrand Thirion³, Emma Robinson¹, David J. Sharp⁴, A. David Edwards⁵, and Daniel Rueckert¹

¹ Department of Computing, Imperial College London, UK

² INSERM U922, Neurospin, CEA Saclay, France

³ Parietal project team, INRIA, Saclay, France

⁴ C3NL, The Division of Experimental Medicine, Imperial College London, UK

⁵ Institute of Clinical Sciences, Imperial College London, UK

Abstract. We present a novel probabilistic framework to learn across several subjects a mapping from brain anatomical connectivity to functional connectivity, i.e. the covariance structure of brain activity. This prediction problem must be formulated as a structured-output learning task, as the predicted parameters are strongly correlated. We introduce a model selection framework based on cross-validation with a parametrization-independent loss function suitable to the manifold of covariance matrices. Our model is based on constraining the conditional independence structure of functional activity by the anatomical connectivity. Subsequently, we learn a linear predictor of a stationary multivariate autoregressive model. This natural parameterization of functional connectivity also enforces the positive-definiteness of the predicted covariance and thus matches the structure of the output space. Our results show that functional connectivity can be explained by anatomical connectivity on a rigorous statistical basis, and that a proper model of functional connectivity is essential to assess this link.

1 Introduction

The brain's power and stability depends critically on its connectional complexity. Recently, it has been suggested that consciousness depends on the brain's ability to integrate information among different thalamo-cortical areas [8]. Disruption in cortical connectivity has been also implicated in a number of disorders and pathologies, such as schizophrenia [3], autism [21], brain trauma [23] and so on. As a result there is a shift of research attention from localised brain function to network organization and dynamics.

A functional network is defined based on connections that reflect temporal dependency between spatially remote neurophysiological events. It is well established that during rest the brain shows spontaneous activity that is highly correlated between multiple brain regions. However, the neuronal basis of resting-state

functional Magnetic Resonance Imaging (r-fMRI) oscillations is not yet fully understood [20]. There is an on-going debate on whether physiological processes, such as respiration and cardiac cycle dominate the resting-state BOLD signal. Support for a neuronal basis comes from several studies that investigate the relationship between structural connectivity and functional brain connectivity [12,15,14]. The underlying hypothesis is that if r-fMRI has a neuronal basis then it must reflect the existence of structural connections that relate functionally linked brain regions.

To this end, significant evidence has emerged that there are strong structural connections between regions that are functionally linked, forming resting-state networks [12,13]. Furthermore, strong functional connectivity has been observed between areas with direct structural link [14]. However, these studies do not provide a systematic framework to investigate localised influences between structural connections and functional links. [30] presents a joint generative model of functional and structural connectivity based on the assumption that they follow Gaussian distributions.

In this work, we are interested in inferring the link between anatomical and functional connectivity in a data-driven way. We adapt an anatomically informed probabilistic model of functional connectivity. We describe subject-level functional connectivity as a multivariate Gaussian process [29] estimated by imposing a common structure based on the structural connectivity matrices. Subsequently, we use statistical prediction to infer functional connectivity from structural connectivity. Compared to discriminant models [14,5], which does not provide an underlying generative process, the use of a probabilistic framework assists in gaining a deeper understanding of the mechanisms underlying functional connectivity.

2 Problem Statement

The aim of our predictive model is to use structural connectivity across subjects to predict functional connectivity. Structural connections in the brain are formed by white matter tracts that directly interconnect large groups of spatially separated neurons via long-distance axons. Here, the structural connectivity \mathbf{A} is described as an undirected weighted graph that shows whether there is or not a link between each pair of ROIs and its strength. In this framework, each brain's connection is treated as a variable, which results in a total of $N = n(n - 1)/2$ variables, where n is the number of ROIs. For each of the functional connection, our goal is to find a subset of structural connections to which it is strongly related.

We consider S subjects, represented by \mathbf{A}^s , the subject specific structural connectivity matrix and the correlation matrices between brain time series extracted from n ROIs: $\{\Sigma^s \in \mathbb{R}^{n \times n}, s = 1, \dots, M\}$.

The current practice in studying functional connectivity, \mathbf{F}^s is based on the estimation of the coefficients of the correlation matrices across subjects. This can be expressed as a univariate additive linear model on the covariance matrix:

$$\Sigma^s = \Sigma^* + \mathbf{d}\Sigma^s \quad (1)$$

where Σ^* is a covariance matrix representative of the group, and $\mathbf{d}\Sigma^s$ encode subject-specific contributions. Very often some coefficients of the covariance matrix are zeroed out by thresholding, which leads to a non-positive definite matrix, hence, it does not yield any consistent signal model. Therefore, it is not possible to generalize the learned model to new subjects [29]. The simplest well-posed alternative is to model the fMRI time series $\mathbf{U} \in \mathbb{R}^{n \times r}$ as centered multivariate Gaussian process [29], hence described by a positive definite covariance.

Let us denote $(y_k)_{k=1 \dots N}$ the strength of functional connections, i.e. model-dependent parameters that characterize the covariance structure. These should be predicted by combinations of an unknown set of structural variables $(x_j)_{j=1 \dots N}$. Penalised linear regression finds a sparse solution of the under-determined linear regression problem.

It is crucial to note here that the application of a multiple regression predictive model provides localised, interpretable results. However, it infers each functional connection independently and thus it does not guarantee that the overall prediction of the covariance matrix would be an SPD matrix. The space of SPD matrices, Sym_p^+ , does not form a vector space: $\mathbf{C}, \mathbf{D} \in \text{Sym}_p^+ \not\Rightarrow \mathbf{C} - \mathbf{D} \in \text{Sym}_p^+$. Another issue arises from the fact that the number of functional connectivity parameters is greater than the number of samples: $n < \frac{1}{2}r(r+1)$. This results in a large estimation error of the sample covariance matrix [16,29].

Parameterizations and loss for Functional connectivity. One of the challenges of predicting functional brain connectivity is that it can be described by many different parameters. Most often, the functional connectivity between a set of regions is described by the correlation matrix, $\Sigma \in \mathbb{R}^{n \times n}$, between the time series, $\mathbf{X} \in \mathbb{R}^{r \times n}$, of the mean activation in the different regions of interest. However, several studies [27,29] have shown that the precision matrix, the inverse of the covariance, $\mathbf{K} = \Sigma^{-1}$, has adequate properties for the estimation of brain functional networks, while Fransson and Marrelec [10] reported that partial correlations identified between-regions interaction better than correlations based on resting-state fMRI. The partial correlation matrix \mathbf{P} corresponds to the precision matrix \mathbf{K} renormalized to have unit diagonal. On our multi-subject dataset (described in section 4), we find that on average, the anatomical connectivity between a pair of regions i and j , $\mathbf{A}_{i,j}$, explains 12.77% of the variance of the corresponding pair-wise functional correlation, $\Sigma_{i,j}$, 13.52% of the precision coefficient $\mathbf{K}_{i,j}$, and 14.36% of the partial correlation $\mathbf{P}_{i,j}$.

In this context, we choose to formalize this learning problem as a structured-output multivariate regression, suitable for predicting interdependent variables. The task is to learn a mapping f from a space \mathcal{A} describing anatomical connectivity matrices to a second space \mathcal{F} describing functional connectivity. It entails the choice of a parameterization for both input and output spaces, as well as a loss function l on the output parameterization, giving a measure of the empirical risk associated with the prediction. Typically, in an energy-based formulation [17], learning from training data is achieved by choosing f to minimize a combination of the loss and a regularization term.

In our case, we use for the input space the anatomical connectivity matrices, that is the space of symmetric matrices $\mathcal{A} = \text{Sym}_p$. For the output space, we work on a parametrization of the space of correlation matrices, that is symmetric definite positive matrices $\mathcal{F} = \text{Sym}_p^+$. Sym_p^+ is not a vector space thus the standard Euclidean distance on matrices, the Frobenius norm, is ill suited to quantify errors. However, Sym_p^+ can be parametrized as a Riemannian manifold using an intrinsic metric [9]:

$$d_{AI}(\mathbf{C}, \mathbf{D})^2 = \text{tr}(\log \mathbf{C}^{-\frac{1}{2}} \mathbf{D} \mathbf{C}^{-\frac{1}{2}})^2 \tag{2}$$

This metric can leads to a full statistical framework on Sym_p^+ [22,19]. Most importantly, it is invariant under affine scaling and inversion of the matrices, and is thus equivalent for a wide range of parameterization for functional connectivity. In particular, if \mathbf{D} is the predicted matrix, and \mathbf{C} the ground truth, it gives the same prediction error on the correlation matrices and on the precision matrices.

Non-structured output predictors, such as independent prediction of each coefficient of the functional connectivity matrix, cannot guarantee that prediction \mathbf{F}^{pred} will be positive definite. To quantify their prediction error in a parametrization-independent way, we develop the metric given in Eq. (2) around the target matrix $\mathbf{F}^{\text{target}}$, that can be a correlation matrix, a precision matrix, or a partial correlation matrix, but is by construction definite positive (note that d is no longer a distance however):

$$d_{AI}(\mathbf{F}^{\text{pred}}, \mathbf{F}^{\text{target}}) \sim d(\mathbf{F}^{\text{pred}}, \mathbf{F}^{\text{target}}) = \|(\mathbf{F}^{\text{target}})^{-1}(\mathbf{F}^{\text{target}} - \mathbf{F}^{\text{pred}})\| \tag{3}$$

$\|\cdot\|$ is the Frobenius norm.

3 From Structure to Functional Correlations: A Probabilistic Framework

Generative model We use multivariate autoregressive models (MAR) to describe the generative process of fMRI time series. If $\mathbf{x} \in \mathbb{R}^n$ is the multivariate vector of observations at a given time r ,

$$\mathbf{x}(r + 1) = \mathbf{A}\mathbf{x}(r) + \mathbf{e}(\mathbf{r} + \mathbf{1}) \tag{4}$$

with $\mathbf{A} \in \mathbb{R}^{n \times n}$ a matrix of the connection between variables and \mathbf{e} additive Gaussian noise between variables with zero mean and identity covariance.

We consider the ongoing brain activity in resting-state as a stationary process. If $\mathbf{U} \in \mathbb{R}^{n \times r}$ is the matrix of the observed time series and $\mathbf{E} \in \mathbb{R}^{n \times r}$ then

$$\mathbf{U} = \mathbf{A}\mathbf{U} + \mathbf{E} \Leftrightarrow \mathbf{U} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{E} \tag{5}$$

where \mathbf{I} is the $n \times n$ identity matrix. Thus, the covariance of the observed time series, Σ , is given by:

$$\Sigma = \frac{1}{r} \mathbf{U}\mathbf{U}^T = (\mathbf{I} - \mathbf{A})^{-1} \text{cov}\mathbf{E}(\mathbf{I} - \mathbf{A})^T = ((\mathbf{I} - \mathbf{A})^T(\mathbf{I} - \mathbf{A}))^{-1} \tag{6}$$

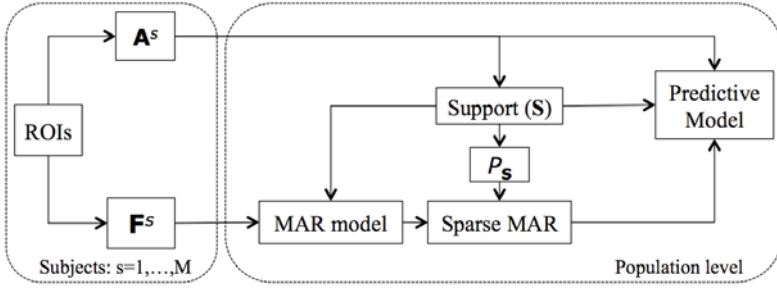


Fig. 1. Probabilistic framework for predicting functional connectivity from structural brain connectivity. \mathbf{A}^s and \mathbf{F}^s represent structural and functional connectivity for subject s , respectively. \mathbf{S} represents the support, the sparsity pattern over all structural connectivity matrices. $P_{\mathbf{S}}$ represents the permutation that provides with a sparser Cholesky factor of the support matrix, \mathbf{S} .

Note that $\text{cov}\mathbf{E} = \mathbf{I}$ and $\mathbf{B} = \mathbf{I} - \mathbf{A}$ is a matrix square root of the inverse covariance, which we call the interaction matrix.

$$\Sigma^{-1} = \mathbf{B}^T \mathbf{B}. \tag{7}$$

Conditional independence between variables is given by the zeros in the precision matrix $\mathbf{K} = \Sigma^{-1}$. In order to reduce the large estimation error of the sample covariance matrix, covariance selection is achieved by imposing a sparse support for the estimated precision matrix [29]. We adopt the sparsity pattern of the structural data \mathbf{S} to represent the support, see Fig. 1.

More precisely, we impose the same support *supp* across individuals, which is the set of supra-threshold t-tests statistics on $\{\mathbf{A}^s, s = 1, \dots, M\}$: $t = \frac{\bar{\mathbf{A}}}{\text{std}(\mathbf{A})\sqrt{M}}$; the choice of the threshold is discussed later. We estimate the non-zero coefficients using the iterative proportional scaling (IPS) algorithm [16].

Instead of predicting the precision matrix Σ^{-1} directly, we aim at recovering the interaction matrix \mathbf{B} . This guarantees that the predicted precision matrix $\hat{\Sigma}^{-1} = \hat{\mathbf{B}}^T \hat{\mathbf{B}}$ is SPD. Here, we use the Cholesky decomposition of the precision matrix Σ^{-1} to estimate the interaction matrix \mathbf{B} . Since the Cholesky decomposition is not invariant to a permutation of the rows and columns of the input matrix Σ^{-1} , we adapt a permutation $P_{\mathbf{S}}(\mathbf{B})$ that tends to yield sparser Cholesky factors LU of the support matrix \mathbf{S} . To find the ordering we construct a matrix \mathbf{M} , such that $\text{supp}(\mathbf{L}^T \mathbf{L}) = \text{supp}(\mathbf{B})$, where *supp* represents the support, i.e. the sparsity pattern of the underlying matrix. To reduce the number of non-zero elements of the Cholesky factor the column approximate minimum degree permutation is utilised. Note that we are interested in predicting correlation and not covariance, thus, we rescale the diagonal of \mathbf{B} to ones by multiplying with the inverse of its diagonal matrix:

$$\tilde{\mathbf{B}} = \mathbf{B} \text{diag}(\mathbf{B})^{-1} \tag{8}$$

Algorithm 1. Inferring functional from structural connectivity based on the sparse MAR model described in Section 3, Fig. 1. Note that it is embedded in a leave-one-subject-out cross-validation loop, in order to measure the performance of the model.

Input: $\{\mathbf{A}^s, s = 1, \dots, M\}$ is the structural connectivity matrix for each subject s ,
 $\{\mathbf{F}^s, s = 1, \dots, M\}$ is the observed functional correlation matrix
Output: Predictive model of the $\{\mathbf{F}^s, s = 1, \dots, M\}$ based on $\{\mathbf{A}^s, s = 1, \dots, M\}$

- 1: Estimate the support \mathbf{S} as the set of supra-threshold t-tests statistics on $\{\mathbf{A}^s, s = 1, \dots, M\}$
- 2: **for** $s=1$ to M **do**
- 3: For all functional correlation matrices estimate the precision matrix \mathbf{K} based on IPS: $\mathbf{K}^s = \text{IPS}(\mathbf{F}^s, \mathbf{S})$
- 4: Find the sparser permutation of the support \mathbf{S} and re-order \mathbf{K}
- 5: Estimate the interaction matrix $\tilde{\mathbf{B}}$ based on the sparse Cholesky decomposition
- 6: **end for**
- 7: **for** $k=1$ to N **do**
- 8: Apply coefficient-wise LASSO to predict $\tilde{\mathbf{B}}_k^s$ from $\{\mathbf{A}^s, s = 1, \dots, M\}$
- 9: **end for**
- 10: The output is $\{\hat{\mathbf{F}}^s = (\hat{\mathbf{B}}^s(\hat{\mathbf{B}}^s)^T)^{-1}, s = 1, \dots, M\}$

This corresponds to fixing the variance of innovation terms of the MAR model.

Statistical Inference. The inference problem takes the form of multiple regression:

$$y_k = b_{k,0} + \sum_{j=1}^N b_{k,j}x_j, \tag{9}$$

where b_0 is the intercept, b_j are the coefficients that encode the relationship between each functional connection y_k and structural connectivity, $(x_j)_{j=1 \dots N}$.

We turn now to the solution of the multiple regression problem Eq. (9), where the quantities to be predicted are the coefficients of \mathbf{B} . We use the Least Absolute Shrinkage and Selective Operator (Lasso), which performs both variable selection and prediction [28]. Over classical least square regression Lasso offers two major advantages that are very useful in modeling brain connectivity: Firstly, it improves prediction by setting some coefficients to zero. This results in removing noisy and irrelevant variables and thus reducing the total variance. Secondly, it allows the selection of the most relevant variables and thus it links each functional connection with a subset of structural connection in a data driven way. The correct predictors are identified with high probability even when the number of variables is higher than the number of observation under the assumption that the true model is sparse [6]. The Lasso estimate is defined as:

$$b_L = \arg \min_b \left\{ \sum_{i=i}^M (y_i^k - b_0 - \sum_j^N b_j x_{i,j})^2 + \lambda \sum_{j=1}^N |b_j| \right\} \tag{10}$$

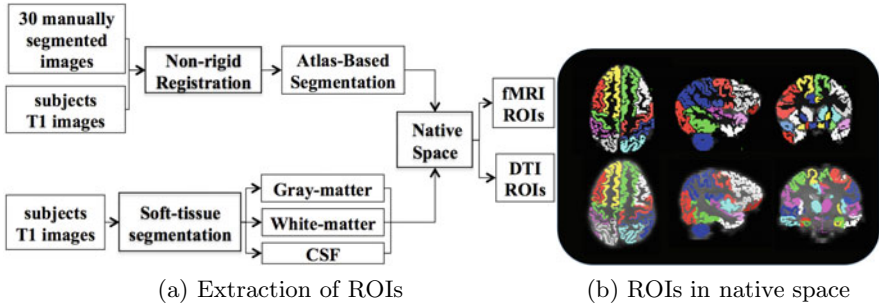


Fig. 2. a) The combination of atlas-based segmentation and soft-tissue segmentation allows the delineation of anatomically sensible ROIs in gray matter. Atlas-based segmentation is based on established work [1], where 30 manual segmentations are propagated to the new subjects by non-rigid registration [25]. SPM has been used for soft-tissue segmentation [11]. b) The resulted 83 ROIs are shown both in diffusion space (top row) as well as in fMRI space (bottom row). Note that here we used only cortical regions.

Because of the ℓ_1 lasso penalty, most of the coefficients are set to zero. We used the LARS implementation of the Lasso algorithm in R statistics [7], which computes the complete lasso solution simultaneously for all values of the shrinkage parameter, λ .

4 Experimental Results: Inter-subject Prediction of Resting-State Functional Connectivity

Brain connectivity analysis was performed in 22 normal adults (11 females, 11 males, mean age 33.4 ± 11.37 years). rs-fMRI: T2*-weighted gradient EPI sequence, TR/TE=2000/30, 31 ascending slices with thickness 3.25mm , gap 0.75mm , voxel size $2.5 \times 2.5 \times 4\text{mm}$, flip angle 90° , FOV $280 \times 220 \times 123\text{mm}$, matrix 112×87 . DWI: 64 non-collinear directions, in 72 slices, slice thickness 2mm , FOV 224mm , matrix 128×128 , voxel size $1.75 \times 1.75 \times 2\text{mm}^3$, b-value 1000 s/mm^2 . High resolution T1-weighted whole-brain structural images were also obtained in all subjects.

FSL was used for image pre-processing of both diffusion weighted (DWI) and fMRI images [26]. This involved eddy current correction of DWI and motion correction as well as spatial smoothing and whitening of fMRI images. Brain extraction was performed originally with FSL and it was manually refined later. Bias correction was applied to T1 and B0 images to improve the robustness of the non-rigid registration tools [25].

BOLD fluctuations are profound in gray matter, while DTI is more reliable in delineating white matter fibers. Hence, we are interested in defining cortical regions of interest (ROIs) that are located in gray matter and they are defined according to widely used anatomical atlas [1]. Cortical parcellation is obtained with the fusion of atlas-based [1] and tissue based segmentation [11], Fig. 2.

Table 1. Prediction performance under different scenarios

Methods	Cor(LW)	Chol Cor(LW)	SP MAR (L)	SP MAR(R)
log-likelihood	NA	194.57	185.13	232.4128
d	38.61	22.32	16.74	8.96

To construct corresponding functional networks the fMRI signal was averaged across voxels within each area. The signal in CSF and white matter has been also averaged and the six motion correction parameters were estimated with FEAT, FSL [26]. All these eight parameters were accounted in the estimation of the covariance matrix.

Tracts between regions are identified using a standard probabilistic algorithm available as part of FSL [26,2]. However, measurements of connection probability are difficult to interpret as the probability measure is very sensitive to noise in the data, as well as the size and separation of the ROIs. Instead, we estimate the local diffusion anisotropy by determining the diffusive transfer between voxels using the orientation distribution function (ODF) [24].

Model selection framework. Here we utilise the probabilistic framework described in section 3 to learn the link between anatomical connectivity and functional brain connectivity. We use leave-one-out cross validation and the algorithm described in Alg. 1 to measure the performance of the model. The results are

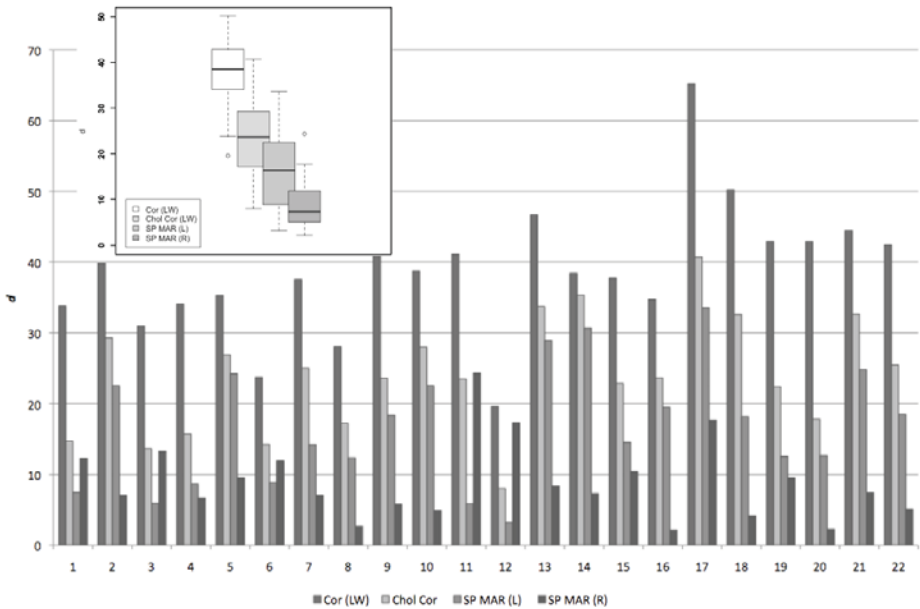


Fig. 3. Performance of the different probabilistic models under leave-one-out validation. Inside the main figure is the Box-and-Whisker diagram over all subjects.

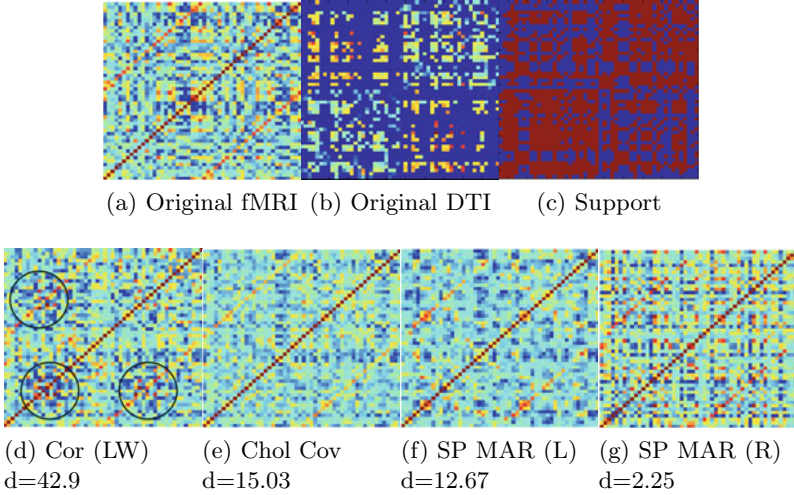


Fig. 4. Qualitative results for one leave-one-out subject when its structural connectivity matrix is used as a predictor. a) Sampled covariance matrix of the original fMRI signal, b) Structural connectivity matrix used as the input to the trained model. c) Sparsity pattern over all other subjects, d) Prediction based on Cor (LW). Black circles point out to relatively large errors in prediction for several connections, e) Prediction based on the Chol cov, f) Prediction based on the SP MAR (L), g) Prediction based on the SP MAR (R).

compared to the univariate linear model described in Eq. (11) with and without the Cholesky decomposition. Quantitative analysis is performed based on both the log-likelihood measure Eq. (12) as well as the d metric described in Eq. (2). The log-likelihood gives the likelihood to observe Σ^{test} given the model Σ^{pred} , and it relies on the underlying model of functional connectivity.

$$L(\Sigma^{\text{pred}}|\Sigma^{\text{test}}) = -\log \det(\Sigma^{\text{pred}}) + \text{tr}((\Sigma^{\text{pred}})^{-1}\Sigma^{\text{test}}) \tag{11}$$

When the prediction model do not use a matrix square root then the output it is not a SPD matrix, Sym_p^+ .

In Table 1, the mean values over all leave-one-out cross validations are given. Cor (LW) corresponds to predicting directly the functional correlation matrix based on Ledoit-Wolf regularisation [18], which provides a more accurate and well-conditioned covariance matrix. Chol Cor (LW) corresponds to predicting the square root matrix of the correlation matrix, Cor (LW). The original correlation matrix has been reordered to be consistent with our methodology (see section 3). Elements on the diagonal are set to one by multiplying with the inverse of the diagonal from the right, Eq. 8. Subsequently, results of our approach, SP MAR (R), as well as a variation, SP MAR (L), where Eq. 8 takes the form ($\tilde{\mathbf{B}} = \text{diag}(\mathbf{B})^{-1} \mathbf{B}$), are presented. Fig. 3 shows a more detailed summary of the results over all subjects. The probabilistic framework described in section 3 outperforms all other approaches. The difference between SP MAR (R) and SP

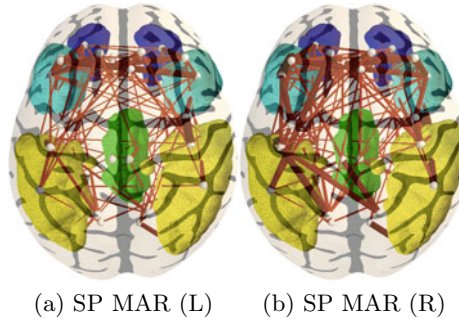


Fig. 5. Identifying structural connections associated with the default mode network. With yellow is represented the lateral parietal cortex, green areas represent the posterior cingulate gyrus (PCC), blue and light blue represent the medial prefrontal and orbito-frontal areas, respectively.

MAR (L) is that one scales the innovation terms of the MAR model, whereas the other scales the relative weights of the observed time series. The innovation terms are not related to connectivity and hence it is not possible to predict them from anatomical connectivity. Fig. 4 shows results distribution obtained in leave-one-out cross-validation scheme. In the simple correlation model, Cor (LW), we use black circles to point out that there is a relatively large error in prediction for several connections.

Finally, in Fig. 5, an application of the predictive framework in identifying structural connections associated with the the default mode network is shown. Connections are plotted when they appear in more than 15 out of 22 folds of the leave-one-subject-out cross validation loop. The diameter of the tubular shapes depends on the average of the absolute value of the coefficient associated with each connection. We note that stronger and more dense connections are identified in the SP MAR (R) relative to the SP MAR (L) model. This indicates that the earlier identifies connections and the associated coefficients more consistently. Also in SP MAR (L) model the PCC is not as well interconnected as in SP MAR (R) model.

5 Conclusions

We have developed a novel probabilistic framework to learn the mapping f from the space of anatomical connectivity \mathcal{A} to the space that describes functional connectivity \mathcal{F} . Functional connectivity is modeled as the covariance structure of functional brain signals. We assume that all the individual precision matrices share the same structure of conditional independence. Cholesky decomposition is used to enforce a SPD, Sym_p^+ , prediction output of the multiple regression. Leave-one-subject-out cross validation demonstrates the efficiency of the technique. We also depict the default mode network and the most influential structural connections associated to it.

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(3), 726–738 (2009)
2. Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S.: Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magnet Reson Med.* 50(5), 1077–1088 (2003)
3. Burns, J.: An evolutionary theory of schizophrenia: Cortical connectivity, metarepresentation, and the social brain. *Behavioral and Brain Sciences* 27(6), 831 (2004)
4. Damoiseaux, J., Greicius, M.: Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Struct. Funct.* 213(6), 525–533 (2009)
5. Deligianni, F., Robinson, E., Beckmann, C., Sharp, D., Edwards, A., Rueckert, D.: Inference of functional connectivity from structural brain connectivity. In: ISBI, pp. 1113–1116 (2010)
6. Donoho, D.: For most large underdetermined systems of linear equations the minimal l_1 -norm solution is also the sparsest solution. *Comm. Pure Appl. Math.* 59(6), 797–829 (2006)
7. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Stat.* 32(2), 407–499 (2004)
8. Ferrarelli, F., Massimini, M., Sarasso, S., Casali, A., Riedner, B., Angelini, G., Tononi, G., Pearce, R.: Breakdown in cortical effective connectivity during midazolam-induced loss of consciousness. *P. Natl. Acad. Sci. Usa* 107(6), 2681–2686 (2010)
9. Förstner, W., Moonen, B.: A metric for covariance matrices. *Qua. Vadis Geodesia*, 113–128 (1999)
10. Fransson, P., Marrelec, G.: The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: Evidence from a partial correlation network analysis. *Neuroimage* 42, 1178–1184 (2008)
11. Friston, K.: *Statistical parametric mapping: the analysis of functional brain images.* Academic Press, London (2007)
12. Greicius, M., Supekar, K., Menon, V., Dougherty, R.: Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* (2008)
13. van den Heuvel, M.P., Mandl, R.C.W., Kahn, R.S., Pol, H.E.H.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping* 30(10), 3127–3141 (2009)
14. Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J., Meuli, R., Hagmann, P.: Predicting human resting-state functional connectivity from structural connectivity. *P. Natl. Acad. Sci. Usa* 106(6), 2035–2040 (2009)
15. Honey, C., Kötter, R., Breakspear, M., Sporns, O.: Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *P. Natl. Acad. Sci. Usa* 104(24), 10240 (2007)
16. Lauritzen, S.: *Graphical models.* Oxford University Press, USA (1996)
17. LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., Huang, F.L.: Energy-based models. In: BakIr, G., Hofmann, T., Schölkopf, B. (eds.) *Predicting Structured Data*, pp. 191–245. MIT Press, Cambridge (2007)

18. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411 (2004)
19. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *J. Math. Imaging Vis.* 25, 423–444 (2006)
20. Morcom, A., Fletcher, P.: Does the brain have a baseline? why we should be resisting a rest. *Neuroimage* 37(4), 1073–1082 (2007)
21. Müller, R.: The study of autism as a distributed disorder. *Ment. Retard. Dev. Disabil. Res.* 13(1), 85–95 (2007)
22. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. *Int. J. Comput. Vision* 66, 41–66 (2006)
23. Pollonini, L., Pophale, S., Situ, N., Wu, M.H., Frye, R., Leon-Carrion, J., Zouridakis, G.: Information communication networks in severe traumatic brain injury. *Brain Topogr.* 23(2), 221–226 (2010)
24. Robinson, E., Hammers, A., Ericsson, A., Edwards, A., Rueckert, D.: Identifying population differences in whole-brain structural networks: a machine learning approach. *Neuroimage* 50(3), 910–919 (2010)
25. Rueckert, D., Sonoda, L., Hayes, C., Hill, D.: Non-rigid registration using free-form deformations: application to breast mr images. *IEEE Trans. Med. Imag.* 18, 712–721 (1999)
26. Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., Luca, M.D., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J., Matthews, P.: Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage* 23, 208–219 (2004)
27. Smith, S., Miller, K., Salimi-Khorshidi, G., Webster, M., Beckmann, C., Nichols, T., Ramsey, J., Woolrich, M.: Network modelling methods for fMRI. *Neuroimage* (2010) (in press)
28. Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B* 58(1), 267–288 (1996)
29. Varoquaux, G., Gramfort, A., Poline, J.B., Thirion, B.: Brain covariance selection: better individual functional connectivity models using population prior. In: *NIPS* (2010)
30. Venkataraman, A., Rathi, Y., Kubicki, M., Westin, C.F., Golland, P.: Joint generative model for fmri/dwi and its application to population studies. *Med. Image Comput. Comput. Assist Interv.* 13(pt. 1), 191–199 (2010)

Optimal Initialization for 3D Correspondence Optimization: An Evaluation Study

Matthias Kirschner^{1,*}, Sebastian T. Gollmer^{2,*},
Stefan Wesarg¹, and Thorsten M. Buzug²

¹ Graphisch-Interaktive Systeme, Technische Universität Darmstadt,
Fraunhofer Straße 5, 64283 Darmstadt, Germany

{matthias.kirschner, stefan.wesarg}@gris.tu-darmstadt.de

² Institute of Medical Engineering, University of Lübeck,
Ratzeburger Allee 160, 23538 Lübeck, Germany

{gollmer, buzug}@imt.uni-luebeck.de

Abstract. The identification of corresponding landmarks across a set of training shapes is a prerequisite for statistical shape model (SSM) construction. We automatically establish 3D correspondence using one new and several known alternative approaches for consistent, shape-preserving, spherical parameterization. The initial correspondence determined by all employed methods is refined by optimizing a groupwise objective function. The quality of all models before and after optimization is thoroughly evaluated using several data sets of clinically relevant, anatomical objects of varying complexity. Correspondence quality is benchmarked in terms of the SSMs' specificity and generalization ability, which are measured using different surface based distance functions.

We find that our new approach performs best for complex objects. Furthermore, all new and previously published methods of our own allow for (i) building SSMs that are significantly better than the well-known SPHARM method, (ii) establishing quasi-optimal correspondence for low and moderately complex objects without additional optimization, and (iii) considerably speeding up convergence, thus, providing means for practical, fast, and accurate SSM construction.

Keywords: Statistical Shape Models, Parameterization, Correspondence, Optimization, Evaluation.

1 Introduction

Statistical Shape Models (SSMs) [3] represent the shape variability of a particular object class on the basis of a representative training population. The strength of SSMs is their ability to adapt elastically to previously unseen, patient specific shape instances under the constraint of statistical plausibility. This makes them applicable to a wide variety of applications in medical imaging, such as image segmentation, shape analysis and shape extrapolation [11].

* Contributed equally.

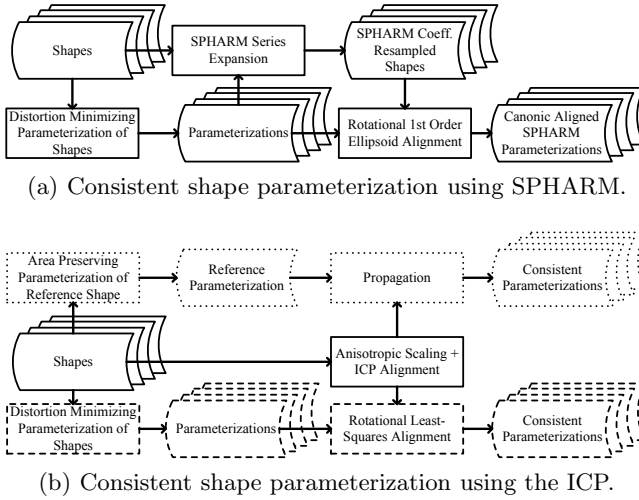


Fig. 1. Algorithms for consistent shape parameterization. Based thereon, an SSM is built using regular icosahedron remeshing and generalized Procrustes analysis.

An SSM integrates the shape variability of a set of training shapes represented by vectors of landmarks with index correspondence. Hence, the elementary prerequisite for SSM construction is to identify corresponding points (so-called landmarks) across all training samples, which creates a strong demand for automatic methods. This is especially the case in the 3D setting, where several thousands of landmarks must be placed consistently. Solving this *correspondence problem* turns out to be the most challenging part of SSM construction. Firstly, a mathematical description is required which formalizes the concept of point correspondence. The Minimum Description Length (MDL) approach [6] establishes correspondence using information theoretical concepts. Its appeal is the definition of optimal *groupwise* correspondence by the properties of the resulting model, thus incorporating the whole shape information contained in the training population. This stands in contrast to the great majority of other approaches, which typically rely on some sort of registration (cf. [11] for an overview).

Secondly, an optimization procedure is required for minimizing the objective function. In order to optimize the model's DL in 3D, the reparameterization approach is employed [5]. To this end, parametric shape representations have to be constructed in the first place, actually initializing the optimization procedure.

Review of Shape Parameterizations. The embedding of the training shapes into a parameter domain (here: unit sphere \mathbb{S}^2) is a prominent possibility for establishing correspondence [14, 16]. The unit sphere is chosen since a large variety of organs have spherical topology. Selecting a suitable parameterization algorithm is crucial for the optimization. Two possibly conflicting aspects must be taken into account: Minimization of area and angle distortion and consistency.

Minimization of distortion is crucial for the reconstruction of shapes from the parameter space. For example, purely conformal (angle-preserving) mapping leads to reconstructions of poor quality when using uniform parameter space sampling [9], which can only be partially alleviated by means of adaptive sampling methods [12]. Therefore, combined minimization of area and angle distortion (as for example in [2]) is the method of choice.

On the other hand, consistency means that similar features on different shapes are mapped to approximately the same spherical regions. As the initial parameterizations define the starting point of the optimization procedure, consistency speeds up convergence or even ensures convergence at all [7]. Consistency can be enforced by rigidly aligning the parameterizations after their construction. To this end, the frequently used SPHARM approach [4,5,14] considers the first order spherical harmonic coefficients of the shapes (Fig. 1(a)) [2]. Alternatively, one can align the shapes using the Iterative Closest Point (ICP) algorithm [1] and use these approximate correspondences to derive a similarity measure. This measure can then be either explicitly optimized as a postprocessing step after all shapes have been parameterized [7], or directly incorporated into the parameterization algorithm [15] (Fig. 1(b), upper path). The latter approach has been shown to be faster and produce better parameterizations [15].

Outline and Contribution. While the latter two approaches allow for establishing consistency non-rigidly, distortions in the parameter domain are possibly biased wrt. the reference parameterization. Therefore, our first contribution is an approach for establishing quasi-distortion-free consistent parameterizations (Sec. 2.2). Based on this, two different parameterizations are constructed. We compare these two consistent parameterization approaches with the SPHARM [2] and the propagation [15] approach in context of correspondence optimization. We quantitatively measure specificity and generalization ability of the SSMS computed from the parameterizations before and after optimization using the MDL function (Sec. 2.3). In this way, we aim to clarify to what extent optimization allows for further improving initially “good” correspondence, which is done by building SSMS for four different data sets comprising populations of varying complexity, size and shape variability. Finally, and in contrast to other evaluation studies [4,5,17], we use surface based metrics in our evaluation, thus adequately accounting for sampling errors in the parameterizations.

2 Materials and Methods

2.1 Statistical Shape Models

Let a set of $n_s \in \mathbb{N}^+$ shapes $\{S_i \in \mathcal{X} \subset \mathbb{R}^3 : i = 1, \dots, n_s\}$ of a certain object class be given, represented as piecewise-linear surfaces consisting of triangles and vertices. Here, \mathcal{X} is the set of all shape instances. The shapes are aligned in a common frame and sampled at corresponding positions using $n_p \in \mathbb{N}^+$ landmarks $\mathbf{x}_{i(j)} \in S_i$, $j = 1, \dots, n_p$. Hence, the i -th shape can be represented by a vector $\mathbf{x}_i \in \mathbb{R}^{3n_p}$. The object’s shape variability is captured by a principal component

analysis (PCA): We compute the mean shape $\bar{\mathbf{x}}$ as well as the $n_m \in \mathbb{N}$ principal eigenvectors $\{\mathbf{p}_m\}$ and associated eigenvalues $\{\lambda_m\}$ of the covariance matrix of the training shapes. Then, the linear model $\mathbf{x}^* = \bar{\mathbf{x}} + \sum_{m=1}^{n_m} \mathbf{p}_m b_m^*$ can be used to approximate a shape instance $S^* \in \mathcal{X}$ with shape parameters $\{b_m^* \in \mathbb{R}\}$.

2.2 Quasi-distortion-Free Consistent Parameterization

Our goal is to compute consistent spherical shape parameterizations with minimal distortion. Let $\omega_i : \mathbb{S}^2 \rightarrow S_i$ be a shape-preserving spherical parameterization, which defines a one-to-one correspondence between the spherical surface $\mathcal{P}_{S_i} \in \mathbb{S}^2$ and S_i . We consider the embedding of a surface mesh S_i , $i = 1, \dots, n_s$ into the space \mathbb{S}^2 as a constrained optimization problem [2]. The variables to be optimized so as to minimize angle and area distortions are the spherical coordinates of the vertices $\tilde{\mathcal{V}}_i = \{\omega_i^{-1}(\mathbf{v}_{i(\iota)}) \in \mathbb{R}^3 \mid \|\omega_i^{-1}(\mathbf{v}_{i(\iota)})\|_2 = 1, \iota = 1, \dots, n_v\}$ of \mathcal{P}_{S_i} , where $\{\mathbf{v}_{i(\iota)} \in \mathbb{R}^3 \mid \iota = 1, \dots, n_v\} := \mathcal{V}_i$ are the $n_v \in \mathbb{N}^+$ vertices of S_i .

To enforce consistency of the parameterizations, they are aligned afterwards. Implicit alignment using the SPHARM method [14] is not always an optimal strategy (cf. Sec. 4). A better approach is to explicitly align the original shapes using the ICP algorithm, and align the parameterizations accordingly [8] (Fig. 1(b), lower path). Because the ICP alignment is poor when the shapes differ in scale, we additionally use the method we presented in [15] for anisotropic scaling: We (a) center all surfaces $\{S_i\}$, (b) select a reference surface S_{ref} , $\text{ref} \in \{1, \dots, n_s\}$, (c) scale the remaining shapes anisotropically such that their extent along their principal axes matches the corresponding axes of S_{ref} , and (d) use the ICP algorithm [1] to align them with S_{ref} . Now, the surfaces $\{\mathcal{P}_{S_i}\}$ are optimally transformed wrt. $\mathcal{P}_{S_{\text{ref}}}$ without changing the individually optimized triangulations: We compute the $n_v \in \mathbb{N}$ closest points $\mathcal{U}_i = \{\mathbf{u}_{i(\iota)} \in \mathbb{R}^3 \mid \iota = 1, \dots, n_v\}$ on the ICP-aligned surface S_i for all vertices of S_{ref} and find $\tilde{\mathcal{U}}_i = \{\omega_i^{-1}(\mathbf{u}_{i(\iota)}) \in \mathbb{R}^3 \mid \|\omega_i^{-1}(\mathbf{u}_{i(\iota)})\|_2 = 1, \iota = 1, \dots, n_v\}$ through barycentric interpolation of \mathcal{P}_{S_i} . Note that, in contrast to [7, 15], \mathcal{U}_i does not necessarily coincide with the mesh vertices of S_i . Finally, \mathcal{P}_{S_i} is rotated so as to match the points $\tilde{\mathcal{U}}_i$ wrt. the vertices $\tilde{\mathcal{V}}_{\text{ref}}$ in the least-squares sense [13].

2.3 Groupwise Correspondence Optimization

The computed parameterizations define an initial correspondence between the shapes. This initial correspondence is further refined by explicit optimization. We use the continuous approximation of the MDL-function [5]

$$\begin{aligned} \mathcal{L}_{\text{mdl}}(\Delta) = f(n_s, R, \Delta) &+ \sum_{i:\lambda_i \geq \lambda_{\min}} (n_s - 2) \log \sqrt{\lambda_i} \\ &+ \sum_{i:\lambda_i < \lambda_{\min}} \left[(n_s - 2) \log \sqrt{\lambda_{\min}} + \frac{n_s \lambda_i}{2\lambda_{\min}} \right], \quad (1) \end{aligned}$$

where Δ is the coding accuracy of the data, R is an upper bound of the spatial range of the data, and $f(n_s, R, \Delta)$ is a function which is constant for a given

data set and given Δ . Moreover, we define $\lambda_{\min} = 4\Delta^2$. We use the method of Davies et al. [5] for computing the objective function, and integrate it over a range $[\Delta_{\min}, \Delta_{\max}]$ of values for Δ (here: $[\Delta_{\min}, \Delta_{\max}] = [0.0001 \text{ mm}, 0.5 \text{ mm}]$).

Our optimization starts by aligning the parameterized input shapes in a common coordinate frame by centering the shapes, scaling them and aligning them using the ICP algorithm [1]. Scaling is handled by normalizing the average root mean square (RMS) distance of the vertices to the origin to 1.

After this initial registration, the pose parameters remain fixed during correspondence optimization. For remeshing landmark vectors from the parameterizations, a set of sampling points is required for each shape. We iteratively subdivide an icosahedron and project its points to the unit sphere. For each shape, a copy of the n_p points of the subdivided icosahedron serves as initial sampling points. In each iteration of the optimization, we select one shape uniformly at random, and manipulate its sampling points in order to optimize our objective function. For manipulating the sampling points, we use Clamped Plate Spline Warps [7]. The optimization stops when the objective function value does not improve by more than a small threshold $\varepsilon \in \mathbb{R}^+$ (here: $\varepsilon = 0.0001$) within n_s iterations.

2.4 Correspondence Quality Measures

Specificity and *generalization ability* [4] are widely used measures [5,12,17] for quantifying the correspondence quality. Both criteria estimate the similarity of different shapes by means of a particular metric $\mathcal{D} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and depend on the number $n_m \in \mathbb{N}$ of shape parameters $\{b_m^* \in \mathbb{R}, m = 1, \dots, n_m\}$.

The specificity of an SSM is measured by randomly drawing a large number of n_r (here: $n_r = 1000$) shapes from the multivariate Gaussian distribution defined by the SSM and estimating the similarity of the reconstructions $\{S_k^*(n_m) \in \mathcal{X}\}$ with the shape samples $\{S_i \in \mathcal{X}\}$. Thus, an SSM with high specificity (low values of \mathcal{S}) always provides valid shape instances of the particular object class, where

$$\mathcal{S}(n_m) = \frac{1}{n_r} \sum_{k=1}^{n_r} \min_{\{i=1, \dots, n_s\}} \mathcal{D}(S_k^*(n_m), S_i). \quad (2)$$

A model with good generalization ability (small values of \mathcal{G}) is able to represent unknown shapes, which can be measured by a series of leave-one-out tests: By omitting the i -th training shape S_i an SSM is built and subsequently used to generate the reconstruction S_i^* of S_i . Averaging over the series of tests yields

$$\mathcal{G}(n_m) = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathcal{D}(S_i^*(n_m), S_i). \quad (3)$$

Landmark based distance functions that can be used as a metric in (2) and (3) are the sum of squared differences between corresponding points [4] or their mean absolute distance [17]. Since landmark based metrics neglect the actual shape similarity and thus potentially falsify the values of \mathcal{S} and \mathcal{G} [9], volume based metrics [12] may be used instead. However, surface based metrics allow for measuring global and local shape differences more precisely. To this end, we

Table 1. Collection of data sets used for SSM construction

	Ventricle	Hippocampus	Liver	Striatum
Origin of shape variability	card. cycle	multi-subj.	multi-subj.	multi-subj.
Isotropic voxel size	1.25 mm	0.50 mm	2.00 mm	0.50 mm
Imaging modality	MRI	MRI	CT	MRI
Population size (# of shapes)	25	42	20	18
Shape complex. (# of vertices)	4244-9630	6790-10380	25246-48352	18316-23116
Observed shape variability	low	medium	high	very high
# of SPHARM degrees/vertices	15/2562	15/4002	18/6252	18/6252
SSM complex. (# of landmarks)	1002	1692	2562	2562

compute values for \mathfrak{S} and \mathfrak{G} wrt. the symmetric RMS surface distance ($\mathcal{D}_{\text{SRMS}}$) and, in order to account for reconstruction errors introduced by distortions in the parameterizations, the symmetric Hausdorff (\mathcal{D}_{SH}) distance following [9]:

$$\mathcal{D}_{\text{SRMS}}(S^*, S_i) = \sqrt{\frac{1}{|S^*|} \int_{\rho^* \in S^*} \delta(\rho^*, S_i)^2 dS^* + \frac{1}{|S_i|} \int_{\rho \in S_i} \delta(\rho, S^*)^2 dS_i}, \quad (4)$$

$$\mathcal{D}_{\text{SH}}(S^*, S_i) = \max(\max_{\rho^* \in S^*} (\delta(\rho^*, S_i)), \max_{\rho \in S_i} (\delta(\rho, S^*))). \quad (5)$$

Thereby, $\delta(\rho', S) = \min_{\rho \in S} \|\rho' - \rho\|_2$ is the distance of an arbitrary point $\rho' \in S'$ to the shape S with surface area $|S|$. Moreover, we define $S^* := S_k^*(n_m)$ and $S_i := S_i^*(n_m)$ when (4) and (5) are inserted into (2) and (3), respectively.

2.5 Data Sets

We built SSMs for four different structures/organs that are of medical relevance (cf. Table I). All of them originate from clinical imaging modalities and have been either segmented manually by expert observers, or, in case of the left heart ventricle, automatically. The segmented volume data were resampled to isotropic voxel size and smoothed using a Gaussian kernel. Surface meshes have been extracted from the preprocessed binary images and moderately smoothed to obtain staircase artifact free, visually eligible shape representations. We used the following publicly available data: Hippocampus dataset of the SPHARM-PDM UNC Toolbox (<http://www.nitrc.org/projects/spharm-pdm/>), liver dataset from the MICCAI liver-segmentation challenge [10] (<http://sliver07.org>), and brain segmentations provided by the Center for Morphometric Analysis (<http://www.cma.mgh.harvard.edu/ibsr/>) to generate the striatum region.

2.6 Experiments

Our experiments are designed to answer the question which parameterizations provide the best compromise between consistency and small area and angle distortion. While we could measure distortions directly on the parameterizations, there is no obvious way to measure consistency in the parameter domain. We therefore measure both properties on the computed SSMs: By using two surface based metrics to compute \mathfrak{S} and \mathfrak{G} (Sec. 2.4), we can coevally measure the correspondence quality and detect the loss of detail caused by distortions.

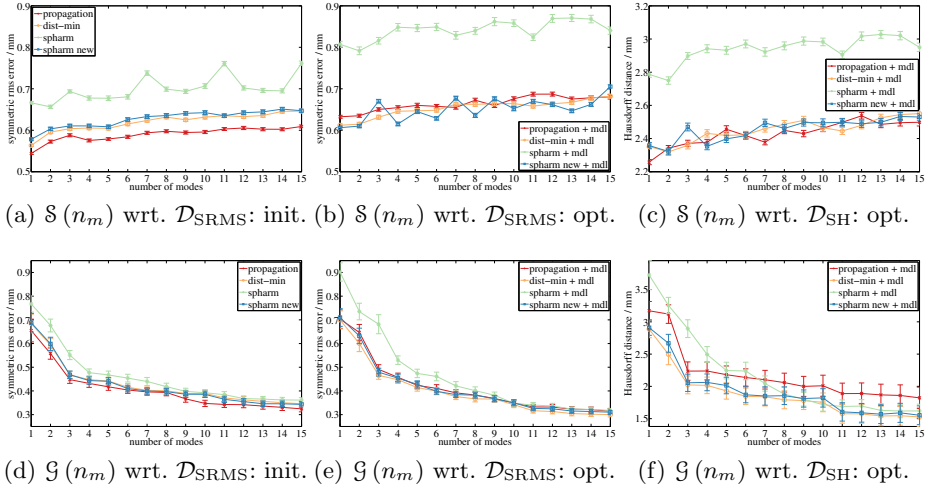


Fig. 2. Correspondence quality of the ventricle SSMs

For each data set, correspondence was established in eight different ways. That is, we parameterized each shape with the distortion minimizing parameterization method of Brechbühler [2] and used the novel approach (Sec. 2.2) in order to establish consistency, which we call *dist-min*. We also computed SPHARM coefficient resampled shapes and parameterizations and aligned them (i) using the first order spherical harmonics coefficients [14] (*spharm*) and (ii) with our new approach (*spharm new*). This allows us to quantify the effect of the regular resampling in the SPHARM approach. Finally, we computed consistent parameterizations with our previously published *propagation* method [15]. Thus, we have four different parameterizations for each shape. For each set of parameterizations, an SSM was constructed by applying uniform sampling with n_p sampling points. Moreover, the positions of the initial sampling points were refined by optimizing the MDL objective function, as described in Sec. 2.3. With the refined sampling points, we again sampled shape representations and built an SSM. The number of sampling points and the degree of SPHARM coefficients have been selected dependent on the complexity of the data sets, such that the original shapes are preserved without significant loss of detail. Our choices are listed in Table 1. Each of the eight resulting SSMs we constructed for a single data set was evaluated using the surface based metrics (4) and (5) (Sec. 2.4).

3 Results

Fig. 2-5 show the quantitative evaluation of the model quality for each data set in terms of \mathcal{S} and \mathcal{G} before and after groupwise correspondence optimization (referred to as “init.” and “opt.” in the caption of the subfigures). The values for \mathcal{S} and \mathcal{G} are illustrated using the metrics \mathcal{D}_{SRMS} (4) and \mathcal{D}_{SH} (5). The maximum error computed using \mathcal{D}_{SH} before correspondence optimization has been omitted

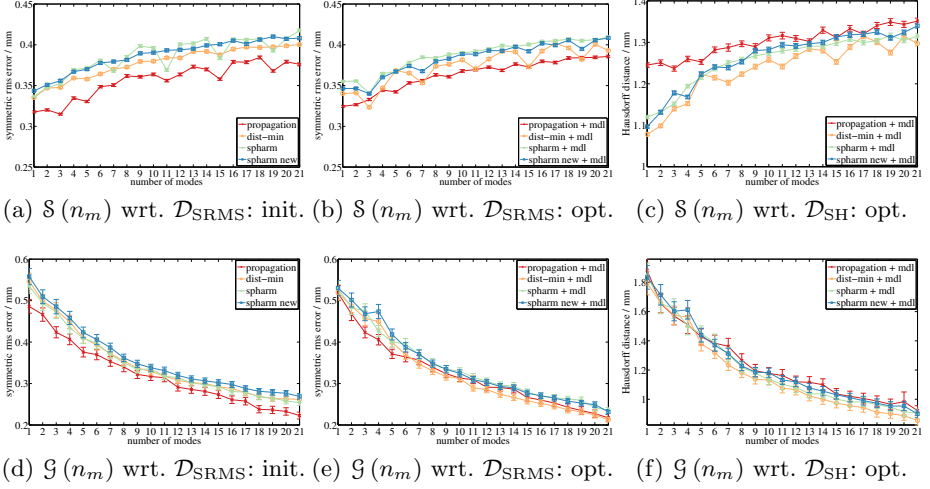


Fig. 3. Correspondence quality of the hippocampus SSMs

Table 2. Correspondence optimization for two exemplary data sets

		propagation	dist-min	spharm	spharm new
Hippo-campus	DL (init.)	-124.36742	-105.02072	-111.39473	-108.60463
	DL (opt.)	-127.10645	-118.04980	-119.90849	-120.67680
	# of iterations	10164	43344	18942	37422
	time / min.	448	1845	637	1248
Liver	DL (init.)	-1.81933	-1.34608	-1.17675	-1.36526
	DL (opt.)	-2.36768	-2.31014	-2.11941	-2.24933
	# of iterations	15760	25500	23660	18960
	time / min.	1447	2096	410	330

for brevity, since it does not provide additional information. Table 2 summarizes figures of the optimization of the DL (1) for two exemplary data sets of varying complexity and Fig. 6 shows shape reconstructions for the complex striatum.

In case the metric $\mathcal{D}_{\text{SRMS}}$ is used for measuring \mathcal{S} (2) and \mathcal{G} (3), we observe:

1. *Dist-min*, *spharm new*, and *propagation* achieve on most data sets smaller values for \mathcal{S} and \mathcal{G} than *spharm* ((a),(b),(d),(e)) in Fig. 2(4/5) whereas the difference is especially large for \mathcal{S} . Only on the hippocampus data set *spharm* performs equally well (Fig. 3(a),(b),(d),(e)).
2. Without optimization *propagation* has the smallest values for \mathcal{S} and \mathcal{G} ((a),(d)) in Fig. 2(3/4/5).
3. After optimization, values for \mathcal{S} are larger on the ventricle for all parameterizations (Fig. 2(b)), for *propagation* on the hippocampus and the liver data set (Fig. 4(b)), but smaller on the striatum data set for all parameterizations (Fig. 5(b)). Values for \mathcal{G} are smaller after optimization (Fig. 2(3/4/5(e))), whereas the effect is way smaller for the *propagation* method.
4. *Dist-min* achieves in most cases smaller values for \mathcal{S} and \mathcal{G} than *spharm new*, but the difference is only minor.

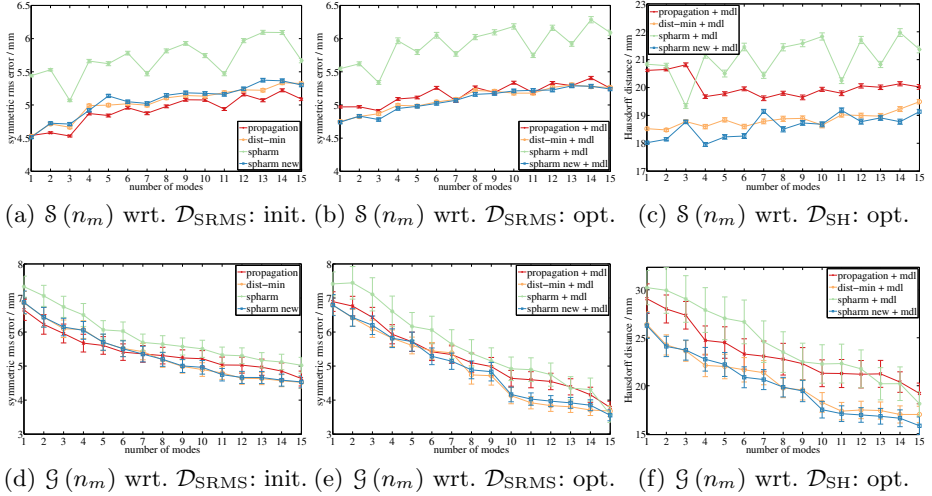


Fig. 4. Correspondence quality of the liver SSMs

Using the metric \mathcal{D}_{SH} for computing eq. (2) and eq. (3), we observe:

5. *Propagation* has well larger values for \mathcal{S} and \mathcal{G} than *dist-min* and *spharm new* (c)(f) in Fig. 2(3/4/5), except for \mathcal{S} of the ventricle data set (Fig 2(c)) and \mathcal{G} of the hippocampus data set (Fig 3(f)).
6. *Dist-min* and *spharm new* have smaller values for \mathcal{S} and \mathcal{G} than *spharm* (c)(f) in Fig. 2(3/4/5) except for the \mathcal{G} of the striatum data set (Fig. 5(f)).

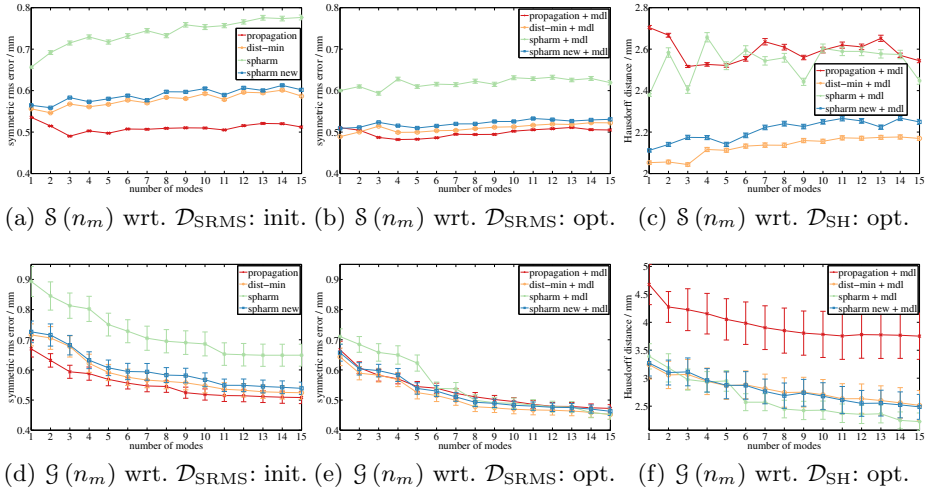


Fig. 5. Correspondence quality of the striatum SSMs

Further observations are:

7. The optimization has a different effect on different parameterization methods. In particular, its influence on the *propagation* method is rather small ((a) vs. (b) and (d) vs. (e) in Fig. 2/3/4/5).
8. The objective function values before and after optimization differ significantly for different parameterization methods (Table 2).
9. *Propagation* starts in all cases with a smaller objective function value and converges more quickly (Table 2).
10. Approaches with less complex parameterizations wrt. the number of vertices (*spharm*, *spharm new*) need significantly less computation time per iteration (Table 2).

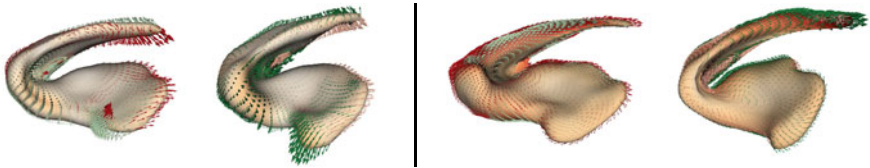


Fig. 6. Shape representations of the striatum generated by varying the 1st mode of the initial *propagation* (left two shapes) and *dist-min* correspondence by $\mp\sqrt{\lambda_1}$. Arrows indicate the landmarks’ moving direction, shape coloring the deviation from \bar{x} .

4 Discussion

We draw the following conclusions from the observations made in Sec. 3 as regards the use of different parameterization methods for correspondence optimization: Firstly (from 1., 6.), our parameterization approaches (*propagation*, *dist-min*, *spharm new*) are better suited for this application than *spharm*. Secondly (from 2., 7., 9.), *propagation* provides quite good initial correspondence. This can be attributed to the fuzzy correspondence that is inherently incorporated from the very beginning through the ICP algorithm. However, on the downside we must note (from 5.) that the heuristics used in the *propagation* approach for minimizing angle and area distortion introduce some artifacts, which manifest especially in distinct local errors. Thirdly (from 4., 10.), the reduction of the parameterizations’ complexity in terms of the number of vertices is a reasonable pre-processing step from a practical point of view, because it allows to speed up correspondence optimization considerably. Fourthly (from 3.), starting from a “good” initial correspondence as *propagation*, *dist-min*, and *spharm new* provide for obvious reasons, subsequent optimization is especially beneficial for objects with distinct shape variability, such as the liver and the striatum in our case, whereas it does not seem to be necessary for simpler object classes like the left ventricle and, to so some extent, the hippocampus.

However, we also note several issues regarding the correspondence optimization itself, which need to be discussed in more detail. Firstly, 5., 7., 9. indicate

that the optimization is hardly able to compensate for initially introduced artifacts. This can also be seen in the extreme case of conformal parameterizations, which must resort to some sort of adaptive sampling [12] to reconstruct shapes reasonably well. Conversely, quasi-distortion-free, consistent parameterizations such as *dist-min* and *spharm new* are especially well suited for optimization.

Furthermore, the 3. and 7. observation suggest that the MDL function is a rather complex objective with many local optima. Moreover (from 8.), SSMs with smaller objective function values do not necessarily perform better wrt. the quality measures \mathcal{S} and \mathcal{G} . For example, *propagation* has the lowest objective function value on the liver data set (Table 2), but *dist-min* and *spharm new* have better \mathcal{S} and \mathcal{G} (Fig. 4(b),(c)). This means that minimizing the DL does not necessarily improve the SSM quality in terms of the measures \mathcal{S} and \mathcal{G} , though (from 3., 7.), the SSM’s generalization ability seems to benefit slightly more in general.

5 Conclusions

In this paper, we devised a new method for computing consistent shape parameterizations for correspondence optimization. We evaluated our new method with the well known SPHARM approach [14] and a previously published method of our own [15] (*propagation*) in context of a state-of-the-art correspondence optimization algorithm. The evaluation shows that all our parameterization approaches are better suited for correspondence optimization than SPHARM. Moreover, the newly devised approach has advantages over *propagation*, because it has lower area and angle distortion. Thus, the maximal reconstruction error is kept small, and the parameterizations are better suited for optimization.

Evaluation results suggest that, starting from reasonable “good” initial correspondence, subsequent minimization of the model’s DL is not necessary for organ classes with relatively low shape variability, but pays off for classes with high complexity. However, we see that SSMs with small DL are not necessarily better in terms of the quality measures specificity and generalization ability. A reason for this might be that the MDL function does not penalize the inevitable loss of detail introduced when landmark vectors are sampled from the parameterizations. Therefore, it may be beneficial to add a distortion metric as a regularization term in the objective to guarantee high quality reconstructions.

The present study has focused on the influence of the parameterization strategy on the optimization results. It has been shown that additional optimization strategies such as using a multi-resolution approach or optimizing the pose parameters allow for further improvement [5]. Noting that there exist several simpler objective functions than the full MDL function, we plan in future to investigate to which degree the quality of the optimized correspondence depends on the choice of (i) one particular objective function, (ii) the parameterization methods and the optimization strategy, and (iii) the parameters to be optimized.

Acknowledgements. Hippocampus data was provided by M. Styner and co-workers, UNC Neuro Image Analysis Laboratory. Original data acquisition was

funded by the Stanley Foundation. Striatum data were generated from manual segmentations of MR brain data sets provided by the Center for Morphometric Analysis at Massachusetts General Hospital.

References

1. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 239–256 (1992)
2. Brechbühler, C., Gerig, G., Kübler, O.: Parametrization of Closed Surfaces for 3-D Shape Description. *Comput. Vis. Image Und.* 61, 154–170 (1995)
3. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active Shape Models – Their Training and Application. *Comput. Vis. Image Und.* 61, 38–59 (1995)
4. Davies, R.H., Twining, C.J., Allen, P.D., Cootes, T.F., Taylor, C.J.: Shape Discrimination in the Hippocampus Using an MDL Model. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 38–50. Springer, Heidelberg (2003), doi:10.1007/b11820
5. Davies, R.H., Twining, C.J., Cootes, T.F., Taylor, C.J.: Building 3-D Statistical Shape Models by Direct Optimization. *IEEE Trans. Med. Imag.* 29, 961–981 (2010)
6. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: A Minimum Description Length Approach to Statistical Shape Modeling. *IEEE Trans. Med. Imag.* 21, 525–537 (2002)
7. Davies, R.H., Twining, C.J., Taylor, C.J.: Consistent Spherical Parameterisation for Statistical Shape Modelling. In: 2006 IEEE ISBI, pp. 1388–1391 (2006)
8. Fripp, J., Bourgeat, P., Mewes, A.J.U.J., Warfield, S.K., Crozier, S., Ourselin, S.: 3D statistical shape models to embed spatial relationship information. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 51–60. Springer, Heidelberg (2005)
9. Gollmer, S.T., Buzug, T.M.: A Method for Quantitative Evaluation of Statistical Shape Models Using Morphometry. In: 2010 IEEE ISBI, pp. 448–451 (2010)
10. Heimann, T., van Ginneken, B., Styner, M., et al.: Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. *IEEE Trans. Med. Imag.* 28, 1251–1265 (2009)
11. Heimann, T., Meinzer, H.P.: Statistical Shape Models for 3D Medical Image Segmentation: A Review. *Med. Image Anal.* 13, 543–563 (2009)
12. Heimann, T., Wolf, I., Meinzer, H.: Automatic Generation of 3D Statistical Shape Models with Optimal Landmark Distributions. *Meth. Inf. Med.* 46, 275–281 (2007)
13. Horn, B.K.P.: Closed-form Solution of Absolute Orientation Using Unit Quaternions. *J. Opt. Soc. Am. A* 4, 629–642 (1987)
14. Kelemen, A., Székely, G., Gerig, G.: Elastic Model-Based Segmentation of 3-D Neuroradiological Data Sets. *IEEE Trans. Med. Imag.* 18, 828–839 (1999)
15. Kirschner, M., Wesarg, S.: Construction of Groupwise Consistent Shape Parameterizations by Propagation. In: Dawant, B., Haynor, D. (eds.) SPIE, vol. 7623 (2010)
16. Lamecker, H., Seebaß, M., Hege, H.C., Deuffhard, P.: A 3D Statistical Shape Model of the Pelvic Bone for Segmentation. In: Fitzpatrick, J., Sonka, M. (eds.) SPIE, vol. 5370, pp. 1341–1351 (2004)
17. Styner, M.A., Rajamani, K.T., Nolte, L.-P., Zsemlye, G., Székely, G., Taylor, C.J., Davies, R.H.: Evaluation of 3D correspondence methods for model building. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 63–75. Springer, Heidelberg (2003)

White Matter Bundle Registration and Population Analysis Based on Gaussian Processes

Demian Wassermann^{1,2,3}, Yogesh Rathi², Sylvain Bouix², Marek Kubicki², Ron Kikinis³, Martha Shenton², and Carl-Fredrik Westin¹

¹ Laboratory of Mathematics in Imaging, Brigham & Women's Hospital, Boston, MA

² Psychiatry and Neuroimaging Lab, Brigham & Women's Hospital, Boston, MA

³ Surgical Planning Lab, Brigham & Women's Hospital, Boston, MA

Abstract. This paper proposes a method for the registration of white matter tract bundles traced from diffusion images and its extension to atlas generation. Our framework is based on a Gaussian process representation of tract density maps. Such a representation avoids the need for point-to-point correspondences, is robust to tract interruptions and reconnections and seamlessly handles the comparison and combination of white matter tract bundles. Moreover, being a parametric model, this approach has the potential to be defined in the Gaussian processes' parameter space, without the need for resampling the fiber bundles during the registration process. We use the similarity measure of our Gaussian process framework, which is in fact an inner product, to drive a diffeomorphic registration algorithm between two sets of homologous bundles which is not biased by point-to-point correspondences or the parametrization of the tracts. We estimate a dense deformation of the underlying white matter using the bundles as anatomical landmarks and obtain a population atlas of those fiber bundles. Finally we test our results in several different bundles obtained from in-vivo data.

Keywords: Diffusion MRI, White Matter Fiber Tracts, Gaussian Processes, Registration.

1 Introduction

The analysis of inter-population brain variability through imaging is an area of extensive study. Within this area, generating a common coordinate space for analyzing several subjects is one of the main issues to be solved. The development of algorithms capable of registering scalar images, like those obtained from anatomical MRI, and volumes, like sub-cortical structures such as the hippocampus, yielded several effective techniques which enabled a wide range of statistical studies. However, existing tools for the registration of curve sets, such as white matter tract bundles obtained from diffusion MRI (dMRI), are still in need of development.

Current approaches to the registration of cerebral white matter tract bundles can be divided in two families: indirect and direct algorithms. The family of indirect methods starts by performing a full-brain registration. Then, the tract bundles are warped using the resulting deformation field obtained from this registration process. These methods use scalar images like the fractional anisotropy (FA) to obtain the deformation fields

[18,10]. A second set of methods within this family use directional information, such as the principal diffusion direction of the estimated tensors, to obtain the displacement vectors [22,21,6]. There are two main issues with this kind of registration when applied to the warping of tracts bundles. First, the continuity of tracts is not explicitly enforced resulting in tracts being cut or warped into unusual shapes. Second, if the continuity is enforced at a voxel (or supra voxel) level, the directional uncertainties within these voxels, produced by partial voluming or limitations of the diffusivity model will lead to a predominance of the most voluminous tracts, like the corona radiata, sectioning or eliminating the small ones like short cortico-cortical fasciculi.

The family of direct methods registers the tracts explicitly. Of these methods, some algorithms require fiber-to-fiber and point-to-point matching between the tracts [16,8]. These approaches are subject to tractography artifacts like discontinuities on tracts and parametrization differences. Other methods register volumetric representations of tracts. Ziyang et al. [23] model the white matter bundles as a voxelized mixture of spatial density functions, bounding his approach to an explicit resolution. Moreover, while registering several clusters of tracts, they only account for an affine transformation per cluster considering each cluster as a geometrical landmark to be affinely registered. This does not account for anatomical variability as in several cases, like uncinata fasciculi, the shape of the tract is largely variable between subjects [15]. Durrleman et al. [9] propose a sound diffeomorphic approach based on currents and LLDDM registration [5] and they incorporate a statistical analysis model. However, their approach strongly relies on the orientation of the tract parametrization. This notion of orientation introduced in their distance metric imposes a constraint to the model which is artefactual: diffusion imaging does not provide information about the orientation of the tracts, only their direction. It is not possible to distinguish between the case of an axonal package going from A to B or from B to A. Hence, it is not possible to calculate the orientation of the tracts, the need for an orientation [9] requires the user to reorient all the tracts to be registered consistently, a task which is not at all trivial.

In this work we propose a new approach to the direct diffeomorphic registration of white matter bundles. Our method has four main advantages: we register tract bundles directly; it does not rely on point-to-point nor fiber-to-fiber correspondences as [16,8]; it is not sensitive to inter-subject total density variations; and it does not depend on the fiber parametrization as [9]. We start by representing the white matter tracts as Gaussian processes (GPs) [20]. This representation associates each tract and each bundle of tracts with a GP mapping each point in space to the density of tracts crossing that point, a tract density map (TDM). It provides a framework where the similarity between two bundles is measured in terms of the mass of common density areas. This metric does not depend on point-to-point correspondences nor on the orientation of the tract parametrization. Moreover, it is calculated from the parameters of the GP without the need for explicit sampling the TDM. We use this similarity to derive a diffeomorphic registration algorithm based on the Log-Euclidean poly-affine framework [3]. Then we use this pairwise registration algorithm along with several desirable properties of our GP framework in order to develop a template estimation algorithm along with a methodology to analyze the characteristic anatomical variations of a population. Finally, we test our algorithms on two different sets of white matter bundles in order illustrate its efficacy.

2 Methods

In this section we introduce our new registration and template estimation methods for white matter bundles. We start by presenting the Gaussian Process-based representation of bundles and their properties which are useful to the development of our algorithms. Then, we develop our registration by using the inner product on our GP-space in combination with a variant of the polyaffine registration algorithm [3]. Finally, we present our template estimation algorithm based on our pairwise registration method and the work of [2].

2.1 Representation of WM Bundles as Tract Density Maps

As emphasized in [7][12][20] a normalized tract density map (TDM) is a convenient non-parametric way to model white matter fiber bundles. For WM tract bundle composed of N tracts, the TDM is a function $y(\mathbf{p}) \triangleq \#t/N$ that associates each point $\mathbf{p} \in \mathbb{R}^3$ with the ratio of number of tracts $\#t$ that are likely to traverse that point over the total number of tracts in the bundle, N . We show examples of TDMs for WM tracts in fig. 1.

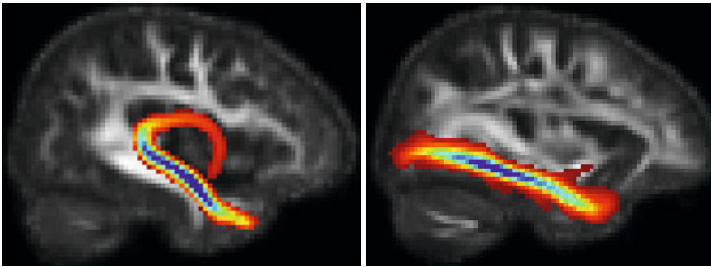


Fig. 1. Tract density map for the fornix and inferior longitudinal fasciculus. The density scales from blue (maximum density) to red (minimum density).

In order to calculate TDMs for WM tracts [4], we use a Gaussian Process (GP) framework [20]. The advantages of this representation are three-fold: first, the TDMs calculated using GPs are continuous functions which can be sampled at any desired resolution; second, being a parametric representation, the GP framework allows us to work robustly on its parameter space instead of performing operations in image space; third, this representation provides us with a vector space of TDMs which we use in order to derive a tract registration algorithm. In the remainder of this section we detail the calculation of TDM for a WM tract: we start by describing the representation of a trajectory within the GP framework; next, we show how to go from the GP representation of a single trajectory to the representation of a bundle of trajectories; and finally, using the GP framework, we describe the calculation of the TDM for a bundle of trajectories.

We model the TDM of an individual trajectory \mathcal{F} as a tract density map $y(\cdot)$, such that, for any point $\mathbf{p} \in \mathcal{F}$, $y(\mathbf{p}) = 1$ and it decays to 0 for points far away from \mathcal{F} at a speed modulated by a parameter R . This smooth TDM $y(\cdot)$ of a single trajectory \mathcal{F} can be written as a GP, $y \sim \mathcal{GP}(y^*(\cdot), c(\cdot, \cdot))$ [20] where $y^*(\cdot)$ is the most

probable TDM for trajectory \mathcal{F} , or $y^*(\cdot) = \mathbb{E}\{y(\cdot)\}$ and $c(\cdot, \cdot)$ is derived from the hypotheses on the shape and smoothness of the TDM and characterizes the variability of suitable TDMs for \mathcal{F} . Then, using the properties of the GPs, the value of $y(\cdot)$ for a trajectory can be characterized at each point in space \mathbf{p} as an univariate Gaussian: $y(\mathbf{p}) \sim \mathcal{G}(y^*(\mathbf{p}), \sigma^2(\mathbf{p}))$. Particularly, sampling a set of points $\mathbf{f} = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ from the trajectory \mathcal{F} , the mean and the variance of this univariate Gaussian are inferred as

$$y^*(\mathbf{p}) = [C_{\mathbf{f}}(\mathbf{p})]^T C_{\mathbf{ff}}^{-1} \mathbf{1} \quad \text{and} \quad \sigma^2(\mathbf{p}) = c(\mathbf{p}, \mathbf{p}) - C_{\mathbf{f}}(\mathbf{p})^T C_{\mathbf{ff}}^{-1} C_{\mathbf{f}}(\mathbf{p}) \quad (1)$$

where

$$[C_{\mathbf{f}}(\mathbf{p})]_i \triangleq [c(\mathbf{f}_i, \mathbf{p})]_i, \quad [C_{\mathbf{ff}}]_{ij} \triangleq [c(\mathbf{f}_i, \mathbf{f}_j)]_{ij}, \quad \mathbf{1} = [1 \dots 1]^T$$

and

$$c(\mathbf{p}, \mathbf{p}') \triangleq \psi(\|\mathbf{p} - \mathbf{p}'\|), \quad \psi(r) = \begin{cases} 2|r|^3 - 3Rr^2 + R^3 & r \leq R \\ 0 & r > R \end{cases}$$

The covariance function used to characterize this GP, $c(\cdot, \cdot)$, has two important properties [20]: first, it ensures that the most probable TDM representing a given fiber minimizes the curvature; second, the resulting TDM has finite support of radius R around each point¹. Later, we will take advantage of the finite support property to specify a finite block coverage of the TDM.

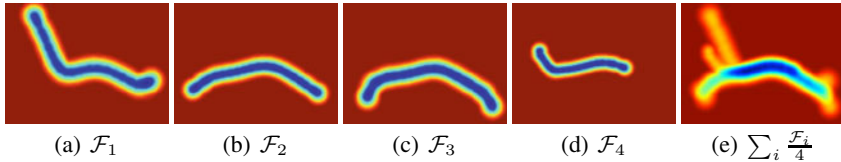


Fig. 2. Tract density map for four fiber tracts (a-d) and for the bundle formed by averaging them according to our framework (e). Color code ranges from blue when it is likely that a voxel belongs to the bundle of fibres to red when it does not belong. Image reproduced from [20].

This GP framework constitutes a **vector space** where $(y_1 + y_2)(\mathbf{p}) = y_1(\mathbf{p}) + y_2(\mathbf{p})$ and $(\lambda y_1)(\mathbf{p}) = \lambda y_1(\mathbf{p})$. Using these operations, obtaining the representation for a bundle of trajectories $\mathcal{B} = \{\mathcal{F}_1, \dots, \mathcal{F}_N\}$ is straightforward: we calculate the mean y^* and variance σ of the Gaussian distribution characterizing the TDM for the bundle \mathcal{B} at a point \mathbf{p} as [20]

$$y_{\mathcal{B}}^*(\mathbf{p}) = \frac{1}{N} \sum_{\mathcal{F} \in \mathcal{B}} y_{\mathcal{F}}^*(\mathbf{p}) \quad \text{and} \quad \sigma_{\mathcal{B}}(\mathbf{p}) = \frac{1}{N^2} \sum_{\mathcal{F} \in \mathcal{B}} \sigma_{\mathcal{F}}^2(\mathbf{p}).$$

The result of the combination of the density functions representing various trajectories into a bundle is illustrated in fig. 2(e).

¹ We take as 2 times the maximum distance between two consecutive points in the tract.

We quantify the similarity between two WM bundles by using the deterministic **inner product** of this vector space,

$$\langle y_1, y_2 \rangle = \mathbb{E} \left\{ \int_{\mathbb{R}^3} y_1(\mathbf{p}) y_2(\mathbf{p}) d\mathbf{p} \right\} = \int_{\mathbb{R}^3} y_1^*(\mathbf{p}) y_2^*(\mathbf{p}) d\mathbf{p} \quad (2)$$

which represents the overlap between the TDMs of the two tracts. Using eq. (1), we calculate $\langle y_1, y_2 \rangle$ without point-to-point matching between bundles (20)

$$\langle y_1, y_2 \rangle = \mathbf{1}^T [C_{\mathbf{f}^1 \mathbf{f}^1}^{-1}]^T \left(\int_{\mathbb{R}^3} C_{\mathbf{f}^1}(\mathbf{p}) [C_{\mathbf{f}^2}(\mathbf{p})]^T d\mathbf{p} \right) C_{\mathbf{f}^2 \mathbf{f}^2}^{-1} \mathbf{1} \quad (3)$$

where \mathbf{f}^1 and \mathbf{f}^2 are the points of to the two bundles. Also, $\langle \cdot, \cdot \rangle$ induces the norm $\|y\|^2 = \langle y, y \rangle$ representing the mass of the tract. Current literature in Gaussian processes (11, 14) shows that the **inner product space** we just presented is, in fact, a **re-producible kernel Hilbert space** (RKHS), allowing to define a set of basis functions whose linear combination spans all TDMs. Later in this paper, we use this RKHS property to characterize the deformation modes of a particular population of WM tracts.

In the above, we have presented a GP-based framework for representing white matter fiber bundles and three operations: the calculation of the normalized tract density map, combination of fibers into a bundle and the quantification of the similarity between two bundles. These tools are fundamental to performing white matter bundle registration, which we develop in the next section.

2.2 Diffeomorphic Bundle Registration with GP-Represented TDMs

Our goal is to align two WM bundles obtained from two different subjects. In other words, given the two bundles $\mathcal{B}_1, \mathcal{B}_2$ and their corresponding GP-represented TDMs y_1, y_2 , we are looking for a dense transformation $s : \mathbf{p} \mapsto \mathbf{s}(\mathbf{p})$ such that it minimizes

$$E(s) = \text{Sim}(s) + \text{Reg}(s) \text{ where } \text{Sim}(s) = 1 - \frac{\langle y_1, y_2 \circ s \rangle^2}{\|y_1\| \|y_2 \circ s\|} \quad (4)$$

where $\text{Sim}(s)$ is the TDM similarity measure and $\text{Reg}(s)$ is a regularization term. In this equation, the transport of the TDM y_2 by the transformation s , $y_2 \circ s$, is carried by the operation $(y_2 \circ s)(\mathbf{p}) = y_2(\mathbf{s}(\mathbf{p})) |\text{jac } s(\mathbf{p})|$. Equation (4) can be regarded as equivalent to minimizing the widely used normalized cross-correlation metric (11) between the images of two TDMs.

In order to develop a diffeomorphic registration algorithm with a reduced number of parameters, we use a variant of the LogEuclidean polyaffine technique (3) to minimize eq. (4) and obtain the desired registration. We start by dividing the domain of y_1, y_2 in a lattice of cubic blocks of volume W . Since y_1 and y_2 are of finite support (section 2.1), it is possible to define a finite set Γ of non-intersecting blocks of volume W which is a total coverage of the two TDMs: $\Gamma = \{\gamma_1, \dots, \gamma_N\} \subset \mathbb{R}^3$. Figure 3(a) illustrates a TDM along with its lattice Γ . Using this set, we define a block-based registration energy which is an upper bound approximation for eq. (4) as

$$\mathbb{E}_\Gamma(s) = \text{Sim}_\Gamma(s) + \text{Reg}(s), \quad \text{Sim}_\Gamma(s) = \sum_{i=1 \dots N} 1 - \frac{\langle y_1, y_2 \circ s \rangle_{\gamma_i}^2}{\|y_1\|_{\gamma_i} \|y_2 \circ s\|_{\gamma_i}} \quad (5)$$

where we define the block-inner product and its induced norm as

$$\langle y_1, y_2 \rangle_\gamma \triangleq \int_\gamma y_1^*(\mathbf{p})y_2^*(\mathbf{p})d\mathbf{p}, \quad \|y\|_\gamma^2 = \langle y, y \rangle_\gamma. \quad (6)$$

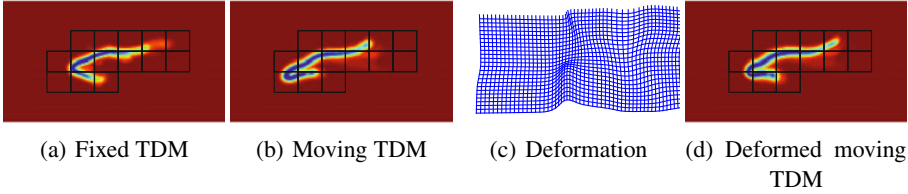


Fig. 3. Illustration of a registration step. Figures (a-b) show the TDMs corresponding to 2 uncinate fasciculi along with their common lattice Γ , each square indicates a block γ_i . We calculate the affine transforms to take each block of the moving image to the fixed one and integrate them into a deformation (c). Finally, the deformation is applied to the moving image (d).

The block-based formulation of the registration energy, eq. (5), enables us solve our registration problem using a block-based polyaffine framework. Let M and \mathbf{t} be the linear and translation components of the affine transform A . Defining the affine transformation of a TDM ($y_2 \circ A$) $\triangleq y_2(M \cdot \mathbf{p} + \mathbf{t})$ renders possible to obtain an affine registration inside of the block γ by minimizing

$$A_\gamma = \operatorname{argmin}_{A'} \left\{ 1 - \frac{\langle y_1, y_2 \circ A' \rangle_\gamma^2}{\|y_1\|_\gamma \|y_2 \circ A'\|_\gamma} \right\} \quad (7)$$

Then, using the LogEuclidean polyaffine framework [3] we calculate the dense diffeomorphic transform s from a set of affine transforms: we compute the velocity field u associated with the affine transforms $A_{\gamma_1}, \dots, A_{\gamma_N}$, as

$$u : \mathbf{p} \mapsto \mathbf{u}(\mathbf{p}), \quad \mathbf{u}(\mathbf{p}) = \sum_{i=1 \dots N} w_{\gamma_i}(\mathbf{p})(L_{\gamma_i} \mathbf{p} + \mathbf{v}_{\gamma_i}), \quad (8)$$

where

$$\begin{pmatrix} L_\gamma & \mathbf{v}_\gamma \\ 0 & 0 \end{pmatrix} = \log(A_\gamma), \quad L_\gamma \in \mathbb{R}^{3 \times 3}, \mathbf{v}_\gamma \in \mathbb{R}^3,$$

and $w_\gamma(\mathbf{p})$ is a smooth weighting function which quantifies the influence that the transform found at block γ has on \mathbf{p} and \log is the matrix logarithm. Particularly, we define

$$w_\gamma(\mathbf{p}) = \frac{w'_\gamma(\mathbf{p})}{\sum_{\gamma \in \Gamma} w'_\gamma(\mathbf{p})}, \quad w'_\gamma(\mathbf{p}) = \exp\left(-\frac{\|\mathbf{p} - \bar{\gamma}\|^2}{\sigma^2}\right)$$

where $\bar{\gamma}$ is the center of block γ and σ controls the rigidity of the global transformation.

We can now present our TDM-based polyaffine white matter bundle registration algorithm: Given two white matter bundles as their TDMs, y_1, y_2 and a block volume W ,

- 1: Calculate the non intersecting block-coverage of y_1 and y_2 : $\Gamma = \{\gamma_1, \dots, \gamma_N\}$
- 2: Set the initial transform $s_0 = Id$
- 3: **repeat**
- 4: $s_{prev} \leftarrow s$
- 5: Find the set of affine transforms $A_{\gamma_i}, i = 1 \dots N$, minimizing $\text{Sim}_\Gamma(s_k)$ (eq. (5))
- 6: Calculate the velocity field u using eq. (8)
- 7: Compose the transforms $c \leftarrow s_{prev} \circ \exp(u)$
- 8: For a diffusion-like regularization $s \leftarrow K_{diff} \star c$, else $s \leftarrow c$
- 9: **until** $E_\Gamma(s_{prev}) - E_\Gamma(s) < \text{threshold}$

where \exp is the exponential of a vector field as defined by [3] and $K_{diff} \star c$ is the convolution of the velocity field c with the Gaussian kernel K_{diff} . Figure 3 depicts a step of the algorithm presented above. Two things should be noted in this algorithm: first, we are using a two-step optimization process as in Vercauteren et al [19]; second, we want to highlight that a multiscale version of this algorithm is easily implementable by adding an outer loop which varies the volume of the blocks in a monotonic decreasing manner.

2.3 Template Construction and Population Analysis

We now extend our pairwise registration algorithm to groupwise registration in order to generate an unbiased template [132].

We consider having M analogous bundles $\mathcal{B}_1, \dots, \mathcal{B}_M$ extracted from rigidly co-registered images of M subjects. Following Allasonniere [2] and Durrleman [9], we consider the bundles the result of a diffeomorphic deformation of a prototype bundle plus a residual. Letting y_i be the GP-based TDM representation associated with bundle \mathcal{B}_i , we formulate the TDM of each bundle in terms of the template as

$$y_i(\mathbf{p}) = (\bar{y} \circ s_i)(\mathbf{p}) + \epsilon_i(\mathbf{p}), \quad \mathbf{p} \in \mathbb{R}^3$$

where $\epsilon_i(\mathbf{p})$ is an image of uncorrelated white noise represented by a zero mean Gaussian process with diagonal covariance function, $c_\epsilon(\mathbf{p}, \mathbf{p}') = \delta(\|\mathbf{p} - \mathbf{p}'\|)$, where $\delta(\cdot)$ is the Dirac delta function. This covariance function stands for absence of correlation between the points in the image or white noise. Then we estimate the template, the deformations and the residues as:

$$\text{argmin}_{\bar{y}, s_i} \left\{ \sum_{i=1}^M \text{Sim}_\Gamma(y_i, \bar{y} \circ s_i) + \text{Reg}(s_i) \right\} \quad (9)$$

We minimize the criterion in eq. (9) using alternate minimization: initially we set $s_i = Id$ and we obtain a first estimate of the template $\bar{y} = \sum_i y_i / M$; then we register \bar{y} to every bundle y_i , and we re-estimate the template $\bar{y} = \sum_i (y_i \circ s_i^{-1}) / M$; finally, we iterate these steps until convergence.

This algorithm eventually yields an unbiased template \bar{y} and the deformations s_i . Taking advantage of the RKHS property of the TDMs y_i and its residuals ϵ_i represented as GPs, section 2.1, we perform PCA analysis of the residuals in order to characterize

the non-diffeomorphic deformations of each bundle. We start by estimating the residual for each bundle \mathcal{B}_i as $\epsilon_i = y_i - \bar{y} \circ s_i$. Then, we calculate the first mode of residual variations on the template at $\pm\theta$, as $m_\epsilon(\theta) = \bar{y} \pm \theta \sum_i \epsilon_i \mathbf{v}_1$ where \mathbf{v}_1 is the eigenvector corresponding to the largest eigenvalue of the covariance matrix $[\langle \epsilon_i, \epsilon_j \rangle]_{ij}$ (see example in Results section).

We have formulated the necessary tools to perform pairwise registration between white matter bundles represented as TDMs in the GP space; to estimate a template from a population and to characterize its residual variation modes. Our model represents the bundles in a population as a diffeomorphic deformation of a template plus a non-diffeomorphic residual as done by Durrleman et al [9]. These tools to perform statistical analysis of the WM bundles are not biased by point-to-point or fiber-to-fiber correspondences. In addition, it is not biased by differences in the orientation of fiber parametrizations as [9]. The separation between the diffeomorphic and residual part is regulated by the tradeoff between the similarity between bundles and the regularization term used in order to find the diffeomorphic transforms in eq. (4). This allows to adjust the deformation and to capture subject-specific anatomical variations. We are now in position to apply all these tools to a bundle population and analyze it.

3 Results

Diffusion-weighted images (DWI) from 43 subjects were acquired on a GE Signa HDxt 3.0T scanner using an echo planar imaging sequence with a double echo option, an 8 Channel coil and ASSET with a SENSE-factor of 2. The acquisition consisted in 51 directions with $b=900 \text{ s/mm}^2$, and 8 images with $b=0 \text{ s/mm}^2$, with scan parameters $TR=17000 \text{ ms}$, $TE=78 \text{ ms}$, $FOV=24 \text{ cm}$, 144×144 encoding steps, 1.7 mm slice thickness. 85 axial slices covering the whole brain were acquired. DWI images were linearly registered. The left uncinate and fronto-occipital fasciculi ROI's were manually drawn by experts using 3D Slicer (www.slicer.org) and fiber tracts were obtained using two-tensor tractography [17].

3.1 Pairwise Registration

In order to test the efficacy of our pairwise registration algorithm, introduced in section 2.2, we randomly picked one sample from each bundle population and registered all others to it. We applied a multiresolution scheme where the cubes of the latices were $100, 50, 20, 10$ and 5 mm^3 each and we set rigidity parameter σ to $1/2$ of the cube width. For the uncinate fasciculi the dice coefficient of the registered TDMs to the fixed TDM was $.93(\pm.03)$ and for the fronto-occipital bundles $.81(\pm.07)$. We applied the transformation s_i to each moving bundle \mathcal{B}_i , and the results for one example of each bundle class are shown in fig. 4. In this figure we show, from left to right, the progress of the registration of two uncinate fasciculi and two fronto-occipital bundles, red indicates the moving bundle and green the fixed one.

3.2 Template Estimation

We applied the template estimation algorithm presented in section 2.3 to both populations of white matter bundles. In fig. 6, we show the convergence of the template

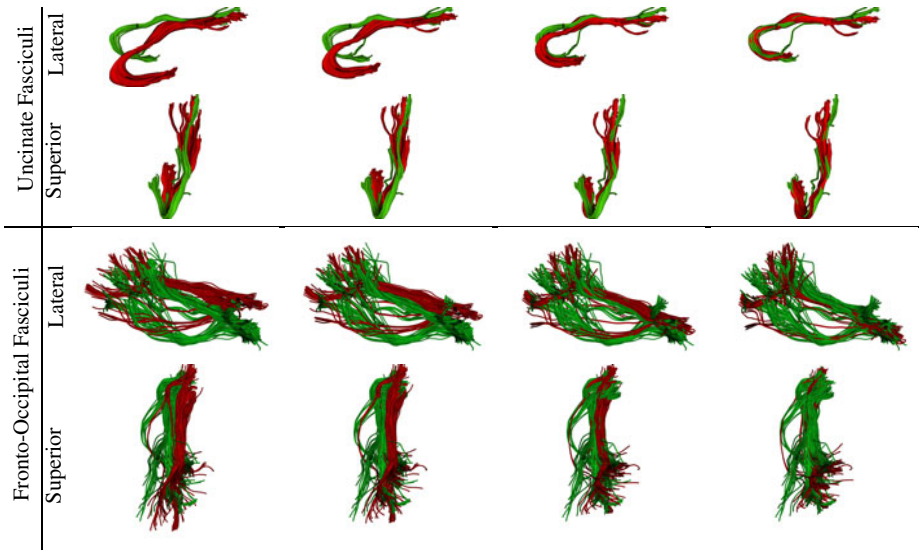


Fig. 4. Result of the pairwise registration of two uncinate fasciculi and two fronto-occipital bundles. Green indicates the fixed bundle and red the moving one. The registration progress is shown from left two right.

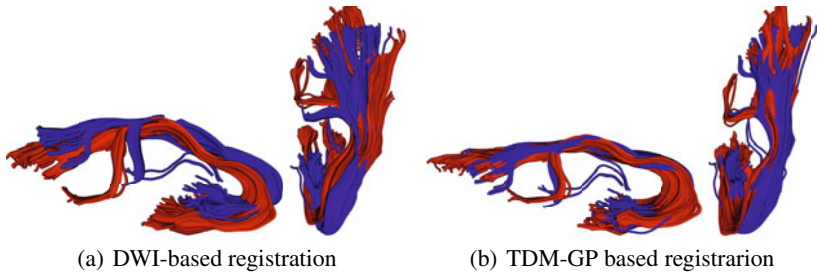


Fig. 5. Comparison of registration methods: two uncinate fasciculi after volumetric registration their DWI images with a state-of-the-art method [6]: (a) and after applying our bundle-based registration (b) . Overlap is better using our TDM-GP approach than by registering the DWI images.

estimation algorithm. The algorithm starts from a set of bundles extracted from rigidly registered DT images and converges to an unbiased template. After the template estimation, we assessed the quality of the results by calculating the Dice coefficient of the population of uncinate fasciculi with the estimated template: $0.97(\pm 0.015)$; the same analysis for the occipito-frontal dataset had a Dice coefficient of $0.85(\pm 0.075)$. To compare our approach with indirect image-based registration, we applied the warps generated from a DWI-based template estimation algorithm [6] to the uncinate fasciculi and we compared them with the results of our bundle-based registration. We extracted two bundles and show the results obtained by applying both methods in fig. 5. Then, we

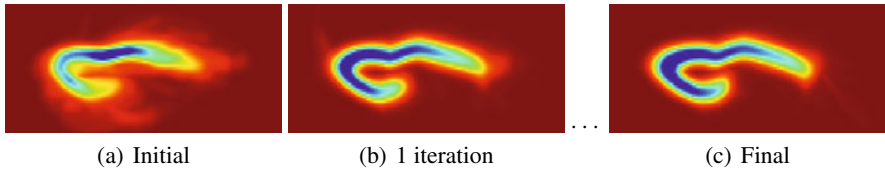


Fig. 6. Template estimation: several steps of the template estimation algorithm presented in section 2.3. The iterations start from a template generated by linear registration of the DWI images and converge to unbiased template generated by our algorithm.

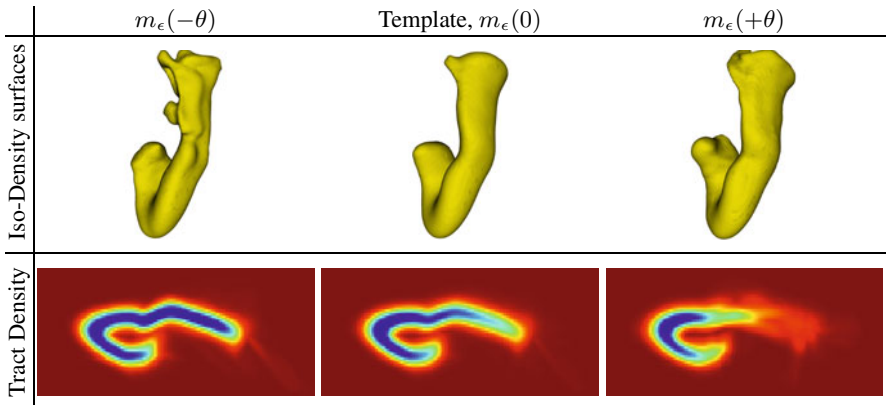


Fig. 7. First mode of the residual variations on the template. We show the mode on the range $-\theta \dots \theta$ where 0 is the actual template. We show the iso-density level at 0.01 from a superior point of view on the top and the maximum intensity projection from a lateral point of view on the bottom. The first mode of variation shows a dispersion on the innervations of the uncinate fasciculus at the orbital cortex and the temporal pole.

calculated the residuals and their variation modes from the template. In fig. 7 we show the first mode of variation for the uncinate bundle.

4 Discussion and Conclusion

In this paper, we introduced a new approach to the direct registration of white matter tract bundles obtained through diffusion MRI tractography. Our method does not rely on point-to-point correspondences nor on the orientation of tract parametrization. Moreover, the GP-based mathematical framework we used enables us to analyze the characteristic deformations of a population. At the heart of our method is the similarity calculation between two bundles which is independent of point-to-point correspondences or orientation of the tract parametrization. Moreover, we parametrically represent these bundles as Gaussian processes, which has several advantages: 1) the similarity is calculated from the parameters of the GP without the need of explicitly

sampling the TDM; 2) the **vector space** of GPs enables us to seamlessly combine bundles, in the intrasubject case where single tracts form a fascicle or the intersubject case where several bundles are combined to form a template 3) The **RKHS space** in which these GPs are embedded provides us with the tools to perform population studies such as PCA. All of these advantages led us to a combined registration and template estimation and population analysis framework.

The results on pairwise registration show that our algorithm effectively produces a deformable registration which puts priority in areas of high fiber density allowing sparse radiations to vary between subjects. This characteristic is important as it enables us to achieve a registration on the main trunk of the tract bundle and leave a degree of freedom for inter subject anatomical variations as those shown by Bürgel et al [7]. When we compared the outcome of our template estimation algorithm with a DWI-based one, ours showed a higher overlap in the bundles. Particularly, we illustrated this by extracting two bundles of the population and showing their overlap. The statistical analysis was evaluated on fibers obtained from 43 subjects. The characteristic deformation shows variability in the innervations of the orbital cortex and the temporal pole which is coherent with histological studies [7].

The method however, raises a main question: how much are we losing by not taking explicitly in account the orientation of the tracts. According to the experiments we performed in this paper, our registration algorithm yields good results even in the absence of orientation features in our model. Since our model based on the density of tracts at each point, the continuity of the density map and the diffeomorphic transforms lead to a registration algorithm which results in the alignment of the most dense areas of the tracts. When the tract has several diverging fibers like the occipito-frontal fasciculi these are allowed to vary between subjects and considered by the template estimation algorithm as non-diffeomorphic transforms. Hence, even if orientation information is not taken in account, our algorithm registers the most dense areas of the tracts effectively.

Further work will focus on the development of discriminative statistics and correlation analyses. Discriminative statistics are fundamental when studying pathologies and comparing populations, the development of a suitable hypothesis testing scheme is then of the utmost importance. Along the same lines, correlation analyses, for instance with scalar indices, will enable to characterize the cognitive or pathological consequences of the anatomical variations unveiled by our population analysis tools.

Acknowledgements. This work was supported by P41RR13218 (RR) funded by NCRR NIH HHS; U41RR019703 (RR) funded by NCRR NIH HHS; U54EB005149 (EB) funded by NIBIB NIH HHS; R01MH5074 funded by NIMH NIH; VA Schizophrenia Center Grant; VA Merit Award; R01MH074794; R01MH092862; 5R01MH82918; and P41RR013218.

References

1. Adler, R.J., Taylor, J.E.: *Random Fields and their Geometry*. Springer, Heidelberg (2007)
2. Allasonnière, S., Amit, Y., Trounev, A.: Towards a coherent statistical framework for dense deformable template estimation. *J. of the Royal Stat. Soc. B* 69(1), 3–29 (2007)

3. Arsigny, V., Commowick, O., Ayache, N., Pennec, X.: A fast and log-euclidean polyaffine framework for locally linear registration. *JMIV* 33(2), 222–238 (2009)
4. Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* 44(4), 625–632 (2000)
5. Beg, M., Miller, M., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61(2), 139–157 (2005)
6. Bouix, S., Rathi, Y., Sabuncu, M.: Building an average population hardi atlas. In: *MICCAI Workshop on Computational Diffusion MRI* (2010)
7. Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J.M., Zilles, K.: White matter fiber tracts of the human brain: Three-dimensional mapping at microscopic resolution, topography and intersubject variability. *NeuroImage* 29(4), 1092–1105 (2006)
8. Corouge, I., Fletcher, P.T., Joshi, S., Gouttard, S., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor mri analysis. *MIA* 10(5), 786–798 (2006)
9. Durrleman, S., Fillard, P., Pennec, X., Trounev, A., Ayache, N.: Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NI* (2010), <http://www.sciencedirect.com/science/article/B6WNP-51K9388-1/2/2094f276b437215fa1da04614514f48b>
10. Goodlett, C.B., Fletcher, P.T., Gilmore, J.H., Gerig, G.: Group analysis of dti fiber tract statistics with application to neurodevelopment. *NeuroImage* 45(1, S1), S133–S142 (2009), <http://www.sciencedirect.com/science/article/B6WNP-4TXF7Y2-3/2/63144928840f25440fbdfac042c2b2b4>
11. Holden, M.: A review of geometric transformations for nonrigid body registration. *TMI* 27(1), 111–128 (2008)
12. Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P.A., Pekar, J.J., van Zijl, P.C.M., Mori, S.: Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *NeuroImage* 39(1), 336–347 (2008), http://www.sciencedirect.com/science?_ob=MImg&imagekey=B6WNP-4PF1WFR-5-N&_cdi=6968&_user=6068170&_orig=search&_coverDate=01/01/2008&_sk=999609998&view=c&wchp=dGLbVlz-zSkzS&md5=de42ccbb978bccd0c2bf38c987ee787e&ie=/sdarticle.pdf
13. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 160, S151–S160 (2004), <http://www.sciencedirect.com/science/article/B6WNP-4DCMGVT-3/2/77a14449f2c349bbfc9c36758645eaac>
14. Kimeldorf, G., Wahba, G.: A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. of Math. Stat.* 41(2), 495–502 (1970)
15. Kubicki, M., Westin, C., Maier, S., Frumin, M., Nestor, P., Salisbury, D., Kikinis, R., Jolesz, F., McCarley, R., Shenton, M.: Uncinate Fasciculus Findings in Schizophrenia: A Magnetic Resonance Diffusion Tensor Imaging Study. *American J Psychiatry* 159(5), 813 (2002)
16. Leemans, A., Sijbers, J., Backer, S.D., Vandervliet, E., Parizel, P.: Multiscale white matter fiber tract coregistration: a new feature-based approach to align diffusion tensor data. *MRM* 55(6), 1414–1423 (2006), <http://dx.doi.org/10.1002/mrm.20898>
17. Malcolm, J.G., Michailovich, O., Bouix, S., Westin, C.F., Shenton, M.E., Rathi, Y.: A filtered approach to neural tractography using the Watson directional function. *MIA* 14, 58–69 (2010)
18. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.J.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NI* 31(4), 1487–1505 (2006), <http://dx.doi.org/10.1016/j.neuroimage.2006.02.024>

19. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* 45(1, S1), S61–S72 (2009), <http://www.sciencedirect.com/science/article/B6WNP-4TW13J3-6/2/519213aac5e0b363bd17fbd0a05e1833>
20. Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., Deriche, R.: Unsupervised white matter fiber clustering and tract probability map generation: Applications of a gaussian process framework for white matter fibers. *NeuroImage* 51(1), 228–241 (2010), <http://www.sciencedirect.com/science/article/B6WNP-4Y5BMD4-2/2/9bba3c64a0443363022999aa7b04918a>
21. Yang, J., Shen, D., Davatzikos, C., Verma, R.: Diffusion Tensor Image Registration Using Tensor Geometry and Orientation Features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part II*. LNCS, vol. 5242, pp. 905–913. Springer, Heidelberg (2008)
22. Zhang, H., Yushkevich, P., Alexander, D., Gee, J.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. *MIA* 10, 764–785 (2006)
23. Ziyang, U., Sabuncu, M.R., Grimson, W.E.L., Westin, C.F.: A robust algorithm for fiber-bundle atlas construction. In: *IEEE MMBIA* (2007)

Personalization of Pictorial Structures for Anatomical Landmark Localization

Vaclav Potesil^{1,2}, Timor Kadir^{1,3}, Günther Platsch², and Sir Michael Brady⁴

¹ Department of Engineering Science, University of Oxford

² Siemens Molecular Imaging, Oxford, United Kingdom

³ Mirada Medical, Oxford, United Kingdom

⁴ Department of Radiation Oncology and Biology, University of Oxford

Abstract. We propose a method for accurately localizing anatomical landmarks in 3D medical volumes based on dense matching of parts-based graphical models. Our novel approach replaces population mean models by jointly leveraging weighted combinations of labeled exemplars (both spatial and appearance) to obtain personalized models for the localization of arbitrary landmarks in upper body images. We compare the method to a baseline population-mean graphical model and atlas-based deformable registration optimized for CT-CT registration, by measuring the localization accuracy of 22 anatomical landmarks in clinical 3D CT volumes, using a database of 83 lung cancer patients. The average mean localization error across all landmarks is 2.35 voxels. Our proposed method outperforms deformable registration by 73%, 93% for the most improved landmark. Compared to the baseline population-mean graphical model, the average improvement of localization accuracy is 32%; 67% for the most improved landmark.

Keywords: landmark localization, parts-based graphical model, model personalization, whole-body CT.

1 Introduction

The automatic localization of anatomical landmarks in medical images is a useful tool not only to facilitate image reading and reporting (e.g. automated placement and referencing of ROIs), and the planning of radiation therapy; but also to initialize more complex processing steps (e.g. the construction of atlases or organ segmentation). This is particularly true in 3D whole-body CT images that contain many different anatomical structures which can be highly variable between patients, especially when pathologies such as tumors are present. This variability makes the automatic localization of anatomical landmarks a challenging task [9]. Nevertheless, the relative position of organs in the human body is not random and shows common structural patterns. A natural way to model such structured objects is to use informative spatial priors based on Parts-based Graphical Models (GM) that have shown attractive properties (e.g. robustness to local image changes) for substructures labeling in single organs [2,7,10,8]. Standard GMs use

population average spatial and appearance models that have shown limitations when large population variations cannot be captured by average statistics (e.g. unimodal Gaussian distributions) representative of no individual, resulting in inferior matching accuracy. In experiments with a standard GM [5], we have observed that the initial landmark match is almost always approximately right (within 4 voxels from ground-truth), but almost never exact.

In this paper, we propose to extend previous work on dense matching of parts-based GM, Pictorial Structures (PS) [4], by personalizing a population model to the test image. We introduce novel jointly personalized appearance and spatial models that directly utilize a subset of labeled training examples adapted to each test image at run-time, while retaining the same framework and the attractive computational complexity of PS. The method described in this paper is applied to the localization of landmarks in 3D CT images (Figure 1) that are used as standard clinical reference points for whole-body PET/CT oncology imaging. Results on clinical data show that our method significantly improves the accuracy of matching highly variable landmarks including skeletal and soft-tissues relative to baseline methods and even more so relative to different methods developed for similar purposes; for example, atlas-based deformable registration and a decision forest approach [3].

There are two streams of related work in computer vision. In *iterative image parsing* [6], an initial match is used to guide model updates at run-time. However, instead of updating the parameters of the initial population average appearance model, our algorithm builds a personalized appearance *and* spatial model that makes direct use of a weighted subset of training exemplars from the annotated training database. Our method is also related to *blind relevance feedback*, in that we build enriched latent models using an algorithm similar in spirit to automated query expansion in visual image retrieval [1]. However, instead of retrieving a set of related images with different viewpoints given a query image, we accurately localize the position of an anatomical landmark *within the query region*, using a query-expansion-like algorithm based on the agreement of landmark localization decisions by the exemplars from the annotated training image corpus.

The structure of this paper is as follows. In Section 2, we briefly summarise the Pictorial Structure (PS) model and describe our implementation of the baseline population average model. In Sections 3.1 - 3.2, we introduce our personalization method that jointly builds enriched models from the annotated spatial and appearance training exemplars, adapted to each test image at run-time. In Section 4, we describe the experimental setup and the significantly improved landmark localization results; then discuss our contribution in Section 5.

2 Parts-Based GM for Anatomical Landmark Localization

We recall the Pictorial Structure GM, referring the reader to [4]. All landmarks share the same representation, which comprises a unary energy term for local part appearance and a set of pair-wise terms for spatial compatibility,

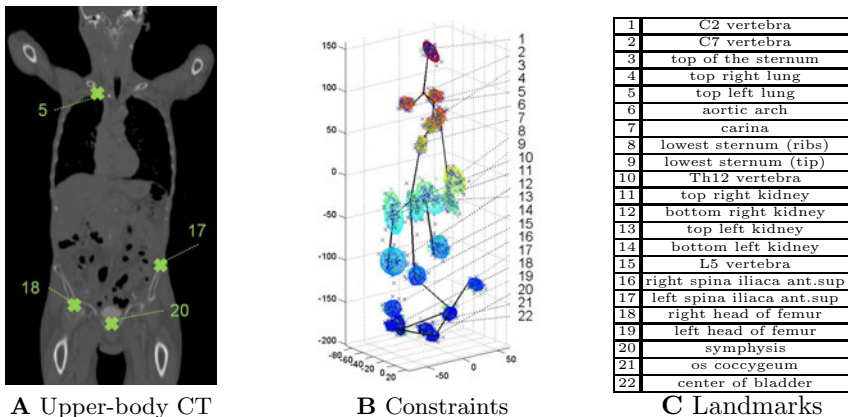


Fig. 1. Landmarks and anatomical constraints. Shapes show valid population mean constraint ranges, plotted at 2σ from the mean. The stick-man in (B) shows the pair-wise relations in a tree-structured GM. The nodes and leaves of the GM correspond to the positions of landmarks (C) in the patient image (A).

reflecting the relational anatomical constraints. Our model takes the form of a tree-structured undirected GM, named G , with parameters Θ

$$p(L | I, \Theta) \propto \prod_{k=1}^n \underbrace{p(I | l_k, u_k)}_{\text{appearance terms}} \prod_{v_k, j \in G} \underbrace{p(l_k, l_j | c_{kj})}_{\text{spatial terms}} \quad (1)$$

At training time, n landmarks v_k , v_j are placed by the user in a set of images of multiple patients at locations \mathbf{l}_k , \mathbf{l}_j , given by their $[x, y, z]$ coordinates. L is the set of \mathbf{l}_k for all n landmarks. Given the quality of match of patch appearance model u_k to the image I , the first term in Eq. 1 represents the cost of placing the landmark v_k at location \mathbf{l}_k . The second term is the spatial compatibility cost of connected pairs of landmarks v_k , v_j placed at \mathbf{l}_k , \mathbf{l}_j , given a set of prior anatomical constraints c_{kj} that restrict their relative positions.

Appearance Models: We model local tissue appearance in terms of local rectangular image patches centred on landmark locations l_k at a scale $s_k(x, y, z)$. Positive and negative examples are used. First, for positives, patches are projected onto a set of n basis functions obtained from the ground-truth patches. The variability of true part appearance is represented by the distribution of the feature responses, $H_k^+ \sim N(\mathbf{M}_k^+, \Sigma_k^+)$ for each part v_k , where \mathbf{M} , Σ are diagonal matrices. $H_k^- \sim N(\mathbf{M}_k^-, \Sigma_k^-)$ is the model for the local “background” i.e. non-landmark tissue. The patch scale $s_k(x, y, z)$ for each part is optimized for landmark localization [5]. In the initial model, patches are projected onto a set of features comprising the mean template and the top eigen-patches obtained from the ground-truth training patches. In the personalized model, the generic features are replaced by a subset of training patches and $H_k^{+,-}$ adapted to each test image at run-time.

Spatial Models: Pair-wise spatial compatibility terms $c_{kj}(\mathbf{l}_k, \mathbf{l}_j)$ penalise part placements outside of the degree of anatomical variability exhibited within the training database. The functional form of this spatial model is approximated as a distribution over $\mathbf{l}_{kj,e} = \mathbf{l}_{k,e} - \mathbf{l}_{j,e}$ (the relative position vector of two parts v_k, v_j in training patient e). $p(\mathbf{l}_k | \mathbf{l}_j) \sim N(\mathbf{l}_{kj}, \mathbf{M}_{kj}^c, \mathbf{\Sigma}_{kj}^c)$ is a Gaussian model with diagonal covariance. The parameters of the GM are learnt following the approach of [4]. In the initial population mean model, distributions are calculated from all examples. In the personalized model, the population means are replaced with local weighted sub-sets of spatial vectors from labeled training exemplars.

Matching: To find the best placements for the parts, we apply the appearance models to the CT volume using a sliding window approach. The log-likelihood ratios are combined using a Naive Bayes classifier to obtain the posterior probability for each landmark and sliding window placements l_i . The min-sum belief propagation is used for an exact, globally optimal fitting using the dense appearance maps and the pair-wise compatibility terms [4].

3 Personalization of Graphical Models

A population mean model introduces a bias in detection due to the averaging effect of aggregating information from all training images into a single appearance and spatial model for each landmark. We address this problem by building personalized spatial and appearance priors that leverage the exemplars directly, while preserving the advantages of the original GM. In pseudo-code, the implementation of the jointly adapted spatial and appearance terms looks as follows. Our key contributions are highlighted in **bold**:

1. Find **local reference frames (LRFs)** f_k for all landmarks v_k (off-line)
2. Match the initial population mean model to a test-image to obtain the posteriors p_k .
3. Estimate the distribution w_k of distances to labeled training exemplars E_k , using p_k and the LRFs f_k .
4. Build **personalized spatial priors** p_k^{SC} using w_k, f_k and the spatial vectors $\mathbf{l}_{ij,e}$ of exemplars $e \in E_k$.
5. Build **personalized appearance models** using p_k^{SC}, p_k and the exemplar patch data of exemplars $e \in E_k$.
6. Use personalized appearance and spatial models in the original GM to obtain improved detections \mathbf{l}_k .

Having presented the pipeline, we will focus on the personalization of the spatial priors in Section 3.1, and appearance priors in Section 3.2.

3.1 Personalized Spatial Models

First, local reference frames (LRF) are found automatically for each landmark at training time. During model fitting, the LRF is used to select the subset of the

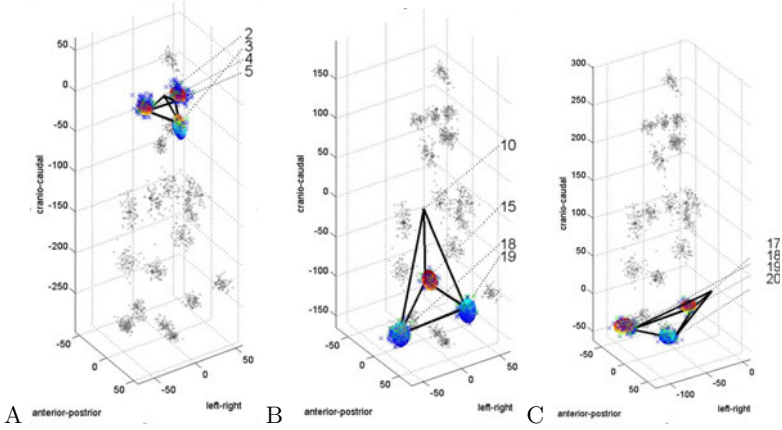


Fig. 2. Local reference frames. We show only 4 landmarks out of 22. A - C7 Vertebra (2).B - Th12 Vertebra (10).C - Left Spina Iliaca (17). For numbering, see Figure [1](#)

training set that will be used to build the personalised spatial prior. The LRFs are determined as sets of landmarks that can predict each others' spatial position. For each landmark v_k , its LRF f_k is composed of the landmark itself plus the top 3 (the minimum number to establish an affine frame in 3D) landmarks according to the following measure

$$Q_{ij}(G | D_E) \propto \sum_{l_i, l_j \in D} -\log p(l_i, l_j | c_{ij}) \quad (2)$$

where p is defined as the spatial term in Eq. [1](#). D_E is the training data i.e. all exemplars and Q_{ij} is a 1-by-n vector of edge compatibilities. This measure is natural to our problem: the relations with low spatial variance are the most informative to predict the location of the landmark of interest, given the positions of other reference landmarks. This is evidenced in the estimated LRFs (Figure [2](#)), which closely resemble clusters of anatomically related landmarks that would be chosen by an expert physician. During model fitting, distance functions measure the local similarity between a test image and the labeled training exemplars, using the LRF. They are calculated using a robust distance measure for each LRF f_k independently, using the following leave-one-out voting

$$w_k(e) \propto \frac{1}{Z} \sum_{m \in F_k} \left(\text{Rank}_{\forall e \in E} \sum_{\forall i, j \in F_k \setminus m} \|l_{ij} - l_{ij}^e\|_2 \right) \quad (3)$$

where Rank is a function that assigns 1..M votes to the top M labeled training exemplars e . The ranking is according to the sum of L^2 distances between position vectors l_{ij} (estimated from the test image) and l_{ij}^e , (ground-truth vectors from labelled training exemplars E_k). i, j are all edges connecting 3 of 4 landmarks $v_{i,j}$ in F_k except v_m (leave-one-out). In each leave-one-out step, the

highest ranked exemplar e (with lowest L^2 score) is assigned M votes and the M^{th} top exemplar 1 vote. In all the experiments reported here, $M = 10$. Votes are summed across the four leave-one-out iterations. Exemplars with no votes are removed from the vector, so obtaining an adaptive number of exemplars according to each landmark's strength of spatial correlations. w_k is normalized in order to obtain a pdf approximation i.e. summing to one. At the end of this step, we have selected a weighted subset of the training exemplars per landmark.

Finally, an efficient algorithm builds personalized spatial priors as weighted combinations of the subset of training exemplars calculated in the previous step, using the LRFs f_k and the inter-patient distances w_k . Specifically, for each landmark v_k , we mobilize its neighbors in the LRF and perform the following propagation in a local graph G_k which connects all landmarks in the local frame f_k with the landmark v_k (root):

$$p_k^{SC} \propto \sum_{\forall e \in E: w_k(e) \neq 0} w_k(e) p_{f_k}(e) \quad (4)$$

where

$$p_{f_k}(e) \propto \min \left(\prod_{\forall i \in G_k \setminus k} p(I | \mathbf{l}_i, u_i) p(\mathbf{l}_i, \mathbf{l}_k | c_{ij}^e) \right) \quad (5)$$

The first term in the (exact) minimization is the negative log-likelihood of the appearance term from Eq. 1 for all landmarks v_i in f_k , excluding the root landmark v_k . The second term is the negative log-likelihood of a personalized spatial model, for an exemplar e parameterized as

$$c_{ij}^e(\mathbf{l}_i, \mathbf{l}_k) \propto p^e(l_i | l_k) \propto N(\mathbf{l}_{ik}, \mathbf{l}_{ik}^e, \eta \Sigma_{ik}^c) \quad (6)$$

This term is similar to the population mean spatial model used in Section 2. However, the (Gaussian) model mean is the spatial vector from landmark v_i to landmark v_k in each labelled training exemplar e in E_k with non-zero weight w_k , as opposed to the population mean M_{ij}^c used in the baseline model. In all experiments reported here, the constant $\eta = 5$. The above formulas are evaluated exactly for tens of exemplars while maintaining the computational complexity of the population mean GM: each pair-wise compatibility term is solved using only one pass of the belief propagation in the LRF of landmark v_j , irrespective of the number of exemplars. Results for all exemplars and states are read out using simple memory indexing shifts of the fast generalized distance transform grid, using translations derived from the annotated training vectors \mathbf{l}_{ik}^e , and summed using the weights w_k .

3.2 Personalized Appearance Models

In this section, we describe how the appearance term is adapted to each test image. In essence, given the spatial prior p_k^{SC} and dense posterior p_k obtained from an initial fit of the population average model, we seek a subset of the

training database that best matches the test image, from which we build an updated (personalized) model and dense appearance maps.

Our method can be seen as a voting process, where each annotated training exemplar e votes for the most likely location l . Given an initial model fit, for each landmark k , search proceeds by defining a local query region r_k (Section 3.2). We apply all annotated exemplar templates to all voxels in r_k to obtain exemplar feature response maps $f_{k,e}$. In this paper, we use normalized cross-correlation, but any other measure can be used. We find the coordinates of the maxima of each $f_{k,e}$ and aggregate the values of their responses. The resulting cost function p_k^{EC} , which we refer to as *exemplar consensus (EC)*, takes account of the intrinsic uncertainty of ground-truth annotations and the limited image resolution. The personalized spatial prior p_k^{SC} introduces spatial context in the scoring, i.e. preference for certain locations l . (Section 3.2). Finally, we find the mode of the cost-function and the sub-set of exemplars that achieves maxima in this mode. To account for the varying spatial uncertainty of the individual exemplars, we learn a positive and negative model for each exemplar in this subset (Section 3.2). The subset of exemplars and models are applied to the test image in order to produce updated dense feature maps and re-fit the model.

In more detail, our method proceeds in three steps.

Step 1: Define Local Query Regions. First, the local query regions r_k , in which we personalize the models for the set of annotated training exemplars E_k , are defined around the modes of both the posterior p_k and of the personalized spatial prior p_k^{SC} for each landmark k in a test image (Figure 3).

$$r_k(l) = \begin{cases} 1 & \text{if } \min \|l - L_{max,p_k}\|_\infty \leq RoiSize \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where L_{max,p_k} is a set containing locations of all modes of the posterior probability for each landmark $p_k(l | I, \Theta)$ that are higher than 50% of the MAP solution; and the location of the mode of the spatial prior p_k^{SC} for that landmark. l are voxels in 3D. *RoiSize* is one half of the patch size for landmark v_k .

Step 2: Exemplars Vote For Landmark Locations. In this step, we obtain exemplar feature response maps $f_{k,e}$, for all annotated exemplar templates e_k in E_k using normalized cross-correlation (NCC) in the region r_k in the test image. The **exemplar consensus map** p_k^{EC} is built for all voxels l where $r_k(l) = 1$, by aggregating the local maxima of all exemplar feature response maps $f_{k,e}$, using the personalized spatial prior p_k^{SC} (Figure 4)

$$p_k^{EC}(l | I, \Theta, E) \propto p_{ex}(I | l, E_k) \times p_k^{SC}(l | I, \Theta) \tag{8}$$

where p_k^{SC} is the personalized spatial prior (Section 3.1) and

$$p_{ex}(I | l, E_k) \propto \sum_{e \in E_k} K_W(p_{k,e}(I | l, e_k)) \tag{9}$$

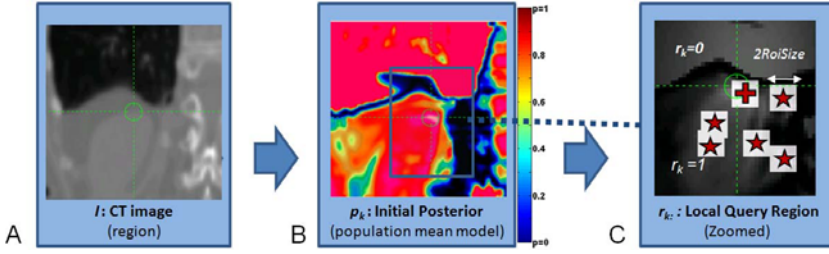


Fig. 3. Define a Local Query Region r_k . Example for the top of right kidney. A - CT image, B - posterior from the initial model, C - local query region r_k (red stars are the local modes of the initial model posterior p_k , cross is the mode of the spatial prior p_k^{SC}). Images show a 16x16x16 cm test image region centered at the ground-truth. For clarity, only coronal slice of the 3D volume is displayed.

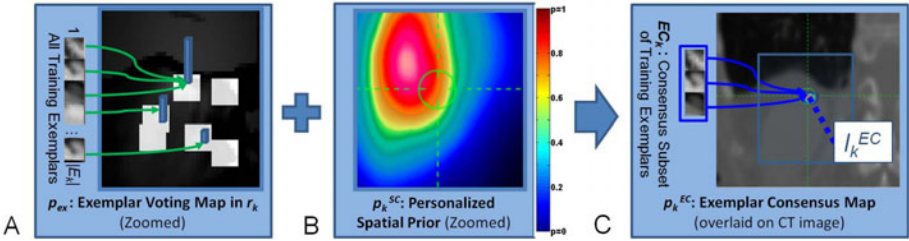


Fig. 4. Exemplars Vote For Landmark Locations. Example for the top of right kidney. A - Exemplar voting map p_{ex} over local query region r_k (white boxes). B - personalized spatial prior p_k^{SC} , used to weight p_{ex} . C - Exemplar consensus map p_k^{EC} . The mode l_k^{EC} is visible at the ground-truth position. Images show a coronal slice of a 16x16x16 cm region centered at the ground-truth location.

K_W is a spatial averaging window, applied around the modes of $p_{k,e}$ which is a pseudo-probability derived from the NCC function: $p_{k,e}(l) \propto \exp -f_{k,e}(l)$.

The location of the strongest mode of exemplar consensus l_k^{EC} is found as

$$l_k^{EC} = \arg \max_{l \in r_k} p_k^{EC}(l | I, \theta, E_k) \tag{10}$$

The set of exemplars in consensus EC_k is defined as all exemplars e whose exemplar feature response maps $f_{k,e}$ peaks at a location $l_{max,e}$ within $Thr = 1$ voxels from l_k^{EC} .

Step 3: Learn Personalized Exemplar Models Models of positive and negative responses $H_{k,e}^+$, $H_{k,e}^-$ are built for each exemplar in EC_k (Figure 5). The aim of this step is to weight all the exemplars associated with the top consensus mode l_k^{EC} according to their localization uncertainty. For example, an exemplar that responds strongly only in location l_k^{EC} and has low response elsewhere, is more favorable than one that has strong response away from the

local maximum. The negative model $H_{k,e}^-$ is learnt as a pdf approximation of feature responses $f_{k,e}$ in the region r_k in test image, for which $W_{k,e}(l_{max,e}) = 0$ i.e. all voxels in r_k outside of the top mode l_k^{EC} . The positive model for each exemplar $H_{k,e}^+$, is learnt as

$$H_{k,e}^+ = aH\{f_e(e, I(l_k^{EC}))\} + (1-a)H\{f_e(e, t) : \forall t \in EC_k\} \quad (11)$$

where the first term is a pdf approximation of the empirical exemplar feature response of the exemplar e in mode l_k^{EC} in the test image. The second term is a pdf approximation of the exemplar feature responses of the exemplar e , scored against all other exemplars t in the set EC_k . At run-time, the second term is evaluated using a pre-stored matrix F_e that contains pair-wise responses of all training exemplar patches. In calculating the matrix F_e , the same quantization using a kernel K_W is used i.e. allowing limited voxel offset of the local maxima between all pairs of training exemplar patches. $a = 0.9$ is a scalar parameter that influences the relative contribution of the empirical and prior p.d.f. approximations (fixed for all landmarks).

The updated model is applied to the test image using the same standard algorithm as in Section 2 with the personalized dense feature maps and updated spatial constraint means. In detail, the sub-set of training exemplar patches in EC_k are correlated with the test image. We use a Naive Bayes classifier with the learnt models $H_{k,e}^+$, $H_{k,e}^-$ in order to obtain updated dense feature maps.

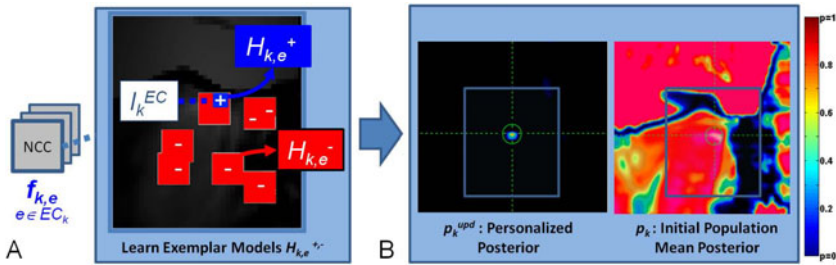


Fig. 5. Learn Personalized Appearance Models. Example for the top of right kidney. A - Learning positive, negative models $H_{k,e}^{+,-}$ of the exemplar feature responses $f_{k,e}$ in the blue (+) region (around the mode l_k^{EC}) and red (-) region. Only exemplars e in EC_k (with maxima near l_k^{EC}) are included in the personalized model. B - The models are re-fitted to produce a personalized posterior. For comparison, we show the initial population model posterior. Images show a coronal slice of a 16x16x16 cm region centered at the ground-truth location.

4 Experimental Results

Our training and test set comprises 63 and 20 patients (split randomly) from a database of lung cancer PET/CT cases. Diagnostic-quality CT scans of the

torso for each patient were acquired using a Siemens Biograph 16 and re-sampled to 2mm isotropic voxels. 22 clinical anatomical landmarks were selected by an expert nuclear medicine physician, according to their utility as reference points for whole-body PET/CT oncology. Ground-truth landmark positions were annotated by a medical trainee, who followed expert guidance. The expert subsequently validated the annotations using a rigorous protocol. To assess the accuracy of the results, we calculated mean RMS error by summarizing the differences between the detected location and the ground-truth over the 20 images.

We compared the new personalized algorithm to a baseline GM with appearance and spatial models built from the entire training population. The average localization error excluding gross mis-detections (defined as the landmarks with RMS over 8 voxels from the ground-truth), improves from 2.77 to 1.94 voxels over all landmarks. Localization accuracy of all landmarks, including gross mis-detections, is shown in Figure 6. M0 refers to the baseline population average model (Section 2) and M1 to the new exemplar model M1. For a number of landmarks, the improvement in localization accuracy is over 50% (e.g. top left kidney, C2 vertebra, femur head, symphysis, ...). These are typically confidently labeled structures which are reliable, but inaccurately, detected with the initial population mean model. The poor accuracy may be caused by high spatial variability and multi-modal appearance, which is averaged out by the (unimodal) population mean model. For example, the top left kidney has intrinsically high spatial variability and multi-modal appearance due to the proximity of the lung, and so benefits strongly from the exemplar model. The least improved landmarks fall broadly into two categories: they are either already rather well localized by the baseline model (e.g. bottom left kidney) or have intrinsically poor discriminability. For example, bottom sternum, defined as the point of attachment to the last rib, is a highly variable structure with ambiguous ground-truth placements and low repeatability even for an expert reader. The same applies to the top of the aortic arch and the center of the bladder. The tip of os coccygeum is often poorly resolvable in our images (2mm resolution) and has low spatial correlation with other landmarks. Finally, the Th12 vertebra is not accurately localized as the currently available constraints are too broad to disambiguate from the adjacent vertebrae that are not included in the model.

Although the primary aim of our method is to improve accuracy of detected landmarks, the exemplar model also improves the reliability of matching. The average maximum error over the 22 landmarks is reduced by 22% (from 9.61 to 7.53 voxels) for the exemplar model compared to the population mean baseline. For completeness, we include results obtained using a deformable registration method¹. The average mean landmark localization error over all 22 landmarks was 8.77 voxels as compared to 2.34 voxels for the proposed method. For most landmarks, our method outperforms deformable registration by over 70%, the smallest improvement is 20% (center of bladder).

¹ Optimized for whole-body CT-CT registration; from a commercial oncology workstation. Each test image was registered to a reference template, selected by finding the nearest training image to the average anatomy

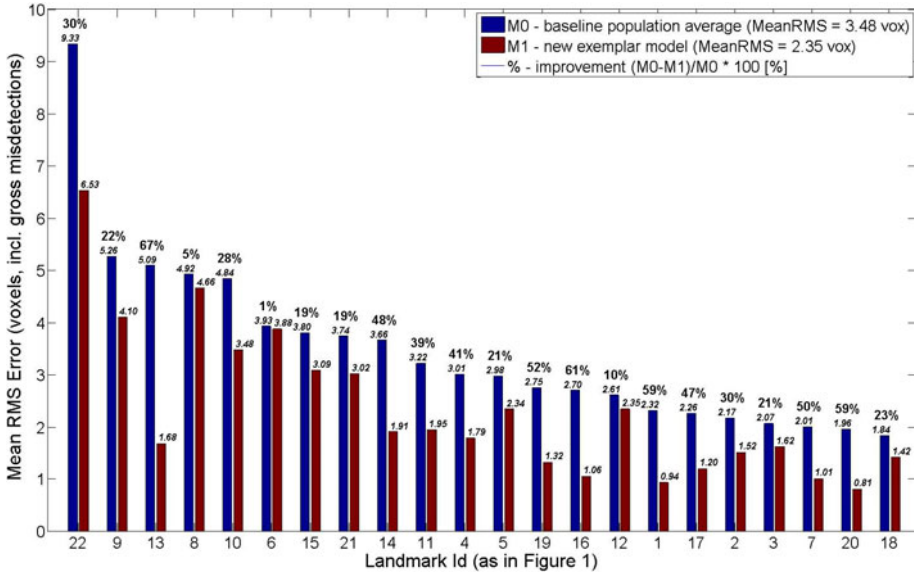


Fig. 6. Average landmark localization accuracy. The numbers in *italic* show mean error (RMS voxels from ground-truth) for the baseline population average (M0) and the new personalized (M1) models for 22 landmarks. Lower error means more accurate localization. % indicate improvement between M1 and M0. Refer to Figure 1 for the list of landmarks (e.g. 20 is symphysis).

5 Discussion

We have presented a personalized method to localize a set of clinical anatomical landmarks which are guaranteed to be present in whole-body oncology images. Our method improves upon both standard deformable registration and population mean GMs by mobilizing directly a subset of the original labeled training exemplars adapted to the test image at runtime (jointly personalizing both the appearance and the spatial term).

We have shown clear improvement in landmark localization using our method (Figures 6). The average mean error over all landmarks is reduced by 32% for the new model compared to the population mean-model and by 73% compared to deformable registration. The highest improvement in average mean error is 67% (top left kidney) compared to GM and 92% (left spina illiaca) compared to deformable registration. The error decreases for all 22 landmarks in our database. It is important to note that the baseline model has been already heavily optimized for localization accuracy in CT e.g. by the selection of features and by learning optimal anisotropic patch sizes for each landmark. Compared to the best performing population average model with isotropic patch size shared by all landmarks, the improvement in average mean error is 38%. Interestingly, the system performance exceeds the intra-observer variance, estimated for our data

by expert physician at 3-5 voxels depending on the landmark; and 10-15 voxels for the bladder, aortic arch and the tip of the sternum. Notably, the free parameters of our method have values shared by all landmarks, despite their widely different patterns of variability. Further improvement could be achieved by optimizing their values for each particular structure.

The inclusion of the personalized spatial priors is essential to making our method perform more accurately. The human body involves many local structures whose appearance is self-similar, so that a spatial prior is essential. The personalized priors we have introduced can be thought of as a Radial Basis Function model, which provides a better approximation to the real anatomical distributions observed in our data than the population mean model. Many of the learnt landmark templates have small patch scales along at least one dimension, which results in chance coincidences as the search region expands away from the true locations if no spatial prior is applied. Furthermore, typically, the local query region is only around 1000 candidate placements and our search can be evaluated in near real-time. As the number of landmarks and training images in the database increases, it would be impractical to evaluate hundreds of templates over 10^8 voxels in a typical CT scan.

6 Conclusion

We have presented an exemplar-based GM for landmark localization that improves over the state of the art. Both appearance and spatial terms leverage directly weighted combinations of labeled exemplar training data, while adhering to the standard GM representation. The current performance of our algorithm that goes beyond inter-expert variability enables a number of clinical oncology applications, including augmented image reading (e.g. automated placement of reference ROIs) and accurate inter-patient localization of arbitrary body regions (e.g. for building normal atlases for radiation therapy planning).

Acknowledgements. We are grateful to Dr. Jean-Marc Peyrat (Siemens) for refining the manuscript, Prof. Daniel Slosman (Clinique Générale-Beaulieu, Geneva, Switzerland) for image database and Mr. Tomas Potesil for help with data annotation.

References

1. Chum, O., Philbin, J., Sivic, J., Isard, M., Zisserman, A.: Total recall: Automatic query expansion with a generative feature model for object retrieval. In: ICCV (2007)
2. Corso, J.J., Alomari, R.S., Chaudhary, V.: Lumbar disc localization and labeling with a probabilistic model on both pixel and object features (chapter). In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 202–210. Springer, Heidelberg (2008)
3. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in ct volumes. In: MICCAI Workshop (2009)

4. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. *IJCV* 61(1), 55–79 (2005)
5. Potesil, V., Kadir, T., Platsch, G., Brady, M.: Improved anatomical landmark localization in medical images using dense matching of graphical models. In: *BMVC* (2010)
6. Ramanan, D.: Learning to parse images of articulated bodies. In: *NIPS* (2007)
7. Schmidt, S., Kappes, J.H., Bergtholdt, M., Pekar, V., Dries, S.P.M., Bystrov, D., Schnörr, C.: Spine detection and labeling using a parts-based graphical model. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007*. LNCS, vol. 4584, pp. 122–133. Springer, Heidelberg (2007)
8. Toews, M., Arbel, T.: A statistical parts-based model of anatomical variability. *IEEE TMI* 26(4), 497–508 (2007)
9. Woerz, S., Rohr, K.: Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models. *Medical Image Analysis* (2006)
10. Zhang, P., Adeshina, S.A., Cootes, T.F.: Automatic learning sparse correspondences for initialising groupwise registration. In: Jiang, T., Navab, N., Plum, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 635–642. Springer, Heidelberg (2010)

Joint Restoration of Bi-contrast MRI Data for Spatial Intensity Non-uniformities

Stathis Hadjidemetriou¹, Martin Buechert², Ute Ludwig¹,
and Juergen Hennig¹

¹ University Medical Center Freiburg, Department of Radiology
60a Breisacher Street, 79106 Freiburg, Germany

² Magnetic Resonance Development and Application Center
60a Breisacher Street, 79106 Freiburg, Germany

Abstract. The reconstruction of MRI data assumes a uniform radio-frequency field. However, in practice the radio-frequency field is inhomogeneous and leads to non-biological intensity non-uniformities across an image. This artifact can complicate further automated analysis of the data. In general, an acquisition protocol provides images of the same anatomic region with multiple contrasts representing similar underlying information, but suffering from different intensity non-uniformities. A method is presented for the joint intensity uniformity restoration of two such images. The effect of the intensity distortion on the auto-co-occurrence statistics of each of the two images as well as on their joint-co-occurrence statistics is modeled and used for their restoration with Wiener filtering. Several regularity constraints for the anatomy and for the non-uniformity are also imposed. Moreover, the method considers an inevitable difference between the signal regions of the two images. The joint treatment of the images can improve the accuracy and the efficiency of the restoration as well as decrease the requirements for additional calibration scans. The effectiveness of the method has been demonstrated extensively with both phantom and real brain anatomic data as well as with real DIXON pairs of fat and water abdominal data.

Keywords: Bi-contrast MRI restoration, MRI bias field correction, generalized co-occurrence statistics, non-stationary Wiener filtering.

1 Introduction

MRI provides high quality three-dimensional data sets that can be used to study anatomy and pathology. The extensive analysis and quantification of the data requires automated methods such as registration and segmentation. This is hampered by the presence of non-biological intensity non-uniformities across an image. This artifact stems from the inhomogeneity of the radio-frequency field due to the coil itself as well as its interaction with the subject that is more pronounced in high field MRI.

There have been attempts to calibrate for the non-uniformity of the radio-frequency field based on its frequency responses to parameterized acquisition

sequences [13]. Also, the combined non-uniformities of the transmission and the receiver coil(s) have been approximated with phantoms. In parallel imaging the sensitivity maps of the coils are also computed. However, the physical correction methods may involve additional acquisitions and are valid only for particular MRI sequences as well as geometries. They do not account for the complicated interaction between the radio-frequency field and the human body. Thus, they are incomplete and they can also be time consuming.

Several post-acquisition image restoration methods have also been proposed [2]. Some of these methods are data driven such as direct low pass filtering [4] and retinex [3]. These methods assume a distinction between lower spatial frequencies corresponding to the non-uniformities and higher ones of the anatomy. Another spatial approach assumes that the image is the piecewise union of uniform regions that can be represented with an integral formulation of image derivatives along their boundaries [10]. The global intensity statistics have also been used for restoration either directly [14] or in combination with tissue intensity priors and tissue spatial priors of an atlas co-registered to the data [9]. The global statistics have been used through their entropy as well [11].

It is common for patient imaging protocols to include multiple sequences. The resulting images suffer from different non-uniformities and post-acquisition methods have been developed for their joint restoration. A variational method enforces smoothness for the non-uniformity fields and preserves the differential structure of the images [6]. The joint restoration has also been implemented with a statistics based method that minimizes the entropy of the joint intensity distribution [17]. Another method for the simultaneous restoration of multiple co-registered images minimizes the sum of the entropy of temporal vectors of intensities that correspond to all spatial image locations [12]. Post-acquisition methods can benefit from regularity properties for the anatomy and for the physical properties of the non-uniformity, are applicable to a range of MRI contrast mechanisms and geometries, as well as decrease the requirements for additional calibration scans.

The method developed in this study performs joint restoration of two images representing the same anatomy with different contrasts. It extends the exclusive use of the auto-co-occurrences of each image [8] to also include the use of the cross- or joint-co-occurrences of the two images. The effect of the intensity distortions on both types of co-occurrence statistics is modeled and used to design Wiener filters for their joint restoration. Several additional constraints have also been formulated and imposed that represent desirable regularities in the data. The inevitable difference between the signal regions of the two images is also considered. The method improves both the efficiency as well as the accuracy of the restoration. Its effectiveness has been demonstrated extensively with Brain-Web phantom images [5], real brain anatomic data sets, as well as abdominal DIXON pairs of fat and water images.

2 Methods

2.1 Data Description

Images from three different data sets have been analyzed. The first consists of seven representative pairs of T_1 and T_2 brain images from the BrainWeb phantom [5]. Both images in a pair have a resolution of $1.0 \times 1.0 \times 1.0 \text{ mm}^3$ and a matrix size of $181 \times 181 \times 217$. They are corrupted with simulated non-uniformities of levels $B = 0\% - 20\% - 40\% - 60\% - 80\% - 100\%$ and noise of $N = 5\%$ as well as one image with $B = 40\%$ and $N = 3\%$. The second brain data set consists of ten real 4.0 *Tesla* pairs of T_1 w and FLAIR images of elderly subjects.¹ The resolution of both images is $1.0 \times 1.0 \times 1.0 \text{ mm}^3$. The matrix size of the T_1 w images is $176 \times 256 \times 256$ and of the FLAIR images is $176 \times 224 \times 256$. The intensity non-uniformity of these images is approximately 2.75 : 1 or 175%. The brain regions for all head images were used exclusively. The BrainWeb database makes it available through the tissue type classification. The brain regions for the real images were computed with the BET tool [15]. The real brain datasets were also co-registered rigidly with the T_1 w image being the reference [16]. The images were also smoothed with a median filter of size $3 \times 3 \times 3 \text{ mm}^3$.

The third group of images are abdominal DIXON data sets. Volunteers participating in an obesity treatment study have been examined on an 1.5 *Tesla* MRI scanner. The images of ten volunteers have been considered. A continuously moving table (CMT) sequence along the axial direction was used. The FOV was adapted to the size of the volunteer and varied between 310 *mm* and 392 *mm* in the phase direction and was 450 *mm* in the read direction. The matrix size is 320×259 . The slice thickness was 5 *mm* without intermediate spacing. Three sets of matrix surface coils were used that resulted in a very high intensity non-uniformity of up to 4 : 1 or 300%. On average 64 slices were acquired to ensure coverage of the abdominal region in the exhaled position. A two-point DIXON reconstruction provided a water and a fat image in the same anatomic space. The abdominal fat image consists of subcutaneous lipids under the skin and visceral lipids in between the organs. The size of the median filter for the DIXON images is $3 \times 3 \text{ voxels}^2$ axially.

2.2 Spatial and Statistical Image Representation

The method involves two images of different contrasts $I_i(\mathbf{x})$, where $i = 0, 1$ and $\mathbf{x} = (x, y, z)$. The images are of the same matrix size and resolution as well as are in the same anatomic space. In general, the two images are corrupted by different spatial intensity non-uniformities B_i , $i = 0, 1$. The B_i represent the combined effect of the transmit and the receive radio-frequency inhomogeneities as well as are assumed to scale the underlying anatomic images $I_{A,i}$, $i = 0, 1$. The anatomic images $I_{A,i}$ are of the same subject and their intensities are related with map g , $I_{A,1}(\mathbf{x}) = g(I_{A,0}(\mathbf{x}))$. Each image is corrupted with additive Rayleigh noise, N_i ,

¹ The images were kindly provided by Prof. Michael W. Weiner, VA, UCSF.

$i = 0, 1$ [7]. The noise is independent and identically distributed. That is, the data model is $I_i = B_i I_{A_i} + N_i$, $i = 0, 1$, where the multiplication between the images is the voxelwise, Hadamard, product.

The non-uniformities, B_i , are expanded in terms of their Taylor series around a voxel \mathbf{x}_0 ,

$$B_i(\mathbf{x}) = B_i(\mathbf{x}_0) + \nabla B_i(\mathbf{x})|_{\mathbf{x}_0}(\mathbf{x} - \mathbf{x}_0) + O(\mathbf{x}^2). \tag{1}$$

The first order term provides a linear approximation of the non-uniformity within a spherical neighborhood \mathcal{N} of radius $\rho = |\mathbf{x} - \mathbf{x}_0|$. The effect of the second and higher order terms is neglected within \mathcal{N} . The statistical representation of images I_0 and I_1 is based on counts of intensities u_0 and u_1 in neighborhood \mathcal{N} to give their co-occurrence statistics as [8,11]:

$$C(I_i, I_j, u_0, u_1) = C_{ij}(u_0, u_1) = \int_{I_i^{-1}(u_0)} \left(\int_{I_j^{-1}(u_1)} (\|\mathbf{x} - \mathbf{x}'\|_2 \leq \rho) d\mathbf{x}' \right) d\mathbf{x}. \tag{2}$$

They give the auto-co-occurrences for $i \equiv j$ and the joint-co-occurrences for $i \neq j$. The different regions of the underlying anatomic images $I_{A,i}$ are assumed to correspond to distinct co-occurrence distributions. An example of the co-occurrences of a pair of T_1 and T_2 BrainWeb phantom images [5] is in figure 1.

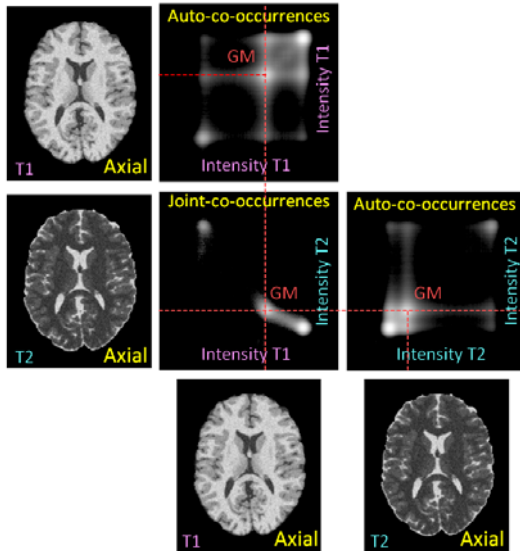


Fig. 1. The auto-co-occurrence statistics and the joint-co-occurrence statistics of the T_1 and T_2 images of the BrainWeb phantom without nonuniformity and noise of 5% [5]. The density in the statistics is in logarithmic scale.

2.3 Effect of Spatial Intensity Non-uniformities on Co-occurrence Statistics

The intensity distortions B_i and the images $I_{A,i}$ are assumed to be generated by independent random variables. Thus, the statistics of the products $B_i I_{A,i}$ correspond to the convolutions of the co-occurrence statistics of the anatomic images $C_{A_i A_i}$ and $C_{A_0 A_1}$ with the point spread functions of the distortions that are non-stationary Gaussian distributions. The effect of the intensity distortion field B in a neighborhood \mathcal{N} around \mathbf{x}_0 can be approximated by the effect of the zero order term $B(\mathbf{x}_0)$ and the first order term $\nabla B|_{\mathbf{x}_0}$ in equation (1) [8]. The zeroth term scales the auto-co-occurrences radially around the origin of $C_{A_i A_i}$ and the first term rotates them around the origin. Thus, the point spread functions effecting $C_{A_i A_i}$ can be more efficiently represented in polar coordinates (r, ϕ) . The standard deviation of the radial scaling is linearly proportional to the radial coordinate $\sigma_r \propto r$. The standard deviation of the rotation σ_ϕ increases with ρ and is largest along the diagonal and zero along the axes. The application of the point spread functions of the intensity distortions to the auto-co-occurrence statistics of the assumed underlying anatomic image given by

$$C_{ii} = C_{A_i A_i} * G(\sigma_{r_i}) * G(\sigma_{\phi_i}), i = 0, 1, \quad (3)$$

where $*$ is convolution, represents the auto-co-occurrences of the distorted image.

The effects of the zero order terms of the non-uniformities B_i of the two images on the joint-co-occurrences $C_{A_0 A_1}$ are also considered. The point spread functions affecting $C_{A_0 A_1}$ are more efficiently represented in Cartesian coordinates. Their sizes are linearly proportional to their distances from the origin $\sigma_{u_i} \propto u_i$. The relation between u_i and r_i is $u_i = r_i/\sqrt{2}$. The application of the point spread functions of the intensity distortions to the joint-co-occurrence statistics of the assumed underlying anatomic images given by

$$C_{01} = C_{A_0 A_1} * G(\sigma_{u_0}) * G(\sigma_{u_1}) \quad (4)$$

represents the joint-co-occurrences of the distorted images.

2.4 Overview of Constrained Restoration

The median filtering removes the high frequency noise N_i . The remaining two products giving I_i , $i = 0, 1$, are separated into B_i and I_{A_i} . The point spread functions of the non-uniformities are non-stationary Gaussians and hence the Wiener deconvolutions are also non-stationary. The relation between the two anatomic images $I_{A1} = g(I_{A0})$ is enforced by restoring their joint statistics. The restored statistics are forced to the images as the first constraint. The fields B_i , $i = 0, 1$, are forced to be smooth with Gaussian filtering to provide the second constraint. The last constraint is to preserve the L_1 norm of the original images, that is $|I_i|_1 = |I_{A_i}|_1$, to also preserve the dynamic range. The constraints are contradictory and the restoration is iterative. At every iteration a coordinate descent is performed for each of the constraints. The estimates of the restoration

fields at iteration t , $W_{i,t} = \frac{1}{B_{i,t}}$, $i = 0, 1$, are the inverses of the non-uniformity distortions $B_{i,t}$. They are applied to $I_{i,t-1}$ to provide updated estimates of the underlying anatomic images $I_{i,t} \equiv I_{A_i,t}$, $i = 0, 1$. The iterations are initialized with the acquired images $I_{A_i,t=0} \equiv I_i$.

2.5 Restoration of the Co-occurrence Statistics

The co-occurrences are restored with Wiener filtering that considers the non-stationary Gaussian point spread functions of the intensity distortions. The Wiener filters for the radial and angular distortions of the auto-co-occurrences introduced in subsection 2.3 are given by

$$f_i(\sigma_{r_i}, \sigma_{\phi_i}) = \frac{G(\sigma_{r_i})}{\|G(\sigma_{r_i})\|_2^2 + \epsilon_1} \times \frac{G(\sigma_{\phi_i})}{\|G(\sigma_{\phi_i})\|_2^2 + \epsilon_1}, \tag{5}$$

where ϵ_1 is a very small constant. The Wiener filter for the joint-co-occurrence statistics of the two images is

$$f_{01}(\sigma_{u_0}, \sigma_{u_1}) = \frac{G(\sigma_{u_0})}{\|G(\sigma_{u_0})\|_2^2 + \epsilon_1} \times \frac{G(\sigma_{u_1})}{\|G(\sigma_{u_1})\|_2^2 + \epsilon_1}. \tag{6}$$

The application of the Wiener filters to the corresponding co-occurrence statistics $C_{ii,t}$ and $C_{01,t}$ provide their restored estimates, $C_{ii,t+1}$ and $C_{01,t+1}$, respectively.

2.6 Enforcement of the Restored Statistics to the Images

The restored statistics are forced back to the images. The restored intensity coordinates in polar form are computed from the auto-co-occurrence statistics with

$$(r'_i, \phi'_i) = \frac{(C_{ii}(r_i, \phi_i) f_i(r_i, \phi_i, \sigma_{r_i}, \sigma_{\phi_i})) * (r_i, \phi_i)}{(C_{ii}(r_i, \phi_i) f_i(r_i, \phi_i, \sigma_{r_i}, \sigma_{\phi_i})) * (1, 1)} \tag{7}$$

and provide the restoration matrices $R_{i,t}^s(r, \phi) = \frac{r'_i}{r_i}$. The restored intensity coordinates considering the joint-co-occurrences are given by

$$(u'_0, u'_1) = \frac{(C_{01}(u_0, u_1) f_{01}(\sigma_{u_0}, \sigma_{u_1})) * (u_0, u_1)}{(C_{01}(u_0, u_1) f_{01}(\sigma_{u_0}, \sigma_{u_1})) * (1, 1)} \tag{8}$$

and provide the restoration matrices $R_{i,t}^b = \frac{u'_i}{u_i}$. The intensity co-occurrences index the corresponding auto-restoration and joint-restoration matrices to provide an initial estimate of the restoration

$$W''_{i,t}(\mathbf{x}) = \frac{1}{2} E_{\Delta \mathbf{x} \in N} (R_{i,t}^s(I_i(\mathbf{x}), I_i(\mathbf{x} + \Delta \mathbf{x})) + R_{i,t}^b(I_0(\mathbf{x}), I_1(\mathbf{x} + \Delta \mathbf{x}))). \tag{9}$$

The estimates W'' are subsequently smoothed with a Gaussian filter to give the restoration fields $W'_{i,t} = W''_{i,t} * G(\sigma_i)$, $i = 0, 1$. The fields $W'_{i,t}$ are rescaled to give $W_{i,t}$, so that the L_1 norm of the restored images is preserved $|I_{it}|_1 = \left| \left(\prod_{t'=0}^t W_{i,t'} \right) I_o \right|_1 = |I_o|_1$.

2.7 Completion of the Iterations

The end condition of the iterations involves the entropy of the auto-co-occurrence statistics $S_{0,t}$ and $S_{1,t}$ as well as of the joint-co-occurrence statistics $S_{01,t}$ for iteration t , that is, $S_{i,t}$, for $i = 0, 1, 01$. The iterations continue as long as all the entropy values continue to decrease. They are terminated when the increase in at least one of the entropy values is considerable, that is,

$\cup_i((S_{i,t+1} - S_{i,t})/S_{i,t} > \epsilon_2)$, where $i = 0, 1, 01$, and $\epsilon_2 < 1$ is a small constant. A maximum number of iterations t_{max} is also imposed. The cumulative restorations $W_{i,t}^{cum} = \prod_{t'=0}^t W_{i,t'}$ are computed for each iteration, t . The field $W_{i,t}^{cum}$ that corresponds to the iteration with minimum entropy $t_{min} = \text{argmin}_{t=0,\dots,t_{max}} S_{01,t}$ provides the restoration field $W_i^{rest} = W_{i,t_{min}}^{cum}$. The restored images are given by $I_i^{rest} = W_i^{rest} I_i$.

2.8 Identification and Restoration of the Valid Image Domain

The non-uniformity is physically present only over the signal regions of an image. In images only partially occupied by signal this region must first be identified. The noise in the non-signal region is Rayleigh and gives the maximum noise intensity or minimum signal intensity I_i^{min} , which is a fraction of the valid dynamic range. The regions with outlying high intensities correspond to artifacts resulting from sources such as blood flow or susceptibility and do not satisfy the non-uniformity assumptions either. Thus, regions corresponding to the low and high intensity ranges are ignored. The remaining regions corresponding to the intermediate intensity ranges $[I_i^{min}, I_i^{max}]$ are back-projected to the respective images. The largest connected components over the back-projections give the valid signal regions, V_i . The image domain over which the non-uniformity is estimated is the union of the valid regions of the two images $V_{01} = V_0 \cup V_1$.

The valid intensity ranges and the corresponding valid signal regions are considered at all steps of the method. The auto-co-occurrences C_{ii}^s in equation (2) are computed over $([I_i^{min}, I_i^{max}], [I_i^{min}, I_i^{max}])$. The joint-co-occurrences C_{01} are computed over the valid intensity ranges in at least one of the two images $([I_0^{min}, I_0^{max}], [0, I_1^{max}])$ for I_0 and $([0, I_0^{max}], [I_1^{min}, I_1^{max}])$ for I_1 . The Wiener filtering for restoration of the statistics in equation (7) is performed in range $([I_i^{min}, I_i^{max}], [I_i^{min}, I_i^{max}])$ for R_i^s . The joint restoration in equation (8) is performed in range $([I_0^{min}, I_0^{max}], [0, I_1^{max}])$ to give R_0^b and in range $([0, I_0^{max}], [I_1^{min}, I_1^{max}])$ to give R_1^b . The Wiener filtering is performed with clipping into the valid intensity ranges as boundary condition. The gain out of the valid ranges is set to unity.

The expectation in equation (9) that provides the initial estimates of the restoration fields W_i'' considers a multiplicative field that is equal to one in the corresponding valid region $V_i(\mathbf{x}) = 1$ and is much less than one $\epsilon_3 \ll 1$ for $V_i(\mathbf{x}) = 0$. The Gaussian smoothing that gives W_i' considers the same multiplicative field. As a result the non-uniformity in regions with opposite validity tends smoothly to unity farther from the valid region and towards the invalid region of the image with $V_i(\mathbf{x}) = 0$. The regions with $V_{01}(\mathbf{x}) = 0$ are set to $W_{i,t}(\mathbf{x}) = 1$.

2.9 Implementation and Efficiency

The restoration being iterative can be time consuming. To improve performance the implementation is in $C++$. Also, the co-occurrence statistics in neighborhood \mathcal{N} are subsampled at regular intervals $\Delta\rho$ and the Gaussian smoothing is performed separably. The dynamic range is also limited to restrict the size of the co-occurrence matrices and the cost of their restoration. The intensity corresponding to the upper 90% percentile of the cumulative histogram, $u_i^{0.9}$, is used. The intensity range up to $1.5 \times u_i^{0.9}$ is preserved, whereas the remaining intensity range is compressed linearly up to $I_i^{max} = 3.0 \times u_i^{0.9}$ to give the range $(1.5 \times u_i^{0.9}, 3.0 \times u_i^{0.9}]$. The minimum intensity I_i^{min} for each image in a pair depends on the standard deviation of the Rayleigh noise σ_{N_i} and its value can be within the range $0.3I_i^{max}$ to $0.45I_i^{max}$. The restoration is accelerated further with an amplification factor $k = 15$ for all four restoration matrices $R_t(u_0, u_1) \leftarrow 1 + k(R_t(u_0, u_1) - 1)$.

The parameters of the co-occurrence statistics are $\rho = 9 \text{ mm}$ and $\Delta\rho = 3 \text{ mm}$. The low value of ρ allows a low angular distortion value, $\sigma_{\phi_i} = 4^\circ$. The variable parameters of the method are $\sigma_{r_1} = \sigma_{r_2}$ and $\sigma_{s_1} = \sigma_{s_2}$. The parameters σ_{r_i} , $i = 0, 1$ are expressed as a fraction of the dynamic range and $\sigma_{u_i} = (1/\sqrt{2})\sigma_{r_i}$. The value of σ_{r_i} is under-estimated and the value of σ_{s_i} is over-estimated. The parameters of the deconvolution filter are $\sigma_{r_i} = 0.02$ for the brain data and $\sigma_{r_i} = 0.05$ for the higher non-uniformities of the DIXON data with $\epsilon_1 = 0.1$. The standard deviation of the Gaussian filter is $\sigma_{s_i} = 140 \text{ pixels}$ for the brain images and $\sigma_{s_i} = 100 \text{ pixels}$ for the lower resolution of the DIXON images with $\epsilon_3 = 0.01$. The parameters controlling the iterations are $\epsilon_2 = 0.2$ and $t_{max} = 100$ for all data sets. To expedite the analysis of the DIXON data only the 15 intermediate axial slices were processed.

3 Experimental Results

Some examples of the analysis of images from all data sets are shown in figure 2, figure 3, figure 4, and figure 5. The figures contain slices from the original and the restored images as well as the co-occurrence statistics in logarithmic scale. The restoration of the BrainWeb phantom with noise 5% and non-uniformity 40% is shown in figure 2. The brain images in figure 2 are shown contrast enhanced by ignoring the dynamic range of the non-signal region. In this example the cerebellum in both the T_1w and the T_2w images becomes brighter and thus its statistics become closer to the mean tissue statistics of the remaining corresponding image. The restoration of a pair of real T_1w and FLAIR images is shown in figure 3. The images in figure 3 are also shown contrast enhanced. In this example the cerebellum becomes brighter and the white matter becomes more uniform in both the T_1w and FLAIR images. The restoration of two example abdominal fat and water DIXON image pairs are in figure 4 and in figure 5. In the DIXON fat images the signal from the subcutaneous fat becomes more uniform and the visceral signal far from the surface coils becomes brighter. The

T_1		T_2		T_1/T_2
Image	Auto-co-occurrences	Image	Auto-co-occurrences	Joint-co-occurrences
		Original		
(a) $I_{0,t=0}$	(b) $C_{00,t=0}$	(c) $I_{1,t=0}$	(d) $C_{11,t=0}$	(e) $C_{01,t=0}$
		Restored		
(f) $I_{0,t_{min}}$	(g) $C_{00,t_{min}}$	(h) $I_{1,t_{min}}$	(i) $C_{11,t_{min}}$	(j) $C_{01,t_{min}}$

Fig. 2. The restoration of a T_1w and a T_2w BrainWeb pair with non-uniformity of 40% and noise of 5%. The images are shown contrast enhanced and the statistics are logarithmic. The cerebellum becomes brighter and the statistics become sharper.

T_1w		FLAIR		$T_1w/FLAIR$
Image	Auto-co-occurrences	Image	Auto-co-occurrences	Joint-co-occurrences
		Original		
(a) $I_{0,t=0}$	(b) $C_{00,t=0}$	(c) $I_{1,t=0}$	(d) $C_{11,t=0}$	(e) $C_{01,t=0}$
		Restored		
(f) $I_{0,t_{min}}$	(g) $C_{00,t_{min}}$	(h) $I_{1,t_{min}}$	(i) $C_{11,t_{min}}$	(j) $C_{01,t_{min}}$

Fig. 3. Example restoration of a T_1w and a FLAIR image pair. The images are shown contrast enhanced and the statistics are logarithmic. The intensity of the white matter in both the T_1w image and the FLAIR images becomes more uniform.

Lipid		Water		Lipid/Water
Image	Auto-co-occurrences	Image	Auto-co-occurrences	Joint-co-occurrences
Original				
(a) $I_{0,t=0}$	(b) $C_{00,t=0}$	(c) $I_{1,t=0}$	(d) $C_{11,t=0}$	(e) $C_{01,t=0}$
Restored				
(f) $I_{0,t_{min}}$	(g) $C_{00,t_{min}}$	(h) $I_{1,t_{min}}$	(i) $C_{11,t_{min}}$	(j) $C_{01,t_{min}}$

Fig. 4. Example restoration of an abdominal fat/water DIXON image pair. The statistics are logarithmic. The subcutaneous fat region becomes more uniform and the visceral region becomes brighter. The uniformity of the water signal also becomes higher.

Lipid		Water		Lipid/Water
Image	Auto-co-occurrences	Image	Auto-co-occurrences	Joint-co-occurrences
Original				
(a) $I_{0,t=0}$	(b) $C_{00,t=0}$	(c) $I_{1,t=0}$	(d) $C_{11,t=0}$	(e) $C_{01,t=0}$
Restored				
(f) $I_{0,t_{min}}$	(g) $C_{00,t_{min}}$	(h) $I_{1,t_{min}}$	(i) $C_{11,t_{min}}$	(j) $C_{01,t_{min}}$

Fig. 5. The restoration of an abdominal fat/water DIXON image pair of an obese subject. The statistics are logarithmic. The intensity uniformity of both the thick subcutaneous and visceral fat regions is significantly improved in (f).

water DIXON images also become more intensity uniform. In all cases both the auto-co-occurrence and the joint-co-occurrence statistics become sharper.

The restoration decreases the entropy of the statistics with the iterations for all brain images. An exception are the phantom brain data without superimposed non-uniformity where the entropy of the statistics is not affected by the restoration. The entropy of the phantom data decreases up to an average of 97 iterations and the entropy of the real data decreases up to an average of 53 iterations. The entropy of the DIXON data is also decreased by the restoration up to an average of 16 iterations. The minimum entropy for the DIXON datasets is achieved with a lower number of iterations due to the larger deconvolution filter.

4 Discussion

Pairs of images from two brain datasets and an abdominal DIXON dataset have been jointly restored for intensity uniformity. The method is able to restore the images despite the high intensity distortions present in many of them. The non-uniformity in the abdominal DIXON images increases with obesity due to the decrease in the coverage of the surface coils. However, obesity also increases the extent of spatial connectedness between the various fat tissues and can thus improve performance. The restoration of images of obese subjects is more clinically relevant in this study. The joint DIXON restoration improves performance compared to the isolated restorations of the fat images. The joint restoration is particularly beneficial to the image in the pair with lower statistical discriminability. In the T_1w and FLAIR pairs this is the FLAIR image, which is also informative for white matter lesion identification.

The method accommodates the inevitable difference between the signal regions of the two images. In the abdominal DIXON fat and water image pairs the signal regions of the two images are approximately complementary. The method uses higher order intensity statistics, the auto- and joint-co-occurrences. They favor the dominant distributions in the data and decrease their variance compared to those of the intensity histogram. Thus, they improve the discriminability between the distributions of the dominant tissues. In brain images the extensive interface between the gray matter and the white matter gives rise to one of the dominant distributions. The global property of the statistics enables the restoration of even disconnected or loosely connected regions such as different brain gray matter regions or different regions of subcutaneous fat. The effect of the distortion field on the co-occurrences is modeled and included in the linear Wiener restoration filters. More sophisticated denoising methods based on Wavelet decompositions or total variation are more appropriate for structured data. The Wiener restoration assumes that the point spread function of the non-uniformity is a unimodal Gaussian. The restoration, however, has been shown to be robust to more general distributions that may result from higher intensity non-uniformities.

In conclusion, a new method has been developed that uses local intensity co-occurrences for effective joint restoration of two anatomic MRI images from intensity non-uniformities. The restoration is robust with respect to contrast mechanisms and subject geometry. Thus, it decreases the need for calibration scans and can reduce acquisition time. The effect of radio-frequency non-uniformity on stimulated echoes can result in contrast non-uniformities as well that are currently ignored, such as in the FLAIR images. The method can be extended in a variety of ways. The performance can be improved by directly using information from calibration scans. Patient examinations provide several images that can be jointly restored by considering all possible image pairs. The method can also be validated quantitatively and used to improve tissue segmentation.

References

1. Arvis, V., Debain, C., Berducat, M., Benassi, A.: Generalization of the cooccurrence matrix for colour images: Application to colour texture classification. *Image Anal. Stereol.* 23, 63–72 (2004)
2. Belaroussi, B., Milles, J., Carme, S., Zhu, Y., Cattin, H.: Intensity non-uniformity correction in MRI: Existing methods and their validation. *Medical Image Analysis* 10, 234–246 (2006)
3. Brainard, D., Wandell, B.: Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A* 3(10), 1651–1661 (1986)
4. Brinkmann, B., Manduca, A., Robb, R.: Optimized homomorphic unsharp masking for MR grayscale intensity correction. *IEEE Trans. on Medical Imaging* 17(2), 161–171 (1998)
5. Cocosco, C., Kollokian, V., Kwan, R.S., Evans, A.: BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage* 5(4-2/4), S425 (1997)
6. Fan, A., Wells, W.M., Fisher, J.W., Çetin, M., Haker, S., Mulkern, R.V., Tempany, C., Willsky, A.S.: A unified variational approach to denoising and bias correction in MR. In: Taylor, C.J., Noble, J.A. (eds.) *IPMI 2003*. LNCS, vol. 2732, pp. 148–159. Springer, Heidelberg (2003)
7. Gudbjartsson, H., Patz, S.: The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine* 34, 910–914 (1995)
8. Hadjdemetriou, S., Studholme, C., Mueller, S., Weiner, M., Schuff, N.: Restoration of MRI data for intensity non-uniformities using local high order intensity statistics. *Medical Image Analysis* 13(1), 36–48 (2009)
9. Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model based bias field correction of MR images of the brain. *IEEE Trans. on Medical Imaging* 18(10), 885–896 (1999)
10. Luo, J., Zhu, Y., Clarysse, P., Magnin, I.: Correction of bias field in MR images using singularity function analysis. *IEEE Trans. on Medical Imaging* 24(8), 1067–1085 (2005)
11. Mangin, J.: Entropy minimization for automatic correction of intensity nonuniformity. In: *Proc. of IEEE Workshop on MMBIA*, pp. 162–169 (2000)
12. Learned-Miller, E.G., Jain, V.: Many heads are better than one: Jointly removing bias from multiple MRIs using nonparametric maximum likelihood. In: Christensen, G.E., Sonka, M. (eds.) *IPMI 2005*. LNCS, vol. 3565, pp. 615–626. Springer, Heidelberg (2005)

13. Noterdaeme, O., Brady, M.: A fast method for computing and correcting intensity inhomogeneities in MRI. In: Proc. of ISBI, pp. 1525–1528 (2008)
14. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. on Medical Imaging* 17(1), 87–97 (1998)
15. Smith, S.: Fast robust automated brain extraction. *Proc. of Human Brain Mapping* 17, 143–155 (2002)
16. Studholme, C.: RView, <http://www.colin-studholme.net/>
17. Vovk, U., Pernus, F., Likar, B.: Intensity inhomogeneity correction of multispectral MR images. *NeuroImage* 32, 54–61 (2006)

The 2D Analytic Signal on RF and B-Mode Ultrasound Images

Christian Wachinger, Tassilo Klein, and Nassir Navab

Computer Aided Medical Procedures (CAMP), TUM, Munich, Germany

Abstract. The fundamental property of the analytic signal is the split of identity, meaning the separation of quantitative and qualitative information in form of the local phase and the local amplitude, respectively. Especially the structural representation, independent of brightness and contrast, of the local phase is interesting for numerous image processing tasks. Recently, the extension of the analytic signal from 1D to 2D, covering also intrinsic 2D structures, was proposed. We show the advantages of this improved concept on ultrasound RF and B-mode images. Precisely, we use the 2D analytic signal for the envelope detection of RF data. This leads to advantages for the extraction of the information-bearing signal from the modulated carrier wave. We illustrate this, first, by visual assessment of the images, and second, by performing goodness-of-fit tests to a Nakagami distribution, indicating a clear improvement of statistical properties. Finally, we show that the 2D analytic signal allows for an improved estimation of local features on B-mode images.

1 Introduction

The analytic signal (AS) enables to extract local, low-level features from images. It has the fundamental property of split of identity, meaning that it separates qualitative and quantitative information of a signal in form of the *local phase* and *local amplitude*, respectively. These quantities further fulfill invariance and equivariance properties [11], allowing for an extraction of structural information that is invariant to brightness or contrast changes in the image. Exactly these properties lead to a multitude of applications in computer vision and medical imaging, such as registration [5,15,20,26,31,33], detection [10,21,25,29], segmentation [11,16,27], and stereo [13]. Phase-based processing is particularly interesting for ultrasound images because they are affected by significant brightness variations [15,16,20,21,26].

For 1D, the local phase is calculated with the 1D analytic signal. For 2D, several extensions of the analytic signal are proposed, with the *monogenic signal* [11] presenting an isotropic extension. The description of the signal's structural information (phase and amplitude) is extended by a geometric component, the *local orientation*. The local orientation indicates the orientation of intrinsic 1D (i1D) structures in 2D images. This already points to the limitation of the monogenic signal; it is restricted to the subclass of i1D signals. Recently, an extension to the monogenic signal, referred to as *2D analytic signal* [28], was proposed that

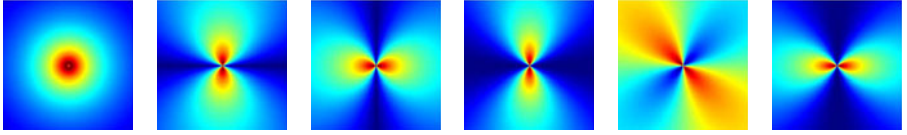


Fig. 1. Magnitude of 2D Hilbert transforms with log-Gabor kernels in frequency domain. From left to right: B , $B \odot H_x^1$, $B \odot H_y^1$, $B \odot H_{xx}^2$, $B \odot H_{yy}^2$, $B \odot H_{xy}^2$.

permits the analysis of i2D signals. Therefore, the 2D signal analysis is embedded into 3D projective space, and a new geometric quantity, the *apex angle*, is introduced. The 2D analytic signal also has the advantage of more accurately estimating local features from i1D signals [28].

In this article, we show the advantages of the calculation of the 2D analytic signal for radio frequency (RF) and B-mode ultrasound images. Instead of performing the demodulation of RF signals for each scan line separately, we perform the demodulation in its 2D context with 2D Hilbert filters of first- and second-order. This leads to advantages in the envelope detection. Since all further processing steps of the creation of the B-mode image are based on the envelope, the improvement of the 2D envelope detection propagates to the quality of the B-mode image. Moreover, the result from the 2D envelope detection bears better statistical properties, as we illustrate with goodness-of-fit tests towards a Nakagami distribution, with its implications to classification and segmentation. Finally, we show the advantages of the 2D analytic signal for estimating local features on B-mode images. All experiments are performed on clinical ultrasound images.

2 2D Analytic Signal

There are various concepts to analyze the phase of signals, such as Fourier phase, instantaneous phase, and local phase [14]. We are primarily interested in the last two. For 1D signals, $g \in L^2(\mathbb{R})$, the instantaneous phase is defined as the argument of the analytic signal, $\arg(g + i \cdot \mathcal{H}\{g\})$, with \mathcal{H} being the Hilbert transform. Since real signals consist of a superposition of multiple signals of different frequencies, the instantaneous phase, although local, can lead to wrong estimates. The signal has to be split up into multiple frequency bands, by means of band-pass filters, to achieve meaningful results, as further described in section 2.2.

Considering 2D signals, $f \in L^2(\mathbb{R}^2)$, the intrinsic dimension expresses the number of degrees of freedom to describe local structures [32]. Intrinsic zero dimensional (i0D) signals are constant signals, i1D signals are lines and edges, and i2D are all other patterns in 2D. The monogenic signal is restricted to i1D signals. The monogenic signal is calculated with the two-dimensional Hilbert transform, also referred to as the Riesz transform. In the frequency domain, the first-order 2D Hilbert transform is obtained with the multiplication of

$$H_x^1(\mathbf{u}) = i \cdot \frac{x}{\|\mathbf{u}\|}, \quad H_y^1(\mathbf{u}) = i \cdot \frac{y}{\|\mathbf{u}\|}, \quad \mathbf{u} = (x, y) \in \mathbb{C} \setminus \{(0, 0)\} \quad (1)$$

with $i = \sqrt{-1}$. For the calculation of the 2D analytic signal, higher order Hilbert transforms are used [28]. The Fourier multipliers of the second-order Hilbert transform [4] are

$$H_{xx}^2(\mathbf{u}) = -\frac{x \cdot x}{\|\mathbf{u}\|^2}, \quad H_{xy}^2(\mathbf{u}) = -\frac{x \cdot y}{\|\mathbf{u}\|^2}, \quad H_{yy}^2(\mathbf{u}) = -\frac{y \cdot y}{\|\mathbf{u}\|^2}, \quad (2)$$

again with $\mathbf{u} = (x, y) \in \mathbb{C} \setminus \{(0, 0)\}$. In contrast to [28], we do not present the formulas of the Hilbert transforms in spatial but in frequency domain, which is more versatile for filtering, see section 2.2. Throughout the article we use upper case letters for filter and signals in frequency domain, and lower case ones for their representation in spatial domain.

2.1 Structural and Geometrical Features

The proposed extension of the 2D analytic signal is obtained by an embedding in 3D projective space. This allows for a differentiation of geometrical features (local orientation, local apex) and structural features (local phase, local amplitude). The filtered signal F_p , the first-order Hilbert transformed signals F_x, F_y , and the second-order Hilbert transformed signals F_{xx}, F_{xy}, F_{yy} are calculated with the bandpass filter B and the pointwise multiplication \odot in frequency domain as

$$\begin{bmatrix} F_p \\ F_x \\ F_y \end{bmatrix} = \begin{bmatrix} B \odot F \\ H_x^1 \odot B \odot F \\ H_y^1 \odot B \odot F \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} F_{xx} \\ F_{xy} \\ F_{yy} \end{bmatrix} = \begin{bmatrix} H_{xx}^2 \odot B \odot F \\ H_{xy}^2 \odot B \odot F \\ H_{yy}^2 \odot B \odot F \end{bmatrix}. \quad (3)$$

We illustrate the Hilbert transforms in frequency domain multiplied with log-Gabor bandpass filters in figure 1. In order to enable an interpretation of second-order Hilbert transformed signals in projective space, an isomorphism between the Hesse matrix and a vector valued representation is used [28], leading to $f_s = \frac{1}{2}[f_{xx} + f_{yy}]$, $f_+ = f_{xy}$, and $f_{+-} = \frac{1}{2}[f_{xx} - f_{yy}]$.

Finally, the local features are calculated as follows. The apex angle α , which differentiates between features of different intrinsic dimensionality, is

$$\alpha = \arccos \frac{\sqrt{f_+^2 + f_{+-}^2}}{\|f_x\|}. \quad (4)$$

With the apex angle, the homogeneous signal component f_h of the signal f_p in projective space is defined as

$$f_h = \sqrt{\frac{1 + \cos \alpha}{2}}. \quad (5)$$

¹ We thank the authors of [28] for discussions and comments on the correct formulas.

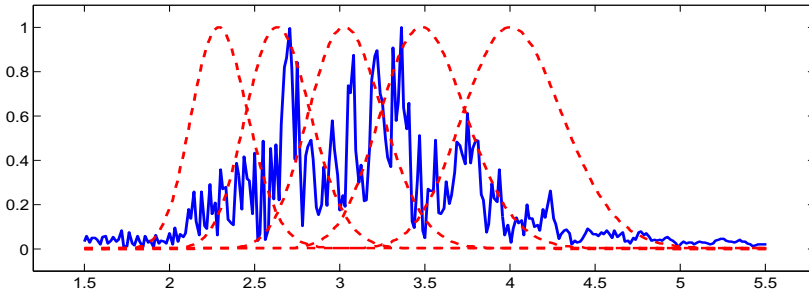


Fig. 2. Log-Gabor filter bank consisting of 5 filters (red) and ultrasound signal spectrum (x-axis: frequency in MHz). Ultrasound acquisition frequency: 3.3 MHz.

The local orientation θ , local phase ϕ , and local amplitude a are calculated with

$$\theta = \frac{1}{2} \arctan \frac{f_+}{f_{+-}}, \quad (6)$$

$$\phi = \text{atan2} \left(\sqrt{[f_h^{-1} f_x]^2 + [f_h^{-1} f_y]^2}, f_p \right), \quad (7)$$

$$a = \frac{1}{2} \sqrt{f_p^2 + [f_h^{-1} f_x]^2 + [f_h^{-1} f_y]^2}. \quad (8)$$

For i1D signals, the homogeneous component is $f_h = 1$, and the formulas above reduce to the ones known from the monogenic signal.

2.2 Frequency Selection

Each signal f can be described with the Fourier series, decomposing the signal into components of different frequencies, each one having its own phase and amplitude. The direct application of the Hilbert transform on the original signal, which presents an accumulation of local signals from different frequencies, does therefore not adequately extract local features. Theoretically, we would need to calculate the analytic signal for infinitely narrow bandwidths, *i.e.*, Dirac deltas in the frequency domain. Following the uncertainty principle this results in filters with global support. Bandpass filters present appropriate approximation for a localization in spatial and frequency domain.

Felsberg *et al.* [11] apply the difference of Poisson kernels for frequency selection. An interesting property of the Poisson filter is that it creates a linear scale-space [12]. Another filter that is commonly applied, especially in ultrasound, is the log-Gabor filter [4, 15, 16, 21, 26]. Also in our analysis on ultrasound images, we achieved better results with the log-Gabor filter so that we concentrate on it in the following. A drawback of the log-Gabor filter is, however, that it has no analytic expression in the spatial domain. This is also the reason why we presented the Hilbert transforms in equations (1) and (2) in frequency and not in spatial domain, as it is done in [28].



Fig. 3. Exemplar ultrasound processing pipeline for RF to B-mode conversion

Important for the design of the filter bank is to create filters, so that the transfer function of each filter overlaps sufficiently with its neighbors, in order to have a uniform coverage of the spectrum. A filter bank with five log-Gabor filters is illustrated in figure 2. A study of alternative bandpass filters is presented in [4]. For the further analysis, it is either possible to focus on the signal at one specific scale, or accumulate all responses from various scales, as it is *e.g.* done for the phase congruency [19].

3 2D Analytic Signal on RF Ultrasound

The pipeline of the RF to B-mode conversion consists of multiple steps, including amongst others demodulation, non-linear intensity mapping, and filtering [17,30], see figure 3. The demodulation is one of the central parts and extracts the information-bearing signal from a modulated carrier wave. In ultrasound processing, the demodulation is commonly performed by an *envelope detection*. Hereby, the amplitude of the analytic signal is calculated for each of the 1D scan lines separately. Interestingly, calculating the amplitude of the 1D analytic signal is equivalent to the *instantaneous amplitude*. In the literature of ultrasound imaging, it is noted that the quality of ultrasound images can be increased by multi-frequency decomposition and compounding of the received signal, simply referred to as *frequency compounding* [6]. This is equivalent to the *local amplitude* estimation. This constitutes an interesting analogy, between the advantages of the frequency compounded signal to the normal one, on the one hand, and the advantage of the local amplitude in comparison to the instantaneous amplitude, on the other hand. We have neither seen this analogy noted in the literature before, nor the application of local amplitude and local phase techniques to RF data.

In contrast to the usually separate processing of each scan line, we consider all scan lines at once and construct the 2D analytic signal to estimate the local amplitude. This enables an improved envelope detection because the signal is analyzed in its 2D context by also considering information in lateral direction. The balance between influence from lateral and axial direction can be adjusted by the bandwidth in each direction of the bandpass filter, where the smaller spacing in axial direction should be considered accordingly.

3.1 Envelope and B-Mode Results

We perform experiments on multiple RF images acquired from the neck with a linear transducer at 3.3 MHz. The sampling frequency of the RF data is 40 MHz. We compare the envelope detection for: i) 1D analytic signal (AS), ii) 1D analytic

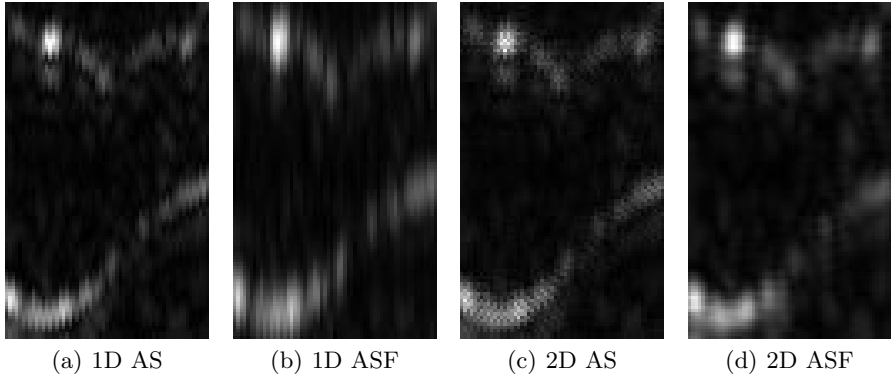


Fig. 4. Magnified region of envelope detected 2D image for various envelopes

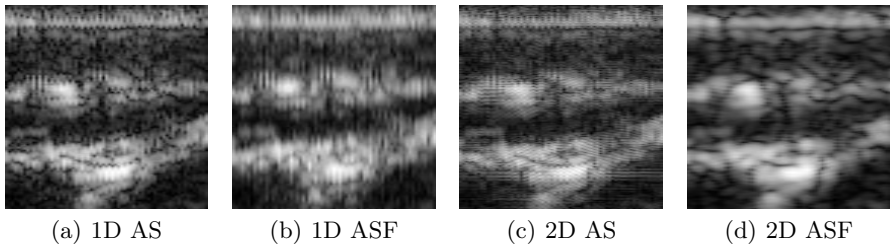


Fig. 5. Magnified regions of images after log-compression

signal with filter bank (ASF), iii) monogenic signal (MS), and iv) monogenic signal with filter bank (MSF), v) 2D AS, and vi) 2D ASF. Exemplarily, we show the frequency spectrum of one scan line together with the log-Gabor filter bank in figure 2. We present magnified regions of the various envelope images in figure 4. Note that we do not show the results of the MS due to limitation of space and better results for 2D AS. However, we include them into the analysis of noise statistics in section 3.2. We can clearly observe that the 2D analytic signal leads to a more accurate and consistent extraction of structures. This becomes particularly clear on the circular structure on the top left, which appears rather ellipsoidal on the estimates from the 1D analytic signal. We also note the positive influence of the filter bank for the estimation of the 2D analytic signal.

We perform an RF to B-mode conversion of local amplitude images a with a log-compression including a translation of 25, $\log(a + 25)$. The results for the various envelopes are presented in figure 5. The B-mode image resulting from the 2D analytic signal clearly shows more consistent structures and less noise. Typically, further filtering steps are applied to the log-compressed image to improve its visual appearance. These processing steps are proprietary to the manufacturer and generally not publicly accessible. Ultrasonix (Redmond, Canada), however, distributes a particular research system with a specific SDK including their post-processing filter, called MUCRO. We apply MUCRO to the

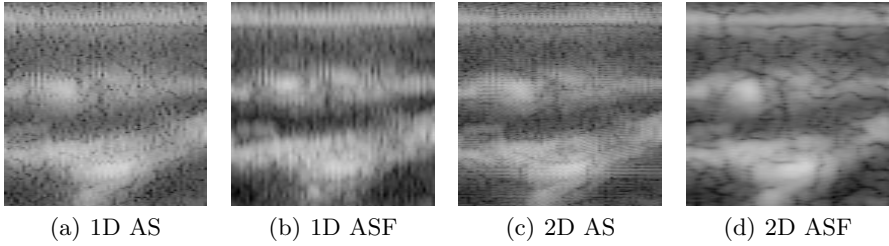


Fig. 6. Magnified regions of images after log-compression and MUCRO

log-compressed images, with the results shown in figure 6. Even after the application of MUCRO, the advantages of the images from the 2D analytic signal are clearly visible. This is not self-evident because the post-processing methods are designed to be applied to 1D envelope detected images, still leaving room for improvement by adapting the post-processing to 2D envelope estimation. Finally, one of the reasons for applying the post-processing filtering is to establish consistency between scan lines, which we already achieve by the 2D envelope detection.

3.2 Analysis of Envelope Statistics

Next to the visual assessment of the 2D envelope detection, we also analyze the statistical properties of the data. Different statistical models have been proposed to model ultrasound data. Among them there is Rician [23], pre-Rician K [18], Generalized K-distribution, homodyned K-distribution as well as Rician Inverse of the Gaussian [9]. Common to all these distributions is the inherent complexity, limiting its practical applicability. In order to address this issue, the Nakagami distribution [22] was proposed, because it admits an explicit analytical expression. It is used in various applications to model backscatter characteristics of US envelope data for segmentation and classification, see [8,24] and references therein. In the following, we analyze the effects of the 2D envelope detection on the speckle statistics choosing the Nakagami model. In particular, we quantify the impact of the 2D analytic signal with goodness-of-fit (GOF) tests, and show the potential for the aforementioned applications based on example images.

The Nakagami distribution $\mathcal{N}(x | \mu, \omega)$ belongs to the exponential family and is controlled by two parameters, μ and ω , specifying shape and scale respectively

$$\mathcal{N}(x | \mu, \omega) = \frac{2\mu^\mu x^{2\mu-1}}{\Gamma(\mu)\omega^\mu} \exp\left(-\frac{\mu}{\omega}x^2\right), \forall x \in \mathbb{R}_+. \quad (9)$$

A goodness-of-fit test evaluates if the data d_1, \dots, d_n , under the assumption of i.i.d. samples, comes from the given theoretical probability distribution p [7]. Note that conventional GOF tests are restricted to the case of single distributions. For inhomogeneous regions in the image, however, a mixture of Nakagami is more appropriate, see figure 8 for an illustration of a misfit of a single Nakagami to mixture Nakagami data as well as a perfect mixture fit. Consequently,

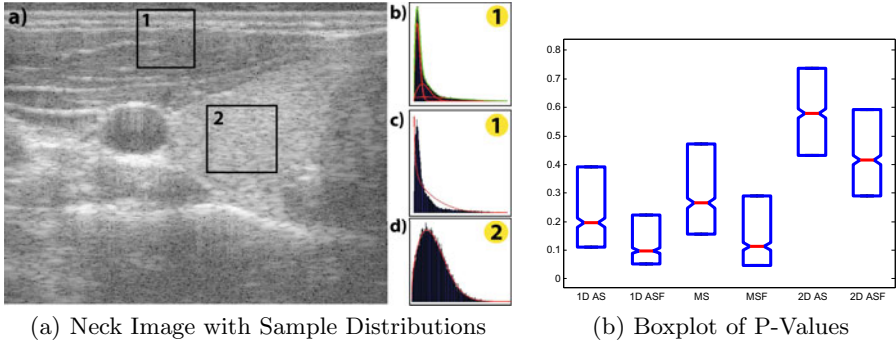


Fig. 7. Left: RF image and distribution estimation of two areas. Region 1 contains a mixture of Nakagami, region 2 a single Nakagami. Whereas MLE can fit nicely in region 2 (d) it expectedly performs poorly in region 1 (c), that can only be represented properly by mixture (b). Right: Box plot of P-values for different envelope detections.

we can only achieve good results with the GOF test on homogeneous image regions. The mixture case has to be further evaluated, with similar results to be expected.

For the GOF test, the range of the data is partitioned into M bins β_i , $i = 1, \dots, M$, with N_i and the number of samples per bin. Moore [4] suggests to divide the data into $M = 2n^{\frac{2}{5}}$ bins. Furthermore, we assume the bins to be equiprobable as suggested in [3]. In this regard, we let p_i be the integral of the distribution in the range β_i given the parameters of the distribution $\theta = \{\mu, \omega\}$

$$p_i = \int_{\beta_i} p(x | \theta) dx. \tag{10}$$

Hence, p_i expresses the likelihood of a sample to be in the bin β_i (identical for all bins). The test statistics underlying the GOF test is the sum of differences between observed and expected outcome frequencies

$$X^2 = \sum_{i=1}^M \frac{N_i - np_i}{np_i}. \tag{11}$$

This yields a quadratic form in N_i that has approximately a χ distribution with $M - N - 1$ degrees of freedom and $N = 2$ the number parameters of the distribution. In order to assess the GOF quantitatively, we employ the P-value bases hypothesis test. The P-value serves as an indicator of how likely the null hypothesis H_0 is true. In our case, H_0 is the hypothesis that the observations are Nakagami distributed, leading to the following calculation of the P-value

$$P = \int_{X^2}^{\infty} \chi^2(M - N - 1) dx, \tag{12}$$

employing equation (11) as the lower bound of integration.

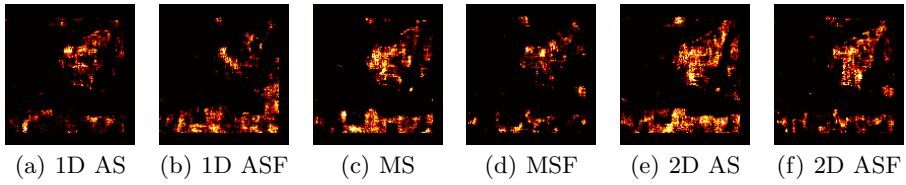


Fig. 8. The P-values are calculated for all patches of an envelope image. Pixel brightness indicates P-value. We perform the calculation for various envelope detection techniques. Comparing the P-value images to the B-mode image in figure 7(a), we see that the bright regions correspond to homogeneous regions in the US image.

3.3 Rao-Robson Statistic

Given the data, we first have to estimate the parameters μ, ω of the Nakagami distribution before the GOF test is performed. This is, however, opposing the general assumption that the parameters of the distribution are a-priori given before the test is performed. Therefore, another quadratic form in N_i has to be used, with the Rao-Robson statistic being one possibility [7]. Considering the parametric form of the distribution $p(x|\theta)$ and the maximum likelihood estimate $\hat{\theta}$, the Rao-Robson statistic is

$$RR = V^\top(\hat{\theta})Q(\hat{\theta})V(\hat{\theta}) \tag{13}$$

with

$$V(\theta) = \frac{N_i - np_i}{(np_i)^{1/2}} \tag{14}$$

$$Q(\theta) = I + D(\theta)[J(\theta) - D^\top(\theta)D(\theta)]^{-1}D^\top(\theta) \tag{15}$$

$$D_{ij}(\theta) = p_i(\theta)^{\frac{1}{2}} \frac{\partial p_i(\theta)}{\partial \theta_j} \tag{16}$$

$J(\theta)$ is the $N \times N$ Fisher information matrix and I is an $M \times M$ identity matrix. The Rao-Robson statistic is χ^2 distributed with $M - N - 1$ degrees of freedom, leading to P-values computed using

$$P = \int_{RR}^{\infty} \chi^2(M - N - 1)dx \tag{17}$$

with the Rao-Robson statistic RR as lower bound of integration.

3.4 Statistical Results

We perform the Rao-Robson GOF test on local patches of size 180×20 , densely throughout the image. Plotting the results for all patches, creates therefore a new image with the intensity values being the P-values. We show these images in figure 8 for the various envelope detections. The brighter the images, the

higher the P-values, and consequently the better for statistical applications because we achieve better fits. We note that the bright regions are corresponding to the homogeneous areas in the ultrasound image because only these areas are appropriately modeled with a single distribution, as discussed previously. Additionally, we calculate the statistics of the P-values, visualized in the box plot of figure 7(b). The red line indicates the median and the box is constructed from the interquartile range. Our results therefore show that the envelope detection without the filter bank produces better fits, which makes sense, because log-Gabor filters influence the distribution. More importantly, however, we note the improvement from 1D AS to MS, and further from MS to 2D AS. This shows on the one hand, the advantage of applying 2D Hilbert transforms in contrast to 1D ones, and on the other hand, the advantage of the 2D analytic signal in contrast to the monogenic signal. This confirms the visually improved results for 2D envelope detection from the previous section.

4 2D Analytic Signal on B-Mode Images

Next to the benefits of the 2D analytic signal for the demodulation of RF data, it also allows for a more accurate estimation of local features on B-mode images [28]. This has the potential to increase the quality of follow-up applications such as registration [15,20,26,33], segmentation [16], and detection [21], that use the local features as input. To demonstrate this, we calculate the local orientation on B-mode images showing a biopsy needle. In figure 9, we illustrate the local orientation that is estimated from the monogenic signal and the 2D analytic signal, both with filtering. The estimation from the monogenic signal shows no consistent orientation information in the region of the needle. In contrast, the improved concept of the 2D analytic signal indicates a consistent result.

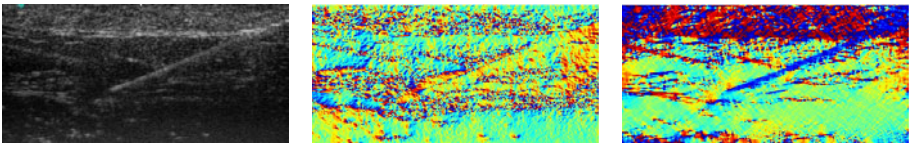


Fig. 9. Ultrasound image with biopsy needle (left). Calculated local orientation for monogenic signal (middle) and 2D analytic signal (right).

5 Conclusion

We demonstrated that the application of the 2D analytic signal has multiple advantageous for RF and B-mode data. The demodulation of RF signals with the 2D analytic signal enables a more consistent extraction of structures, because the signal is analyzed in its natural 2D context. We further showed that the improved envelope detection enables the creation of B-mode images of enhanced quality. To validate this, we applied a proprietary post-processing filtering for

ultrasound on the log-compressed images and compared the result of 1D and 2D analytic signal. Moreover, we validated the improved statistical properties of envelope data resulting from the 2D analytic signal by performing goodness-of-fit tests to a Nakagami distribution. Finally, the advanced signal model of the 2D analytic signal leads to benefits in the estimation of local features in B-mode images, as we have illustrated for the case of needle detection.

For the demodulation, we focused on scans from a linear transducer. For curved linear transducers, the application of 2D Hilbert transforms without a previous scan conversion can be achieved with the polar Fourier transform [2].

Acknowledgement. This work was partly funded by the European Commission. We thank Lennart Wietzke for active discussions.

References

1. Ali, R., Gooding, M., Christlieb, M., Brady, M.: Advanced phase-based segmentation of multiple cells from brightfield microscopy images. In: ISBI (2008)
2. Averbuch, A., Coifman, R., Donoho, D., Elad, M., Israeli, M.: Fast and accurate polar Fourier transform. *Applied and Computational Harmonic Analysis* 21(2), 145–167 (2006)
3. Bock, R.K., Krischer, W.: *The Data Analysis Briefbook*, 1st edn. Springer-Verlag New York, Inc., Secaucus (1998)
4. Boukerroui, D., Noble, J., Brady, M.: On the Choice of Band-Pass Quadrature Filters. *Journal of Mathematical Imaging and Vision* 21(1), 53–80 (2004)
5. Carneiro, G., Jepson, A.D.: Phase-based local features. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2350, pp. 282–296. Springer, Heidelberg (2002)
6. Cincotti, G., Loi, G., Pappalardo, M.: Frequency decomposition and compounding of ultrasound medical images with wavelet packets. *IEEE Transactions on Medical Imaging* 20(8), 764–771 (2001)
7. D’Agostino, R., Stephens, M.: *Goodness-of-fit techniques*, p. 560. Marcel Dekker, Inc., New York (1986)
8. Destremes, F., Meunier, J., Giroux, M.F., Soulez, G., Cloutier, G.: Segmentation in ultrasonic b-mode images of healthy carotid arteries using mixtures of nakagami distributions and stochastic optimization. *IEEE Trans. Med. Imaging* 28(2), 215–229 (2009), <http://dx.doi.org/10.1109/TMI.2008.929098>
9. Dutt, V., Greenleaf, J.F.: Ultrasound echo envelope analysis using a homodyned k-distribution signal model. *Ultrason. Imag.* 16(4-5), 265–287 (1994)
10. Estépar, R.S.J., Washko, G.G., Silverman, E.K., Reilly, J.J., Kikinis, R., Westin, C.-F.: Accurate airway wall estimation using phase congruency. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 125–134. Springer, Heidelberg (2006)
11. Felsberg, M., Sommer, G.: The monogenic signal. *IEEE Transactions on Signal Processing* 49(12), 3136–3144 (2001)
12. Felsberg, M., Sommer, G.: The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and Vision* 21(1), 5–26 (2004)
13. Fleet, D., Jepson, A., Jenkin, M.: Phase-based disparity measurement. *CVGIP: Image Understanding* 53(2), 198–210 (1991)

14. Granlund, G.H., Knutsson, H.: Signal Processing for Computer Vision. Kluwer Academic Publishers, Dordrecht (1995)
15. Grau, V., Becher, H., Noble, J.: Registration of multiview real-time 3-d echocardiographic sequences. *IEEE Transactions on Medical Imaging* 26(9), 1154–1165 (2007)
16. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A.J., Rohling, R.: Bone segmentation and fracture detection in ultrasound using 3D local phase features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 287–295. Springer, Heidelberg (2008)
17. Hedrick, W.R., Hykes, D.L., Starchman, D.E.: *Ultrasound Physics and Instrumentation*, 4th edn., Mosby (2004)
18. Jakeman, E., Pusey, P.N.: A model for non-rayleigh sea echo. *IEEE Trans. Antennas Propag.* 24(4-5), 806–814 (1976)
19. Kovesi, P.: Image features from phase congruency. *Videre: Journal of Computer Vision Research* 1(3) (1999)
20. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. *Medical Image Analysis* 9(4), 330–343 (2005)
21. Mulet-Parada, M., Noble, J.: 2D+ T acoustic boundary detection in echocardiography. *Medical Image Analysis* 4(1), 21–30 (2000)
22. Nakagami, N.: The m-distribution, a general formula for intensity distribution of rapid fading. In: Hoffman, W.G. (ed.) *Statistical Methods in Radio Wave Propagation*, Pergamon, Oxford (1960)
23. Shankar, P.M., Reid, J.M., Ortega, H., Piccoli, C.W., Goldberg, B.B.: Use of non-rayleigh statistics for the identification of tumors in ultrasonic b-scans of the breast. *IEEE Trans. Med. Imag.* 12(4-5), 687–692 (1993)
24. Shankar, P., Dumane, V., Reid, J., Genis, V., Forsberg, F., Piccoli, C., Goldberg, B.: Classification of ultrasonic b-mode images of breast masses using nakagami distribution. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* 48(2), 569–580 (2002)
25. Szilágyi, T., Brady, S.M.: Feature extraction from cancer images using local phase congruency: a reliable source of image descriptors. In: ISBI, pp. 1219–1222 (2009)
26. Wachinger, C., Navab, N.: Alignment of viewing-angle dependent ultrasound images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 779–786. Springer, Heidelberg (2009)
27. Wang, P., Kelly, C., Brady, M.: Application of 3d local phase theory in vessel segmentation. In: ISBI, pp. 1174–1177 (2009)
28. Wietzke, L., Sommer, G., Fleischmann, O.: The geometry of 2d image signals. In: CVPR, pp. 1690–1697 (2009)
29. Xiaoxun, Z., Yunde, J.: Local Steerable Phase (LSP) Feature for Face Representation and Recognition. In: CVPR, vol. 2 (2006)
30. Zagzebski, J.: *Essentials Of Ultrasound Physics*, 1st edn., Mosby (1996)
31. Zang, D., Wietzke, L., Schmaltz, C., Sommer, G.: Dense optical flow estimation from the monogenic curvature tensor. In: *Scale Space and Variational Methods in Computer Vision*, pp. 239–250 (2007)
32. Zetsche, C., Barth, E.: Fundamental limits of linear filters in the visual processing of two dimensional signals. *Vision Research* (30) (1990)
33. Zhang, W., Noble, J.A., Brady, J.M.: Spatio-temporal registration of real time 3D ultrasound to cardiovascular MR sequences. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 343–350. Springer, Heidelberg (2007)

A Compressed Sensing Approach for MR Tissue Contrast Synthesis

Snehashis Roy, Aaron Carass, and Jerry Prince

Image Analysis and Communication Laboratory,
Dept. of Electrical and Computer Engg.,
The Johns Hopkins University, USA
{snehashisr, aaron_carass, prince}@jhu.edu
<http://iac1.ece.jhu.edu/Prince>

Abstract. The tissue contrast of a magnetic resonance (MR) neuroimaging data set has a major impact on image analysis tasks like registration and segmentation. It has been one of the core challenges of medical imaging to guarantee the consistency of these tasks regardless of the contrasts of the MR data. Inconsistencies in image analysis are attributable in part to variations in tissue contrast, which in turn arise from operator variations during image acquisition as well as software and hardware differences in the MR scanners. It is also a common problem that images with a desired tissue contrast are completely missing in a given data set for reasons of cost, acquisition time, forgetfulness, or patient comfort. Absence of this data can hamper the detailed, automatic analysis of some or all data sets in a scientific study. A method to synthesize missing MR tissue contrasts from available acquired images using an atlas containing the desired contrast and a patch-based compressed sensing strategy is described. An important application of this general approach is to synthesize a particular tissue contrast from multiple studies using a single atlas, thereby normalizing all data sets into a common intensity space. Experiments on real data, obtained using different scanners and pulse sequences, show improvement in segmentation consistency, which could be extremely valuable in the pooling of multi-site multi-scanner neuroimaging studies.

Keywords: compressed sensing, magnetic resonance imaging (MRI), image synthesis, phantom, standardization, segmentation, intensity normalization, histogram matching, histogram equalization.

1 Introduction

Magnetic resonance (MR) imaging (MRI) is a noninvasive imaging modality that is the gold standard for imaging the brain. MR image processing, particularly segmentation of brain structures, has been used to further the understanding of normal aging or the progression of diseases such as multiple sclerosis, Alzheimer's disease, and schizophrenia. Large multi-site and multi-center studies are often used to gather more data across a broader population or to carry out follow-up

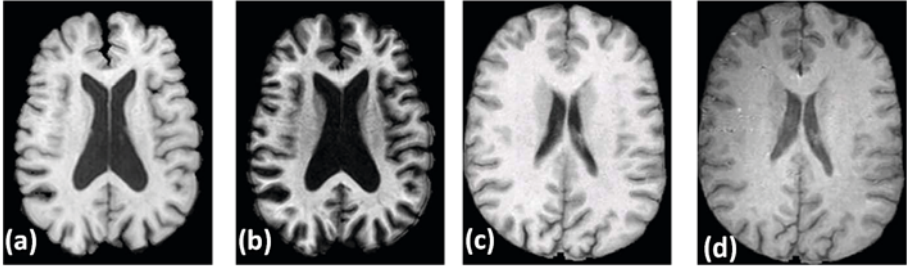


Fig. 1. Data acquired under different pulse sequences or different scanners: (a) An acquisition of a Spoiled Gradient Recalled (SPGR) sequence in GE 3T scanner [16], (b) same subject with Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence in GE 3T scanner, (c) another subject [12], SPGR, GE 1.5T scanner, (d) same subject with MPRAGE sequence in a Siemens 3.0T scanner. Evidently, tissue contrast is dependent on the choice of pulse sequences, as well as on the choice of scanner. The contrast is intrinsically dependent on the choice of MR acquisition parameters, like flip angle, repetition time, echo time *etc.*

imaging in longitudinal studies [12,16]. Because the intensities in conventional MR images do not have specific numerical units (unlike those in computed tomography (CT) images), there is a major concern about the consistency of quantitative results obtained in these studies due to the different pulse sequences [5] and scanners [11] that are used. In fact, because the intensity scale and the tissue contrast are dependent on the MR acquisition pulse sequence as well as the underlying T_1 , T_2 , T_2^* , P_D values in the tissue (cf. Fig. 1), any image processing task that is carried out on these data cannot normally be expected to behave quite the same across subjects. In this work, we focus on the consistency of image processing tasks—particularly on the task of image segmentation—using conventionally acquired MR data.

Numerous methods have been proposed to find the segmentation of cortical and sub-cortical structures of the human brain. Some of these methods assume a statistical model [17] for the probability distribution function (pdf) of the image intensity and generate a segmentation rule based on a maximum likelihood (ML) estimation [8] or maximum a-priori (MAP) estimation of the pdf parameters. Another popular class of segmentation method is based on Fuzzy C-Means (FCM) [2] and its modifications [1]. But all of these methods are intrinsically dependent on the tissue contrast of the image, which is dependent on the acquisition parameters. Accordingly, the same data acquired under different pulse sequences or different scanners, having different contrasts, yielding inconsistent segmentations [5].

One way of reducing the segmentation inconsistency is intensity normalization. Some popular ways to do this are histogram equalization or histogram matching by non-rigid registration [15], intensity scaling based on landmarks

[18], intensity correction by a global piecewise linear [19] or a polynomial mapping [14]. A common problem with these methods is that the histogram matching is never perfect with discrete valued images. Also landmark based methods are mostly initialized by manually chosen landmarks, which are time consuming to create and lack robustness. Another way to normalize involves the use of the peak intensities of white matter (WM) or gray matter (GM) [4] to match the image histogram to a target histogram by information theoretic methods [25]. Histograms of the segmented sub-cortical structures can also be matched to the individual histograms of the sub-cortical structures of a registered atlas [13]. It has been shown to produce consistent sub-cortical segmentation over datasets acquired under different scanners. In spite of their efficiency, all these methods are intrinsically dependent on the accuracy of the underlying segmentation of the image and the atlas-to-subject registration method. Another group of techniques have been proposed to include the MR acquisition physics into the segmentation methodology [21][11] to normalize all intensity images by their underlying T_1 , T_2 , T_2^* and P_D values. They suffer from the fact that many images are to be acquired with precise acquisition parameters, which is often not possible in large scale studies.

In this paper, we propose an MR image example-based contrast synthesis (MIMECS) method, that addresses the problem of intensity standardization over pulse sequences or scanners. We build upon the concepts of compressed sensing [3][9] to develop a patch matching approach, that uses patches from an atlas to synthesize different MR contrast images for a given subject. The primary purpose of MIMECS synthesis—and therefore the application for evaluation of its efficacy—is consistent segmentation across pulse sequences and scanners.

There are a few key differences with previous methods of intensity normalizations [19][18][25][4][13][11][21]. MIMECS is a pre-processing step to the segmentation. It neither needs to estimate T_1 , T_2 , P_D values, nor does it need the intrinsic MR acquisition parameters, like echo time (TE) or repetition time (TR). It also does not need landmarks or any atlas-to-subject registration, thus being fully automatic and independent of the choice of registration methods used in any subsequent processing. We use it to normalize a dataset which is acquired under different pulse sequences and on different scanners [12], thus having diverse contrasts. On this dataset, we show that MIMECS can normalize all the images to a particular contrast, that produce more consistent cortical segmentations compared to their original contrasts.

The paper is organized as follows. First, we briefly summarize compressed sensing in Sec. 2.1. Then we describe the imaging model in Sec. 3.1 and the contrast synthesis algorithm is explained in Sec. 3.3. A validation study with phantoms is described in Sec. 4.1. Then, we use MIMECS to simulate alternate pulse sequences and show its applicability as an intensity normalization and histogram matching process on a set of multi-site and multi-scanner data in Sec. 4.2.

2 Background

2.1 Compressed Sensing

We use the idea of compressed sensing in our MIMACS approach. Compressed sensing exactly recovers sparse vectors from their projections onto a set of random vectors [9,3] that need not form a basis. The idea behind compressed sensing comes from the fact that most of the signals that we observe are usually sparse, thus it is better not to observe the full signal, but a part of it, and reconstruct the whole signal from those small number of measurements.

Suppose we want to reconstruct a signal $\mathbf{x} \in \mathbb{R}^n$ which is s -sparse, i.e. has at most s non-zero elements. We want to observe another vector $\mathbf{y} \in \mathbb{R}^d$, $s < d < n$, such that each element of \mathbf{y} can be obtained by an inner product of \mathbf{x} and another vector from \mathbb{R}^n . Then, compressed sensing can be used to reconstruct $\mathbf{x} \in \mathbb{R}^n$ exactly from $\mathbf{y} \in \mathbb{R}^d$, with $\mathbf{y} = \Phi \mathbf{x}$, $d < n$, \mathbf{x} being s -sparse, $\Phi \in \mathbb{R}^{d \times n}$. Thus, compressed sensing is also a way to reduce the dimension of the observed sparse data in a lossless way.

One approach for finding \mathbf{x} is to solve

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_0 \quad \text{such that } \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 < \epsilon_1, \quad (1)$$

where ϵ_1 is the noise in the measurement and $\|\cdot\|_0$ indicates the number of non-zero elements in the vector. Although this approach provides some simple conditions on Φ [10], it is an NP-hard problem. Another approach is to solve

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_1 \quad \text{such that } \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 < \epsilon_2, \quad (2)$$

where $\|\mathbf{x}\|_1$ is the L_1 norm of a vector. This is a convex problem and can be transformed into a linear program that can be solved in polynomial time. If ϵ_2 is unknown, Eqn. 2 can be written in the following form,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{\|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1\}, \quad (3)$$

where λ is a weighing factor. The sparsity on $\hat{\mathbf{x}}$ increases as λ increases.

It has been shown that if Φ follows the global restricted isometry property (RIP) [3], then the solutions to Eqn. 1 and Eqn. 2 are identical and the optimal solution can be obtained by an L_1 minimization problem using Eqn. 3. This result is interesting because it has been shown that *random* subsets of incoherent matrices satisfy the RIP [24]. Thus, to reconstruct \mathbf{x} , it is possible to observe its projections onto a set of previously chosen incoherent vectors. We use this idea to find a sparse vector for each patch and then use that sparse vector as an index into the atlas.

3 Method

3.1 Imaging Model

Consider two MR images \mathbf{Y}_1 and \mathbf{Y}_2 of the same person, acquired at the same time, but having two different MR contrasts, C1 and C2, labeled as 1 and 2. E.g.

\mathbf{Y}_1 and \mathbf{Y}_2 can be either T1-w, T2-w or PD-w images. They could be either 2D slices or 3D volumes. The imaging equations can be written as,

$$\mathbf{Y}_1 = \mathcal{W}_1(T_1, T_2, P_D, T_2^*; \Theta_1) + \eta_1 \quad (4)$$

$$\mathbf{Y}_2 = \mathcal{W}_2(T_1, T_2, P_D, T_2^*; \Theta_2) + \eta_2 \quad (5)$$

where Θ_1 and Θ_2 are the intrinsic imaging parameters, like TR, TE, flip angle etc for that particular acquisition, \mathcal{W}_1 and \mathcal{W}_2 are imaging equations [7] corresponding to the contrast or the pulse sequence used, η_1 and η_2 are random noise. T_1 , T_2 and P_D are relaxation times and the proton density maps of the tissues. They could be 2D or 3D maps, according to \mathbf{Y}_1 and \mathbf{Y}_2 .

Ideally, if a number of C1 contrast images of the subject is acquired, with the Θ 's known for all the acquisitions, then an estimate of the underlying T_1 , T_2 and P_D maps can be obtained by either directly inverting \mathcal{W}_1 in Eqn. 4 [11] or by a least square estimate [21]. Then \mathbf{Y}_2 can directly be synthesized using Θ_2 and the estimates of T_1 , T_2 and P_D 's using Eqn. 5.

There are several drawbacks for this strategy. Θ_1 and Θ_2 are often not known, \mathcal{W}_1 and \mathcal{W}_2 are difficult to model accurately or multiple acquisitions are not taken. Therefore, it is almost impossible to reconstruct \mathbf{Y}_2 from \mathbf{Y}_1 using the straight-forward approach. We will try to synthesize \mathbf{Y}_2 from \mathbf{Y}_1 using an atlas.

3.2 Atlas Description

Define an atlas \mathcal{A} as a pair of images, $\mathcal{A} = \{\phi_1, \phi_2\}$, where ϕ_1 and ϕ_2 are C1 and C2 contrasts of the same subject having the same resolution. We assume that ϕ_1 and ϕ_2 are co-registered. Also assume that ϕ_1 and ϕ_2 are made of $p \times q \times r$ 3D patches. For convenience, we assume that each of the 3D patches is stacked into a 1D vector of size $d \times 1$, $d = p \times q \times r$. The patches are then denoted by $d \times 1$ vectors $\phi_1(i)$ and $\phi_2(i)$, $i \in \Omega_\phi$. Ω_ϕ is the image domain of both ϕ_1 and ϕ_2 , as they are co-registered. Then we define the C1 and C2 contrast dictionaries Φ_1 and $\Phi_2 \in \mathbb{R}^{d \times N}$, where the columns of Φ_1 and Φ_2 are patches $\phi_1(i)$ and $\phi_2(i)$, $i \in \Omega_\phi$ from the atlas and $N = |\Omega_\phi|$ is the number of patches from \mathcal{A} . Clearly, a column of Φ_1 corresponds to the same column in Φ_2 .

3.3 Contrast Synthesis

Now, given a subject image \mathbf{Y}_1 of contrast C1, we want to generate its C2 contrast using Φ 's. \mathbf{Y}_1 is first decomposed into $d \times 1$ patches $\mathbf{y}_1(j)$, $j \in \Omega_Y$, Ω_Y is the input image domain. The primary idea of MIMCECS is that each subject patch $\mathbf{y}_1(j)$ can be matched to one or more patches from the dictionary Φ_1 , because they are of the same C1 contrast. The matching patches have their own C2 contrast counterparts in Φ_2 , which are then used to synthesize the C2 contrast version of $\mathbf{y}_1(j)$. As a result, the need of any atlas to subject registration is eliminated. This idea of patch matching can be explained efficiently using the idea of sparse priors [26] in a compressed sensing paradigm.

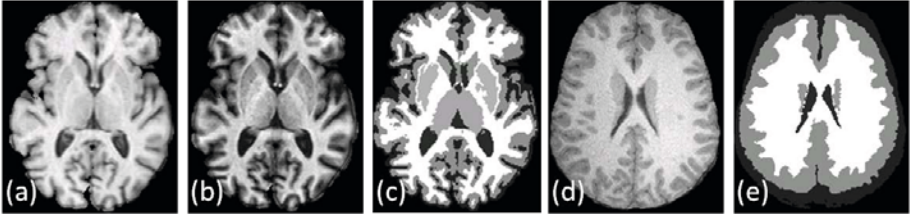


Fig. 2. Modified dictionary using segmentation: (a) An example SPGR T1-w contrast atlas image ϕ_1 [16] from a GE 3T scanner, (b) its MPRAGE T1-w contrast ϕ_2 , (c) their TOADS [1] segmentation \mathbf{S}_ϕ , used to generate the sub-dictionaries $\Phi_1^{(l)}$ and $\Phi_2^{(l)}$ according to Eqn. 10, (d) An SPGR T1-w subject image \mathbf{Y}_1 [12], (e) an approximate hard segmentation \mathbf{S}_Y (obtained from TOADS), used to choose a reduced sub-dictionary

From Eqn. 4 and Eqn. 5, if \mathcal{W}_1 and \mathcal{W}_2 are linear operators of $\Theta = [T_1, T_2, T_2^*, P_D, \theta_1, \theta_2]$, then a pseudo-inverse on Eqn. 4 will provide \mathbf{Y}_2 from \mathbf{Y}_1 . Ignoring noise, the reconstruction of \mathbf{Y}_2 can be written as,

$$\mathbf{Y}_2 = \mathbf{G}\mathbf{Y}_1, \quad \mathbf{G} = \mathcal{W}_2(\mathcal{W}_1^T \mathcal{W}_1)^{-1} \mathcal{W}_1^T \mathbf{Y}_1. \tag{6}$$

This inverse problem is ill-posed because \mathbf{G} is almost always never known. The problem is regularized by assuming that any patch $\mathbf{y}_1(j)$ can be found from a rich and over-complete dictionary Φ_1 , having the same contrast as $\mathbf{y}_1(j)$, by

$$\mathbf{y}_1(j) \approx \Phi_1 \mathbf{x}_1(j), \text{ for some } \mathbf{x}_1(j) \in \mathbb{R}^N, \text{ with } \|\mathbf{x}_1(j)\|_0 \ll N, j \in \Omega_Y. \tag{7}$$

$\|\cdot\|_0$ is the L_0 norm, denoting number of non-zero elements in a vector. Intuitively, the sparsest representation $\mathbf{x}_1(j)$ denotes an index of Φ_1 , such that $\mathbf{y}_1(j)$ is matched to a particular column of Φ_1 . This regularized problem is written in a compressed sensing paradigm following Eqn. 3.

$$\mathbf{x}_1(j) = \arg \min_{\alpha} [\|\mathbf{y}_1(j) - \Phi_1 \alpha\|_2^2 + \lambda \|\alpha\|_1], \quad \alpha \geq \mathbf{0}, j \in \Omega_Y, \tag{8}$$

where λ is a weighing factor, as defined in Eqn. 3. A positivity constraint on the sparse representation $\mathbf{x}_1(j)$ is enforced to impose a valid anatomical meaning on the elements of $\mathbf{x}_1(j)$, such that an element of $\mathbf{x}_1(j)$ denotes a positive weight of how much a column of Φ_1 contributes in reconstructing $\mathbf{y}_1(j)$. This positivity constraint was previously explored in Lasso [23].

With the sparse representation of $\mathbf{y}_1(j)$, we simply reconstruct the C2 contrast patch as,

$$\hat{\mathbf{y}}_2(j) = \Phi_2 \mathbf{x}_1(j). \tag{9}$$

Then the C2 contrast image \mathbf{Y}_2 is reconstructed by combining all the $\hat{\mathbf{y}}_2(j)$'s thus obtained.

3.4 Contrast Synthesis with Modified Dictionary

Typically for a $256 \times 256 \times 198$ image volume, number of example patches $N \approx 100,000$. It is computationally very intensive to work with Φ_1 that is of the order of $27 \times 100,000$. To reduce the computation overhead, we use a dictionary selection procedure that uses an approximate segmentation of the input image. The rationale can be seen from the fact that if a patch $\mathbf{y}_1(j)$ is known to come from a certain tissue type, e.g., pure white matter (WM), the computation of Eqn. 8 can be reduced by choosing a subset of Φ_1 that contains only pure WM patches.

We break down Φ_1 , as well as Φ_2 , into several sub-dictionaries of separate tissue classes, which are obtained from a topology preserving anatomy driven segmentation, called TOADS [11]. The atlas ϕ_1 is segmented into 6 tissue classes, namely cerebro-spinal fluid (CSF), ventricles, gray matter (GM), white matter (WM), basal ganglia and thalamus, labeled $l \in L = \{1, \dots, 6\}$, respectively. The segmentation of ϕ_1 , denoted as \mathbf{S}_ϕ , is similarly decomposed into patches $\mathbf{s}_\phi(i), i \in \Omega_\phi$. The sub-dictionaries $\Phi_1^{(l)}$ and $\Phi_2^{(l)}$ for each tissue class l are generated as,

$$\forall i \in \Omega_\phi, \phi_1(i) \in \Phi_1^{(l)}, \phi_2(i) \in \Phi_2^{(l)}, \text{ if } l \text{ is an element of } \mathbf{s}_\phi(i). \quad (10)$$

Fig. 2(a)–(c) show atlas SPGR and MPRAGE contrasts (ϕ_1 and ϕ_2) and their segmentation (\mathbf{S}_ϕ).

In our experiments, typically, $|\Phi_1^{(l)}| \approx 10,000$ for a $256 \times 256 \times 198$ image. With the smaller atlas sub-dictionaries, we modify Eqn. 8 so as to reduce the search space for $\mathbf{y}_1(j)$'s. An approximate segmentation of \mathbf{Y}_1 , denoted as \mathbf{S}_Y , is computed using one iteration of TOADS (e.g., Fig. 2(d)–(e)). \mathbf{S}_Y is again decomposed into patches $\mathbf{s}_Y(j), j \in \Omega_Y$, so that the information about the tissue classes for $\mathbf{y}_1(j)$ is obtained from $\mathbf{s}_Y(j)$. Now, the contrast synthesis algorithm described in Eqn. 8–9 is modified with the inclusion of the sub-dictionaries,

1. Divide the atlases ϕ_1, ϕ_2 and the segmentation \mathbf{S}_ϕ into $d \times 1$ patches, and generate the sub-dictionaries $\Phi_1^{(l)}$ and $\Phi_2^{(l)}$ according to Eqn. 10.
2. Find an approximate segmentation of the input C1 contrast \mathbf{Y}_1 as \mathbf{S}_Y . Divide the \mathbf{Y}_1 and \mathbf{S}_Y into $d \times 1$ patches $\mathbf{y}_1(j)$'s and $\mathbf{s}_Y(j)$'s, $j \in \Omega_Y$.
3. For each j , find all the tissue classes l that $\mathbf{s}_Y(j)$ contains, $l \in L$.
4. For each j , generate patch specific dictionaries $\Phi_1(j)$ by concatenating all $\Phi_1^{(\ell)}$'s and $\Phi_2(j)$ by concatenating $\Phi_2^{(\ell)}$'s, $\ell \in L$. Thus $\Phi_1(j)$ contains all the potential classes that $\mathbf{y}_1(j)$ could belong to, according to its atlas based approximate segmentation. Clearly, $\Phi_1(j), \Phi_2(j) \in \mathbb{R}^{d \times N_j}$ with $N_j \ll N$.
At this point, if $\Phi_1(j)$ becomes too large, we randomly choose a $d \times N_0$ random subset of $\Phi_1(j)$ and the corresponding subset of $\Phi_2(j)$ for further analysis to minimize computational overhead, $N_0 < N_j$.
5. Solve Eqn. 8 with $\mathbf{y}_1(j)$ and $\Phi_1(j)$ to get $\mathbf{x}_1(j)$.
6. Solve Eqn. 9 with $\Phi_2(j)$ and $\mathbf{x}_1(j)$ thus obtained.
7. Repeat for every $j \in \Omega_Y$.

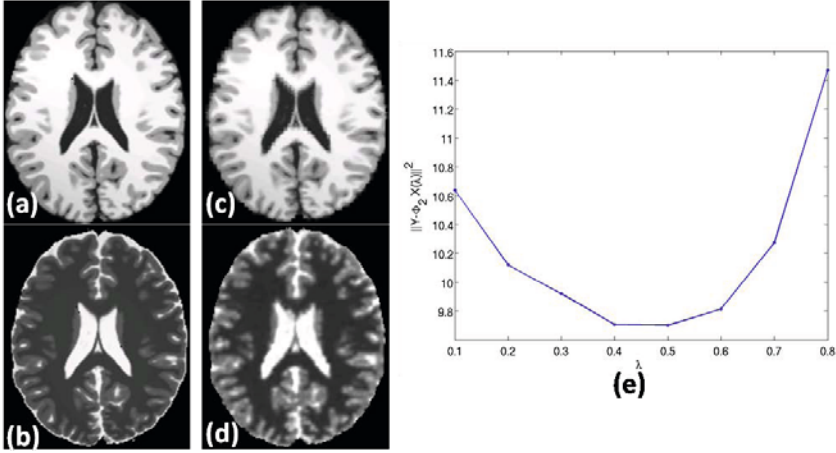


Fig. 3. Optimal λ : (a) A T1-w Brainweb phantom ϕ_1 [6], (b) original T2-w version ϕ_2 , they are used as atlas. (c) Reconstructed T1-w image $\Phi_1 \mathbf{X}(\hat{\lambda})$, from Eqn. (11). (d) Reconstructed T2-w image ($\hat{\mathbf{Y}}_2 = \Phi_2 \mathbf{X}(\hat{\lambda})$) using $\mathbf{Y}_1 = \phi_1$, (e) plot of λ vs. the reconstruction error $\|\mathbf{Y}_2 - \hat{\mathbf{Y}}_2\|_2^2$.

4 Results

4.1 Estimation of λ

We use $3 \times 3 \times 3$ patches in all our experiments. All the images are normalized so that their peak WM intensities are unity. The optimal value of λ is obtained using a cross-validation study on the Brainweb phantoms [6] following the idea of homotopy methods [20]. We use the atlas C1 contrast as a T1-w (ϕ_1) and C2 as a T2-w (ϕ_2) phantom having 0% noise, shown in Fig. 3(a)–(b). Now, using these two phantoms as atlas, we use the same ϕ_1 phantom as the test image \mathbf{Y}_1 and try to reconstruct its T2-w contrast $\hat{\mathbf{Y}}_2$, while the true T2-w contrast is already known as $\phi_2 = \mathbf{Y}_2$. The optimal λ is obtained by,

$$\begin{aligned} \hat{\lambda} &= \arg \min_{\lambda} \|\mathbf{Y}_2 - \hat{\mathbf{Y}}_2\|_2^2, \quad \hat{\mathbf{Y}}_2 = \Phi_2 \mathbf{X}(\lambda) \\ \text{such that } \mathbf{X}(\lambda) &= \arg \min_{\alpha} [\|\mathbf{Y}_1 - \Phi_1 \alpha\|_2^2 + \lambda \|\alpha\|_1], \end{aligned} \quad (11)$$

while Φ_1 and Φ_2 are obtained from ϕ_1 and ϕ_2 as defined earlier. λ is varied from $[0.1, 0.8]$ and the optimal $\lambda = 0.5$ from Fig. 3(d). Using this λ , we generate a T2 contrast of the phantom shown in Fig. 3(c). Ideally, if we are to use all the patches from Φ_1 instead of creating sub-dictionary $\Phi_1^{(l)}$, $\hat{\lambda} \approx 0$. However, the use of a reduced sized dictionary gives improvement in computational efficiency with a sub-optimal performance, as seen from the reconstructed image (Fig. 3(c)). We use this lambda for all the subsequent experiments.

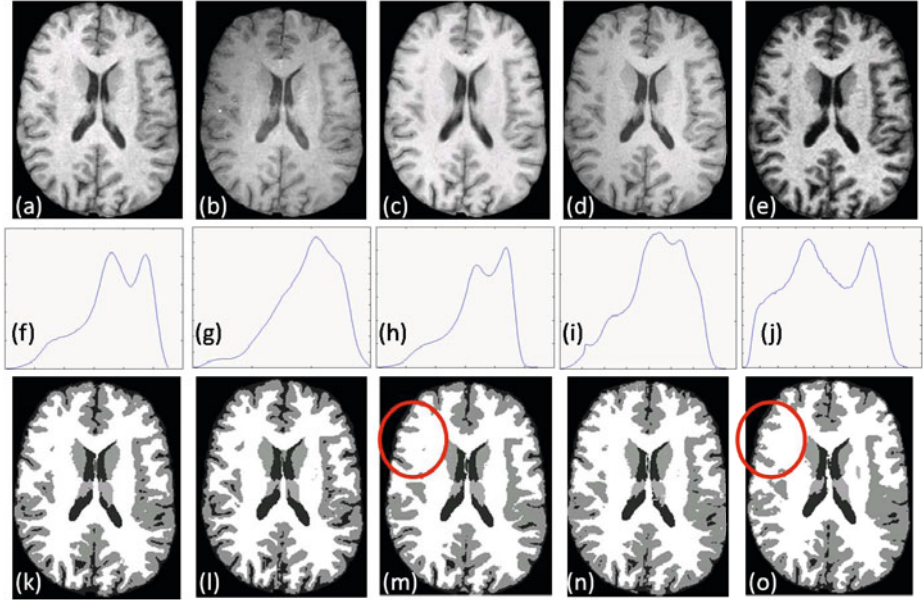


Fig. 4. BIRN data: T1-w SPGR contrasts from (a) GE 1.5T, (b) GE 4T, (c) Philips 1.5T, (d) Siemens 3T scanner, (e) T1-w MPRAGE contrast from Siemens 1.5T. Their histograms are shown in (f)–(j) and the hard segmentations [\[1\]](#) are shown in (k)–(o). The histograms are quite different which is reflected on the difference in the segmentations.

4.2 Experiment on BIRN Data

There has been concerns over consistency of the segmentations arising from scanner variability. To show that synthetic contrasts can produce consistent segmentations, we use the “traveling subject” data [\[12\]](#) that consists of scans of the same person under different pulse sequences and different scanners. These data were downloaded from the BIRN data repository, project accession number 2007-BDR-6UHZ1. We use the scans of 3 subjects, each consisting of one SPGR acquisitions from 1.5T GE, 4T GE, 1.5T Philips and 3T Siemens scanners each and one MPRAGE acquisition from a 1.5T Siemens scanner, each of them having $0.86 \times 0.86 \times 1.2$ mm resolution. One of the subjects is shown in Fig. [4](#)(a)–(e). Their histograms are quite different (cf. Fig. [4](#)(f)–(j)), and the Kullback-Leibler distance (KL distance) between each pair of the normalized histograms (Table [1](#)) affirms this fact. This difference in histograms affect the consistency in the segmentations, shown in Fig. [4](#)(k)–(o).

With the aim of having consistent segmentation of these images, we want to normalize the intensities of all the images to a particular target contrast. The purpose of choosing MPRAGE as the target C2 contrast is to have better delineation of cortical structures, because the GM-WM contrast is very poor on

Table 1. KL distances between $\text{original/synthesized}$ image histograms for images acquired under 5 different scanners, averaged over 3 subjects

	GE 1.5T	GE 4T	Philips 1.5T	Siemens 3T	Siemens 1.5T
GE 1.5T	.	0.28/0.04	0.39/0.06	0.30/0.02	0.49/0.03
GE 4T	0.72/0.04	.	1.22/0.02	0.11/0.06	0.91/0.07
Philips 1.5T	0.48/0.05	0.85/0.23	.	0.88/0.04	0.07/0.07
Siemens 3T	0.53/0.01	0.07/0.04	0.98/0.03	.	0.88/0.06
Siemens 1.5T	0.62/0.03	1.01/0.06	0.05/0.01	1.04/0.08	.

SPGR acquisitions. We use a GE 3T T1-w SPGR and its MPRAGE acquisitions from the BLSA study [16] as atlas ϕ_1 and ϕ_2 , shown in Fig. 2(a)–(b). The synthesized images are shown in Fig. 5(a)–(e). It is seen from the images that they have more similar contrasts than the original images, and this is visually confirmed by the histograms in Fig. 5(f)–(j). Also, Table 1 shows that the histograms of the synthetic MPRAGEs are more similar, having the KL distances being an order of magnitude less.

To show the improvement in segmentation consistency, we compare our method with a registration based approach, where we deformably register the SPGR atlas ϕ_1 to the SPGR input image Y_1 by ABA [22] and use that transformation

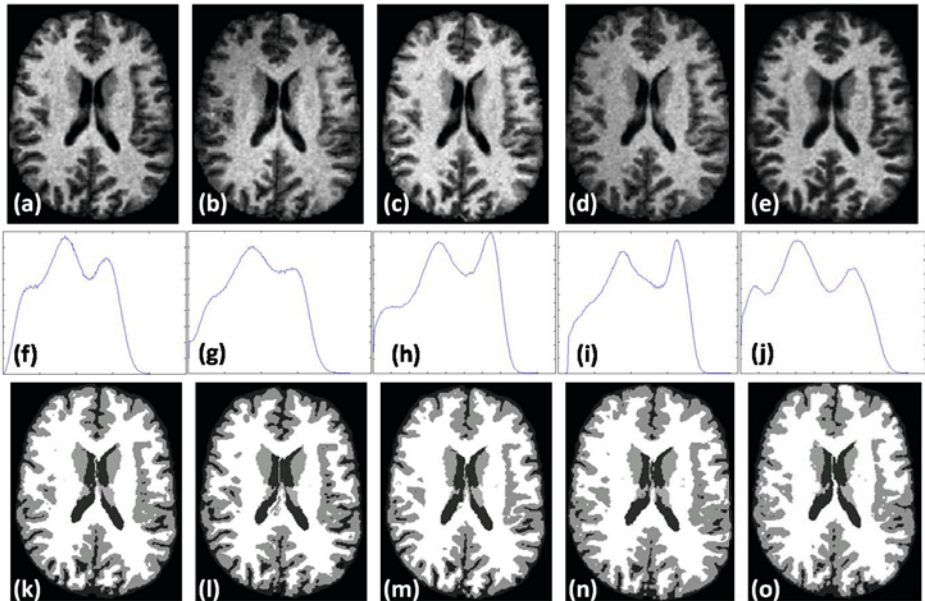
**Fig. 5.** Synthesizing MPRAGE contrast from BIRN data: Synthetic MPRAGE contrasts of the data shown in Fig. 4(a)–(e) using the atlas shown in Fig. 2(a)–(c). The histograms of the synthetic MPRAGEs are shown in (f)–(j) and the hard segmentations [1] are shown in (k)–(o).

Table 2. Consistency comparison: Segmentation consistency is measured on original 4 SPGR and 1 MPRAGE scans, their MPRAGE registered and transformed versions (described in Sec. 4.2) and synthesized MPRAGES. Mean Dice coefficients, averaged over CSF, GM and WM, are reported between the segmentations of (original,registered,synthetic) images acquired under 5 different scanners/pulse sequences (GE 1.5T SPGR , GE 4T SPGR, Philips 1.5T SPGR, Siemens 3T SPGR and Siemens 1.5T MPRAGE) averaged over 3 subjects.

	GE 1.5T	GE 4T	Philips 1.5T	Siemens 3T	Siemens 1.5T
GE 1.5T	.	0.81, 0.78, 0.84	0.77, 0.83, 0.85	0.86, 0.87, 0.89	0.82, 0.80, 0.87
GE 4T	.	.	0.70, 0.71, 0.79	0.79, 0.77, 0.84	0.76, 0.77, 0.82
Philips 1.5T	.	.	.	0.78, 0.83, 0.84	0.74, 0.75, 0.84
Siemens 3T	0.78, 0.81, 0.85

on ϕ_2 to generate the \hat{Y}_2 . Then the segmentation is performed on the registered images. If Y_1 is MPRAGE (e.g. in Siemens 1.5T scanner), ϕ_2 is simply registered to Y_1 to get \hat{Y}_2 . The Dice coefficients for the original, registered and transformed images and the synthetic images are shown in Table 2. Original images usually show poor consistency, especially between SPGR and MPRAGE acquisitions, which was already reported in literature [5]. Table 2 also shows that the Dice between SPGR and MPRAGE sequences is improved between the synthetic MPRAGE versions of the images compared to their original scans.

5 Conclusion

We have developed a compressed sensing based method, called MIMECS, that uses multiple contrast atlases to generate multiple contrasts of MR images. It is a patch-based method, where a patch in the subject C1 contrast image is matched to an atlas to generate a corresponding C2 contrast patch. An application of MIMECS is to normalize intensities between MR images taken from different scanners and different pulse sequences to generate synthetic images, while the synthetic images produce more consistent segmentation. In all our experiments, we have used only one pair of images as atlas to generate patch dictionaries. Also we have not particularly taken into account the topology of the structures. Our future work includes inclusion of features like topology and labels as a matching criteria, instead of only using intensities. Also we would like to explore the possibility of including more atlases and more contrasts.

Acknowledgment

Some data used for this study were downloaded from the Biomedical Informatics Research Network (BIRN) Data Repository (<http://fbirnbdr.nbirn.net:8080/>)

BDR/), supported by grants to the BIRN Coordinating Center (U24-RR019701), Function BIRN (U24-RR021992), Morphometry BIRN (U24-RR021382), and Mouse BIRN (U24 - RR021760) Testbeds funded by the National Center for Research Resources at the National Institutes of Health, U.S.A.

This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. We are grateful to Dr. Susan Resnick for providing the data and all the participants of the Baltimore Longitudinal Study on Aging (BLSA), as well as the neuroimaging staff for their dedication to these studies.

References

1. Bazin, P.L., Pham, D.L.: Topology-preserving tissue classification of magnetic resonance brain images. *IEEE Trans. on Medical Imaging* 26(4), 487–496 (2007)
2. Bezdek, J.C.: A Convergence Theorem for the Fuzzy ISO-DATA Clustering Algorithms. *IEEE Trans. on Pattern Anal. Machine Intell.* 20(1), 1–8 (1980)
3. Candes, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Comm. on Pure and Appl. Math.* 59(8), 1207–1223 (2006)
4. Christensen, J.D.: Normalization of brain magnetic resonance images using histogram even-order derivative analysis. *Mag. Res. Imaging* 21(7), 817–820 (2003)
5. Clark, K.A., Woods, R.P., Rottenber, D.A., Toga, A.W., Mazziotta, J.C.: Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *NeuroImage* 29(1), 185–202 (2006)
6. Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Evans, A.C.: BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage* 5(4), S425 (1997), <http://www.bic.mni.mcgill.ca/brainweb/>
7. Deichmann, R., Good, C.D., Josephs, O., Ashburner, J., Turner, R.: Optimization of 3-D MP-RAGE Sequences for Structural Brain Imaging. *NeuroImage* 12(3), 112–127 (2000)
8. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of Royal Stat. Soc.* 39, 1–38 (1977)
9. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inf. Theory* 52(4), 1289–1306 (2006)
10. Elad, M., Bruckstein, A.M.: A Generalized Uncertainty Principle and Sparse Representation in Pairs of Bases. *IEEE Trans. Inf. Theory* 48(9), 2558–2567 (2002)
11. Fischl, B., Salat, D.H., van der Kouwe, A.J.W., Makris, N., Segonne, F., Quinn, B.T., Dale, A.M.: Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23(1), 69–84 (2004)
12. Friedman, L., Stern, H., Brown, G.G., Mathalon, D.H., Turner, J., Glover, G.H., Gollub, R.L., Lauriello, J., Lim, K.O., Cannon, T., Greve, D.N., Bockholt, H.J., Belger, A., Mueller, B., Doty, M.J., He, J., Wells, W., Smyth, P., Pieper, S., Kim, S., Kubicki, M., Vangel, M., Potkin, S.G.: Test-Retest and Between-Site Reliability in a Multicenter fMRI Study. *Human Brain Mapping* 29(8), 958–972 (2008)
13. Han, X., Fischl, B.: Atlas Renormalization for Improved Brain MR Image Segmentation Across Scanner Platforms. *IEEE Trans. Med. Imag.* 26(4), 479–486 (2007)
14. He, R., Datta, S., Tao, G., Narayana, P.A.: Information measures-based intensity standardization of MRI. In: *Intl. Conf. Engg. in Med. and Biology Soc.*, pp. 2233–2236 (August 2008)

15. Jager, F., Nyul, L., Frericks, B., Wacker, F., Hornegger, J.: Whole Body MRI Intensity Standardization. In: Bildverarbeitung für die Medizin 2008. Informatik aktuell, ch. 20, Springer, Heidelberg (2007)
16. Kawas, C., Gary, S., Brookmeyer, R., Fozard, J., Zonderman, A.: Age-specific incidence rates of Alzheimer's disease: the Baltimore Longitudinal Study of Aging. *Neurology* 54(11), 2072–2077 (2000)
17. Leemput, K.V., Maes, F., Vandermeulen, D., Suetens, P.: Automated Model-Based Tnumber Classification of MR Images of the Brain. *IEEE Trans. on Med. Imag.* 18(10), 897–908 (1999)
18. Madabhushi, A., Udupa, J.K.: New methods of MR image intensity standardization via generalized scale. *Med. Phys.* 33(9), 3426–3434 (2006)
19. Nyul, L.G., Udupa, J.K.: On Standardizing the MR Image Intensity Scale. *Mag. Res. in Medicine* 42(6), 1072–1081 (1999)
20. Osborne, M.R., Presnell, B., Turlach, B.A.: A new approach to variable selection in least squares problems. *IMA J. Numerical Analysis* 20(3), 389–403 (2000)
21. Prince, J.L., Tan, Q., Pham, D.L.: Optimization of MR Pulse Sequences for Bayesian Image Segmentation. *Medical Physics* 22(10), 1651–1656 (1995)
22. Rohde, G.K., Aldroubi, A., Dawant, B.M.: The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Trans. on Med. Imag.* 22, 1470–1479 (2003)
23. Tibshirani, R.: Regression Shrinkage and Selection via the Lasso. *J. Royal Stat. Soc.* 58(1), 267–288 (1996)
24. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* 53, 4655–4666 (2007)
25. Weisenfeld, N.I., Warfield, S.K.: Normalization of Joint Image-Intensity Statistics in MRI Using the Kullback-Leibler Divergence. In: Intl. Symp. on Biomed. Imag (ISBI), vol. 1, pp. 101–104 (April 2004)
26. Yang, J., Wright, J., Huang, T., Ma, Y.: Image Super-Resolution Via Sparse Representation. *IEEE Trans. Image. Proc.* 19(11), 2861–2873 (2010)

Restoring DIC Microscopy Images from Multiple Shear Directions

Zhaozheng Yin¹, Dai Fei Elmer Ker², and Takeo Kanade¹

¹Robotics Institute, ²Department of Biological Sciences
Carnegie Mellon University, Pittsburgh, US

Abstract. Differential Interference Contrast (DIC) microscopy is a non-destructive imaging modality that has been widely used by biologists to capture microscopy images of live biological specimens. However, as a qualitative technique, DIC microscopy records specimen's physical properties in an indirect way by mapping the *gradient* of specimen's optical path length (OPL) into the image intensity. In this paper, we propose to restore DIC microscopy images by quantitatively estimating specimen's OPL from a collection of DIC images captured from multiple shear directions. We acquire the DIC images by rotating the specimen dish on the microscope stage and design an Iterative Closest Point algorithm to register the images. The shear directions of the image dataset are automatically estimated by our coarse-to-fine grid search algorithm. We develop a direct solver on a regularized quadratic cost function to restore DIC microscopy images. The restoration from multiple shear directions decreases the ambiguity among different individual restorations. The restored DIC images are directly proportional to specimen's physical measurements, which is very amenable for microscopy image analysis such as cell segmentation.

1 Introduction

Under a traditional brightfield microscope, living specimens such as cells are colorless and transparent because they are predominantly phase objects that absorb and scatter little illumination light. That is, cells do not significantly alter the amplitude of the light waves passing through them and as a result, produces little or no contrast when viewed under a brightfield microscope. For tissue culture cells, a cell's optical path length (OPL, product of its refractive index and geometric thickness) is normally different from that of the surrounding medium (about $0.125\mu\text{m}$ or a quarter wavelength of green light). This optical path difference induces a small phase difference between the light waves passing through cells and those traversing the surrounding medium. Since human eyes are sensitive to amplitude differences between light waves as opposed to phase differences, Differential Interference Contrast (DIC) microscopy technique was invented in 1950s to convert these minute phase variations to intensity changes that can be easily detected by human eyes (see textbook [12]).

The DIC microscope works by splitting a polarized illumination light wave into two component waves that are spatially displaced (sheared) along a specific

direction, and then recombining the two waves after they travel through adjacent locations on the specimen plate. The recombination (interference) is sensitive to phase variations between the two component waves. An adjustable bias (bias retardation) can be added into the phase variation. Because the phase variation between the two waves is caused by OPL difference at two adjacent locations, this microscopy imaging technique is then called “differential interference,” and the observed intensity in DIC images is proportional to the OPL *gradient* along the shear direction. The relief-like images generated by DIC microscopy have the pseudo 3D shadow-cast effects as if the specimens are illuminated from an oblique lighting resource (e.g. Fig. 1(a,b)), but this artifact only indicates the orientation of a specimen’s OPL gradient rather than the real topographical structure.

1.1 Related Work

Since the intensity of a DIC image is not a linear mapping of specimen’s inherent properties such as refractive index, thickness or OPL, this has triggered strong research interest in reconstructing the original physical properties of specimens from DIC images. We summarize the related work in three aspects:

(1) hardware-related techniques. Arnison et al. [1] proposed a hardware extension to the conventional differential interference by inserting an extra quarter wave plate in the optical layout of a DIC microscope, and restored the phase objects by varying bias setting and using geometric phase-shift techniques. Shribak et al. [15] developed an orientation-independent DIC microscopy by adding liquid crystal devices in the common DIC microscopes. The setup of these new optical configurations might be complicated and inaccessible to the common biology labs.

(2) reconstruction from a single DIC image. Noticing the gradient interpretation of DIC images, line integration methods were developed to reconstruct DIC images [8]. The line-by-line integration along shear direction introduces new streaking artifacts in reconstructed images and it is sensitive to gradient noise, thus Hilbert transform [2] and other ad hoc techniques such as low-pass filtering [7] were explored to reduce the streaking artifacts to a certain degree. General image processing algorithms such as deconvolution by Wiener filter [7,11] or by Landweber iterations [6] have been applied to reconstruct optical path length from DIC images. A preconditioning approach was recently proposed in [10] where the DIC image is reconstructed by minimizing a nonnegative mixed-norm constrained cost function. We reimplemented these three types of approaches and applied them on a pair of DIC images of the same specimens captured from two different shear directions. As shown in Fig. 1(c) and (d), we can observe the streak artifacts by line integration. Fig. 1(e) and (f) show the unsatisfactory restoration results by Wiener filtering with 1% noise-to-signal power ratio of the additive noise. The deconvolution performance depends on the prior knowledge of various hardware parameters (such as shear directions and bias setting) and image noise models. Fig. 1(g) and (h) show the reconstruction results by the

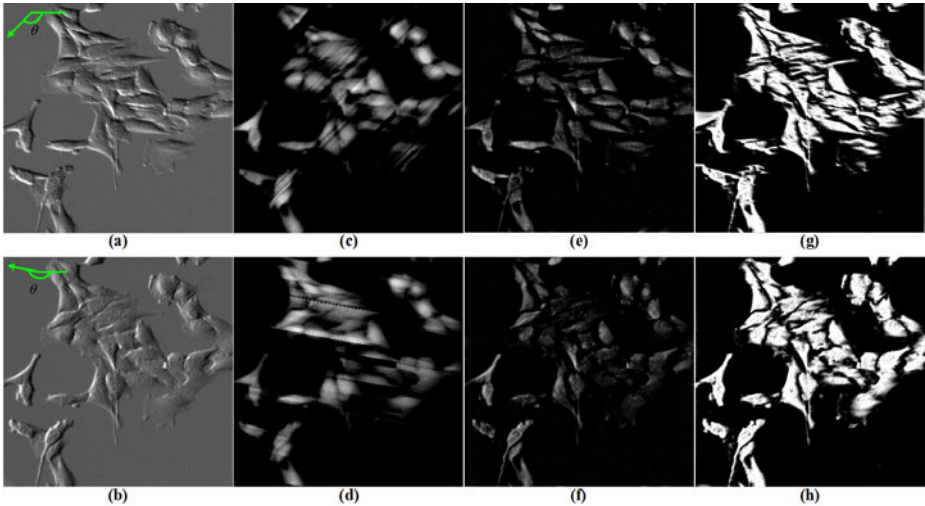


Fig. 1. Reconstructing optical path length from DIC images. (a,b) Two DIC image of the same specimens captured from two different shear directions (the arrow denotes the shear direction θ); (c,d) Reconstruction by line integration enhanced by low-pass filtering; (e,f) Reconstruction by deconvolution (Wiener filtering); (g,h) Reconstruction by preconditioning.

preconditioning method. It is time-consuming to estimate the direct measurement on specimens by the iterative preconditioning method.

From Fig. 1 we have a common observation that the reconstructions of the image pair (Fig. 1(c,d), Fig. 1(e,f), and Fig. 1(g,h)) are not the same for the same specimens. That is, when biologists analyze specimens, they will obtain different measurements on the specimen’s optical path length according to different shear directions. This is very undesirable because biologist don’t know which direction-specific reconstruction unveils the real properties of specimens.

(3) reconstruction using multiple DIC images. A few approaches have been proposed to restore specimen’s properties from multiple DIC images [6,9,13,15]. These approaches either rotate prisms, change bias settings or step the shear azimuth to capture multiple DIC images, and they require at least two images captured from a pair of orthogonal shear directions. Without specially-designed hardware, it is hard to rotate the specimen dish or prism manually by exact 90 degrees to satisfy the orthogonal requirement.

1.2 Our Proposal

We propose a novel approach to restore DIC microscopy images captured from multiple shear directions without the strict orthogonal requirement. In Section 2, we derived a closed-form solution for the restoration. Since the DIC images were captured by manually rotating the dish on the stage of a common DIC microscope, there are Euclidean transformation (rotation and translation) among

captured DIC images. We designed an Iterative Closest Point (ICP) algorithm to register the image dataset (Section 3). Rather than measuring the shear directions of the DIC images manually, we propose a coarse-to-fine grid search algorithm to find the shear directions automatically (Section 4). We show our experiment results in Section 5 with the conclusion followed in Section 6.

2 Problem Formulation and Restoration Method

Based on the gradient interpretation of DIC images, we have the following simplified DIC imaging model

$$\mathbf{g} = \nabla_{\theta} \mathbf{f} \quad (1)$$

where $\mathbf{g}(u, v)$ is an observed DIC image¹, ∇_{θ} is the gradient operator along the shear direction θ and $\mathbf{f}(u, v)$ is the DIC image to be restored. This imaging model is also used by some other DIC reconstruction methods such as the iterative preconditioning method [10] and deconvolution by Landweber iteration [6]. More accurate and complicated DIC imaging models can be referred to [14].

Based on Eq. 1, the DIC image can be restored by applying line-by-line integration on the observed gradient \mathbf{g} along the shear direction θ

$$\mathbf{f} = \int \mathbf{g} \, d\mathbf{x}^{\theta} \quad (2)$$

where \mathbf{x}^{θ} denotes a location on the lines along the shear direction. For a line on the specimen plate with M locations, we can observe $M - 1$ gradient values by DIC microscopy. The line integration method reverses the differential problem by estimating the OPL values at M locations from observed $M - 1$ gradient values, thus there are more unknowns than available equations. The under-constrained equation system plus the image acquisition noise may make the restored signal inconsistent when we perform line integration along different shear directions. As shown in Fig. 2(a), at a specimen location, three gradient signals are extracted from three DIC images along their shear directions. When we integrate the three gradient signals independently (Fig. 2(b)), the three reconstructed signals do not intersect in the OPL space - they have different restoration values on \mathbf{f} at the same specimen location! To avoid the ambiguity and achieve the consensus among different restorations, we propose to estimate the true \mathbf{f} in a least-square sense. In other words, the real \mathbf{f} at that location should have the minimum total distance to all integrated signal curves (Fig. 2(c)). Thus, we are looking for an image \mathbf{f} to minimize

$$\sum_{i=1}^K \int_{\mathcal{R}^2} (\mathbf{f} - \int \mathbf{g}_i \, d\mathbf{x}^{\theta_i})^2 \, d\mathbf{x} \quad (3)$$

where i indexes the K DIC images captured from different shear directions θ_i on the same specimens, $\mathbf{x} = (u, v)$ is a pixel location on the 2D Euclidean

¹ We drop the 2D location indices (u, v) in all the equations for concise expressions.

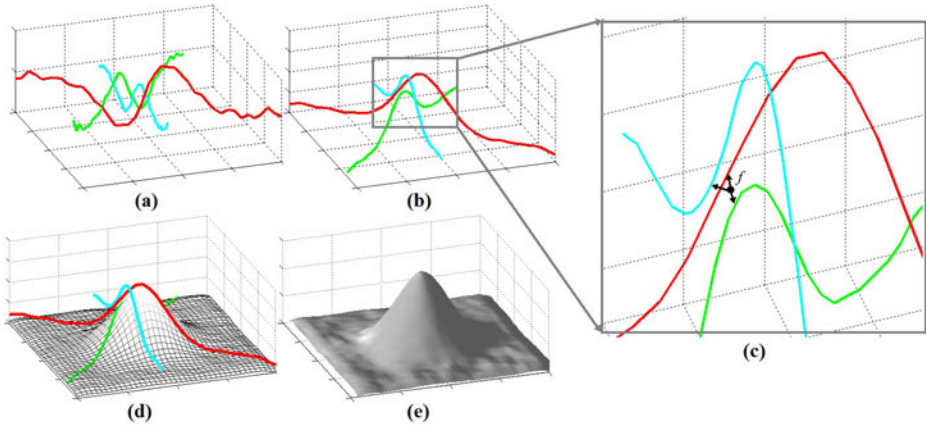


Fig. 2. Estimate the optical path length in DIC images. (a) Three observed gradient signals along different shear directions; (b) The three signals after integration do not intersect in the OPL space (i.e. there is no consensus among the restored signals); (c) We propose to restore DIC images (i.e. estimate the optical path length) by minimize the total distance to all integrated signals; (d) The spatial smooth constraint (gray mesh) is considered during the restoration; (e) The restoration result using three DIC images.

space \mathcal{R}^2 . The minimization on Eq. 3 needs to carry out the line integration $\int \mathbf{g}_i d\mathbf{x}^{\theta_i}$ explicitly for each DIC image. However, the line integration itself is not a satisfactory restoration as we see in Fig. 1(c) and (d). Instead, we propose to restore \mathbf{f} by minimizing a cost function in the gradient domain directly

$$\sum_{i=1}^K \int_{\mathcal{R}^2} (\nabla_{\theta_i} \mathbf{f} - \mathbf{g}_i)^2 d\mathbf{x}. \tag{4}$$

The goal is to compute an \mathbf{f} whose gradients along different shear directions are as close as possible to the corresponding given gradients, \mathbf{g}_i 's. After mapping the pixel location from 2D Euclidean space to a new surface defined by the K shear directions

$$\{\mathbf{x} = (u, v), \mathbf{x} \in \mathcal{R}^2\} \rightarrow \{\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K), \boldsymbol{\theta} \in \Theta\} \tag{5}$$

and using the commutativity of \sum and \int operations, the cost function (Eq. 4) is converted into

$$\int_{\Theta} \sum_{i=1}^K (\nabla_{\theta_i} \mathbf{f} - \mathbf{g}_i)^2 d\boldsymbol{\theta}. \tag{6}$$

Eq. 6 only measures the fidelity of the restoration to all the observed data. We enhances the data fidelity with smooth and sparse regularizations and propose the following objective function for restoration

$$O(\mathbf{f}) = \int_{\theta} \left[\left(\sum_{i=1}^K (\mathbf{d}_{\theta_i} * \mathbf{f} - \mathbf{g}_i)^2 \right) + \omega_s (\mathbf{a} * \mathbf{f})^2 + \omega_r \mathbf{f}^2 \right] d\theta \quad (7)$$

where \mathbf{d}_{θ} is a differential kernel along the shear direction θ , “*” is the convolution operation, $\mathbf{d}_{\theta_i} * \mathbf{f}$ is equivalent to $\nabla_{\theta_i} \mathbf{f}$, \mathbf{a} is a kernel for local smooth, ω_s and ω_r are weighting coefficients for the smooth and sparse regularizations, respectively. \mathbf{d}_{θ_i} can be defined by a directional first-derivative-of-Gaussian kernel [10]. The smooth constraint encourages nearby pixels to have the same restoration values (Fig. 2(d)). For example, we can regularize a restored pixel value to be close to the average of its neighboring pixels (i.e. $\mathbf{a} = [1 \ 1 \ 1; 1 \ -8 \ 1; 1 \ 1 \ 1]/8$ for 8-connected neighborhood). The l_2 sparse regularization penalizes large \mathbf{f} values and enforces the restored background pixels (with equal OPL at adjacent locations) to be close to zero. A stronger sparse regularization is using l_1 norm but there is no closed-form solution for that. More discussions on the regularizations can be referred to the rich research work on compressive sensing [3].

The solution that minimizes Eq. 7 must satisfy the Euler-Lagrange equation

$$\frac{\partial \mathbf{E}}{\partial \mathbf{f}} - \sum_{i=1}^K \frac{\partial}{\partial \theta_i} \frac{\partial \mathbf{E}}{\partial \mathbf{f}_{\theta_i}} = 0 \quad (8)$$

where \mathbf{f}_{θ_i} is a shorthand notation of $\mathbf{d}_{\theta_i} * \mathbf{f}$ and \mathbf{E} is the integrand inside Eq. 7

$$\mathbf{E} = \left(\sum_{i=1}^K (\mathbf{d}_{\theta_i} * \mathbf{f} - \mathbf{g}_i)^2 \right) + \omega_s (\mathbf{a} * \mathbf{f})^2 + \omega_r \mathbf{f}^2. \quad (9)$$

Substituting \mathbf{E} into Eq. 8 for the differentiating, we have

$$2\omega_s \mathbf{a} * \mathbf{a} * \mathbf{f} + 2\omega_r \mathbf{f} - 2 \sum_{i=1}^K \mathbf{d}_{\theta_i} * (\mathbf{d}_{\theta_i} * \mathbf{f} - \mathbf{g}_i) = 0. \quad (10)$$

Now, applying Fourier transform, \mathcal{F} , on both sides of this equation, we obtain

$$2\omega_s \mathbf{A}^2 \cdot \mathbf{F} + 2\omega_r \mathbf{F} - 2 \sum_{i=1}^K \mathbf{D}_{\theta_i}^2 \cdot \mathbf{F} + 2 \sum_{i=1}^K \mathbf{D}_{\theta_i} \cdot \mathbf{G}_i = 0 \quad (11)$$

where $\mathbf{A} = \mathcal{F}\{\mathbf{a}\}$, $\mathbf{F} = \mathcal{F}\{\mathbf{f}\}$, $\mathbf{D}_{\theta_i} = \mathcal{F}\{\mathbf{d}_{\theta_i}\}$, $\mathbf{G}_i = \mathcal{F}\{\mathbf{g}_i\}$, “ \cdot ” denotes the element-wise production and $\mathbf{D}_{\theta_i}^2 = \mathbf{D}_{\theta_i} \cdot \mathbf{D}_{\theta_i}$. Solving Eq. 11 for \mathbf{F} , we have

$$\mathbf{F} = - \left(\sum_{i=1}^K \mathbf{D}_{\theta_i} \cdot \mathbf{G}_i \right) ./ \left(\omega_s \mathbf{A}^2 + \omega_r - \sum_{i=1}^K \mathbf{D}_{\theta_i}^2 \right). \quad (12)$$

where “ $./$ ” denotes the element-wise division. \mathbf{f} is then restored by $\mathbf{f} = \mathcal{F}^{-1}\{\mathbf{F}\}$. Fig. 2(e) shows a restored result using three images with different shear directions.

For a single DIC image with shear direction θ , the direct solution is

$$\mathbf{F} = -(\mathbf{D}_{\theta} \cdot \mathbf{G}) ./ (\omega_s \mathbf{A}^2 + \omega_r - \mathbf{D}_{\theta}^2). \quad (13)$$

If without regularizations ($\omega_s = \omega_r = 0$), Eq. 13 is degraded into an inverse filtering

$$\mathbf{F} = \mathbf{G}./\mathbf{D}_\theta. \quad (14)$$

However, the simple inverse filtering can not restore a correct DIC image (Fig. 3(b)), which justifies the needs of regularization. As a comparison, our restoration by Eq. 13 with $\omega_s = 0.1$ and $\omega_r = 0.001$ is shown in Fig. 3(c) that is much better than the inverse filtering. Please note that the restoration from a single shear direction (Eq. 13) contains ambiguity to measure the real specimen property thus we have derived solution (Eq. 12) to restore DIC images from multiple shear directions. Eq. 13 is only used in Algorithm 2 (Section 4) for estimating the shear direction of each individual DIC image.

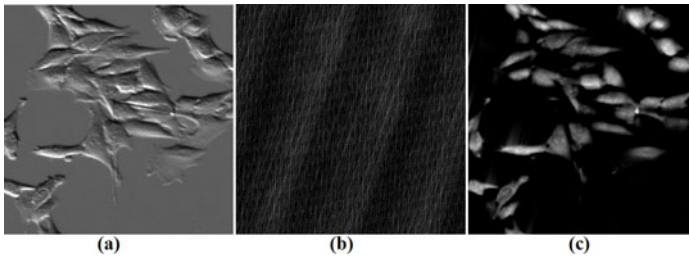


Fig. 3. Restoring a DIC image. (a) An input image; (b) Restoration by inverse filtering; (c) Our restoration by minimizing a regularized cost function.

3 Register a Collection of DIC Images

When capturing a collection of DIC images with different shear directions by rotating and translating dishes, we must register those images such that the same pixel location in the image dataset represents the same specimen sample in the world. Note: this registration step can be waived if there is a DIC microscope with rotatable prisms at hand. We revised the Iterative Closest Point (ICP) idea 4 to register two DIC images. Harris corner detector 5 and local non-max suppression are used to locate corners as feature points for matching in the ICP algorithm. The corners are tolerant to appearance changes in DIC images from different shear directions. However, the ICP algorithm can converge to the optimum only when the initialization (rotation \mathbf{R}_0 and translation \mathbf{T}_0) is close to the optimum. To find the correct \mathbf{R} and \mathbf{T} to register images, we uniformly sample the entire search space of all possible initializations (e.g. every 30 degrees of rotation and every 100 pixels of translation) and run the ICP algorithm from these initializations to find the global optimal \mathbf{R} and \mathbf{T} . The new designed ICP algorithm is summarized below. Fig. 4 shows an registration example using this ICP algorithm.

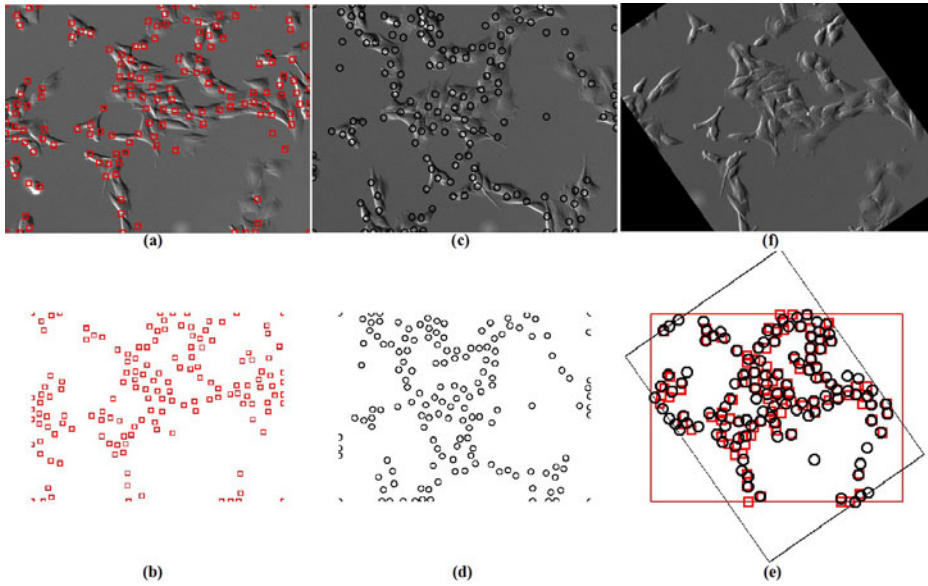


Fig. 4. Register two DIC images by Iterated Closest Point algorithm. (a,b) Image 1 and its corner points (red squares); (c,d) Image 2 and its corner points (black circles); (e) The two groups of corner points are matched with the least total distance cost; (f) The registered image 2 regarding to image 1 based on the Euclidean transformation computed from matched corner points.

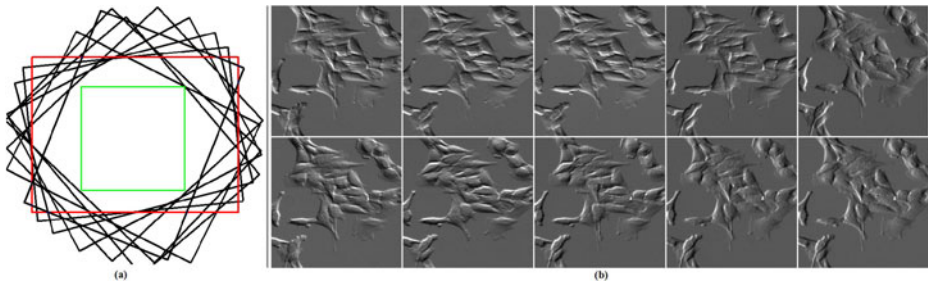


Fig. 5. Register a collection of images and extract the sub-images within the overlapped regions for our analysis. (a): Nine images are registered regarding to the first image. (b) The ten sub-images within the overlapped square regions.

For a collection of DIC images, we randomly pick a reference image and register all the others regarding to it. The registered images overlap in a polygon area and we crop the largest square sub-images from the overlapped region for our analysis (Fig. 5).

Algorithm I: ICP Algorithm to Register Two DIC Images

Extract two groups of corner points $\{\mathbf{Q}_j\}$ and $\{\mathbf{P}_i\}$ from image 1 and image 2, respectively. Compute the centroids: $\bar{\mathbf{P}} = \frac{1}{|\mathbf{P}|} \sum_i \mathbf{P}_i$, and $\bar{\mathbf{Q}} = \frac{1}{|\mathbf{Q}|} \sum_j \mathbf{Q}_j$.

Update $\mathbf{P}_i \leftarrow \mathbf{P}_i - \bar{\mathbf{P}}$, $\mathbf{Q}_j \leftarrow \mathbf{Q}_j - \bar{\mathbf{Q}}$.

Initialize $\mathbf{R} = \mathbf{R}_0$, $\mathbf{T} = \mathbf{T}_0$, and $c = 0$.

Repeat the following steps until there is no change on c .

1. Bi-directional Matching: $\forall \mathbf{P}_i$, find the closest \mathbf{Q}_j in the Euclidean space. For \mathbf{Q}_j , find the closest \mathbf{P}_k . If $i = k$, the two corner points are matched, and update $c += |\mathbf{P}_i - \mathbf{Q}_j|$.

2. Transformation: For all matched corner points, compute $\mathbf{W} = \sum_{\langle i,j \rangle} \mathbf{P}_i \mathbf{Q}_j^T$. Take the singular value decomposition (SVD) of matrix \mathbf{W} , $\mathbf{W} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$. Compute the rotation matrix as $\mathbf{R} = \mathbf{V} \mathbf{U}^T$, and the translation vector as $\mathbf{T} = \bar{\mathbf{Q}} - \mathbf{R} \bar{\mathbf{P}}$. $\forall \mathbf{P}_i$, update $\mathbf{P}_i \leftarrow \mathbf{R} \mathbf{P}_i + \mathbf{T}$.

Use the final matched points to compute \mathbf{R}^* and \mathbf{T}^* , rotate and translate image 2 regarding to image 1 accordingly.

4 Estimate the Shear Directions

After registration, the shear direction difference between the first image and the other $K - 1$ registered images are actually the rotation angle θ_r^i ($i = 1 \dots K$ and $\theta_r^1 = 0$). In other words, if the shear direction of the first image is θ_s , the shear directions of the rest images are $\theta_s - \theta_r^i$. Since θ_r^i 's are already known from the registration step, we only need to estimate a single unknown variable θ_s . As shown in Fig. 6(g), when correct shear directions are estimated for a pair of DIC images, the difference between the two restorations reaches the minimum. We use this fact to estimate the shear direction over the collection of registered images by a fast coarse-to-fine grid search algorithm.

Algorithm II: Coarse-to-Fine Grid Search for Shear Direction θ_s

Initialize $lb = 0$, $ub = 360$ and $\delta_\theta = 30$.

while $\delta_\theta > 1$

 for $\theta_s = lb; \theta_s < ub; \theta_s = \theta_s + \delta_\theta$

 for $i = 1; i < K; i ++$

 Solve \mathbf{f}_i with $\theta = \theta_s - \theta_r^i$ through Eq. 13;

$cost[\theta_s] = \sum_{i,j=1 \dots K, i \neq j} |\mathbf{f}_i - \mathbf{f}_j|$

$\theta_s^* \leftarrow \arg \min cost[\theta_s]; \delta_\theta \leftarrow \delta_\theta / 4; lb \leftarrow \theta_s^* - 2\delta_\theta; ub \leftarrow \theta_s^* + 2\delta_\theta;$

Return θ_s^* ;

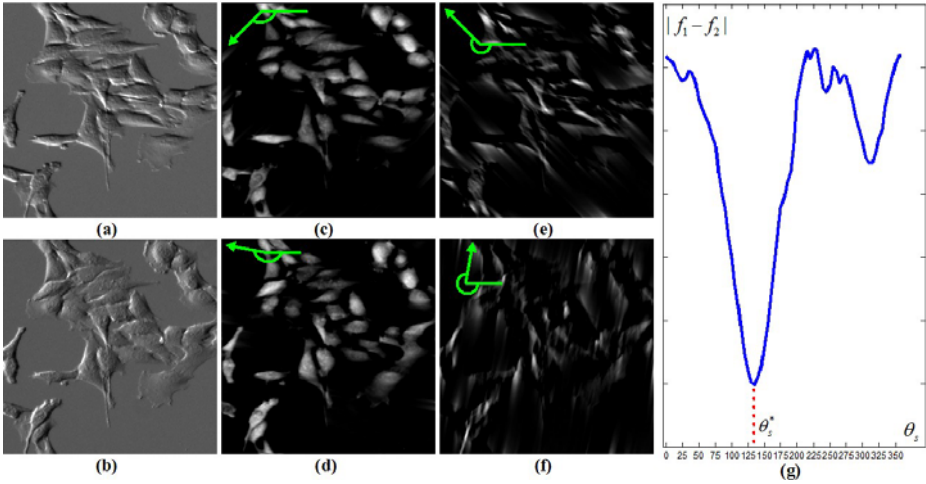


Fig. 6. Estimate the shear direction. (a,b) Image 1 and 2; (c,d) The restorations corresponding to the correct shear directions; (e,f): The restorations corresponding to wrong shear directions; (g) The difference between two restored images is a function of the shear direction.

5 Experiment Results

We captured DIC images from different shear directions by manually rotating the specimen dish on the stage. The collected images are registered automatically by our ICP algorithm and the shear directions of each image is estimated by our coarse-to-fine grid search algorithm.

First, we qualitatively evaluate our restoration results. Fig. 7(a) shows an image (700*700 pixels) on two cells with high magnification. When we apply the direct solver on the captured images individually, we observe different restoration results (Fig. 7(c-h)). The ensemble restoration on the entire collection of images reveals the cells' optical path length much better and has less noise left on the restoration (Fig. 7(b)). Fig. 8 show the other three collections of DIC images we acquired with low magnification and their image restorations. Compared to the independent restorations (e.g. the last two columns of Figure 8), the jointly restored image by an ensemble of DIC images are closer to the physical properties of cells. There is less ambiguity to compute the optical path length in joint restoration than the rotation-variant independent restoration.

To quantitatively evaluate the effect of our restoration results on microscopy image analysis, we apply it onto the cell segmentation task. As we see in the third column of Fig. 8, since the pixel values in the restored images represent the specimen's inherent properties, there is no pseudo 3D shadow-cast effect in the images. In the restored images, cell pixels are always positive against approximately zero background, thus images can be easily segmented by straightforward thresholding for further applications of cell counting, tracking etc., which is not

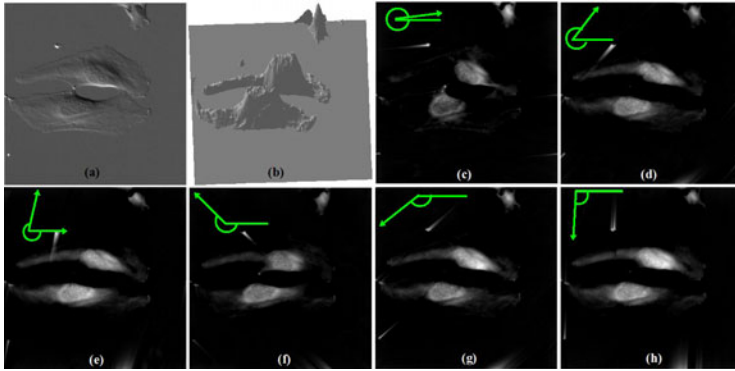


Fig. 7. Reconstructing DIC images. (a) One DIC image of the image collection; (b) The reconstructed OPL from a collection of DIC images (3D surface view); (c-i) The independent restoration from individual DIC images (green arrows denote the shear directions).

easily possible with original DIC images. We manually label all cell boundaries in each image collection as ground truth masks. After thresholding the restored DIC image to obtain the cell mask, we compare it with the manually-labelled ground truth mask using the accuracy as the evaluation measure

$$\text{ACC} = (|\text{TP}| + |\text{N}| - |\text{FP}|) / (|\text{P}| + |\text{N}|) \quad (15)$$

where cell and background pixels are denoted as positive (P) and negative (N) respectively, true positive (TP) stands for those cell pixels correctly labelled by both human annotator and computer algorithms, and false positive (FP) are those cell pixels classified by computer algorithms mistakenly. We use the same technique in [16] to learn the best threshold for segmentation.

With K DIC images, we pick k ($k = 1 \dots K$) of them as a collection and run our restoration algorithm on this collection. For example, when $k = 1$ we run the restoration algorithm on each single DIC image. When $k = 5$, we pick 5 images from the K DIC images and run the restoration. We exhaustively tested all the possible combinatorial choices, and computed the mean and standard deviation on the accuracy regarding to different image collection sizes. As we see in Fig. 9, there are more ambiguities (the vertical red bar in Fig. 9) among different restoration results when the image collection size k is small. As more images are added into the image collection, the restoration accuracy increases and the ambiguity decreases. The accuracy curve levels off when enough DIC images from different shear directions are included into the image collection. Overall, we achieve the segmentation accuracy of 81% – 95% on the three collections of DIC images. Difficulties were encountered during the restoration of predominantly flat cells that consequently had low gradient values in the observed DIC images. Less sparsity regularization on these regions may overcome the challenge. We leave the spatially-adaptive regularization in the future work to explore.

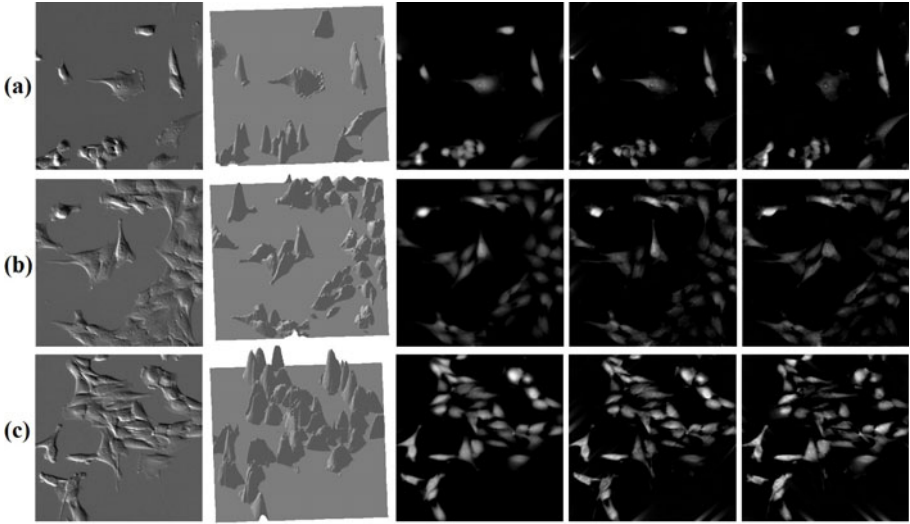


Fig. 8. Restoring DIC images. Column 1: one sample image in each collection; Column 2: the jointly reconstructed OPL (3D surface view); Column 3: The reconstructed DIC images whose intensity values correspond to the reconstructed OPL values; Column 4 and 5: Two independent restorations of DIC images with different shear directions, as comparisons to the ensemble restoration.

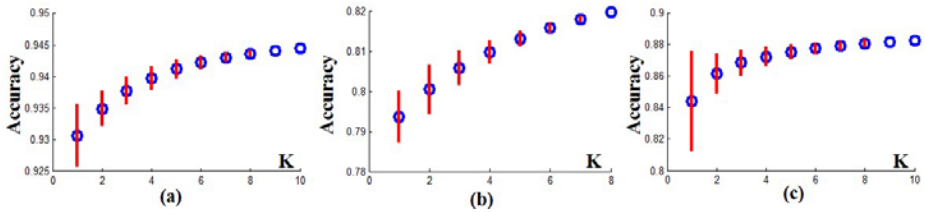


Fig. 9. The mean (blue circle) and stand deviation (red bar) of the segmentation accuracy when using different image collection size k . (a,b,c) are evaluation results from three image datasets corresponding to Figure 8(a,b,c), respectively.

6 Conclusions

In DIC microscopy images, the intensity values are proportional to the gradient of specimen's optical path length (OPL). To quantitatively measure specimen's physical properties directly, we propose to restore DIC images from multiple shear directions. The specimen dish is manually rotated to acquire a collection of DIC images with different shear directions. An Iterative Closest Point algorithm is designed to register these images, and the shear directions of the image dataset are automatically estimated by our coarse-to-fine grid search algorithm.

We formulate the restoration problem by minimizing a regularized cost function with a closed-form solution.

Compared to the reconstruction methods based on a single DIC image which may reconstruct different optical path lengths according to different shear directions, our method is orientation-invariant. Without the strict requirement of at least two orthogonal shear directions as needed by the previous multiple-image-based reconstruction techniques, our approach can restore DIC images from various shear directions. As qualitatively and quantitatively proved by our experiments, restoration from multiple shear directions can decrease the ambiguity among different individual restorations. The restored DIC images are directly proportional to specimen's physical measurements without the pseudo 3D effect, which is very amenable for microscopy image analysis such as cell segmentation.

Acknowledgements

We would like to thank Dr. Frederick Lanni for useful discussions and assistance with DIC microscopy. This work was supported by NIH grants RO1EB004343 & RO1EB007369.

References

1. Arnison, M., Larkin, K.G., Sheppard, C.J.R., Smith, N.I., Cogswell, C.J.: Linear phase imaging using differential interference contrast microscopy. *J. Microsc.* 214, 7–12 (2003)
2. Arnison, M.R., Cogswell, C.J., Smith, N.I., Fekete, P.W., Larkin, K.G.: Using the hilbert transform for 3d visualization of differential interference contrast microscope images. *J. Microsc.* 199(1), 79–84 (2000)
3. Baraniuk, R.: Compressive sensing. *IEEE Sig. Proc. Mag.* 24(4), 118–121 (2007)
4. Besel, P., Mckay, N.: A method for registration of 3-d shapes. *IEEE Trans. on Pattern Recognition and Machine Intelligence* 14(2), 239–256 (1992)
5. Harris, C., Stephens, M.: A combined corner and edge detector. In: *The 4th Alvey Vision Conference* (1988)
6. Heise, B., Arminger, B.: Some aspects about quantitative reconstruction for differential interference contrast (dic) microscopy. In: *ICIAM 2007* (2007)
7. Heise, B., Snnleitner, A., Klement, E.: Dic image reconstruction on large cell scans. *Microscopy Research and Techniques* 66, 312–320 (2005)
8. Kam, Z.: Microscopic differential interference contrast image processing by line integration (lid) and deconvolution. *Bioimaging* 6, 166–176 (1998)
9. King, S., Libertun, A., Piestun, R., Cogswell, C.: Quantitative phase microscopy through differential interference imaging. *J. BioMed. Opt.* 13, 24020 (2008)
10. Li, K., Kanade, T.: Nonnegative mixed-norm preconditioning for microscopy image segmentation. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) *IPMI 2009*. LNCS, vol. 5636, pp. 362–373. Springer, Heidelberg (2009)
11. Munster, E., Vliet, L., Aten, J.: Reconstruction of optical path length distributions from images obtained by a wide-field differential interference contrast microscope. *J. Microsc.* 188, 149–157 (1997)

12. Murphy, D.: *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley, Chichester (2001)
13. Preza, C.: Rotational-diversity phase estimation from differential-interference-contrast microscopy images. *J. of the Opt. Soc. of America A* 17(3), 415–424 (2000)
14. Preza, C., Snyder, D., Conchello, J.A.: Theoretical development and experimental evaluation of imaging models for differential interference contrast microscopy. *J. Opt. Soc. Am. A* 16(9), 2185–2199 (1999)
15. Shribak, M., LaFountain, J., Biggs, D., Inoue, S.: Orientation-independent differential interference contrast (dic) microscopy and its combination with orientation-independent polarization system. *J. Biomed. Opt.* 13(1), 14011 (2008)
16. Yin, Z., Li, K., Kanade, T., Chen, M.: Understanding the optics to aid microscopy image segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 209–217. Springer, Heidelberg (2010)

Identifying Nuclear Phenotypes Using Semi-supervised Metric Learning

Shantanu Singh¹, Firdaus Janoos¹, Thierry Pécot¹, Enrico Caserta²,
Gustavo Leone², Jens Rittscher³, and Raghu Machiraju¹

¹ Dept. of Computer Science and Engg., The Ohio State University, U.S.A.

² Dept. of Molecular Genetics, The Ohio State University, U.S.A.

³ General Electric Global Research Center, U.S.A.

Abstract. In systems-based approaches for studying processes such as cancer and development, identifying and characterizing individual cells within a tissue is the first step towards understanding the large-scale effects that emerge from the interactions between cells. To this end, nuclear morphology is an important phenotype to characterize the physiological and differentiated state of a cell. This study focuses on using nuclear morphology to identify cellular phenotypes in thick tissue sections imaged using 3D fluorescence microscopy. The limited label information, heterogeneous feature set describing a nucleus, and existence of sub-populations within cell-types makes this a difficult learning problem. To address these issues, a technique is presented to learn a distance metric from labeled data which is *locally adaptive* to account for heterogeneity in the data. Additionally, a label propagation technique is used to improve the quality of the learned metric by expanding the training set using unlabeled data. Results are presented on images of tumor stroma in breast cancer, where the framework is used to identify fibroblasts, macrophages and endothelial cells – three major stromal cells involved in carcinogenesis.

Keywords: Nuclear Morphometry, Metric Learning, Microscopy.

1 Introduction

Recent years have seen an increasing role of quantitative imaging in systems biology research [12]. With the rapid advances made in fluorescence microscopy [10], it is now possible to image tissue sections within the 3D context of the biological microenvironment. The vast array of fluorescent proteins with distinct spectral properties and their ability to label different cell-types using transgenic animal models have opened up a whole new arena [14], where complex phenomena such as the interaction between cells in their tissue microenvironment can be studied within a high-content, 3D spatial context. A crucial first step in these studies is identifying and characterizing the cells within the tissue. This paper focuses, in specific, on the characterization of *cell nuclei*. The morphology of the cell nucleus is frequently used as a proxy for characterizing the physiological and differentiated state of a cell. In cancer, modifications in nuclear structure have been

shown to correlate with different types and stages of cancer [23]. The present work addresses the analysis of cell nuclei in the *stromal regions*¹ of the tumor microenvironment in breast cancer. This is motivated by findings [18,1] that have revealed the surprising role of stromal cells in the initiation and progression of cancer. This paper is aimed at the automatic identification of fibroblasts, macrophages and endothelial cells based on nuclear morphology and spatial positioning in 3D fluorescence microscopy images of thick tissue sections, and the characterization of the diversity of these cell-types by identifying cellular sub-populations that may exist. In these images, the ground truth information about cell-type is available through the expression of fluorescent proteins in transgenic mice. There are several challenges posed by this problem. *First*, quantitative representation of the cell nucleus requires a heterogeneous set of features corresponding to shape, texture and spatial properties. There exists no methodical way of comparing two nuclei given these features. *Second*, the number of labeled examples that can be collected in these experiments is restricted due to various factors (a) only one cell-type specific fluorescent protein is expressed per tissue section; (b) the signal quality of the transgenic fluorescence is often degraded due to background fluorescence, thus requiring an expert to validate the labeled data for each experiment; (c) stromal cells are sparsely distributed in the tissue resulting in few cells being visible in a given field of view.

To address these problems, first, an approach is proposed to learn an appropriate distance metric to compare nuclei. *Metric learning* techniques use prior information, typically the label information, to learn an optimal metric from the data for a specific task. Most techniques learn a global metric through a constrained optimization problem. In this paper, a method is proposed for learning *local metrics* in order to deal with the heterogeneity that exists within cell-types in the microenvironment. The method constructs a locally adaptive metric by extending an existing global technique by using a hierarchical Bayesian model. Next, to address the problem of small size of the training set, unlabeled data is used to enhance the quality of this metric by using label propagation on a graph constructed using the learned metric. The rest of the paper is organized as follows. Related work in cell analysis literature and metric learning is reviewed in Section 2. The metric learning framework is presented in Section 3. Results on 3D fluorescence microscopy images of nuclei in the stromal regions of mammary tissue are presented in Section 4. We conclude with a discussion in Section 5.

2 Related Work

Nuclear Analysis. A method to compare nuclear morphologies was presented in [13] that uses a combination of large deformation metric mapping and multi-dimensional scaling to derive a metric in an unsupervised manner. A method to perform spherical mapping of the nuclear volume is presented in [7] based on a mechanical model of deformation to produce a normalized representation of the

¹ The stroma forms the supporting structure in mammary tissue. Fibroblasts, macrophages and endothelial cells are the main constituents of breast tumor stroma.

nucleus, thereby enabling the comparison of sub-cellular structures between two cells. An expansive set of features to characterize subcellular structures in a cell is proposed in [2]; this set is frequently used in image-based cellular analysis to quantify cellular phenotypes in fluorescence microscopy images. In cell analysis literature to date, there has been no work in the direction of *learning* an optimal metric from data to compare nuclei. To the best of our knowledge, this is the first formal analysis of this problem.

Metric learning. Among the earliest pieces of work on this topic was presented by Xing et. al. [21], where a metric is learned using prior knowledge about the similarity or dissimilarity of a set of points. Neighborhood component analysis [8] was used to learn a metric optimized for nearest-neighbor based classification. An approach to estimate a Mahalanobis metric using an efficient optimization methodology has been proposed in [5]. These approaches infer a *global* metric from the data. Among local techniques, an approach to learn a point-wise metric was proposed in [20], where the method requires solving expensive optimization problems for each data point. In this paper, a local metric learning approach is proposed that extends the method presented in [5] by casting the problem in a hierarchical Bayesian framework.

3 Methods

In this section, a *hierarchical Bayesian approach* is proposed for learning a locally adaptive distance metric. Next, a graph is constructed based on this metric to improve the quality of the learned metric by expanding the size of the training set through *label propagation* [22]. Finally, the nuclear features used to represent shape, appearance and spatial neighborhood are described.

The Global Metric Model

Consider the set of nuclei in an image $\mathcal{X} \triangleq (\mathcal{X}_l, \mathcal{X}_u) \triangleq \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l, \mathbf{x}_{l+1}, \dots, \mathbf{x}_n\}$ where \mathcal{X}_l and \mathcal{X}_u corresponds to the labeled and unlabeled set respectively, and \mathbf{x}_i is the representation of the nucleus in feature-space of dimension M . For each $\mathbf{x}_i \in \mathcal{X}_l$ there exists a label $y_i \in \{1, \dots, c\}$, where c is the number of cell-types that have fluorescent markers identifying them. The sets $\mathcal{S} = \{(i, j) | y_i = y_j\}$ and $\mathcal{D} = \{(i, j) | y_i \neq y_j\}$ correspond to pairs of points that have the same or different label respectively.

The distance between two nuclei i, j is given by the metric $d_{\mathbf{A}_0}(\Delta \mathbf{x}_{i,j}) \triangleq \Delta \mathbf{x}_{i,j}^\top \mathbf{A}_0 \Delta \mathbf{x}_{i,j}$ with \mathbf{A}_0 the metric tensor and $\Delta \mathbf{x}_{i,j} = \mathbf{x}_i - \mathbf{x}_j$. This global metric tensor is assumed to be distributed according to a Wishart distribution [6] $\mathcal{W}(\mathbf{A})$ on the space of positive semi-definite (PSD) matrices, defined as:

$$p(\mathbf{A}_0 | \mathbf{A}) \propto |\mathbf{A}^{-1} \mathbf{A}_0|^{M/2} \exp \left\{ -\frac{1}{2} \text{Tr}(\mathbf{A}^{-1} \mathbf{A}_0) \right\}.$$

The PSD matrix \mathbf{A} expresses the prior belief about the metric tensor \mathbf{A}_0 and is typically initialized to the $M \times M$ identity matrix implying an equal and independent contribution of all the features in measuring the distance. In order

that the metric on this data keep the pairs in \mathcal{S} close and the points in \mathcal{D} apart, the data likelihood is given by:

$$p(\Delta \mathbf{x}_{i,j} | y_i = y_j, A_0) \propto \exp \left\{ -\frac{\lambda}{2} |d_{A_0}(\Delta \mathbf{x}_{i,j}) - u|_+ \right\}$$

and

$$p(\Delta \mathbf{x}_{i,j} | y_i \neq y_j, A_0) \propto \exp \left\{ -\frac{\eta}{2} |l - d_{A_0}(\Delta \mathbf{x}_{i,j})|_+ \right\},$$

where $u \geq 0, l \gg u, |c|_+ = \min(0, c)$ and $\lambda \geq 0, \eta \geq 0$ are non-zero constants. This form of the likelihood function requires that similar points be less than a distance u of each other and dissimilar point be a large distance l apart. Maximizing the posterior distribution $p(A_0 | \mathbf{x}_i, \mathbf{x}_j, y_i, y_j)$ with respect to A_0 results in the following constrained optimization problem [5]:

$$\begin{aligned} & \min_{A_0} \text{LogDiv}(A, A_0) + \lambda \sum_{i,j \in \mathcal{S}} \zeta_{i,j} + \eta \sum_{i,j \in \mathcal{D}} \vartheta_{i,j} \\ & \text{subject to } d_{A_0}(\Delta \mathbf{x}_{i,j}) \leq u + \zeta_{i,j} \quad (i, j) \in \mathcal{S}, \\ & \quad \quad \quad d_{A_0}(\Delta \mathbf{x}_{i,j}) \geq l - \vartheta_{i,j} \quad (i, j) \in \mathcal{D}, \\ & \quad \quad \quad A_0 \succeq 0, \quad \zeta_{i,j} \geq 0 \quad \text{and} \quad \vartheta_{i,j} \geq 0. \end{aligned}$$

where $\text{LogDiv}(A, A_0) \triangleq \text{Tr}(A^{-1}A_0) - \log |A^{-1}A_0| - d$ is the log-determinant divergence. This optimization problem can be efficiently computed using the technique of Bregman projections [9], involving rank-one updates of A_0 by projecting onto each of the constraints iteratively. The Bregman updates preserve the PSD property of A_0 in each iteration, thereby guaranteeing that the solution is a metric without explicitly handling the PSD condition.

The Local Metric Model

As shown in Section 4.1 a single global metric tensor fails to capture the heterogeneity of the differences between nuclei, suggesting the concept of an ensemble $\mathcal{A} = \{A_k, k = 1 \dots K\}$ of K local metrics that vary according to nucleus type. Each local metric is characterized by the metric tensor A_k , which is again assumed to have a Wishart prior $\mathcal{W}(A_0)$. Also, for each pair of cell nuclei i, j we associate a hidden variable $z_{i,j} \in \{1, \dots, K\}$ indicating the metric active on that pair. Specifically, if $z_{i,j} = k$, then $d_{A_k}(\Delta \mathbf{x}_{i,j}) \triangleq (\Delta \mathbf{x}_{i,j})^T_{A_k} (\Delta \mathbf{x}_{i,j})$. The marginal probability of a particular nucleus \mathbf{x}_i belonging to the local support of a metric A_k is modeled by a Gaussian distribution $p(\mathbf{x}_i | z_{i,j} = k, \mu_k, \Sigma_k) = \mathcal{N}(\mu_k, \Sigma_k)$ where $\mu \triangleq \mu_1 \dots \mu_k, \Sigma \triangleq \Sigma_1 \dots \Sigma_K$. Let $p(z_{i,j} = k) \triangleq \pi_k$ represent the marginal probability of a pair of nuclei belonging to the support of A_k , such that $0 \leq \pi_k \leq 1, \forall k = 1 \dots K$ and $\sum_{k=1}^K \pi_k = 1$. Therefore, $\theta \triangleq \{\mu, \Sigma, \pi\}$ are the parameters of the Gaussian mixture model (GMM) of the support of the local metric ensemble \mathcal{A} . As before, each local metric tensor A_k is adapted such that for all pairs of cell nuclei $(\mathbf{x}_i, \mathbf{x}_j)$ with labels y_i, y_j which belong to its local support, it tries to keep similar nuclei less than distance $u > 0$ of each other and dissimilar ones at a large distance $l \gg u$ apart as per:

$$p(\Delta \mathbf{x}_{i,j} | \mathbf{x}_i, \mathbf{x}_j, y_i = y_j, z_{i,j} = k, \mathbf{A}_k) \propto \exp \left\{ -\frac{\lambda}{2} |d_{\mathbf{A}_k}(\Delta \mathbf{x}_{i,j}) - u|_+ \right\}$$

and $p(\Delta \mathbf{x}_{i,j} | \mathbf{x}_i, \mathbf{x}_j, y_i \neq y_j, z_{i,j} = k, \mathbf{A}_k) \propto \exp \left\{ -\frac{\eta}{2} |l - d_{\mathbf{A}_k}(\Delta \mathbf{x}_{i,j})|_+ \right\}.$

Therefore for any pair of data-points, the probability is given by :

$$p(\Delta \mathbf{x}_{i,j}, \mathbf{x}_i, \mathbf{x}_j, z_{i,j}, \mathcal{A} | y_i, y_j, \theta, \mathbf{A}_0) = \left[\prod_{k=1}^K p(\mathbf{A}_k | \mathbf{A}_0) \right] p(z_{i,j} | \pi)$$

$$p(\mathbf{x}_i | z_{i,j}, \mu, \Sigma) p(\mathbf{x}_j | z_{i,j}, \mu, \Sigma) p(\Delta \mathbf{x}_{i,j} | \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, z_{i,j}, \mathcal{A}).$$

Assuming pair-wise independence of the cell nuclei, the complete data log probability is

$$\ln p(\Delta \mathcal{X}, \mathcal{X}, \mathcal{Z}, \mathcal{A} | \mathcal{Y}, \theta, \mathbf{A}_0) = \sum_{k=1}^K [\ln p(\mathbf{A}_k | \mathbf{A}_0) + \ln \pi_k]$$

$$+ \sum_{i=1}^N \left[\ln p(\mathbf{x}_i | z_i, z_j, \mu, \Sigma) + \sum_{j>i}^N p(\Delta \mathbf{x}_{i,j} | \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, z_i, z_j, \mathcal{A}) \right],$$

where $\Delta \mathcal{X} \triangleq \{\Delta \mathbf{x}_{i,j}, \forall i = 1 \dots N - 1, i < j \leq N\}$, $\mathcal{Y} \triangleq \{y_i, \forall i = 1 \dots N\}$ and, $\mathcal{Z} \triangleq \{z_{i,j}, \forall i = 1 \dots N - 1, i < j \leq N\}.$

Local Metric Estimation

The optimal estimate of the GMM parameters θ^* and metric tensor ensemble \mathcal{A}^* is obtained using expectation maximization (EM) which involves iterating the following two steps until convergence:

E-step: $\mathcal{Q}((\mathcal{A}, \theta), (\mathcal{A}, \theta)^n) = \mathbb{E}_{p(\mathcal{Z} | \Delta \mathcal{X}, \mathcal{X}, \mathcal{Y}, \mathcal{A}^n, \theta^n, \mathbf{A}_0)} [\ln p(\Delta \mathcal{X}, \mathcal{X}, \mathcal{Z}, \mathcal{A} | \mathcal{Y}, \theta, \mathbf{A}_0)]$

M-step: $\arg \max_{\mathcal{A}, \theta} \mathcal{Q}((\mathcal{A}, \theta), (\mathcal{A}, \theta)^n),$

where $p(\mathcal{Z} | \Delta \mathcal{X}, \mathcal{X}, \mathcal{Y}, \mathcal{A}^n, \theta^n, \mathbf{A}_0) = \prod_{i=1}^{N-1} \prod_{j>i}^N p(z_{i,j} | \Delta \mathbf{x}_{i,j}, \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, \mathcal{A}^n, \theta^n, \mathbf{A}_0).$ We define the responsibility term $\gamma_{i,j}(k) \triangleq p(z_{i,j} = k | \Delta \mathbf{x}_{i,j}, \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, \mathcal{A}^n, \theta^n, \mathbf{A}_0),$ which is evaluated as:

$$\gamma_{i,j}(k) = \frac{p(\Delta \mathbf{x}_{i,j}, \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, z_{i,j} = k, \mathcal{A}^n | \theta^n, \mathbf{A}_0)}{\sum_{z_{i,j}=1}^K p(\Delta \mathbf{x}_{i,j}, \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, z_{i,j}, \mathcal{A}^n | \theta^n, \mathbf{A}_0)}.$$

Denoting the term $N_k \triangleq \sum_{i=1}^{N-1} \sum_{j>i}^N \gamma_{i,j}(k),$ the M-Step for the GMM parameters results in the following closed-form updates:

$$\pi_k^{(n+1)} = \frac{2N_k}{N(N-1)}, \quad \mu_k^{(n+1)} = \frac{1}{N_k} \sum_{i=1}^{N-1} \sum_{j>i}^N \gamma_{i,j}(k) \mathbf{x}_i,$$

and $\Sigma_k^{(n+1)} = \frac{1}{N_k} \sum_{i=1}^{N-1} \sum_{j>i}^N \gamma_{i,j}(k') (\mathbf{x}_i - \mu_k^{(n+1)}) (\mathbf{x}_i - \mu_k^{(n+1)})^\top.$

The M-Step for the metric tensor A_k is

$$A_k^{(n+1)} = \arg \max_{A_k} \sum_{i=1}^{N-1} \sum_{j>i}^N \gamma_{i,j}(k) \ln p(\Delta \mathbf{x}_{i,j} | \mathbf{x}_i, \mathbf{x}_j, y_i, y_j, z_{i,j} = k, A_k) + \ln p(A_k | A_0),$$

which is again equivalent to:

$$\begin{aligned} & \min_{A_k} \text{LogDiv}(A_0, A_k) + \lambda \sum_{i,j \in \mathcal{S}} \gamma_{i,j}(k) \zeta_{i,j} + \eta \sum_{i,j \in \mathcal{D}} \gamma_{i,j}(k) \vartheta_{i,j} \\ & \text{subject to } d_{A_0}(\Delta \mathbf{x}_{i,j}) \leq u + \zeta_{i,j} \quad (i, j) \in \mathcal{S}, \\ & \quad d_{A_0}(\Delta \mathbf{x}_{i,j}) \geq l - \vartheta_{i,j} \quad (i, j) \in \mathcal{D}, \\ & \quad A_0 \succeq 0, \quad \zeta_{i,j} \geq 0 \quad \text{and} \quad \vartheta_{i,j} \geq 0. \end{aligned}$$

Inducing a Global Metric from the Local Metric Ensemble

Finally, under this local metric model, the global distance between any two cell nuclei i, j in feature space is then defined as :

$$d_{\mathcal{A}}(\Delta \mathbf{x}_{i,j}) \triangleq \sum_{k=1}^K \gamma_{i,j}(k) \Delta \mathbf{x}_{i,j}^{\top} A_k \Delta \mathbf{x}_{i,j}.$$

In general, the construction of this metric does not ensure that the triangular inequality holds. In order to construct a globally consistent metric on the data, the distance between two points is defined as the geodesic distance on the graph where the edge weights are given by $d_{\mathcal{A}}(\Delta \mathbf{x}_{i,j})$. The geodesic distance is obtained using Floyd's algorithm [4] to compute the shortest paths between all pairs of vertices on this graph.

Expanding the training set using label propagation

The training data used in learning metric described above was restricted to the labeled set \mathcal{X}_l . Since the quantity of labeled data available per experiment is limited, an approach is formulated to increase the training set by incrementally adding new points from the unlabeled data for which the labels can be reliably estimated. The label propagation technique [22] provides a method to learn a *label function* using both labeled and unlabeled points. This method is used to add new points to the training set, re-estimate the local metric based on this expanded set, and repeat this process until no new points can be added. The details of this process are described as follows. To begin, a weighted graph \mathcal{G} is constructed on the points in \mathcal{X} represented by the $n \times n$ weight matrix $W_{ij} \triangleq \exp(-d_{\mathcal{A}}(\Delta \mathbf{x}_{i,j})/\sigma^2)$, where kernel bandwidth σ is a tunable parameter. A *label function* $f : \mathcal{X} \rightarrow \{0, 1\}^c$ maps data points to their labels (using a *1-of-c* representation for labels). The *prediction matrix* $F \in \mathbb{R}^{n \times c}$ represents the values of the label function on \mathcal{X} . Define a split $F = (F_u F_l)$ where the two matrices correspond to the unlabeled and labeled points respectively. The label propagation estimates F_u given W and F_l by minimizing the quadratic energy function $E(F) = \sum_{i,j} W_{ij} \|F(\mathbf{x}_i) - F(\mathbf{x}_j)\|_2^2$ keeping F_l fixed. The minimization is obtained in a closed form solution $F_u = (D_{uu} - W_{uu})^{-1} W_{ul} F_l$ where D is the diagonal matrix $D_{ii} = \sum_{j=1}^n W_{ij}$. Here, the matrix W is split into four sub-matrices defined as $W = (W_{uu} W_{ul}; W_{lu} W_{uu})$ where W_{uu} corresponds to pairs of

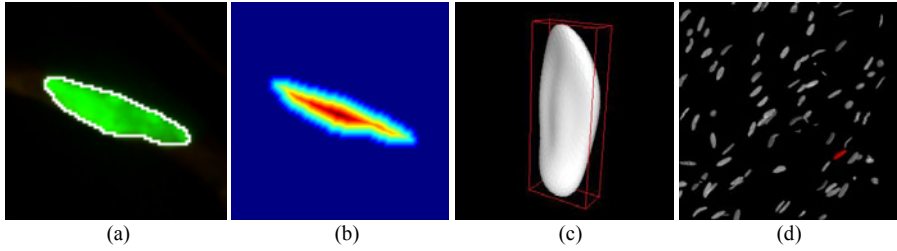


Fig. 1. Different views of a cell nucleus. (a) Maximum intensity projection of cell nucleus with the automatic segmentation outlined. (b) Distance transform of nuclear mask to generate radial profile. (c) Spherical harmonics reconstruction of the nuclear surface. (d) Iso-surface of the tissue image with the selected nucleus in red.

points from the product set $\mathcal{X}_u \times \mathcal{X}_u$. W_{ul} , W_{lu} and W_{ll} are constructed similarly. The matrix D is split similar to W . The points for which the labels are predicted with high confidence in F_u (as measured through label entropy) are added to the training set. A new metric is now learned using this updated set and the process is repeated until no new points can be added.

3.1 Nuclear Features

In the present experiments, three sets of nuclear features were used corresponding to shape, appearance and spatial neighborhood as described below. Note that the proposed metric learning approach is independent of the choice of these features – they can be chosen to suit the specific phenotyping questions pertinent to the study.

Spherical harmonic based shape features. The shape of a nucleus is modeled using a spherical harmonic representation of its surface [3]. The surface of the object is mapped on to a sphere through a transformation that minimizes area distortion. The mapping results in a representation of the original surface through a set of three coordinate functions $(x(\phi, \theta), y(\phi, \theta), z(\phi, \theta))$ on the unit sphere. Each function maps a point on the sphere (given by ϕ and θ) to the corresponding coordinate on the surface. A representation of these coordinate functions is obtained in terms of their projection onto spherical harmonic basis functions. The function $\mathbf{x}(\phi, \theta)$ is given by $x(\phi, \theta) = \sum_{i=0}^{\infty} \sum_{m=-l}^l c_{i,m}^x Y_{l,m}(\phi, \theta)$ where $Y_{l,m}(\phi, \theta)$ are the spherical harmonic basis functions of degree l and order m [3]. Similarly for $\mathbf{y}(\phi, \theta)$ and $\mathbf{z}(\phi, \theta)$. Fig. 1(c) shows the spherical harmonics reconstruction of a nucleus using $l = 5$.

Radial profile. The appearance of the nuclear volume is modeled using the radial distribution of the density of the DNA stain as a function of its distance from the nuclear surface. This feature can be efficiently computed using the distance transform on the foreground mask of the nucleus (Fig. 1(b)). The distances

are scaled to lie in the range $[0, 1]$, binned at intervals of fixed length and the average DNA intensity for each bin is computed to give the radial profile.

Spatial distribution. The spatial characteristic of a nucleus is represented by the density of cells in the local neighborhood of the cell. The local density is computed by identifying nuclear centers in the image and computing the kernel density estimate of the distribution of points in the 3D volume. The size of the kernel bandwidth σ_d is set as a certain factor of the average nuclear diameter.

4 Experimental Results

The methods were applied to a study of the tumor microenvironment in murine breast cancer [18] that seeks to understand the role of stromal cells in the initiation and progression of cancer. Previous studies using knockout models have shown that fibroblasts, the primary constituents of the extra-cellular matrix, play a role in tumor suppression in breast cancer. Other cells in the microenvironment – most importantly the macrophages and endothelial cells – are also known to participate in the cancer process. In these experiments, we focus on identifying the different cell-types in the microenvironment of a normal mouse mammary gland based on nuclear morphology and spatial positioning within the tissue, and further, on identifying sub-populations that may exist within these cell-types.

Data collection. Tissue sections were collected from two-month wild-type mice. The cell-types are identified using fluorescent proteins that are endogenously expressed in the transgenic mouse lines. In a given specimen in our experiments, endogenous fluorescence is present in only one of the three stromal cells (macrophages, fibroblasts or endothelial cells). To identify cell nuclei, the tissue sections were stained with DAPI, a fluorescent marker for DNA. An Olympus FV1000 confocal microscope with an objective of $40x/1.3NA$ was used to collect images. The images were acquired at an in-plane resolution $0.31\mu m$ and axial resolution of $0.5\mu m$, and have field of view $317\mu m \times 317\mu m$ and depth of $40-70\mu m$.

Image Processing. The confocal images consist of two channels corresponding to the nuclear stain (represented in green) and the cell-specific fluorescent protein (represented in red). The channels were denoised and the nuclear channel was segmented using a standard processing pipeline [15]. Segmentation errors were manually corrected in most cases, while a small fraction of poor segmented nuclei were discarded. The fluorescence protein channel was then used to detect nuclei corresponding to the cell-type identified by the protein. The set of nuclei selected through this process were manually validated. These operations were executed on images for each cell-type and constituted the labeled examples for the study. The nuclei that were not identifiable by the fluorescent protein were used as unlabeled examples.

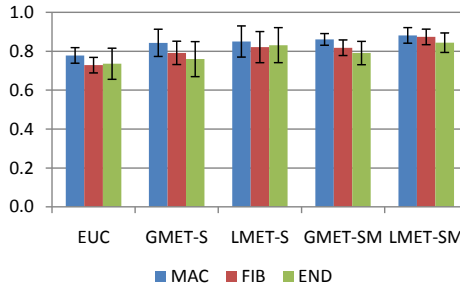


Fig. 2. Classification rates for the three cell-types

Feature Extraction. The spherical harmonics coefficients were computed based on the SPHARM-PDM method [17] using a degree of $l = 5$ resulting in a 108-dimensional shape feature. The radial profile was computed by discretizing into 10 bins, thereby resulting in a 10-dimensional appearance feature. The spatial density of nuclei was computed by first identifying nuclear centers from the segmentations and then computing the kernel density estimate using a Gaussian kernel with bandwidth of $20\mu m$. This results in a scalar-valued density feature. The dimensionality of the shape and appearance features were reduced independently using PCA, maintaining 95% of the data variance. Metric learning was performed on the resulting in a set of features.

4.1 Cell-Type Classification

Results were generated on a dataset of 984 nuclei, of which 229 were labeled, consisting of 102 macrophages, 95 fibroblasts and 32 endothelial cells. In the following, the semi-supervised extension refers to the use of the label propagation technique described in Section 3. Results on cell-type classification were obtained using k -nearest neighbors classifier. Classification rates were compared for five different metrics – the standard Euclidean metric (EUC), the global metric [5] with and without semi-supervised extension (GMET-SM and GMET-S respectively) and the proposed local metric with and without semi-supervised extension (LMET-SM and LMET-S respectively). An entropy threshold of 0.01 was used for the semi-supervised extensions in all cases. In LMET-SM and LMET-S, the number of GMMs was set to 3, equal to the number of classes in the data set. The comparative results are summarized in Fig. 2. Classification rates were averaged over 5 runs.

The rates for both GMET-S and LMET-S show an average improvement over EUC. However the error variance in both cases was high because of the small size of the training set. The variance decreases significantly in GMET-SM and LMET-SM with the use of unlabeled data that provides a more robust estimate of the metrics. In all cases, the classification accuracy for macrophages was higher than for the other two cell-types. This is explained by the fact that in many cases fibroblasts and endothelials have similar morphologies, making the task

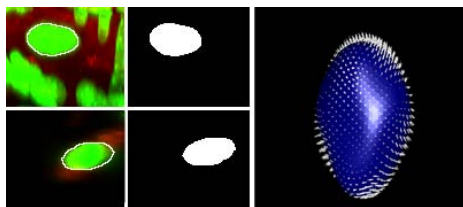


Fig. 3. Spatial neighborhood used to discriminate between cell-types in the absence of a significant shape differences. (Top) endothelial cell; (Bottom) fibroblast; (Right) visualization of shape difference. Surface of endothelial nucleus is in blue, arrows indicate the difference between nuclear shapes.

of discriminating between them difficult. These error rates decrease in LMET-SM, due to the local metric being able to learn the appropriate weighting of the spatial feature as shown in Fig. 3. In this figure, while the nuclei come from two different cell-types, their morphologies are very similar. This is seen in the 2D segmentation masks as well as in the 3D visualization of the difference between the shapes.

However, endothelial cells lie in closely packed regions surrounding blood vessels while fibroblasts are typically scattered in the stroma as seen in the left panel of Fig. 3. Thus the spatial neighborhood is a good feature to discriminate between these two nuclei. The learned metric was able to accurately identify this difference by appropriately weighting the spatial feature to discriminate between these samples.

4.2 Identifying Cellular Sub-populations

There is significant genetic and physiological heterogeneity observed within cell-types [11]. The problem of identifying patterns of cellular heterogeneity has been studied in *cell culture* to understand the effects of perturbations in a system [16]. In the following, we attempt to identify cellular subtypes *within the tissue sections* by using the locally adaptive metrics to induce a clustering on the set of nuclei. For each cell-type, a weighted graph was constructed from the data, with the vertices as nuclei and edge weights given by the distance under the learned metric. By employing a graph clustering method [19], subgroups within each cell-type were identified by manually choosing a scale parameter that resulted in meaningful clusters based on visual examination. The subgroups that were discovered through this process are shown in Fig. 4. For fibroblasts, it was observed that the three nuclear subgroups vary primarily in the degree of flatness of the nuclear shape. The macrophage subgroups were observed to differ in size as well as in the internal appearance. For example, the MAC1 group consists of nuclei that have low intensity regions in the interior, most likely due to the macrophages undergoing phagocytosis [11] (engulfing a foreign substance). Nuclei in groups MAC1 and MAC3 were observed to be similar in shape and size, both being significantly smaller than MAC2. The differences

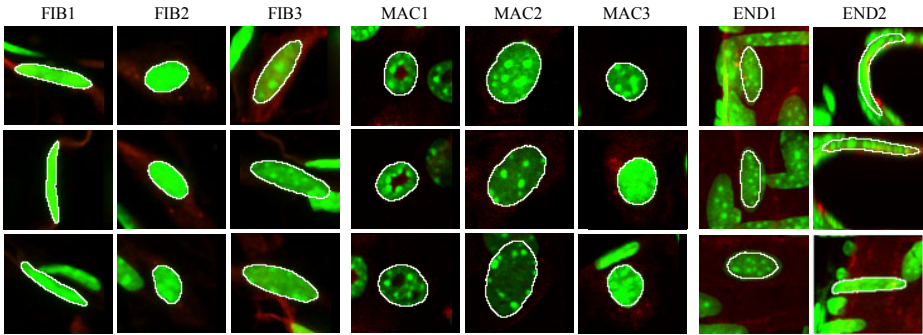


Fig. 4. Nuclear subtypes within fibroblasts (FIB 1-3), macrophages (MAC 1-3) and endothelial cells (END 1-3). Rows correspond to different examples of the nuclear subtypes. Nuclei are colored in green, the cell-specific fluorescence in red.

between the groups END1 and END2 appear to be that of elongatedness and curvedness of their shapes. The most likely explanation for this difference is the heterogeneity of cell-types that line blood vessels – endothelial cells form the inner lining of blood vessels, and a related cell-type, *pericytes*, wrap around the vessels forming an outer sheath. Since these cells are in very close proximity it is difficult to distinguish between them using fluorescent markers. The cluster analysis however is able to distinguish between these two types based on their nuclear morphology. This result is an indicator that morphological analysis of nuclei can be used to supplement traditional staining techniques or endogenous fluorescent proteins in distinguishing between cell-types.

5 Conclusions

In this paper, an approach was presented for analyzing nuclear phenotypes to identify cell-types in 3D images of tissue sections. This problem poses several challenges – (a) a diverse set of features is required to describe the nuclear phenotype (b) training data available per experiment is limited, and (c) the existence of cellular subpopulations results in phenotypic heterogeneity within cell-types. An approach of constructing a *locally adaptive* distance metric learning framework was adopted to address this problem, which provided a principled way of comparing a pair of nuclei by learning the structure of the metric locally in a supervised framework. In addition, a label propagation scheme was employed that incrementally expands the training set by selecting previously unlabeled points which can be confidently labeled, thereby circumventing the constraint of limited training examples. In addition to demonstrating results on the problem of cell-type identification, the learned metric was also used to identify cellular *subtypes* using a clustering strategy. Through the latter, we demonstrated the existence of phenotypic heterogeneity of nuclei within cell-types.

Acknowledgments. This work was supported by National Institutes of Health Grants UL1RR025755 and 5 P01 CA 097189-05.

References

1. Bissell, M.J., Radisky, D.: Putting tumours in context. *Nature Reviews. Cancer* 1(1), 46–54 (2001)
2. Boland, M.V., Markey, M.K., Murphy, R.F.: Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry* 33(3), 366–375 (1998)
3. Brechbühler, C., Gerig, G., Kibler, O.: Parametrization of closed surfaces for 3-d shape description. *Computer Vision and Image Understanding* 61(2), 154–170 (1995)
4. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: *Introduction to Algorithms*. The MIT Press, New York (2001)
5. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: *Proceedings of the 24th International Conference on Machine Learning, ICML 2007*, pp. 209–216 (2007)
6. Eaton, M.L.: *Multivariate statistics: a vector space approach*. Lecture notes-monograph series, vol. 53. Institute of Mathematical Statistics (1983)
7. Gladilin, E., Goetze, S., Mateos-Langerak, J., Van Driel, R., Eils, R., Rohr, K.: Shape normalization of 3D cell nuclei using elastic spherical mapping. *Journal of Microscopy* 231(Pt 1), 105–114 (2008)
8. Goldberger, J., Roweis, S., Hinton, G.: Neighbourhood components analysis. In: *Advances in Neural Information Processing Systems* (2004)
9. Kulis, B., Sustik, M., Dhillon, I.: Learning low-rank kernel matrices. In: *Proceedings of the 23rd International Conference on Machine Learning*, pp. 505–512. ACM, New York (2006)
10. Lichtman, J.W.: Fluorescence microscopy. *Nature Methods* 2(12) (2005)
11. Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., Matsudaira, P.: *Molecular Cell Biology*, 6th edn. W.H. Freeman, New York (2007)
12. Megason, S.G., Fraser, S.E.: Imaging in Systems biology. *Cell* 130(5), 784–795 (2007)
13. Rohde, G.K., Ribeiro, A.J.S., Dahl, K.N., Murphy, R.F.: Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells. *Cytometry. Part A: the Journal of the International Society for Analytical Cytology* 73(4), 341–350 (2008)
14. Shaner, N.C., Steinbach, P.A., Tsien, R.Y.: A guide to choosing fluorescent proteins. *Nature Methods* 2(12), 905–909 (2005)
15. Singh, S., Raman, S., Caserta, E., Leone, G., Ostrowski, M., Rittscher, J., Machiraju, R.: Analysis of Spatial Variation of Nuclear Morphology in Tissue Microenvironments. In: *2010 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, Los Alamitos* (2010)
16. Slack, M.D., Martinez, E.D., Wu, L.F., Altschuler, S.J.: Characterizing heterogeneous cellular responses to perturbations. *Proceedings of the National Academy of Sciences of the United States of America* 105(49), 19306–19311 (2008)
17. Styner, M., Oguz, I., Xu, S., Levitt, J.J., Shenton, M.E., Gerig, G.: Framework for the Statistical Shape Analysis of Brain Structures using SPHARM-PDM. *Insight Journal*, 1–20 (2006)

18. Trimboli, A.J., Cantemir-Stone, C.Z., Li, F., Wallace, J.A., Merchant, A., Creasap, N., Thompson, J.C., Caserta, E., Wang, H., Chong, J.L., Naidu, S., Wei, G., Sharma, S.M., Stephens, J.A., Fernandez, S.A., Gurcan, M.N., Weinstein, M.B., Barsky, S.H., Yee, L., Rosol, T.J., Stromberg, P.C., Robinson, M.L., Pepin, F., Hallett, M., Park, M., Ostrowski, M.C., Leone, G.: Pten in stromal fibroblasts suppresses mammary epithelial tumours. *Nature* 461(7267), 1084–1091 (2009)
19. Van Dongen, S.: A cluster algorithm for graphs (2000)
20. Weinberger, K.Q., Saul, L.K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification. *Journal of Machine Learning Research* 10, 207–244 (2009)
21. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning, with application to clustering with side-information. In: *Advances in Neural Information Processing Systems* (2003)
22. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using gaussian fields and harmonic functions. In: *ICML*, vol. 20, p. 912 (2003)
23. Zink, D., Fischer, A.H., Nickerson, J.A.: Nuclear structure in cancer cells. *Nature reviews. Cancer* 4(9), 677–687 (2004)

Actin Filament Segmentation Using Dynamic Programming

Hongsheng Li, Tian Shen, and Xiaolei Huang

Department of Computer Science & Engineering, Lehigh University, USA

Abstract. We introduce a novel algorithm for actin filament segmentation in 2D TIRFM image sequences. This problem is difficult because actin filaments dynamically change shapes during their growth, and the TIRFM images are usually noisy. We ask a user to specify the two tips of a filament of interest in the first frame. We then model the segmentation problem in an image sequence as a temporal chain, where its states are tip locations; given candidate tip locations, actin filaments' body points are inferred by a dynamic programming method, which adaptively generates candidate solutions. Combining candidate tip locations and their inferred body points, the temporal chain model is efficiently optimized using another dynamic programming method. Evaluation on noisy TIRFM image sequences demonstrates the accuracy and robustness of this approach.

1 Introduction

Actin is one of the most abundant proteins in cells. It has the ability to polymerize into long filaments out of monomers (single proteins). These filaments form a fibrous network within cells called the “cytoskeleton”, which provides cells with mechanical integrity and shape. Through the fast assembly and disassembly of the cytoskeleton fibers, cells can change shape, move and divide. One way to study the cytoskeleton is to analyze actin filaments' polymerization *in vitro* using Total Internal Reflection Fluorescence Microscopy (TIRFM) [5], [6] (see Fig. 1(a-b)). Biologists are interested in measuring actin filaments' elongation rates (by measuring their length differences between frames) and rigidity (by measuring the curvature along them). Therefore, actin filaments need to be segmented before the analysis.

There has been related work on tracking or segmentation of filamentous and microtubule (MT) structures in biological images. Hadjidemetriou *et al.* [7] developed a method to automatically track microtubules which are characterized locally using consecutive level sets segments. In [12], after MT tips are detected in the first frame, they are tracked by looking for the closest ones in the subsequent frames. Active contours are then used to extract MT bodies based on obtained tip locations. In [13], MT tips are located using second order derivative of Gaussian filtering. Geodesic paths are then iteratively calculated to segment MT bodies. In [9], Stretching Open Active Contour (SOAC) is used to segment and track actin filaments in TIRFM images frame by frame. In the above methods, movements of the structures are assumed small, which is not always valid in our TIRFM images.

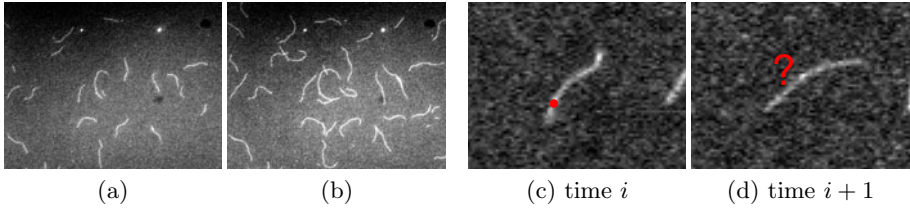


Fig. 1. (a-b) Two frames in a TIRFM sequence. (c-d) It is difficult to judge which points at time $i + 1$ should be the temporal correspondences of the red point at time i .

More complex movement models can be taken into consideration in Bayesian tracking frameworks. In [8] and [14], Particle Filters (PF) are used to track the locations of tip parts of MT. The MT bodies can then be segmented in a subsequent step using obtained tip locations [8]. Li *et al.* [10] combined the SOAC and PF to simplify the modeling of PF in a one-dimensional state space while integrating filament body constraints into tip estimation. However, biological sequences are usually taken prior to data analysis and therefore do not need real time performance. The above methods only consider temporal information up to the current frame and thus ignore all available information after it. In [11], a 2D time-lapse image sequence is treated as a 3D image volume, and a spatiotemporal active surface model is proposed to segment an actin filament in all frames simultaneously. The assumption of this method is that an actin filament's body remains static across time. However, this assumption does not apply to general actin filament data, where an actin filament may grow, drift, and change shape at the same time (see Fig. 1).

In this paper, we propose a novel method to segment actin filaments in the general data while fully utilizing all available temporal information. We treat the segmentation of an actin filament across time as a 1D temporal chain model. Tip locations are used as the state of each temporal node because tips are usually blurry in the images, and their localization is most difficult. The temporal chain can be globally optimized using dynamic programming. Filament bodies are inferred in the same process using another dynamic programming method, which adaptively generates candidate solutions. Dynamic programming has been used to optimize tree models in [3] or chain models in [1]. Compared with existing methods [10], [11], which have some strong assumptions and constraints, our proposed method is more flexible. Every constraint in the framework can be freely modified or deleted, and new constraints can be easily added in.

2 Methodology

2.1 Problem Formulation

The problem of actin filament segmentation in a single image can be viewed as optimizing a 1D spatial chain model as shown in Fig. 2(a), where each node's state represents the position of a point on the actin filament [9]. Compared with

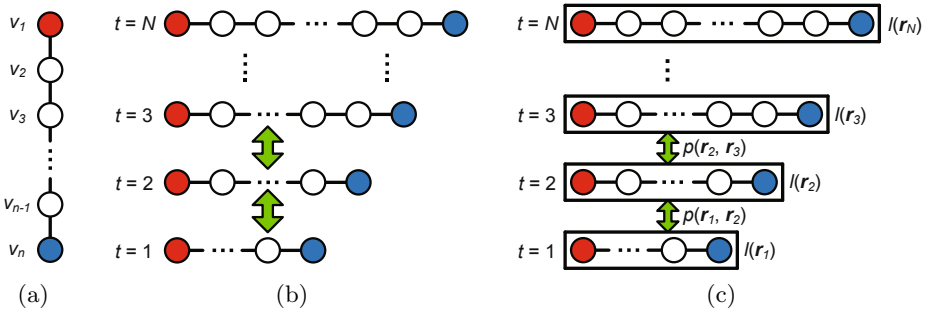


Fig. 2. (a) Illustration of the 1D spatial chain in a single frame. (b) Illustration of a series of spatial chains. (c) Illustration of the temporal chain model. Tip nodes of spatial chains are marked in red and blue. Spatial relations are shown as solid lines. Temporal relations are shown as green arrows.

other pixel-wise segmentation methods, one major advantage of the 1D chain model is that it has clear topology. Its results are ready for further biological analysis, while other methods need post-processing to reconstruct the topology.

If segmenting an actin filament in a time-lapse sequence, temporal information would be helpful, for example, to recover information on an intermediate frame that is unclear. The problem of actin filament segmentation in a time-lapse sequence can then be defined as optimizing a series of 1D spatial chain models while considering their temporal relations between each other. Fig. 2(b) illustrates the segmentation model for a time-lapse sequence. However, formulating this model is challenging on three aspects. (i) The number of nodes in each 1D chain is different. To represent different actin filaments with the same resolution, the number of points on a chain should be proportional to its length. (ii) Each node's temporal correspondences are difficult to define because an actin filament may move its entire body during its growing process (see Fig. 1(c-d)). (iii) The desired energy functions measuring how well a 1D spatial chain represents an actin filament (likelihood terms) or how well two chains are temporally constrained (pairwise terms) may have very complex forms.

To address the difficulties, we partially bypass the second difficulty because there is no convincing way to establish temporal correspondences between nodes at two chains (see Fig. 1(c-d)). By focusing on addressing the first and third difficulties, we are able to propose a simplified model aiming at jointly recovering the best tip points in all time frames while considering inferred filament body points' temporal relations. The entire segmentation problem is modeled as a 1D temporal chain as shown in Fig. 2(c), where the i th node's state represents the starting (red) and ending (blue) points of the actin filament in the i th frame. Jointly using image information in all time frames and temporal information between every two frames, we locate the best tip points, $\{\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \dots, \hat{\mathbf{r}}_N\}$, in the image domain Ω by minimizing the following energy function:

$$f(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \left(\sum_{i=2}^N l(\mathbf{r}_i) + \sum_{i=1}^{N-1} p(\mathbf{r}_i, \mathbf{r}_{i+1}) \right), \quad (1)$$

where N is the number of frames in the input sequence, $\mathbf{r}_i = [r_{i,1}, r_{i,2}]^T \in \Omega \times \Omega$ is the state of the i th node (*i.e.*, the locations of the two tip points, $r_{i,1}$ and $r_{i,2}$, in the i th frame). For instance, if $r_{i,1}$ has two candidate locations $\{v_1, v_2\}$, and $r_{i,2}$ has two candidate locations $\{v_3, v_4\}$, \mathbf{r}_i 's candidate states would then include $[v_1, v_3]^T$, $[v_1, v_4]^T$, $[v_2, v_3]^T$, and $[v_2, v_4]^T$. $l(\mathbf{r}_i)$ is the likelihood term measuring how well the 1D spatial chain inferred by the two tip points, \mathbf{r}_i , represents an actin filament, and $p(\mathbf{r}_i, \mathbf{r}_{i+1})$ is the pairwise term measuring how well the i th and $i+1$ th spatial chains inferred by the tip points, \mathbf{r}_i and \mathbf{r}_{i+1} , follow the given temporal constraints. The objective function does not calculate the likelihood cost in the 1st frame, $l(\mathbf{r}_1)$, because the user was asked to specify the correct tip locations of the actin filament of interest in the 1st frame. The choices of $l(\cdot)$ and $p(\cdot)$ are very flexible. We introduce our implementation in Section 3. However, they can be freely modified and deleted to fit any new constraints.

2.2 Efficient Optimization of the Temporal Chain Using Dynamic Programming

Taking advantages of the 1D structure of the temporal chain, the objective function (1) can be efficiently optimized using dynamic programming if the solution set is discrete. We restrict tip points to only be on image pixel locations. The dynamic programming algorithm involves generating a sequence of ‘‘optimal-value’’ functions of one discrete variable, $\{s_i\}_{i=1}^{N-1}$. To obtain each s_i , a minimization is performed over a single variable. (1) can then be iteratively optimized as follows:

$$s_1(\mathbf{r}_2) = p(\mathbf{r}_1, \mathbf{r}_2) + l(\mathbf{r}_2), \quad (2)$$

$$s_2(\mathbf{r}_3) = \min_{\mathbf{r}_2} [s_1(\mathbf{r}_2) + p(\mathbf{r}_2, \mathbf{r}_3)] + l(\mathbf{r}_3), \quad (3)$$

⋮

$$s_{N-1}(\mathbf{r}_N) = \min_{\mathbf{r}_{N-1}} [s_{N-2}(\mathbf{r}_{N-1}) + p(\mathbf{r}_{N-1}, \mathbf{r}_N)] + l(\mathbf{r}_N), \quad (4)$$

$$\min_{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N} f(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \min_{\mathbf{r}_N} s_{N-1}(\mathbf{r}_N). \quad (5)$$

In the 1st iteration, for every possible \mathbf{r}_2 , the algorithm calculates $p(\mathbf{r}_1, \mathbf{r}_2) + l(\mathbf{r}_2)$ and assign this value to the function $s_1(\mathbf{r}_2)$. No minimization is performed in this iteration since \mathbf{r}_1 is fixed. In the 2nd iteration, for every possible \mathbf{r}_3 , the algorithm searches for a \mathbf{r}_2 that minimizes $s_1(\mathbf{r}_2) + p(\mathbf{r}_2, \mathbf{r}_3)$. Then all such $(\mathbf{r}_3, \mathbf{r}_2)$ pairs are stored in a table, and the function values of (3) are assigned to $s_2(\mathbf{r}_3)$; The remaining iterations are performed similarly. As shown in (5), the minimum of $f(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ can be obtained as the minimal element in the table $s_{N-1}(\mathbf{r}_N)$. The optimal tip locations in all frames are then recovered by back-tracking the minimization-correspondence tables from the last one to the first.

If there are m possible pairs of starting and ending points in each frame, the overall time complexity of the above dynamic programming algorithm is $O(Nm^2)$ (see Section 3.5 for actual computation time).

2.3 Inference of Actin Filament Body Points

The calculation of both the likelihood terms, $\{l(\mathbf{r}_i)\}_{i=2}^N$, and the pairwise terms, $\{p(\mathbf{r}_i, \mathbf{r}_{i+1})\}_{i=1}^{N-1}$, requires body points to be efficiently and accurately inferred by given tip points $\{\mathbf{r}_i\}_{i=1}^N$, *i.e.*, optimizing the 1D spatial chain model in Fig. 2(a). The problem can be viewed as a shortest path problem with given starting and ending points. We propose a novel algorithm to efficiently solve for the shortest path, $\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n\}$, by minimizing an active contour model energy function:

$$E(v_1, v_2, \dots, v_n) = \alpha \sum_{j=1}^{n-1} |v_{j+1} - v_j|^2 + \beta \sum_{j=2}^{n-1} |v_{j+1} - 2v_j + v_{j-1}|^2 + \sum_{j=1}^n E_{ext}(v_j), \quad (6)$$

where $v_j \in \Omega$ denotes the location of the j th point's on the path, n is the number of points on the path, α and β are the weights of the first-order and second-order smoothness terms. The first two internal energy terms keep the path smooth. The last external energy term, $E_{ext}(\cdot)$, makes the path fit the bright ridges in the image. The smaller $E_{ext}(v_j)$ is, the better v_j represents a point on bright ridges. If n is given and $\{v_j\}_{j=1}^n$ are constrained to only be on pixel locations, (6) can be efficiently minimized by dynamic programming [1]:

$$\begin{aligned} e_1(v_2, v_3) &= \alpha |v_2 - v_1|^2 + \beta |v_3 - 2v_2 + v_1|^2 + E_{ext}(v_2), \\ e_2(v_3, v_4) &= \min_{v_2} [e_1(v_2, v_3) + \alpha |v_3 - v_2|^2 + \beta |v_4 - 2v_3 + v_2|^2] + E_{ext}(v_3), \\ &\vdots \\ e_{n-2}(v_{n-1}, v_n) &= \min_{v_{n-2}} [e_{n-3}(v_{n-2}, v_{n-1}) + \alpha |v_{n-1} - v_{n-2}|^2 + \beta |v_n - 2v_{n-1} + v_{n-2}|^2] \\ &\quad + E_{ext}(v_{n-1}), \\ e_{n-1}(v_n) &= \min_{v_{n-1}} [e_{n-2}(v_{n-1}, v_n) + \alpha |v_n - v_{n-1}|^2], \\ \min_{v_1, v_2, \dots, v_n} E(v_1, v_2, \dots, v_n) &= e_{n-1}(v_n), \end{aligned} \quad (7)$$

The above equations have similar meanings as those of (2)–(5) except that the optimal-value functions $\{e_j\}_{j=1}^{n-1}$ now have two discrete variables. This is because the calculation of second-order smoothness terms involves three neighboring spatial nodes. Minimizations in the first and last iterations are not needed because the starting and ending points are given. The time complexity of this algorithm is $O(nk^4)$ if every point can take k different locations. We further improve the above algorithm by proposing a scheme to adaptively generate candidate locations, and by decreasing minimization-search range for each candidate point.

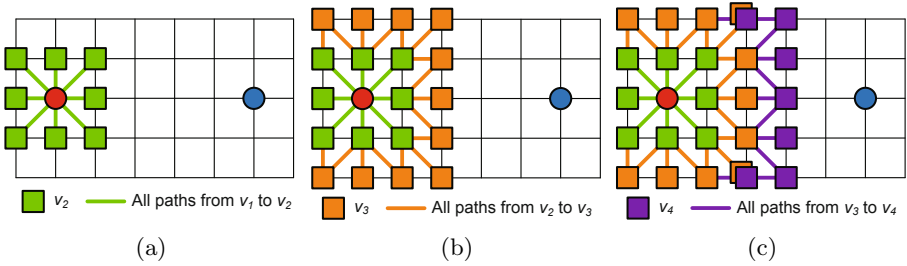


Fig. 3. (a) Generating candidate locations of v_2 . (b) Generating candidate locations of v_3 . (c) Generating candidate locations of v_4 . Red and blue dots represent the starting point (v_1) and the ending point (v_n) of a path. Colored lines represent possible paths.

In the first iteration, our method generates candidate locations of v_2 as 8 neighbors of the starting point on the image grid. Candidate locations of v_3 are generated as neighbors of v_2 with some path constraints (see the next paragraph). Similarly, candidate locations of v_j are generated as the 8-connection neighbors of v_{j-1} on the image grid. 3 iterations of our method are illustrated in Fig. 3.

To decrease the number of candidate locations in every iteration, we do not allow any consecutive path segments to have turning angles greater than 90 degrees. In Fig. 3, all possible paths are shown as colored lines. When calculating $e_j(v_{j+1}, v_{j+2})$, our method only searches for v_j that does not generate sharp turns on any path $v_j \rightarrow v_{j+1} \rightarrow v_{j+2}$. Obviously, for any $e_j(v_{j+1}, v_{j+2})$ entry, the algorithm searches for at most three locations of v_j . In this way, we are able to decrease the time complexity of each iteration from $O(k^4)$ to $O(k^2)$.

To prevent too many loops on a path, we let any grid location can only be used as a candidate location for at most n_v times. To exclude obvious short-cuts and wrong paths, our method stops continuing a path if a consecutive n_{thres} of points on the path have intensity below I_{thres} . This operation explicitly excludes possible short-cuts and further improves the efficiency of our method. The algorithm terminates the iterations if (i) the ending point is visited for more than $10n_v$ times or (ii) the maximum iteration number, n_{max} is reached. We empirically set $n_{max} = 200$ in our application, since no actin filament is longer than 200 pixel-long in our dataset. During the iterations, all paths ended with the given ending point are recorded. After the iterations terminate, the path with the smallest energy value per node is chosen as the best one. In other words, the proposed method does not penalize length and therefore would not favor “short-cuts”.

Compared with other shortest path methods, such as the global minimal path method [2] and the graph-based method, our method has three advantages. (i) It avoids the common short-cut problems. Two operations, excluding paths with too many consecutive points with low intensity and choosing the path with minimal unit energy value, contribute to this advantage. Fig. 4 shows comparison results with other methods using real TIRFM images and a synthetic image. (ii) This method is efficient. Our MATLAB MEX implementation takes less than

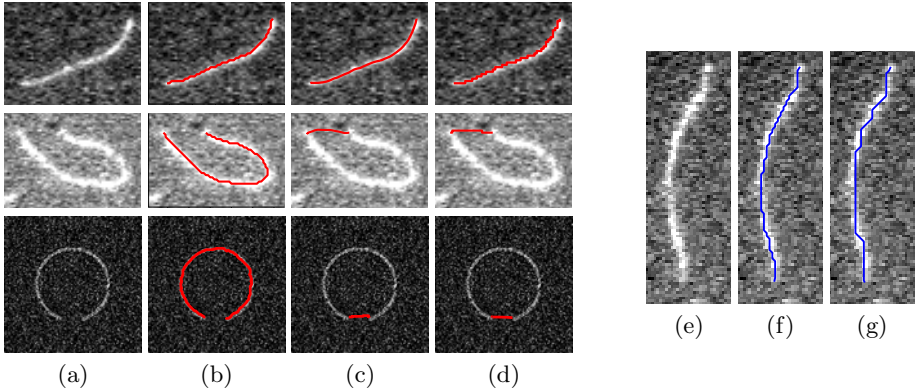


Fig. 4. (Left) Comparison between our proposed method, the global minimal path method [2], and the shortest-path-on-the-graph method. (a) Original images. (b) Results of our proposed method with $\alpha = 1, \beta = 0.1$. (c) Results of the global minimal path method with weight equaling 1. (d) Results of the shortest-path-on-the-graph method. (Right) The influence of the second-order smoothness term. (e) The original image. (f) $\alpha = 1, \beta = 0.1$. (g) $\alpha = 1, \beta = 0.5$.

0.02s to optimize a 200 pixel-long path with a complex external energy term (9) on an Intel E6850 3.0GHz CPU. (iii) Our method is able to take the second-order smoothness terms into consideration and therefore is more versatile to different applications, while other methods only consider first-order smoothness terms. Fig. 4 illustrates paths with different β settings.

3 Implementation

3.1 Image Preprocessing

The input TIRFM images are usually noisy and sometimes have biased intensity fields. We use the Frangi vessel enhancement filter [4] to enhance filamentous structures and to suppress noises. Let I_F and $I_{F,i}$ represent the vessel-enhanced results of a general image and the i th frame of a sequence, respectively. $I(v)$ represents the intensity value of image I at position $v \in \Omega$. If v falls between pixel grid locations, its intensity is obtained by bilinear interpolation.

3.2 Implementation of the External Energy E_{ext} in (6)

The smaller $E_{ext}(v_j)$ is, the better v_j represents a point on an actin filament. In [4], the Gaussian-second-order-derivative kernel shows its effectiveness in detecting vessel-like structures. The 1D kernel with scale σ , $G''_\sigma(x)$, measures the contrast between the regions inside and outside the range $(-\sigma, \sigma)$ in the direction of the derivative:

$$G''_\sigma(x) = \frac{(x^2 - \sigma^2)}{\sqrt{2\pi}\sigma^5} e^{-\frac{x^2}{2\sigma^2}} \quad (8)$$

We adopt this 1D kernel and calculate $E_{ext}(v_j)$ as its correlation with the image at location v_j . This 1D correlation has the lowest response if the filter's direction is perpendicular to the path's direction at location v_j . To calculate the path's direction at v_j , the external energy term needs three neighboring points, v_{j-1} , v_j , and v_{j+1} , we re-define the external energy term as

$$\tilde{E}_{ext}(v_{j-1}, v_j, v_{j+1}) = \sum_{x=-4\tilde{\sigma}}^{4\tilde{\sigma}} I_F(v_j + x\mathbf{d}) \cdot G''_{\tilde{\sigma}}(x) \quad (9)$$

where $\mathbf{d} \perp \left(\frac{v_{j+1} - v_j}{2\|v_{j+1} - v_j\|_2} + \frac{v_j - v_{j-1}}{2\|v_j - v_{j-1}\|_2} \right)$, $\|\mathbf{d}\|_2 = 1$.

\perp denotes that the unit vector \mathbf{d} is perpendicular to the path's direction at v_j , and $\tilde{\sigma} = 1.5$, which is half of the width of actin filaments. Consequently, the optimal-value functions $\{e_j\}_{j=2}^{n-1}$ in (7) should be re-defined as

$$e_j(v_{j+1}, v_{j+2}) = \min_{v_j} [e_{j-1}(v_j, v_{j+1}) + \alpha|v_{j+1} - v_j|^2 + \beta|v_{j+2} - 2v_{j+1} + v_j|^2 + \tilde{E}_{ext}(v_j, v_{j+1}, v_{j+2})], \quad (10)$$

3.3 Implementation of the Likelihood Term $l(\mathbf{r}_i)$ in (II)

Given a pair of starting and ending points, \mathbf{r}_i , in the i th frame, a spatial chain $\{v_{i,1}, \dots, v_{i,n_i}\}$ is optimized to represent the best location of the actin filament with \mathbf{r}_i as its tip points. The likelihood of this spatial chain representing an actual actin filament, $l(\mathbf{r}_i)$, is defined as a summation of a tip likelihood term, $l_t(\mathbf{r}_i)$, and a body likelihood term, $l_b(\mathbf{r}_i)$:

$$l(\mathbf{r}_i) = \lambda_t l_t(\mathbf{r}_i) + \lambda_b l_b(\mathbf{r}_i), \quad (11)$$

where λ_t and λ_b are the weights of the two likelihood terms. The tip likelihood term is calculated as correlations between a 9×9 template tip patch I_t with the rotated enhanced image $I_{F,i}$ at tip locations $v_{i,1}$ and v_{i,n_i} . The template tip patch is the mean of 50 manually labeled actin filament tip patches. The rotation directions are the two directions $\mathbf{d}_{i,1}$ and \mathbf{d}_{i,n_i} pointing outward the resulting spatial chain from the two tip points, $v_{i,1}$ and v_{i,n_i} , respectively:

$$\mathbf{d}_{i,1} = \frac{v_{i,1} - v_{i,5}}{\|v_{i,1} - v_{i,5}\|_2}, \quad \mathbf{d}_{i,n_i} = \frac{v_{i,n_i} - v_{i,n_i-4}}{\|v_{i,n_i} - v_{i,n_i-4}\|_2}. \quad (12)$$

The tip likelihood term is defined as

$$l_t(\mathbf{r}_i) = \frac{1}{81} \sum_{s=-4}^4 \sum_{t=-4}^4 I_t(s, t) \cdot \left[1/2 I_{F,i}(v_{i,1} + s\mathbf{d}_{i,1} + t\mathbf{d}_{i,1}^\perp) + 1/2 I_{F,i}(v_{i,n_i} + s\mathbf{d}_{i,n_i} + t\mathbf{d}_{i,n_i}^\perp) \right], \quad (13)$$

where $\mathbf{d}_{i,1}^\perp$ and \mathbf{d}_{i,n_i}^\perp are unit vectors perpendicular to $\mathbf{d}_{i,1}$ and \mathbf{d}_{i,n_i} , respectively. The body likelihood term is calculated as the average intensity value on the path:

$$l_b(\mathbf{r}_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_{F,i}(v_{i,j}). \quad (14)$$

3.4 Implementation of the Pairwise Term $p(\mathbf{r}_i, \mathbf{r}_{i+1})$ in (11)

We consider the pairwise relation of two adjacent spatial chains, $\{v_{i,1}, \dots, v_{i,n_i}\}$ and $\{v_{i+1,1}, \dots, v_{i+1,n_{i+1}}\}$, which are inferred by \mathbf{r}_i and \mathbf{r}_{i+1} , respectively. The relation is constrained by the desired growing speed and differences of tip directions, *i.e.*,

$$p(\mathbf{r}_i, \mathbf{r}_{i+1}) = \lambda_g p_g(\mathbf{r}_i, \mathbf{r}_{i+1}) + \lambda_d p_d(\mathbf{r}_i, \mathbf{r}_{i+1}), \quad (15)$$

where λ_g and λ_d weight the growing speed term $p_g(\cdot)$ and the tip direction difference term $p_d(\cdot)$. The growing speed term is defined as

$$p_g(\mathbf{r}_i, \mathbf{r}_{i+1}) = \begin{cases} 0 & |L(\mathbf{r}_{i+1}) - L(\mathbf{r}_i) - 5| < 3, \\ \left\| \|\mathbf{L}(\mathbf{r}_{i+1}) - L(\mathbf{r}_i) - 5\| - 3 \right\|^2 & \text{otherwise,} \end{cases} \quad (16)$$

where $L(\mathbf{r}_i) = \sum_{j=2}^{n_i} \|v_{i,j} - v_{i,j-1}\|_2$ is length of the spatial chain inferred by \mathbf{r}_i . $p_g(\cdot)$ does not penalize growing length in the range $[2, 8]$ pixel-long (around $[0.24, 1.36] \mu m$). Note that this range is set according to our data at hand and can be changed for data with other growing speed priors. $p_d(\cdot)$ is calculated as the differences between the pair of spatial chains' tip directions at the two tips:

$$p_d(\mathbf{r}_i, \mathbf{r}_{i+1}) = 1/2 \|\mathbf{d}_{i+1,1} - \mathbf{d}_{i,1}\|_2 + 1/2 \|\mathbf{d}_{i+1,n_{i+1}} - \mathbf{d}_{i,n_i}\|_2. \quad (17)$$

3.5 Generating the Solution Set of \mathbf{r}_i and Body Points Inference from \mathbf{r}_i

The dynamic programming algorithm (2)-(5) guarantees a global minimum on a discrete solution set. If the starting and ending points in the i th frame, \mathbf{r}_i , can be set to any grid location in that image, the size of solution set for each \mathbf{r}_i would be tremendous but also unnecessary. We threshold the vessel-enhanced image $I_{F,i}$ and perform the skeleton morphological operation to obtain the candidate locations for \mathbf{r}_i . In our application, we set the threshold to 0.2. We further constrain \mathbf{r}_i within the domain $[\tilde{r}_{i-1,1} - D, \tilde{r}_{i-1,1} + D] \times [\tilde{r}_{i-1,2} - D, \tilde{r}_{i-1,2} + D]$, where $[\tilde{r}_{i-1,1}, \tilde{r}_{i-1,2}] = \tilde{\mathbf{r}}_{i-1} = \arg \min_{\mathbf{r}_{i-1}} s_{i-2}(\mathbf{r}_{i-1})$. $\tilde{\mathbf{r}}_{i-1}$ represents the best tip locations on the $i-1$ th frame based on the information of only the first $i-1$ frames (see Fig. 5(a) and 5(b)). Note that it may not equal the final optimized result $\hat{\mathbf{r}}_{i-1}$. In this way, we usually obtain no more than 30 candidate points for each tip point with $D = 20$. Therefore, the size of the solution set for \mathbf{r}_i is usually no more than $30 \times 30 = 900$.

Each iteration of (2)-(5) takes most time on inferring body points given m pairs of different \mathbf{r}_i . Although an exhaustive calculation usually takes no longer

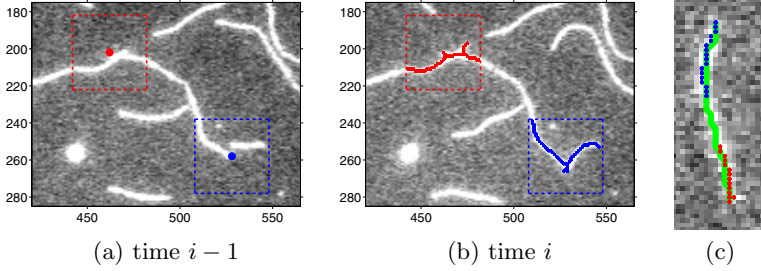


Fig. 5. (a) The best tip locations (red and blue) calculated on the $i-1$ th frame based on the first $i-1$ th frames, $\hat{\mathbf{r}}_{i-1} = \arg \min_{\mathbf{r}_{i-1}} s_{i-2}(\mathbf{r}_{i-1})$. The dashed-line boxes represent the range specified by $D = 20$ for the i th frame. (b) The solution set of \mathbf{r}_i . Red and blue points represent the candidate locations of the two tip points. (c) One snake calculated for all starting (red) and ending (blue) points.

than $0.02s \times 900 = 18s$, it can be further shortened by calculating only several paths. Given candidate tip points, inferred bodies usually coincide each other. Therefore, we calculate the path inferred by the tip pair with farthest distance first, and pairs with smaller distance later. In details, given two tip points v_1 and v_n , let $\{v_1, \dots, v_i, \dots, v_j, \dots, v_n\}$ represent its inferred body points. If $[v_i, v_j]^T$ is or close to a possible solution of \mathbf{r}_i , we use the path $\{v_i, \dots, v_j\}$ as the inferred body points of that possible solution. In this way, we usually calculate no more than 40 paths for each \mathbf{r}_i . An example is shown in Fig. 5(c), where only one path is calculated for all possible tip locations.

4 Experiments

We used 7 TIRFM image sequences from [6]. In these experiments, polymerization of muscle Mg-ADP-actin was monitored in the presence of varying concentrations of inorganic phosphate (Pi) and actin monomers. 30% of the actin was labeled on lysine side chains with Alexa green. Methylcellulose was used to suppress lateral Brownian movements of the actin filaments. Images were captured with 500 ms exposure time using a 488-nm laser on an Olympus IX71 microscope. Photobleaching is minimal in these sequences. The resolution was $0.17 \mu\text{m}/\text{pixel}$. There were 15-25 frames in each sequence. The time interval between frames was 10 sec or 30 sec.

For the spatial chains, we set $\alpha = 1, \beta = 0.1, n_v = 10, n_{thres} = 5, I_{thres} = 0.15$. For the temporal chain, we set $\lambda_t = 1, \lambda_b = 0.5, \lambda_g = 0.5, D = 20$ for all sequences and empirically chose λ_d from $\{0.1, 0.5, 1.5\}$ for each sequence. These parameters are empirically set according to experiments on a small set of 10 randomly sampled filaments. The errors of our segmentation method are among two types: (i) tips being inaccurately located, and (ii) body points being wrongly inferred to include several actin filaments. A biologist selected 62 actin filaments of interest in the 7 sequences (which consist of 1104 individual 2D segmentation

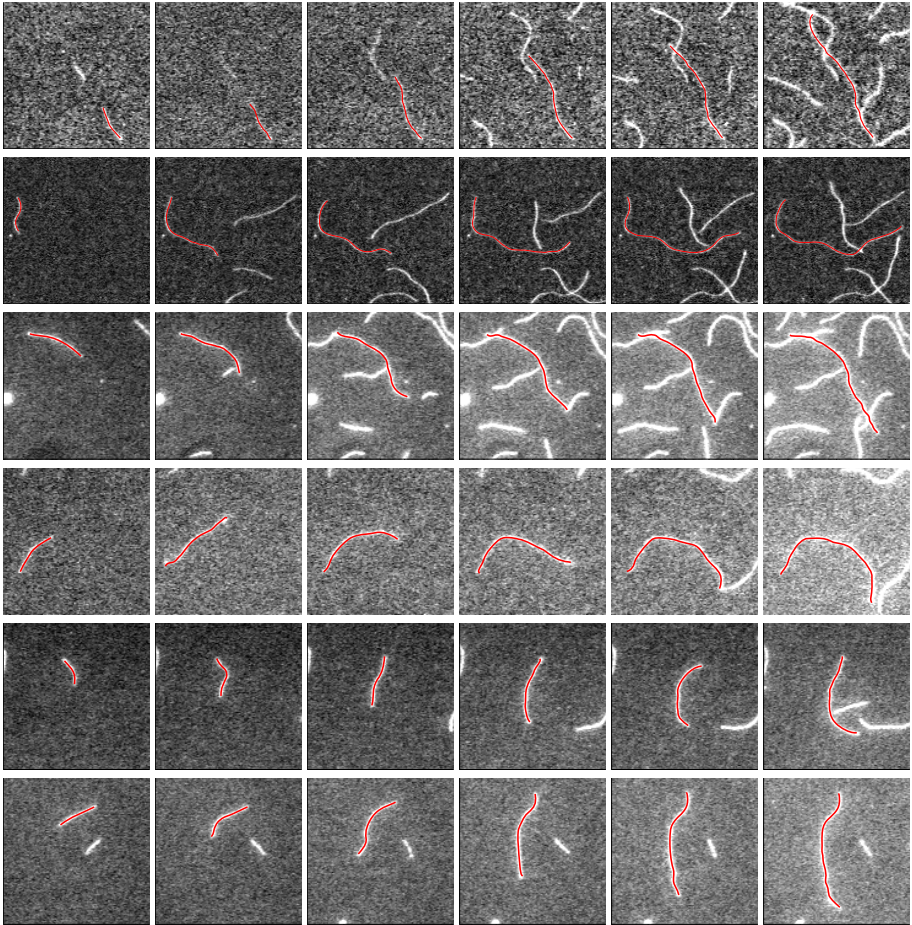


Fig. 6. Six example results of our proposed method. Each row shows segmentation results of one actin filament polymerizing in a period of time.

cases), and marked the actin filaments' correct tip locations in each frame as ground truth. We evaluated our algorithm's performance by calculating the L_2 distances between the ground truth and our method's results, and by counting how many times several actin filaments are mis-segmented as one. The statistics of our method is recorded in Table 1. The two previous methods [9] and [10] assumes a filament's body mostly remains static across time. However, in the general actin filament data that we tested, the actin filaments drift, grow, and change shapes at the same time. The previous methods failed in at least 1/3 of the cases because they require a part of the filament remain static over time for successful initialization.

We analyzed the proposed method's segmentation results and found two factors contribute most to large errors: (i) the Frangi filter enhances filaments'

Table 1. Tip tracking error statistics of successful segmentation cases and the number of mis-segmentation cases (Unit: pixel)

Sequence	# of Selected Filaments	Mean	Maximum	Standard Deviation	# of Mis-segmentation
I	8	1.2845	3.3642	0.8968	6 out of 135
II	7	1.0931	3.8613	1.0961	0 out of 108
III	14	1.0273	4.4414	0.8194	4 out of 212
IV	6	1.1092	5.7203	1.0658	12 out of 108
V	13	1.5557	3.4653	0.6634	4 out of 321
VI	5	1.7516	6.4418	1.3195	2 out of 100
VII	9	0.7515	1.8070	0.5781	9 out of 120

bodies robustly but may generate large artifacts around tip locations if the image has heavy noises. Such artifacts degenerate the performance of the tip likelihood term and result in most inaccurate tip localizations. To alleviate this factor, pre-processing filter need to be re-designed to generate less artifacts around tip parts. A more robust likelihood term also need to be considered. (ii) Our shortest-path method may infer a path across several actin filaments' bodies mostly if several filaments are tangled together in the image. This problem may be solved by jointly segmenting all actin filaments of interest in a sequence simultaneously. In this way, a temporal chain can exchange information from others and thus has better knowledge to avoid segmenting several actin filaments together.

5 Conclusion

In this paper, we proposed a novel algorithm to segment actin filaments in time-lapse sequences. The overall problem is modeled as a 1D temporal chain model which is efficiently optimized using dynamic programming. All available temporal information is integrated in this model. The body points are inferred in the process using another dynamic programming method which adaptively generates candidate solutions. Each likelihood term or pairwise term can be freely modified or deleted, and new terms representing new constraints can easily be added into this framework. Therefore, the proposed method is very flexible. Extensive experiments demonstrated the effectiveness and robustness of our proposed method.

Acknowledgments. This work was supported by NIH grant R21GM083928.

References

1. Amini, A.A., Weymouth, T.E., Jain, R.C.: Using dynamic programming for solving variational problems in vision. *IEEE TPAMI* 12, 855–867 (1990)
2. Cohen, L.D., Kimmel, R.: Global minimum for active contour models: A minimal path approach. *IJCV* 24, 57–78 (1997)

3. Felzenszwalb, F., Huttenlocher, P., Efficient, D.P.: matching of pictorial structures. In: Proc. CVPR, vol. 2, pp. 66–73 (2000)
4. 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)
5. Fujiwara, I., Takahashi, S., Tadakuma, H., Funatsu, T., Ishiwata, S.: Microscopic analysis of polymerization dynamics with individual actin filaments. *Nat. Cell Biol.* 4, 666–673 (2002)
6. Fujiwara, I., Vavylonis, D., Pollard, T.D.: Polymerization kinetics of ADP- and ADP-Pi-actin determined by fluorescence microscopy. *Proc. Natl. Acad. Sci. USA* 104, 8827–8832 (2007)
7. Hadjidemetriou, S., Toomre, D., Duncan, J.: Motion tracking of the outer tips of microtubules. *Medical Image Analysis* 12, 689–702 (2008)
8. Kong, K., Marcus, A., Giannakakou, P., Wang, M.: Using particle filter to track and model microtubule dynamics. In: Proc. ICIP, vol. 5, pp. 517–520 (2007)
9. Li, H., Shen, T., Smith, M., Fujiwara, I., Vavylonis, D., Huang, X.: Automated actin filament segmentation, tracking and tip elongation measurements based on open active contour models. In: Proc. ISBI (2009)
10. Li, H., Shen, T., Vavylonis, D., Huang, X.: Actin filament tracking based on particle filters and stretching open active contour models. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 673–681. Springer, Heidelberg (2009)
11. Li, H., Shen, T., Vavylonis, D., Huang, X.: Actin filament segmentation using spatiotemporal active-surface and active-contour models. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 86–94. Springer, Heidelberg (2010)
12. Saban, M., Altinok, A., Peck, A., Kenney, C., Feinstein, S., Wilson, L., Rose, K., Manjunath, B.: Automated tracking and modeling of microtubule dynamics. In: Proc. ISBI, vol. 1, pp. 1032–1035 (2006)
13. Sargin, M.E., Altinok, A., Kiris, E., Feinstein, S.C., Wilson, L., Rose, K., Manjunath, B.S.: Tracing microtubules in live cell images. In: Proc. ISBI (2007)
14. Smal, I., Draegestein, K., Galjart, N., Niessen, W., Meijering, E.: Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis. *IEEE TMI* 27, 789–804 (2008)

Multimodal Registration via Spatial-Context Mutual Information

Zhao Yi and Stefano Soatto

University of California, Los Angeles, USA
{zyi,soatto}@ucla.edu

Abstract. We propose a method to efficiently compute mutual information between high-dimensional distributions of image patches. This in turn is used to perform accurate registration of images captured under different modalities, while exploiting their local structure otherwise missed in traditional mutual information definition. We achieve this by organizing the space of image patches into orbits under the action of Euclidean transformations of the image plane, and estimating the modes of a distribution in such an orbit space using affinity propagation. This way, large collections of patches that are equivalent up to translations and rotations are mapped to the same representative, or “dictionary element”. We then show analytically that computing mutual information for a joint distribution in this space reduces to computing mutual information between the (scalar) label maps, and between the transformations mapping each patch into its closest dictionary element. We show that our approach improves registration performance compared with the state of the art in multimodal registration, using both synthetic and real images with quantitative ground truth.

1 Introduction

Image registration is a basic operation that plays a key role in many applications of medical imaging, motion correction, growth quantification, just to mention a few. There are many instances of registration algorithms, depending on what assumptions one is willing to make on the relation between the two images (e.g., changes of viewpoint, different sensors), and the nature of the scene (e.g., static geometry, growth effect). To each of these assumptions there correspond different criteria for registration, none “correct” or “incorrect” per se, just catering to different assumptions.

Assuming that the data are sampled from a joint density, mutual information (MI) [14,5,23] has emerged as an appealing criterion for multimodal registration. Without imposing a specific data formation model, one seeks for the registration that maximizes the MI between the distributions that generate the measurements. The key question then becomes *distributions of what?*

The first, and most naive, choice is to consider an image itself to be a *distribution of pixel values*. This is done without regards to the structure of the image, as one could scramble the pixels around and obtain an identical sample, and corresponds to a *first-order statistic* of the data formation process. More appealing

would be to consider *higher-order statistics*, for instance the joint distribution of pixels in a patch, that can capture the structure and spatial relations between image values. While this approach is appealing in theory, and straightforward in concept, the challenge lies in the curse of dimensionality: Even considering relatively small patches (say 11×11) would force us to estimate joint densities in a 121-dimensional space, a rather tall order. Most density estimators suffer from *sample impoverishment* phenomena in high-dimensional spaces, and therefore we can expect the estimate of MI to be rather crude in addition to being computationally expensive.

Our contribution in this paper is to provide an *efficient way to compute mutual information in a high-dimensional space of image patches avoiding sample impoverishment*. This enables us to perform registration of multimodal images using their local structures, improving the state of the art in MI-based registration. We do so by first organizing the space of patches into orbits under the action of the Euclidean group of the plane. This enables us to learn not dictionaries of individual patches, but of their equivalence classes, akin to “hyper-columns” in primate visual cortex. The result is that MI between image patches can be reduced to MI between their (scalar) label maps, and MI between the transformations mapping each patch into a dictionary element. This is achieved *regardless of the dimension of the patch*.

1.1 Prior Related Work

A number of extensions have been explored to include spatial context since the introduction of MI. The simplest is to add local features as a third channel, such as region labeling by Studholme et al. [20], image gradient by Pluim et al. [13], or spatial location by Studholme [19], Leockx [10], and Toews [21]. Though generally promising, feature selection itself is a tricky problem and the selected features have no guarantee to form sufficient statistics. An alternative strategy is to consider neighboring pixels instead of a single pixel in defining joint statistics, such as second-order MI by Rueckert et al. [16], region MI by Russakoff et al. [17], or high-dimensional normalized MI by Bardera et al. [1]. Such algorithms have to make simplifying assumptions (usually not true in most cases) about the underlying high-dimensional statistics in order to estimate them from limited number of sample points.

Our work adopts a similar idea to [17,1], but achieving a totally different analytical solution. We utilize the fact that though images live in a very high dimensional space, their intrinsic dimension is actually much lower; since the basic elements of images, e.g., edge, ridge, texture patterns, are limited and repetitive. Therefore, dimensionality reduction can be performed on local image patches which makes the computation of MI in a high-dimensional space naturally convert to a much lower-dimensional space, as shown in the next section.

2 Methodology

Let $I_j : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$; $x \mapsto I_j(x)$ be a (scalar-valued) image, where $j = 1, 2$ indicate the template and target images respectively. These are different images of the same scene or object (e.g. brain, heart, liver), obtained under different viewing conditions or at different times or from different persons/modalities. Now consider a neighborhood of size $\epsilon > 0$ of a location $x \in D$, $\mathcal{B}_\epsilon(x) = \{y \in D \mid d(x, y) \leq \epsilon\}$. We neglect border effects by discarding an ϵ -band around the perimeter of D (this can be done by zero-padding the image). When the domain of the image is discretized $D \subset \mathbb{Z}^2$ and we choose d to be the ℓ^1 distance, \mathcal{B} represents a “patch”. We call $S = \#\mathcal{B}_\epsilon$ the cardinality of the patch (the number of pixels in \mathcal{B}), which is a squared integer $(2\epsilon + 1)^2$ (e.g. $\epsilon = 5, S = 11^2 = 121$). The map that associates to each location $x \in D$ the vectorized intensity in the patch centered at x can be thought of as an “augmented” image:

$$\bar{I} : D \subset \mathbb{Z}^2 \rightarrow \mathbb{R}^S; x \mapsto \bar{I}(x) \doteq \{I(y) \mid \forall y \ \|x - y\| \leq \epsilon\}. \quad (1)$$

2.1 Mutual Information as a Registration Criterion

The MI between two random variables u, v that have joint distribution $dP(u, v)$ with density $p(u, v)$ and marginals $p(u)$ and $p(v)$ is defined as

$$\mathcal{I}(u, v) \doteq \int \log \frac{p(u, v)}{p(u)p(v)} dP(u, v). \quad (2)$$

It measures the “degree of independence” between the random variables, in the sense that MI is zero when u and v are independent, and is maximal when they are identical. Several authors have suggested using MI as a criterion for registration, by choosing w such that the aligned images, I_1 and $I_2 \circ w$, are “maximally dependent”. Therefore, it is an attractive prospect to look for $w = \arg \min \mathcal{I}(I_1, I_2 \circ w) + \lambda \|\nabla w\|$. However, images are not probability distributions, which leaves the question open as to what distributions to pull from the images in order to compute \mathcal{I} .

By far the most common assumption [14,5,23,3] is that the image itself is an (un-normalized) distribution of pixel values. That is (assuming that the set of possible values u taken by the image I at a pixel x is quantized into a finite set, say $[0, 255]$)

$$p(I = u) \simeq \frac{\#\{x \in D \mid I(x) = u\}}{\#D}. \quad (3)$$

This choice of *first-order statistic* corresponds to considering the image as a “bag of pixels” where only the *value* taken by the image at one pixel matters, but not *where that pixel is*. In other words, *there is no spatial context* taken into account on a first-order statistic.

A natural way to take spatial context into account is to consider *higher-order statistics*. That is, instead of considering $p(I = u)$, one considers

$$p(\bar{I} = \bar{u}) \simeq \frac{\#\{x \in D \mid I(y_i) = u_i, \forall y_i \in \mathcal{B}_\epsilon(x)\}}{\#D}. \quad (4)$$

This is a S -th order statistic, in the sense that it looks at the joint distribution of $S = (2\epsilon + 1)^2$ different values taken by the image in the neighborhood $\mathcal{B}_\epsilon(x)$. The attractive aspect of this approach is that the spatial structure of images is now fully captured in the model. Unfortunately, populating a joint distribution (histogram) from samples in a S -dimensional space is a notoriously difficult problem that suffers from the “curse of dimensionality” [2]. For this reason, the potential of using higher-order statistics to capture spatial context in MI-based image registration has so far been largely untapped. This is the problem we tackle in the next section. For now, we recall that the meaning of the formal notation $p(\bar{I}_1, \bar{I}_2)$ stands for

$$p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v}) \simeq \frac{\#\{x \in D \mid \bar{I}_1(y_i) = u_i, \bar{I}_2(y_i) = v_i, \forall y_i \in \mathcal{B}_\epsilon(x)\}}{\#D}, \quad (5)$$

and by $\mathcal{I}(I_1, I_2)$ we mean

$$\mathcal{I}(I_1, I_2) \doteq \sum_{\bar{u}, \bar{v} \in \mathbb{R}^S} p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v}) \cdot \log \frac{p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v})}{p(\bar{I}_1 = \bar{u})p(\bar{I}_2 = \bar{v})}. \quad (6)$$

Ideally, we want to compute, for a small $\delta > 0$, the best (δ -dependent) estimate of the registration field

$$\hat{w} = \arg \max_w \mathcal{I}(I_1, I_2 \circ w) \quad \text{s. t.} \quad \|\nabla w\|_{\mathbb{L}^1(D)} \leq \delta. \quad (7)$$

Solving this problem is conceptually simple, being a straightforward generalization of (2). However, the devil is in the computation, in that generating a reasonable estimate of MI becomes increasingly difficult in high-dimensional space. In the next two sections we address this problem and show that, regardless of the dimension S of the data space, we can reduce the problem to the computation of MI between labels (modes of the marginal densities), and their transformation parameters.

2.2 Learning Patch Dictionaries by Euclidean-Invariant Vector Quantization

One of the most common approaches to deal with the curse of dimensionality is to attempt dimensionality reduction [6]. If one thinks of the patch $\mathcal{B}_\epsilon(x)$ as a high-dimensional “signature” attached to the pixel x [11, 18, 22], one could reduce the dimensionality of the signature by clustering similar signatures around “centers”, each with a label, and then attaching the label to the pixel x . Such dimensionality reduction is usually performed using vector quantization (VQ) [8], and the cluster centers are the modes of the distribution in high-dimensional space, also known as “dictionary elements”.

For the case of images, where a set of (overlapping) patches is obtained by considering the neighborhood of each pixel, it is important to avoid representing

the same patch multiple times via different dictionary elements that are translated versions of each other. Therefore, we organize the space of patches so that transformed versions of the same patch are represented by the same dictionary element. In particular, we consider planar Euclidean transformations, whereby each “base patch” generates an *orbit* under the action of the Euclidean group of planar rotations and translations:

$$[\bar{I}(x)] \doteq \{\bar{I}(Rx + T) \mid R \in SO(2), T \in \mathbb{R}^2\}, \quad (8)$$

where $SO(2)$ denotes 2×2 orthogonal matrices with positive determinant. All patches in this orbit are equivalent, and any one element of the orbit can be chosen as a representative. When comparing two patches $\bar{I}_1(x)$ and $\bar{I}_2(x)$, we have to compare their corresponding orbits; if we denote with $w_E(x) \doteq Rx + T$ a Euclidean transformation, we have that

$$d([\bar{I}_1(x)], [\bar{I}_2(x)]) \doteq \min_{w_E} \|\bar{I}_1(x) - \bar{I}_2 \circ w_E(x)\|, \quad (9)$$

where the fact that we minimize with respect to a transformation only (as opposed to one per each orbit) comes from the fact that $\|\bar{I}_1 \circ w_1 - \bar{I}_2 \circ w_2\| = \|\bar{I}_1 - \bar{I}_2 \circ w_2 \circ w_1^{-1}\|$, since $SE(2)$ is isometric and by the group properties, $w_E \doteq w_2 \circ w_1^{-1}$ is also a Euclidean transformation.

Now, VQ can be applied to the space of patches, relative to the norm (9), in a conceptually straightforward manner. In practice this is a costly operation, so affinity propagation [7] is used as an approximate mode seeking algorithm. The output of the algorithm is a collection of “cluster centers” $\{C^k\}_{k=1}^{K_c}$ per each image ensemble $\{I_i\}_{i=1}^N$ taken from the same modality, say MR-T1. Naturally, a separate dictionary learning procedure has to be undertaken for each different modality, say CT, MR-T2, etc. The residual, or the distance between each patch and its closest cluster center,

$$e(x) \doteq d(\bar{I}(x), [C^{k(x)}]) \quad (10)$$

can be used to measure the effectiveness (covering) of the clustering algorithm. The results of learned dictionary elements using affinity propagation relative to a Euclidean-invariant distance and a nearest-neighbor criterion is shown in Fig. 11. Note that MR-T2 has the biggest residual among all three modalities due to its high contrast of edges, i.e., a small misalignment will lead to a big $e(x)$. To illustrate the importance of Euclidean invariance in dictionary learning, we compare the representative power of dictionary elements learned using different levels of invariance in Fig. 2. It is clear that when Euclidean invariance is considered, the residual image shows least structures.

2.3 Computing Spatial-Context Mutual Information

For simplicity of reference, without claiming any paternity to the semantics of the word “context”, we call the quantity in (6) “spatial-context mutual information” (SCMI) so long as $S > 1$. It is just a generalization of MI to patches. As we have

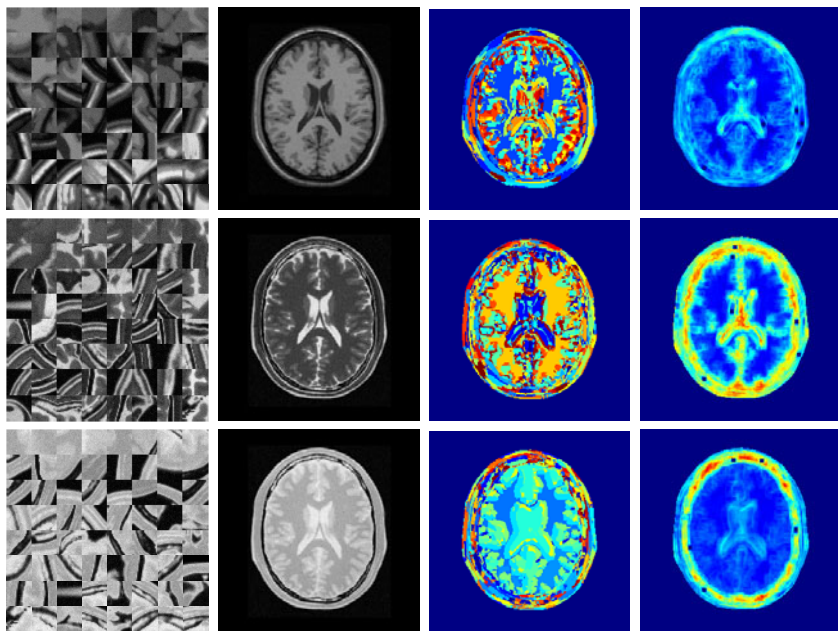


Fig. 1. VQ results on MR-T1 (top), MR-T2 (middle), and MR-PD (bottom) modalities: (left) learned dictionary elements, (middle-left) an example training image, (middle-right) its label map in pseudocolor, and (right) the residual image with blue/red representing low/high residuals

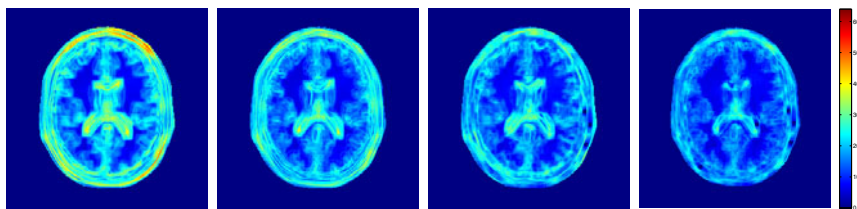


Fig. 2. Residual images under different levels of invariance for MR-T1 modality: (left) no invariance, (middle-left) only rotation invariance, (middle-right) only translation invariance, and (right) Euclidean invariance

mentioned, the challenge is not to define this quantity, but to compute it. Note that if we consider patches of size (cardinality) S , for instance $S = 11 \times 11$, the ambient space where the distributions $p(\bar{I})$ live is \mathbb{R}^{121} , a rather large space.

The VQ procedure described in the previous section partitions this space into K disjoint subsets each represented by a dictionary element C^k , $k = 1, \dots, K$. We denote with $L : \mathbb{R}^S \mapsto [1, \dots, K]$ the nearest-neighbor operator that associates a patch $\bar{I}(x)$ with its dictionary element C^k , then the partition is given by $k = \arg \min_{\hat{k}} d(\bar{I}(x), [C^{\hat{k}}])$. The map $D \rightarrow [1, \dots, K]; x \mapsto L(\bar{I}(x)) = k$ is called

the *label map*. Since $\bar{I} = \bar{u} \Rightarrow L(\bar{I}) = L(\bar{u})$, we have that, neglecting the spatial argument x ,

$$\begin{aligned} p(\bar{I} = \bar{u}) &= p(\bar{I} = \bar{u}, L(\bar{I}) = L(\bar{u})) \\ &= p(\bar{I} = \bar{u} \mid L(\bar{I}) = L(\bar{u}))p(L(\bar{I}) = L(\bar{u})) \end{aligned} \quad (11)$$

and a similar relation holds for for the joint density. Using this simple fact, in Appx. B we show that the SPMI can be computed as follows:

$$\mathcal{I}(I_1, I_2) = \mathcal{I}(L(\bar{I}_1), L(\bar{I}_2)) + \sum_{k,m=1}^{K,M} p(L(\bar{I}_1) = k, L(\bar{I}_2) = m) \cdot \mathcal{I}(\bar{I}_1, \bar{I}_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m). \quad (12)$$

Note that here we have dropped the subscripts for notation simplicity: $L(\bar{I}_1), L(\bar{I}_2)$ actually mean $L_1(\bar{I}_1), L_2(\bar{I}_2)$, since different images I_1, I_2 may correspond to different dictionaries $\{C_1^k\}_{k=1}^K, \{C_2^m\}_{m=1}^M$. The first term in (12) is the MI between two label maps, which is easy to calculate as L is a scalar-valued function; the second term is the weighted sum of the conditional SCMI on labels, which depends on the structure of the two patch clusters under given labels.

We are going to assume that the modes of the distribution have been faithfully captured, and that data points cluster close to their respective modes so that the quantization error is negligible. This can always be achieved by increasing the number of cluster centers in the VQ procedure. Also, we are going to make the assumption that the cluster centers, or “dictionary elements,” are not Euclidean invariants, lest the entire orbit collapses to a singleton, and that patch cannot be matched, as we show in Appx. A. Under these assumptions, we have that

$$\mathcal{I}(\bar{I}_1, \bar{I}_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m) = \mathcal{I}(w_1, w_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m), \quad (13)$$

where $\bar{I}_1 = C_1^k \circ w_1, \bar{I}_2 = C_2^m \circ w_2$. Finally, calling for ease of notation $L_1 \doteq L(\bar{I}_1)$ and similarly for L_2 , we have

$$\boxed{\mathcal{I}(I_1, I_2) = \mathcal{I}(L_1, L_2) + \sum_{k,m=1}^{K,M} p(L_1 = k, L_2 = m) \cdot \mathcal{I}(w_1, w_2 \mid L_1 = k, L_2 = m)} \quad (14)$$

This expression is now easy to compute since $w_1, w_2 \in SE(2)$ can be represented by local coordinates in \mathbb{R}^3 (one angular parameter and two translational parameters).

2.4 Building Patch Correspondence from Joint Statistics

Now we already have a complete derivation of SCMI and one may apply it directly to multimodal registration. Furthermore, we can perform prior learning of corresponding labels from joint statistics in the patch space. Given pre-aligned training images, we define the matching score for each pair of labels (k, l) as

$$\mathcal{M}(k, l) \doteq p(L(\bar{I}_1) = k, L(\bar{I}_2) = m) \cdot \mathcal{I}(\bar{I}_1, \bar{I}_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m), \quad (15)$$

where the first term is the occurrence of such labels, and the second term is the conditional SCMI computed by equation (13). Higher matching scores mean more occurrence and bigger SCMI for chosen label pairs, and thus higher likelihood for corresponding patches to be matched. Such scores can actually be used as a *joint landmark detector between different modalities*. Fig. 3 shows the top seven pairs of corresponding patches ranked by (15) on pre-aligned training images.

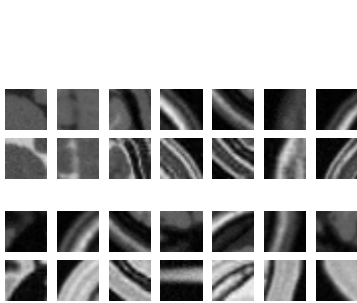


Fig. 3. Top seven pairs of corresponding patches between MR-T1 (top) and MR-T2 (middle-top), MR-T1 (middle-bottom) and MR-PD (bottom) modalities.

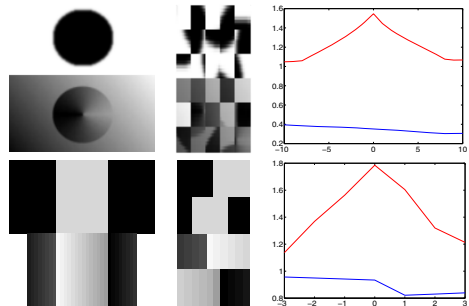


Fig. 4. Registration results on synthetic phantoms: (left) input images, (middle) learned dictionary elements, and (right) comparison between MI (blue) and SCMI (red) values over horizontal translation.

3 Experimental Validation

The synthetic phantoms shown in Fig. 4 are similar to those used in [12,15], where the top ones contain an elliptical region and the bottom ones contain a ridge structure. In both cases, the given image pair is already aligned. It is clear that the SCMI correctly exploits the local structures near the edge, therefore can successfully capture the stage of alignment, while the standard MI is adversely affected by the strong intensity gradation in the target images and leads to a spurious registration. This illustrates the advantage of our approach to accommodate severe shading artifacts during registration.

We validate the accuracy of our approach on simulated MR brain images, obtained from the BrainWeb MR simulator [4] with slice thickness $1mm$, noise level 3% , and intensity non-uniformity 20%. Triplets of pre-registered T1/T2/PD images are selected, where we artificially transform all T1 images and plot the changes of MI values over different transformation parameters (representative cases in Fig. 5). Though standard MI, local MI [9], and SCMI all reach highest values when aligned, SCMI tends to have less local maxima than the standard MI and local MI when nuisances present (mostly interpolation and inhomogeneity in our case). To have a better estimate of the robustness, we perform such testing for 12 pairs of testing images and calculate the total number of spurious

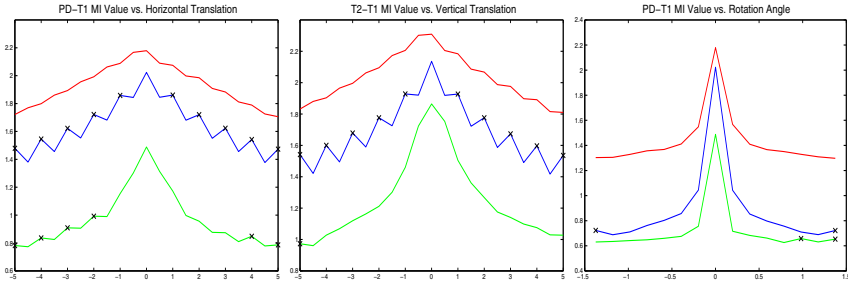


Fig. 5. Registration results on simulated images. Each plot shows the comparison between MI (blue), local MI (green), and SCMI (red) values over transformation parameters with black cross marking spurious local maxima.

Table 1. Reduction in number of spurious maxima over rigid transformation parameters. Note that our approach has the least number of local maxima.

Cases	T2-T1			PD-T1		
	Rotation	X Translation	Y Translation	Rotation	X Translation	Y Translation
MI	20	118	118	22	120	120
Local MI	16	25	16	19	46	27
SCMI	9	9	8	3	17	17

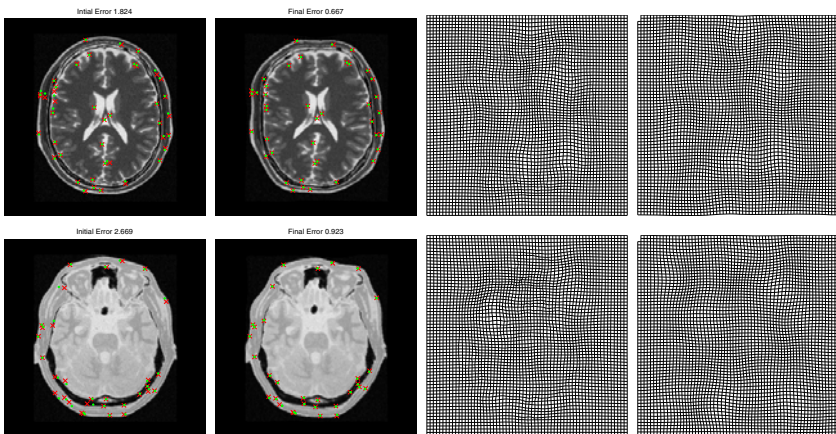


Fig. 6. Landmark detection results between T2-T1 (top) and PD-T1 (bottom) modalities: (left) initial landmarks detected, (mid-left) final landmarks refined, (mid-right) recovered deformation, and (right) ground truth. The red cross represent the detection, while the green dots are the projection.

maxima for each approach. Tab. 3 shows the comparison between our approach, the standard MI approach, and the local version of MI approach.

We further incorporate the priori learned patch correspondence in Fig. 3 as a joint landmark detector between different modalities for non-rigid registration [9]. The detected corresponding landmarks are not only used to guide the registration process with the intensity statistical measures, but also gradually refined as the deformation approaching ground truth. Fig. 6 shows an example of the landmark detection results. Note that the detection error is reduced from 1.8 to 0.7 pixels for T2-T1, and from 2.7 to 0.9 pixels for PD-T1 registration.

4 Discussion

We propose extending mutual-information based registration beyond “bags of pixels” to higher-dimensional representations that capture spatial context in images. This can be achieved assuming that the joint density of image patches is well approximated by a sparse collection of modes, whose representatives (the “cluster centers”, or “dictionary elements”) can undergo Euclidean transformations. Under these conditions, MI can be computed by simply computing the MI between label maps (that are scalar), and by the corresponding transformation parameters (three-dimensional). This enables us to solve curse of dimensionality, and make high-dimensional MI-based registration practical. As expected, we verify that our model improves the performance of the state of the art.

References

1. Bardera, A., Feixas, M., Boada, I., Sbert, M.: High-dimensional normalized mutual information for image registration using random lines. In: Workshop on Biomedical Image Registration, pp. 264–271 (2006)
2. Bellman, R.: Adaptive Control Processes: A Guided Tour. Princeton University Press, Princeton (1961)
3. Chung, A.C.S., Wells III, W.M., Norbash, A., Grimson, W.E.L.: Multi-modal image registration by minimising kullback-leibler distance. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2489, pp. 525–571. Springer, Heidelberg (2002)
4. Cocosco, C., Kollokian, V., Kwan, R., Evans, A.: Brainweb: online interface to a 3D MRI simulated brain database. In: Int. Conf. Functional Mapping of the Human Brain, p. S427 (1997)
5. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multimodality medical image registration using information theory. In: Int. Conf. Information Processing in Medical Imaging, pp. 263–274 (1995)
6. Fodor, I.: A survey of dimension reduction techniques. Tech. rep., U.S. Department of Energy (2002)
7. Frey, B., Dueck, D.: Clustering by passing messages between data points. *Science* 315, 972–976 (2007)
8. Gray, R.: Vector Quantization. *IEEE ASSP Magazine* (1984)
9. Hermosillo, G., Faugeras, O.: Dense image matching with global and local statistical criteria: a variational approach. In: IEEE Conf. Computer Vision and Pattern Recognition, pp. 73–78 (2001)

10. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 725–737. Springer, Heidelberg (2007)
11. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision* 60, 91–110 (2004)
12. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. *Medical Image Analysis* 9, 330–343 (2005)
13. Pluim, J., Maintz, J., Viergever, M.: Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Medical Imaging* 19, 809–814 (2000)
14. Pluim, J., Maintz, J., Viergever, M.: Mutual information based registration of medical images: a survey. *IEEE Trans. Medical Imaging* 22, 986–1004 (2003)
15. Roche, A., Malandain, G., Pennec, X., Ayache, N.: The correlation ratio as a new similarity measure for multimodal image registration. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 1115–1124. Springer, Heidelberg (1998)
16. Rueckert, D., Clarkson, M., Hill, D., Hawkes, D.: Non-rigid registration using higher-order mutual information. In: *Proc. SPIE Image Processing: Medical Imaging*, pp. 438–447 (2000)
17. Russakoff, D.B., Tomasi, C., Rohlfing, T., Maurer Jr., C.R.: Image similarity using mutual information of regions. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 596–607. Springer, Heidelberg (2004)
18. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: *textronBoost*: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)
19. Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Trans. Medical Imaging* 25, 626–639 (2006)
20. Studholme, C., Hill, D., Hawkes, D.: Automated 3D registration of MR and CT images of the head. *Medical Image Analysis* 1, 163–175 (1996)
21. Toews, M., Wells III, W.M.: Bayesian registration via local image regions: Information, selection and marginalization. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 435–446. Springer, Heidelberg (2009)
22. Varma, M., Zisserman, A.: Texture classification: are filter banks necessary. In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. II–691–8 (2003)
23. Viola, P., Wells, W.: Alignment by maximization of mutual information. *Int. Journal of Computer Vision* 24, 137–154 (1997)

A Invariance and Matchability

In Sect. 2.3 we have argued that cluster centers C^k that are Euclidean invariant should be eliminated from the representation, for they do not help the matching process. First, we say that a statistic $\phi(I)$ (a deterministic function of the data, for instance a patch $\phi(I) = I_{|B}$) is *invariant* to the group $g \in SE(2)$ if $\phi(I \circ g) = \phi(I)$ for all $g \in SE(2)$. Therefore, by the implicit function theorem, the Jacobian determinant of ϕ is singular at $g(x)$; however, the Jacobian of ϕ has the gradient of the image as one of its factor. The fact that the Jacobian is

zero implies that the infinitesimal generators of w are not defined at $g(x)$ by the registration constraint $I_1(x) - I_2(w(x)) = 0$, and therefore $w(x)$ is undetermined. Therefore, there is no loss of generality in assuming that the cluster centers are not degenerate.

In order to check whether C^k is rotation invariant, a necessary condition is

$$\nabla C^k(y_i) \cdot (Gy_i) = 0 \quad \forall y_i \in \mathcal{B}_\epsilon(x) \quad (\text{A-1})$$

where $G \doteq [0, 1; -1, 0]$ is the group generator for rotation matrix. (A-1) means that the position vector x cannot be always perpendicular to the gradient direction ∇C^k within $\mathcal{B}_\epsilon(x)$, otherwise C^k is rotation invariant (e.g., circle or constant patches). In order to check whether C^k is translation invariant, a necessary condition is

$$\nabla C^k(y_i) \parallel \nabla C^k(y_j) \quad \forall y_i, y_j \in \mathcal{B}_\epsilon(x). \quad (\text{A-2})$$

(A-2) means that the gradient direction ∇C^k cannot be homogeneous within $\mathcal{B}_\epsilon(x)$, otherwise C^k is translation invariant in its normal direction.

B Calculation of SPMI

In the following we show how to compute (6). Recall the nearest-neighbor map L_1, L_2 which partitions \mathbb{R}^S into K, M disjoint subsets, denoted by $R^S = \cup_{k=1}^K \Omega_k = \cup_{m=1}^M \Lambda_m$. Substitute (11) into (6), we have

$$\begin{aligned} & \mathcal{I}(\bar{I}_1, \bar{I}_2) \\ &= \sum_{\bar{u}, \bar{v} \in R^S} p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v} \mid L(\bar{I}_1) = L(\bar{u}), L(\bar{I}_2) = L(\bar{v})) \cdot p(L(\bar{I}_1) = L(\bar{u}), L(\bar{I}_2) = L(\bar{v})) \\ & \left(\log \frac{p(L(\bar{I}_1) = L(\bar{u}), L(\bar{I}_2) = L(\bar{v}))}{p(L(\bar{I}_1) = L(\bar{u}))p(L(\bar{I}_2) = L(\bar{v}))} + \log \frac{p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v} \mid L(\bar{I}_1) = L(\bar{u}), L(\bar{I}_2) = L(\bar{v}))}{p(\bar{I}_1 = \bar{u} \mid L(\bar{I}_1) = L(\bar{u}))p(\bar{I}_2 = \bar{v} \mid L(\bar{I}_2) = L(\bar{v}))} \right) \\ &= \sum_{k, m=1}^{K, M} p(L(\bar{I}_1) = k, L(\bar{I}_2) = m) \cdot \log \frac{p(L(\bar{I}_1) = k, L(\bar{I}_2) = m)}{p(L(\bar{I}_1) = k)p(L(\bar{I}_2) = m)} \\ & \cdot \sum_{\bar{u} \in \Omega_k, \bar{v} \in \Lambda_m} p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v} \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m) + \sum_{k, m=1}^{K, M} p(L(\bar{I}_1) = k, L(\bar{I}_2) = m) \cdot \\ & \cdot \sum_{\bar{u} \in \Omega_k, \bar{v} \in \Lambda_m} p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v} \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m) \cdot \log \frac{p(\bar{I}_1 = \bar{u}, \bar{I}_2 = \bar{v} \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m)}{p(\bar{I}_1 = \bar{u} \mid L(\bar{I}_1) = k)p(\bar{I}_2 = \bar{v} \mid L(\bar{I}_2) = m)} \\ &= \mathcal{I}(L(\bar{I}_1), L(\bar{I}_2)) + \sum_{k, m=1}^{K, M} p(L(\bar{I}_1) = k, L(\bar{I}_2) = m) \cdot \mathcal{I}(\bar{I}_1, \bar{I}_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m). \end{aligned} \quad (\text{A-3})$$

Assume the VQ error is negligible, i.e., $e(x) \simeq 0 \Rightarrow \bar{I}(x) \simeq C^{k(x)} \circ w(x)$, we have

$$\mathcal{I}(\bar{I}_1, \bar{I}_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m) = \mathcal{I}(C_1^k \circ w_1, C_2^m \circ w_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m) \quad (\text{A-4})$$

When C_1^k, C_2^m are not rotation or translation invariant, i.e.,

$$C_1^k \circ w_1 \neq C_1^k, \quad C_2^m \circ w_2 \neq C_2^m \quad \forall w_1, w_2 \quad (\text{A-5})$$

(A-4) reduces to

$$\mathcal{I}(\bar{I}_1, \bar{I}_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m) = \mathcal{I}(w_1, w_2 \mid L(\bar{I}_1) = k, L(\bar{I}_2) = m). \quad (\text{A-6})$$

Generalized Partial Volume: An Inferior Density Estimator to Parzen Windows for Normalized Mutual Information

Sune Darkner and Jon Sporring

eScience Center, Department of Computer Science, University of Copenhagen
Universitetsparken 1, DK-2100 Copenhagen, Denmark
{darkner,sporring}@diku.dk

Abstract. Mutual Information (MI) and normalized mutual information (NMI) are popular choices as similarity measure for multimodal image registration. Presently, one of two approaches is often used for estimating these measures: The Parzen Window (PW) and the Generalized Partial Volume (GPV). Their theoretical relation has so far been unexplored. We present the direct connection between PW and GPV for NMI in the case of rigid and non-rigid image registration. Through step-by-step derivations of PW and GPV we clarify the difference and show that GPV is algorithmically inferior to PW from a model point of view as well as w.r.t. computational complexity. Finally, we present algorithms for both approaches for NMI which is comparable in speed to Sum of Squared Differences (SSD), and we illustrate the differences between PW and GPV on a number of registration examples.

Keywords: Similarity measure, registration, normalized mutual information, density estimation, scale space, locally orderless images.

1 Introduction

Mutual information (MI) and its normalized version (NMI) are considered state of the art for image registration. MI and NMI are particularly useful for registering Magnetic Resonance Images (MRI) to MRI as well as for multimodal image registration in general. MI and NMI are entropy based measures and hence rely on intensity distributions. Intensity distributions are most often approximated by discrete histograms, which poses a challenge to gradient based methods. Today, one of two approaches is often used for estimating these measures: The Parzen Window (PW) [21] and the Generalized Partial Volume (GPV) [10, 3]. Empirical comparisons have previously been presented, but their theoretical connection has so far been unexplored despite the fact that both are used in the same context. We present the derivations of both PW and GPV in a joint theoretical context. This gives novel insight to the relation between PW and GPV enabling a theoretical evaluation and comparison of the two. We show that PW and GPV are special cases of histogram registration using Locally Orderless Images (LOI) [7] on NMI. The concept of LOI allow us to treat derivatives in

measurement, integration, and intensity space in a well-posed manner, as well as offer a scale-space formulation of these spaces. From this we formulate algorithms, analytically compare their speed, and discuss the different approaches.

We evaluate PW and GPV in the setting of LOI by a series of simple experiments with numerically the exact same prerequisites, interpolation, regularization, and optimization on publicly available data: 3D cardiac MRI [1], and 3D T1 brain MRI [12] for rigid intra- and inter-patient registration as well as non-rigid intra patient registration. For each registration we report standard deviation and mean as a function of method parameters, and we conclude that there are differences between the proposed algorithms, and differences reported originates from approximations and smoothing in different spaces: GPV in namely the isophote domain and in PW the image and intensity domain. This leads us to conclude that PW is more attractive and in general superior density estimator to GPV for MI and NMI.

1.1 Previous Work

The use of Mutual Information (MI) for image registration was originally proposed by [4, 21]. An extensive overview was given in [17]. Normalized Mutual Information (NMI) was introduced as a more robust alternative especially designed for multi modal image registration [19]. The first implementations relied on Powell's method [10], hill climbing [19], or similar methods without gradients, which were accurate but slow. A GPU speedup was suggested in [13]. Today, state-of-the-art implementations are gradient based methods and group in two algorithm types: The first type is based on Parzen Windows (PW) [21] and relies on the fact that the marginal and the joint histograms are made continuous by using different kernels, e.g., Gaussian or B-splines [20]. The second type is based on Generalized Partial Volume (GPV), where the distribution is sampled from the image directly [10]. Analytical derivatives of this method were presented in [11] and a generalization using B-splines was presented in [3]. A variational method relating to LOI [7] for MI (and other measures) was presented in [6]. GPV and PW was compared numerically in [8] concluding that PW is precise and GPV has a larger convergence radius. MI and NMI are notorious for their local minima and difficulty of implementation and the choice of interpolation scheme greatly influence the smoothness of the objective function. Some investigations into this can be found in [16, 5]. An alternative approach is the Conditional Mutual Information [9].

In this article, we investigate PW and GPV for NMI, using differential calculus in a thorough step-by-step presentation. The derivations may be seen as an alternative to the variational approach in [6], our's holds the same generality, but lead to much faster algorithms, since numerical issues become obvious, and our derivation lead to a direct comparison between PW and GVP. The remainder of this article is organized as follows: In Section 2 the general registration framework is described. In Section 3 we discuss LOI as a basis for analyzing GPV and PW as well as demonstrate their similarity and derive fast algorithms. In Section 5

we compare the methods experimentally, and finally, in Section 6 we give our conclusions.

2 Image Registration

Image registration is the process of transforming one image $\tilde{I} : \Omega \rightarrow \Gamma$, where $\Omega \subseteq \mathbb{R}^N$ and $\Gamma \subseteq \mathbb{R}$, w.r.t. a reference image $R : \Omega \rightarrow \Gamma$ such that some functional $\mathcal{F}(\tilde{I}, R)$ is minimized. We consider diffeomorphic transformation of NM parameters, $\phi : \Omega \times \mathbb{R}^{NM} \rightarrow \Omega$, and for brevity we write $I = \tilde{I} \circ \phi$. The general form of \mathcal{F} is,

$$\mathcal{F} = \mathcal{M}(I, R) + \mathcal{S}(\phi), \tag{1}$$

where \mathcal{M} is a (dis)similarity measure between the images and $\mathcal{S}(\phi)$ is a regularization term. Regularization is almost always required in order to obtain a unique solution, since image registration is generally ill-posed [14]. We use Riemannian Elasticity [15],

$$\mathcal{S}(\phi) = \frac{\mu}{4} \sum_i \log^2 \epsilon_i + \frac{\lambda}{8} \left(\sum_i \log \epsilon_i \right)^2, \tag{2}$$

where ϵ_i is the i 'th squared eigenvalue of the Jacobian of the transformation ϕ , μ and λ are regularization parameters.

The focus of this paper is the similarity measure, and we consider NMI for \mathcal{M} [19],

$$\mathcal{M}_{\text{NMI}} = \frac{\mathcal{H}_I + \mathcal{H}_R}{\mathcal{H}_{I,R}}, \tag{3}$$

where \mathcal{H} denotes the marginal and the joint entropy of the intensity distribution [18], specifically,

$$\mathcal{H}_I = - \int_{\Gamma} p_I(i) \log p_I(i) di, \quad \mathcal{H}_R = - \int_{\Gamma} p_R(i) \log p_R(i) di, \tag{4}$$

$$\mathcal{H}_{I,R} = - \int_{\Gamma^2} p_{I,R}(i, j) \log p_{I,R}(i, j) di \wedge dj, \tag{5}$$

using the natural logarithm for convenience, and the intensity and joint intensity distributions, $p(\cdot) : \Gamma \rightarrow \mathbb{R}_+$ and $p(\cdot, \cdot) : \Gamma^2 \rightarrow \mathbb{R}_+$, are estimated by the histogram and joint histogram of the intensity values. NMI has proven to be very powerful for registration of medical images in general.

Finally, we consider Uniform B-splines [2] for coordinate transformation,

$$\mathbf{x} = \phi(\tilde{\mathbf{x}}, \Phi) = \sum_{m=0}^M \phi_m \prod_{n=1}^N f_{m,t}(\tilde{x}_n), \tag{6}$$

where $\Phi = [\phi(\tilde{\mathbf{x}}_1), \dots, \phi(\tilde{\mathbf{x}}_M)] \in \Omega^M$ is a matrix of values of the transformation at regular grid coordinates $\tilde{\mathbf{x}}_m$, $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_N]^T$ is the evaluation point, and where the coordinate wise interpolation function, $f_{m,t} : \mathbb{R} \rightarrow \mathbb{R}$, is given by the Cox-de Boor recursion formula [2].

In the following, we will study the implication of the 3 independent scales of the local histogram on the registration problem, and we derive the PW and GPV for NMI in the context of LOI for usage in a quasi-Newton method.

3 Locally Orderless Images (LOI)

To fully understand the connection between PW and GPV we first shortly review the concept of locally orderless images (LOI), the theoretical foundation of histogram generation.

The critical element for NMI based registration methods is the generation of the density distribution and efficiently calculating the gradient of (1), especially the gradient of (3), and in parts the gradient of ϕ . We study NMI registration through LOI [7], which extends the concept of histograms with the 3 fundamental scales, the amount of spatial smoothing of the images (image smoothing), the amount of histogram smoothing (intensity smoothing), and the size of the window (the partial volume) for calculating local histograms. Thus, a local histogram is written as,

$$h_I(i, \mathbf{x}, \Phi, \alpha, \beta, \sigma) = P(I(\mathbf{x}, \Phi, \sigma) - i, \beta) * W(\mathbf{x}, \alpha), \tag{7}$$

$$I(\psi, \Phi, \sigma) = I(\mathbf{x}) * K(\mathbf{x}, \sigma), \tag{8}$$

where $d\psi_N$, $i \in \Gamma$, P is a Parzen window of intensity or tonal scale $\beta \in \mathbb{R}_+$, K is a spatial measurement kernel of scale $\sigma \in \mathbb{R}_+$, W is an integration window of scale $\alpha \in \mathbb{R}_+$ and located at \mathbf{x} , $\cdot * \cdot$ is the convolution operator taken w.r.t. the variable \mathbf{x} , and Φ denotes the parameters for the transformation. The histogram h_R is defined similarly independently of Φ . In [7] it is proposed to use $P(i, \beta) = e^{-i^2/(2\beta^2)}$, and $K(\mathbf{x}, \sigma) = W(\mathbf{x}, \sigma) = e^{-\mathbf{x}^T \mathbf{x}/(2\sigma^2)}/(2\pi\sigma^2)^{N/2}$, and this structure is called the Locally Orderless Image. The distributions are obtained by normalizing to unity,

$$p_I(i|\mathbf{x}, \Phi, \alpha, \beta, \sigma) \simeq \frac{h_I(i, \mathbf{x}, \Phi, \alpha, \beta, \sigma)}{\int_{\Gamma} h_I(j, \mathbf{x}, \Phi, \alpha, \beta, \sigma) dj}, \tag{9}$$

$$p_I(i|\Phi, \alpha, \beta, \sigma) = \frac{1}{|\Omega|} \int_{\Omega} p_I(i|\mathbf{x}, \Phi, \alpha, \beta, \sigma) d\mathbf{x}, \tag{10}$$

where we have assumed (conditional) independence and uniformity such that $p_I(i, \mathbf{x}|\Phi, \alpha, \beta, \sigma) = p_I(i|\mathbf{x}, \Phi, \alpha, \beta, \sigma)/|\Omega|$. The density p_R is defined in a similar manner. As [6], we extend the concept to the joint distributions as follows,

$$h_{I,R}(i, j, \mathbf{x}, \Phi, \alpha, \beta, \sigma) = (P(I(\mathbf{x}, \Phi, \sigma) - i, \beta)P(J(\mathbf{x}, \sigma) - j, \beta)) * W(\mathbf{x}, \alpha), \tag{11}$$

$$p_{I,R}(i, j | \mathbf{x}, \Phi, \alpha, \beta, \sigma) \simeq \frac{h_{I,R}(i, j, \Phi, \mathbf{x}, \alpha, \beta, \sigma)}{\int_{\Gamma^2} h_{I,R}(k, l, \mathbf{x}, \alpha, \beta, \sigma) dk \wedge dl}, \tag{12}$$

$$p_{I,R}(i, j | \Phi, \alpha, \beta, \sigma) = \frac{1}{|\Omega|} \int_{\Omega} p_{I,R}(i, j | \Phi, \mathbf{x}, \alpha, \beta, \sigma) d\mathbf{x}, \tag{13}$$

where we also have assumed (conditional) independence and uniformity such that $p_{I,R}(i, j, \mathbf{x} | \Phi, \alpha, \beta, \sigma) = p_{I,R}(i, j | \mathbf{x}, \Phi, \alpha, \beta, \sigma) / |\Omega|$.

3.1 First Order Structure

In order to use quasi-Newton methods for optimization, we need to derive the gradient of (II) w.r.t. the parameters of the uniform cubic b-spline, Φ . We use the notation of differentials, $dg(x) = Dg(x) dx$, where D is the partial derivative operator. Note that $d\mathbf{x}$ is a vector of differentials, not the wedge product of its elements, when used in relation to differentiation. Further, we will only write up non-zero terms that depend on $d\Phi$. The differential of (II) is,

$$d\mathcal{F} = d\mathcal{M} + d\mathcal{S}, \tag{14}$$

where arguments have been omitted for brevity. Ignoring the regularization term we focus on the differential of the similarity measure. The differential of (3) is,

$$d\mathcal{M}_{\text{NMI}} = \frac{(d\mathcal{H}_I + d\mathcal{H}_R)\mathcal{H}_{I,R} - (\mathcal{H}_I + \mathcal{H}_R)d\mathcal{H}_{I,R}}{\mathcal{H}_{I,R}^2}. \tag{15}$$

The entropy, \mathcal{H}_R , is independent of ϕ , hence $d\mathcal{H}_R = 0$. Further,

$$d\mathcal{H}_I = - \int_{\Gamma} dp_I (\log p_I + 1) di, \tag{16}$$

$$d\mathcal{H}_{I,R} = - \int_{\Gamma^2} dp_{I,R} (\log p_{I,R} + 1) di \wedge dj. \tag{17}$$

Using Leibniz integration rule, the differentials of the distributions are given as

$$dp_I(i, \Phi) = \frac{1}{|\Omega|} \int_{\Omega} dp_I(i | \mathbf{x}, \Phi) d\mathbf{x}, \tag{18}$$

$$dp_I(i | \mathbf{x}, \Phi) \simeq \frac{dh_I(i, \mathbf{x}, \Phi)}{\int_{\Gamma} h_I(j, \mathbf{x}, \Phi) dj} - \frac{h_I(i, \mathbf{x}, \Phi) \int_{\Gamma} dh_I(j, \mathbf{x}, \Phi) dj}{(\int_{\Gamma} h_I(j, \mathbf{x}, \Phi) dj)^2}, \tag{19}$$

$$dh_I(i, \mathbf{x}, \Phi) = (dP(I(\mathbf{x}, \Phi, \sigma) - i, \beta) * W(\mathbf{x}, \alpha)), \tag{20}$$

where irrelevant arguments have been omitted for brevity. Likewise, we have:

$$dp_{I,R}(i, j) = \frac{1}{|\Omega|} \int_{\Omega} dp_{I,R}(i, j|\mathbf{x}) d\mathbf{x}, \tag{21}$$

$$dp_{I,R}(i, j|\mathbf{x}) \simeq \frac{dh_{I,R}(i, j, \mathbf{x})}{\int_{\Gamma^2} h_{I,R}(k, l, \mathbf{x}) dk \wedge dl} - \frac{h_{I,R}(i, j, \mathbf{x}) \int_{\Gamma^2} dh_{I,R}(k, l, \mathbf{x}) dk \wedge dl}{\left(\int_{\Gamma^2} h_{I,R}(k, l, \mathbf{x}) dk \wedge dl\right)^2}, \tag{22}$$

$$dh_{I,R}(i, j, \mathbf{x}) = (dP(I(\boldsymbol{\psi}, \boldsymbol{\Phi}, \sigma) - i, \beta) P(J(\boldsymbol{\psi}, \sigma) - j, \beta)) * W(\mathbf{x} - \boldsymbol{\psi}, \alpha). \tag{23}$$

3.2 The Parzen Window (PW)

Originally MI was proposed in [21] using PW as density estimator. In the following we will examine this special case of LOI, often used in the literature. Consider (7) and let $\alpha \rightarrow \infty$. In that case, the window h_I simplifies as,

$$h_I(i, \mathbf{x}, \boldsymbol{\Phi}, \alpha, \beta, \sigma) \rightarrow \text{const.} \int_{\Omega} P(I(\boldsymbol{\psi}, \boldsymbol{\Phi}, \sigma) - i, \beta) d\boldsymbol{\psi}, \tag{24}$$

$$p_I(i|\boldsymbol{\Phi}, \alpha, \beta, \sigma) \rightarrow \frac{\int_{\Omega} P(I(\boldsymbol{\psi}, \boldsymbol{\Phi}, \sigma) - i, \beta) d\boldsymbol{\psi}}{\int_{\Gamma} \int_{\Omega} P(I(\boldsymbol{\psi}, \boldsymbol{\Phi}, \sigma) - j, \beta) d\boldsymbol{\psi} \wedge dj}. \tag{25}$$

Choosing

$$P(i, \beta) = e^{-i^2/(2\beta^2)}, \tag{26}$$

we find that

$$\int_{\Gamma} \int_{\Omega} P(I(\boldsymbol{\psi}, \boldsymbol{\Phi}, \sigma) - j, \beta) d\boldsymbol{\psi} \wedge dj = |\Omega| \sqrt{2\pi\beta^2}, \tag{27}$$

and

$$p_I(i|\boldsymbol{\Phi}, \alpha, \beta, \sigma) \rightarrow \frac{1}{|\Omega| \sqrt{2\pi\beta^2}} \int_{\Omega} e^{-\frac{(I(\mathbf{x}, \boldsymbol{\Phi}, \sigma) - i)^2}{2\beta^2}} d\mathbf{x}. \tag{28}$$

Likewise, we have

$$p_{I,R}(i, j|\boldsymbol{\Phi}, \alpha, \beta, \sigma) \rightarrow \frac{1}{|\Omega| 2\pi\beta^2} \int_{\Omega} e^{-\frac{(I(\mathbf{x}, \boldsymbol{\Phi}, \sigma) - i)^2 + (R(\mathbf{x}, \sigma) - j)^2}{2\beta^2}} d\mathbf{x}. \tag{29}$$

This is precisely the Parzen window method using a Gaussian kernel with infinite support [21]. Similar results are obtained for any integrable $P(i, \beta)$. The PW can be interpreted as a globally orderless image, as \mathbf{W} defining the locality extends globally. Further, since both (28) and (29) obey the diffusion equation w.r.t. $\beta^2/2$, we may use Green's theorem and write,

$$p_I(i|\sqrt{\beta_0^2 + \beta^2}) = p_I(i|\beta_0) * G(i, \beta), \tag{30}$$

$$p_{I,R}(i, j|\sqrt{\beta_0^2 + \beta^2}) = p_{I,R}(i, j|\beta_0) * G([i, j]^T, \beta), \tag{31}$$

for fast computation of a range of Parzen window sizes.

3.3 Generalized Partial Volume (GPV)

Shortly after the introduction of PW partial volume (PV) was introduced in [10] and extended to GPV in [3]. In the context of LOI GPV can be derived as follows: Considering (7), the change in local histogram for I may conveniently be considered in the original coordinate system through an inverse transformation of the integration window W . In general, for diffeomorphic transformations $x = \phi(\tilde{x})$,

$$f(x) * g(x) = \int_{\mathbb{R}} f(\alpha)g(x - \alpha) d\alpha = \int_{\mathbb{R}} f(\phi(\tilde{\alpha}))g(x - \phi(\tilde{\alpha}))|D_{\tilde{\alpha}}\phi(\tilde{\alpha})|d\tilde{\alpha} \quad (32)$$

$$\simeq f(\phi(\tilde{x})) * (|D_{\tilde{x}}\phi(\tilde{x})|g(\phi(\tilde{x}))), \quad (33)$$

where the approximation is good, when the determinant of the Jacobian of ϕ is approximately constant within the extend of g . Further,

$$d(f(x) * g(x)) \simeq f(\phi(\tilde{x})) * d(|D_{\tilde{x}}\phi(\tilde{x})|g(\phi(\tilde{x}))) \quad (34)$$

$$\simeq \tilde{f}(\tilde{x}) * (|D_{\tilde{x}}\phi(\tilde{x})|D_{\phi}g(\phi(\tilde{x}))d\phi) \quad (35)$$

w.r.t. the transformation parameters, and writing $\tilde{f} = f \circ \phi$. Thus,

$$dh_I = d(P(I(\mathbf{x}, \Phi, \sigma) - i, \beta) * W(\mathbf{x}, \alpha)) \quad (36)$$

$$\simeq P(I(\tilde{\mathbf{x}}, \Phi, \sigma) - i, \beta) * (|D_{\tilde{\mathbf{x}}}\phi(\tilde{\mathbf{x}})|D_{\mathbf{x}}W(\mathbf{x}, \alpha)d\phi), \quad (37)$$

and

$$d\phi = \left[\prod_{n=1}^N f_1(\tilde{x}_n), \dots, \prod_{n=1}^N f_M(\tilde{x}_n) \right] d\text{vec}(\Phi), \quad (38)$$

where $\text{vec}(\Phi) = [\phi(\tilde{\mathbf{x}}_1)^T, \dots, \phi(\tilde{\mathbf{x}}_M)^T]^T$.

For the joint histograms we have a similar but approximated result by the Hölder inequality,

$$\begin{aligned} h_{I,R}(i, j, \mathbf{x}, \alpha, \beta, \sigma) &= \int_{\Omega} P(I(\psi, \sigma) - i, \beta)P(J(\psi, \sigma) - j, \beta)W(\mathbf{x} - \psi, \alpha) d\psi \quad (39) \end{aligned}$$

$$\leq \int_{\Omega} P(I(\psi, \sigma) - i, \beta)W(\mathbf{x} - \psi, \alpha) d\psi \int_{\Omega} P(J(\psi, \sigma) - j, \beta) d\psi \quad (40)$$

$$\begin{aligned} &= \int_{\Omega} P(I(\phi(\tilde{\psi}), \sigma) - i, \beta)W(\mathbf{x} - \phi(\tilde{\psi}), \alpha) \left| D_{\tilde{\psi}}\phi(\tilde{\psi}) \right| d\tilde{\psi} \\ &\quad \int_{\Omega} P(J(\psi, \sigma) - j, \beta) d\psi \quad (41) \end{aligned}$$

$$= ((P(I(\phi(\tilde{\mathbf{x}}), \sigma) - i, \beta) |D_{\tilde{\mathbf{x}}}\phi(\tilde{\mathbf{x}})|) * W(\phi(\tilde{\mathbf{x}}), \alpha)) \int_{\Omega} P(J(\psi, \sigma) - j, \beta) d\psi. \quad (42)$$

Hence, the differential w.r.t. the transformation and in the original coordinates is,

$$\begin{aligned}
 dh_{I,R}(i, j, \mathbf{x}, \alpha, \beta, \sigma) & \leq ((P(I(\phi(\tilde{\mathbf{x}}), \sigma) - i, \beta) |D_{\tilde{\mathbf{x}}}\phi(\tilde{\mathbf{x}})|) * (D_{\phi}W(\phi(\tilde{\mathbf{x}}), \alpha) d\phi)) \\
 & \int_{\Omega} P(J(\psi, \sigma) - j, \beta) d\psi, \tag{43}
 \end{aligned}$$

and equivalently

$$\begin{aligned}
 D_{\text{vec}(\Phi)}h_{I,R}(i, j, \mathbf{x}, \alpha, \beta, \sigma) & \leq \left((P(I(\phi(\tilde{\mathbf{x}}), \sigma) - i, \beta) |D_{\tilde{\mathbf{x}}}\phi(\tilde{\mathbf{x}})|) \right. \\
 & \left. * (D_{\phi}W(\phi(\tilde{\mathbf{x}}), \alpha)D_{\Phi}\mathbf{x}) \right) \int_{\Omega} P(J(\psi, \sigma) - j, \beta) d\psi. \tag{44}
 \end{aligned}$$

Set $P(I(\psi, \Phi, \sigma) - i, \beta)$ to a boxcar function

$$P(I(\psi, \Phi, \sigma) - i, \beta) = \begin{cases} 1 & \text{if } -\frac{\beta}{2} \leq \tilde{I}(\psi, \Phi, \sigma) - i < \frac{\beta}{2}, \\ 0 & \text{otherwise,} \end{cases} \tag{45}$$

where β is chosen such that $I(\psi, \Phi, \sigma)$ is mapped into non-coinciding isophotes curves. The motivation for this is that all isophotes can be evaluated at \tilde{x} simultaneously. This is the generalized partial volume (GPV) scheme when integrating over the entire domain Ω . Thus GPV is small local histograms integrated to form the globally orderless image as in the PW approach.

4 Fast Implementations

This section presents a complexity analysis of the compact pseudo code required to implement SSD, PW and GPV. The purpose is to show that these measures can be implemented as very fast algorithms with only slightly more computations than SSD. We compare the computational complexity to SSD using identically interpolation scheme. All kernels used in implementations are 3rd order uniform B-splines as well as boxcar functions to reduce computational complexity. The code assumes 3D images, the use cubic B-splines for all kernels and K bins in the histograms. The pseudo code can be seen in Figure 1 including the analysis of the computational complexity. We assume that today’s processors have equal processing time of, e.g., sum, log, sin etc. From the pseudo code in Figure 1 and the complexity we see that the Parzen window can be considerable larger without being inferior to GPV in performance. This is of cause a rough estimate and implementations may vary, but the amount of computations for NMI using either GPV or PW are comparable in computational complexity to SSD using B-splines. W.r.t. memory, GPV requires $192 \times N \times 8$ bytes of memory to obtain the speed, where the PW only requires $8 \times N \times 8$ bytes (on 64-bit, double precision).

```

# Given 2 images, I and R, and the determinant of the
# transformation, det, as a function of space,
# calculate PW and GPV for NMI and SSD, based on N
# image evaluation points, and K marginal and K^2 joint
# histogram bins. Flops are based on cubic splines

FOR N evaluation points
  calculate image spline coeff. (60 flops)
  IF(SSD || PW)
    calculate derivative of image spline coeff. (48 flops)
  FOR 64 combinations of image spline coeff.
    IF(SSD || PW)
      update image at evaluation point (4 flops)
      update image gradient at evaluation point (12 flops)
    IF(GPV)
      update histograms (4 flops)
  IF(SSD)
    update residual (2 flops)
  IF(PW)
    calculate histogram spline coeff. (20 flops)
  FOR 16 histogram spline coeff.
    update histograms (2 flops)
IF(PW || GPV)
  calculate NMI and derivative on histograms (9*K^2+6K flops)
FOR N evaluation points
  IF(GPV)
    calculate derivative of image spline coeff. (48 flops)
  FOR 64 combinations of image spline coeff.
    update derivative of histogram (16 flops)
  IF(PW)
    FOR 16 histogram spline coeff.
      update derivative of histogram (9 flops)
  update derivatives (3 flops)
# Total flop usage:
#   SSD: 1134N flops
#   PW: 1331N +9K^2 +6K flops
#   GPV: 1383N +9K^2 +6K flops

```

Fig. 1. Pseudo code for SSD, PW and GPV

5 Experiments

For ease of comparison we restrict ourselves primarily to rigid registration examples with a few non-rigid included as well. It has previously been shown that PW is superior to GPV in accuracy, and thus we quantitatively evaluate the difference in estimated parameters with the prerequisites listed in Section 4. As the results in Table 1 show, there are more or less pronounced differences numerically between the two methods. In some of the cases both PW or GPV got

Table 1. The difference in registration results using PW and GPV. The table contains Inter and intra subject rigid registration difference and standard deviation in pixels for T1-weighted MRI brain data and cardiac CINE-MRI data. Furthermore the difference in parameter estimates over 1800 parameters in 18 non-rigid intra patient T1-weighted MRI registrations in voxels. As seen the two methods produces significantly different results especially between subjects. This originates from the smoothing in different domains and the use of the Hölder inequality. The differences are measured in voxels and radians.

param	rot x	rot y	rot z	tx	ty x	tz
Between subject (Brain)						
mean	0.0236	-0.0420	-0.0195	2.1120	1.7486	-0.4984
std	0.0556	0.2689	0.0781	7.5876	3.3922	1.6482
Intra subject (Brain)						
mean	0.0005	0.0001	-0.0004	-0.1304	-0.0326	0.1685
std	0.0016	0.0032	0.0008	0.7604	0.0825	0.7033
Intra subject (Cardiac)						
mean	0.0005	-0.0002	-0.0007	0.0629	0.0473	0.0142
std	0.0006	0.0005	0.0009	0.0883	0.0729	0.0406
Intra subjects non-rigid (Brain)						
	Difference					
mean	-0.0182					
std	2.446					

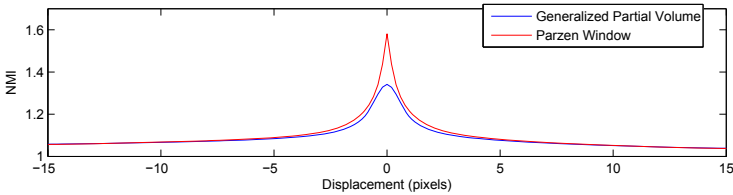


Fig. 2. The difference in NMI estimate for PW and GPW evaluated for 1 dimensional translational displacement. The observed difference originates from the smoothing in different domains and the use of the Hölder inequality.

stuck in a local minima often simultaneously but with different minimas. Such problem can be solved using scale space.

6 Discussion and Conclusion

The differences between the PW and GPV can be explained from a theoretical point of view: From (29) it is clear that GPV and PW smooth in 2 different spaces, thus the comparisons are interesting computation wise but less meaningful w.r.t. accuracy. Nevertheless, the approximation made by GPV substantiate that PW as reported in [8] is more accurate than GPV. In addition The Hölder inequality also accounts for some of the difference in value of the functional as

does the choice of Parzen window. Our results also indicate that smoothing in the isophote domain \mathbf{W} influences the objective function more than smoothing in the image domain I . The exact effects of smoothing in the 3 domains is left for future investigations.

From (32) we see that the Jacobian locally scales the integration compensating for the deformation. In some schemes this is omitted, reformulating the transformation to a re-sampling. The choice of approach should be based on the specific application; however, for density distributions the Jacobian should be present. The local scaling can cause instability in non-rigid, non-regularized settings with entropy based measures, as it tends to concentrate the mass at a single bin in the histogram, a global optimum for both MI and NMI. This effect is particularly pronounced in GPV, whereas the PW diffuses this effect over several bins in the \mathbf{R} direction and not only the \mathbf{I} direction. Some schemes overcome the missing Jacobian implicitly by composing small deformations and re-sampling.

To conclude, PW operates directly on intensity distributions and is therefore the natural choice for information theoretical measures such as NMI. In this paper we have shown that GPV and PW are special cases of LOI extended to joint probability distributions. From the perspective of LOI we have derived the PW and GPV, and shown how GPV makes a series of approximations and use of some special kernels with locality assumptions to achieve its low computational cost. The PW only rely on transforming LOI to globally orderless images, whereas the GPV rely approximations using the Hölder inequality to achieve computability, the use of a boxcar window as Parzen estimator and no image interpolation as a part of the model. The global entropy comes from integration over the entire domain. The approximations made in GPV along with smoothing in different spaces account for the difference in functional and in value of similarity measure makes PW far more attractive theoretically as well as in practice. In addition we have shown that PW is faster than the GPV and we substantiated the result from [8] that PW is more accurate. We therefore conclude that PW is a superior estimator for NMI compared to GPV.

References

1. Andreopoulos, A., Tsotsos, J.: Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. *Medical Image Analysis* 12(3), 335–357 (2008)
2. Boor, C.d.: *A Practical Guide to Splines*. Springer, Heidelberg (1978)
3. Chen, H., Varshney, P.: Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. *IEEE Transactions on Medical Imaging* 22(9), 1111–1119 (2003)
4. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multi-modality image registration based on information theory. In: *Information Processing in Medical Imaging*, pp. 263–274 (1995)
5. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multi-modal images. *Methods of information in medicine* 46(3), 292–299 (2007)

6. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. *International Journal of Computer Vision* 50(3), 329–343 (2002)
7. Koenderink, J., Van Doorn, A.: The structure of locally orderless images. *International Journal of Computer Vision* 31(2), 159–168 (1999)
8. Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Comparison between parzen window interpolation and generalised partial volume estimation for nonrigid image registration using mutual information. In: *Biomedical Image Registration*, pp. 206–213 (2006)
9. Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007. LNCS*, vol. 4584, pp. 725–737. Springer, Heidelberg (2007)
10. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* 16(2), 187–198 (1997)
11. Maes, F., Vandermeulen, D., Suetens, P.: Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis* 3(4), 373–386 (1999)
12. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience* 19(9), 1498–1507 (2007)
13. Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S.: Fast free-form deformation using graphics processing units. In: *Computer Methods and Programs in Biomedicine* (2009)
14. Modersitski, J.: *Numerical Methods for Image Registration*. Oxford University Press, Oxford (2004)
15. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian Elasticity: A Statistical Regularization Framework for Non-linear Registration. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005. LNCS*, vol. 3750, pp. 943–950. Springer, Heidelberg (2005)
16. Pluim, J., Antoine Maintz, J., Viergever, M.: Interpolation artefacts in mutual information-based image registration. *Computer vision and image understanding* 77(2), 211–232 (2000)
17. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging* 22(8), 986–1004 (2003)
18. Shannon, C.: A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423, 623–656 (1948)
19. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* 32(1), 71–86 (1999)
20. Thevenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing* 9(12), 2083–2099 (2000)
21. Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis* 1(1), 35–51 (1996)

Large Deformation Diffeomorphic Metric Mapping of Orientation Distribution Functions

Jia Du¹, Alvina Goh², and Anqi Qiu^{1,3}

¹ Division of Bioengineering, National University of Singapore, Singapore

² Department of Mathematics, National University of Singapore, Singapore

³ Clinical Imaging Research Center, National University of Singapore, Singapore

Abstract. We propose a novel large deformation diffeomorphic registration algorithm to align high angular resolution diffusion images (HARDI) characterized by Orientation Distribution Functions (ODF). Our proposed algorithm seeks an optimal diffeomorphism of large deformation between two ODF fields in a spatial volume domain and at the same time, locally reorients an ODF in a manner such that it remains consistent with the surrounding anatomical structure. We first extend ODFs traditionally defined in a unit sphere to a generalized ODF defined in \mathbb{R}^3 . This makes it easy for an affine transformation as well as a diffeomorphic group action to be applied on the ODF. We then construct a Riemannian space of the generalized ODFs and incorporate its Riemannian metric for the similarity of ODFs into a variational problem defined under the large deformation diffeomorphic metric mapping (LDDMM) framework. We finally derive the gradient of the cost function in both Riemannian spaces of diffeomorphisms and the generalized ODFs, and present its numerical implementation. Both synthetic and real brain HARDI data are used to illustrate the performance of our registration algorithm.

Keywords: Diffeomorphic mapping, HARDI, Riemannian manifold of ODFs.

1 Introduction

Diffusion weighted magnetic resonance imaging (DW-MRI) is a unique *in vivo* imaging technique that allows us to visualize the three-dimensional architecture of neural fiber pathways in the human brain. Several techniques may be used to reconstruct the local orientation of brain tissue from DW-MRI data. Diffusion tensor imaging (DTI) [5], which characterizes the diffusivity profile by a single oriented 3D Gaussian probability distribution function (PDF), has been demonstrated to be valuable for studying brain white matter development in children and detecting abnormalities in patients with neuropsychiatric disorders and neurodegenerative diseases. Nevertheless, one shortcoming of DTI is that it can only reveal one dominant fiber orientation at each location, when between one and two thirds of the voxels in the human brain white matter are thought to contain multiple fiber bundle crossings [6]. High angular resolution diffusion

imaging (HARDI) [25] addresses this well-known limitation of DTI. HARDI measures diffusion along n uniformly distributed directions on the sphere and can characterize more complex fiber geometries. Several reconstruction techniques can be used to characterize diffusion based on the HARDI signals. One class is based on higher-order tensors [3,15] and leverage prior work on DTI. Another method is Q-ball Imaging, which uses the Funk-Radon transform to reconstruct the orientation distribution function (ODF). The model-free ODF is the angular profile of the diffusion PDF of water molecules and has been approximated using different sets of basis functions such as spherical harmonics (SH) [22,11,13,18,1]. Such methods are relatively fast to implement because the ODF is computed analytically. By quantitatively comparing fiber orientations retrieved from ODFs against histological measurements, [20] shows that accurate fiber estimates can be obtained from HARDI data, further validating its usage in brain studies.

An open challenge in the analysis of mathematically complex HARDI data is registration. Unlike scalar-based image registration algorithms, HARDI registration algorithms need to define two key transformations: a transformation to spatially align anatomical structures between two brains in a 3D volume domain, and a transformation to align the local diffusivity defined at each voxel of two brains. Several HARDI registration algorithms have been recently proposed under a specific model of local diffusivity. [9] takes the approach of representing HARDI by Gaussian mixture fields (GMF) and assumes a thin-plate spline deformation. The L^2 metric of GMFs is minimized, and reorientation is performed on the individual Gaussian components, each representing a major fiber direction. [4] uses a 4th order tensor model and assumes a region-based nonrigid deformation. The rotationally invariant Hellinger distance is considered and an affine tensor reorientation, which accounts for rotation, scaling and shearing effects, is applied. [14] performs a diffeomorphic registration is performed with the L^2 metric on ODFs represented by spherical harmonics. Reorientation is done by altering the SH coefficients in a manner similar to the “Finite Strain (FS)” method in DTI where only the rotation is extracted and applied. [7] performs alignment of ODF fields by using a multi-channel diffeomorphic demons registration algorithm on rotationally invariant feature maps and uses the FS scheme in reorientation. [26] uses the SH-based ODF representation and proposes a hierarchical registration scheme, where descriptors are extracted at each level and the alignment is updated by using features extracted from the increasing order of the SH representation. Reorientation is done by tilting the gradient directions via multiplying with the local affine transform and normalizing.

Paper Contributions. We present a novel registration algorithm for HARDI data represented by ODFs. Previous studies [21] suggest that the transformation from one brain to another can be really large and therefore small deformation models may not be enough. Therefore, we define our algorithm in the setting of large deformation diffeomorphic metric mapping (LDDMM). Our proposed algorithm seeks an optimal diffeomorphism of large deformation between two ODF fields across a spatial volume domain and at the same time, locally reorients an ODF in a manner that it remains consistent with the surrounding anatomical

structure. To do so, we first extend ODFs traditionally defined in a unit sphere \mathbb{S}^2 to a generalized ODF defined in \mathbb{R}^3 . The generalized ODF makes it easy to apply an affine transformation and a diffeomorphic group action. We construct a Riemannian space of the generalized ODFs, similar to that of traditional ODFs proposed in [17,10]. The Riemannian metric for the similarity of ODFs is then incorporated into a variational problem in LDDMM. Finally, we derive the gradient of the cost function in both Riemannian spaces of diffeomorphisms and the generalized ODFs and present its numerical implementation.

2 Methods

2.1 The Riemannian Manifold of ODFs Extended to \mathbb{R}^3

As mentioned previously, HARDI measurements can be used to reconstruct the orientation distribution function (ODF), the angular profile of the diffusion PDF of water molecules. The ODF is actually a PDF defined on the 2-sphere \mathbb{S}^2 (of radius 1) and its space is defined as

$$\mathcal{P} = \{\mathbf{p} : \mathbb{S}^2 \rightarrow \mathbb{R}^+ | \forall \mathbf{s} \in \mathbb{S}^2, \mathbf{p}(\mathbf{s}) \geq 0; \int_{\mathbf{s} \in \mathbb{S}^2} \mathbf{p}(\mathbf{s}) d\mathbf{s} = 1\},$$

where the gradient direction \mathbf{s} is $(\theta, \varphi) \in \mathbb{S}^2, \theta \in [0, \pi], \varphi \in [0, 2\pi]$ expressed in spherical coordinates. For the convenient computation of affine transformation on this ODF (see details in §2.2), we generalize \mathbf{p} to a PDF in \mathbb{R}^3 , as follows,

$$\mathbf{q}(\mathbf{s}) = \mathbf{q}(r, \theta, \varphi) = \frac{\mathbf{p}(\theta, \varphi)}{r^2} \delta(r - 1), \quad (1)$$

where the gradient direction \mathbf{s} is $(r, \theta, \varphi) \in \mathbb{R}^3, r \in [0, \infty), \theta \in [0, \pi], \varphi \in [0, 2\pi]$ expressed in spherical coordinates. We refer to \mathbf{q} as a *generalized ODF* (gODF) throughout this paper. The space of \mathbf{q} can be defined as

$$\mathcal{Q} = \{\mathbf{q} : \mathbb{R}^3 \rightarrow \mathbb{R}^+ | \forall \mathbf{s} \in \mathbb{R}^3, \mathbf{q}(\mathbf{s}) \geq 0; \int_{\mathbf{s} \in \mathbb{R}^3} \mathbf{q}(\mathbf{s}) d\mathbf{s} = 1\}. \quad (2)$$

It is well-known from the field of *information geometry* [2] that the space of PDFs forms a Riemannian manifold, also known as the statistical manifold. [23] introduces the Riemannian structure formed by the statistical manifold whose elements are probability density functions and shows that the *Fisher-Rao* metric determines a Riemannian metric. [8] shows that the Fisher-Rao metric is the *unique intrinsic metric* on the statistical manifold \mathcal{Q} , therefore it is invariant to re-parameterizations of the functions. There are many different parameterizations of PDFs that are equivalent but with different forms of the Fisher-Rao metric, leading to the Riemannian operations having different computational complexity. In this paper, we choose the square-root representation, which is used recently in ODF processing [17,10]. The square-root representation is one of the most efficient representations found to date as the various Riemannian

operations such as geodesics, exponential maps, logarithm maps are available in closed form.

The *generalized square-root ODF* ($\sqrt{g\text{ODF}}$) is defined as $\psi(\mathbf{s}) = \sqrt{q(\mathbf{s})}$, where $\psi(\mathbf{s})$ is assumed to be non-negative to ensure uniqueness. Again, the domain of $\psi(\mathbf{s})$ is \mathbb{R}^3 , unlike in [17,10] where \mathbb{S}^2 is used. The space of such functions is defined as

$$\Psi = \{ \psi : \mathbb{R}^3 \rightarrow \mathbb{R}^+ \mid \forall \mathbf{s} \in \mathbb{R}^3, \psi(\mathbf{s}) \geq 0; \int_{\mathbf{s} \in \mathbb{R}^3} \psi^2(\mathbf{s}) d\mathbf{s} = 1 \}. \tag{3}$$

We see that from Eq. (3), the functions ψ lie on the positive orthant of a unit Hilbert sphere, a well-studied Riemannian manifold. It can be shown [24] that the Fisher-Rao metric is simply the \mathbb{L}^2 metric, given as

$$\langle \xi_j, \xi_k \rangle_{\psi_i} = \int_{\mathbf{s} \in \mathbb{R}^3} \xi_j(\mathbf{s}) \xi_k(\mathbf{s}) d\mathbf{s}, \tag{4}$$

where $\xi_j, \xi_k \in T_{\psi_i} \Psi$ are tangent vectors of ψ at ψ_i . The geodesic distance between any two functions $\psi_i, \psi_j \in \Psi$ on a unit Hilbert sphere is the angle

$$\text{dist}(\psi_i, \psi_j) = \|\log_{\psi_i}(\psi_j)\|_{\psi_i} = \cos^{-1} \langle \psi_i, \psi_j \rangle = \cos^{-1} \left(\int_{\mathbf{s} \in \mathbb{R}^3} \psi_i(\mathbf{s}) \psi_j(\mathbf{s}) d\mathbf{s} \right), \tag{5}$$

where $\langle \cdot, \cdot \rangle$ is the normal dot product between points in the sphere under the \mathbb{L}^2 metric. For the sphere, the *exponential map* has the closed-form formula

$$\exp_{\psi_i}(\xi) = \cos(\|\xi\|_{\psi_i}) \psi_i + \sin(\|\xi\|_{\psi_i}) \frac{\xi}{\|\xi\|_{\psi_i}}, \tag{6}$$

where $\xi \in T_{\psi_i} \Psi$ is a tangent vector at ψ_i and $\|\xi\|_{\psi_i} = \sqrt{\langle \xi, \xi \rangle_{\psi_i}}$. By restricting $\|\xi\|_{\psi_i} \in [0, \frac{\pi}{2}]$, we ensure that the exponential map is bijective. The *logarithm map* from ψ_i to ψ_j has the closed-form formula

$$\overrightarrow{\psi_i \psi_j} = \log_{\psi_i}(\psi_j) = \frac{\psi_j - \langle \psi_i, \psi_j \rangle \psi_i}{\sqrt{1 - \langle \psi_i, \psi_j \rangle^2}} \cos^{-1} \langle \psi_i, \psi_j \rangle. \tag{7}$$

2.2 Affine Transformation on Generalized Square-Root ODFs

In this section, we discuss the reorientation of the $\sqrt{g\text{ODF}}$, $\psi(\mathbf{s})$, when an affine transformation A is applied. We denote the transformed $\sqrt{g\text{ODF}}$ as $\widehat{\psi}(\widehat{\mathbf{s}}) = A\psi(\mathbf{s})$, indicating a change in magnitude of ψ as well as gradient directions of \mathbf{s} , due to the affine transformation. In addition, we denote $\widehat{\mathbf{s}}$ as $(\widehat{r}, \widehat{\theta}, \widehat{\varphi})$ in spherical coordinates and $(\widehat{r} \sin \widehat{\theta} \cos \widehat{\varphi}, \widehat{r} \sin \widehat{\theta} \sin \widehat{\varphi}, \widehat{r} \cos \widehat{\theta})$ in Cartesian coordinates, and \mathbf{s} as (r, θ, φ) in spherical coordinates and $(r \sin \theta \cos \varphi, r \sin \theta \sin \varphi, r \cos \theta)$ in Cartesian coordinates. We assume that the change of the gradient directions due to affine transformation A is

$$(\widehat{r} \sin \widehat{\theta} \cos \widehat{\varphi}, \widehat{r} \sin \widehat{\theta} \sin \widehat{\varphi}, \widehat{r} \cos \widehat{\theta}) = A^{-1} (r \sin \theta \cos \varphi, r \sin \theta \sin \varphi, r \cos \theta),$$

alternatively, $\widehat{\mathbf{s}} = A^{-1} \mathbf{s}.$ (8)

The change of the magnitude due to A is denoted as $\widehat{\psi}(\cdot) = A\psi(\cdot)$. Similar to [19], we assume that the volume fraction of fibers oriented toward a small patch must remain the same after the patch is transformed. This ensures that the transformed ODF remains consistent with the surrounding anatomical structure. Therefore, the following property for gODFs $(\psi(\cdot))^2$ and $(\widehat{\psi}(\cdot))^2 = (A\psi(\cdot))^2$ must hold

$$(A\psi(r, \theta, \varphi))^2 r^2 \sin(\theta) d\theta d\varphi = (\psi(\widehat{r}, \widehat{\theta}, \widehat{\varphi}))^2 \widehat{r}^2 \sin(\widehat{\theta}) d\widehat{\theta} d\widehat{\varphi}. \tag{9}$$

As a consequence of this assumption, it can be shown that the following theorem is true (we omit the proof for the sake of brevity).

Theorem 1. Reorientation of ψ . *Let $A\psi(\mathbf{s})$ be the result of an affine transformation A acting on a $\sqrt{\text{gODF}}$ $\psi(\mathbf{s})$. The following analytical equation holds true,*

$$A\psi(\mathbf{s}) = \sqrt{\frac{\det A^{-1} \|\mathbf{s}\|}{\|A^{-1}\mathbf{s}\|}} \psi(A^{-1}\mathbf{s}), \tag{10}$$

where $\|\cdot\|$ is the norm of a vector. With a slight abuse of notation, we assume

$$\psi(A^{-1}\mathbf{s}) = \sqrt{\frac{\mathbf{p}(\widehat{\theta}, \widehat{\varphi})}{\widehat{r}^2}} \delta(r - 1)$$

such that Eq. (10) is the action of the affine transformation on the $\sqrt{\text{gODF}}$ defined on a unit sphere.

By construction, $A\psi(\mathbf{s})$ fulfills the definition of the $\sqrt{\text{gODF}}$.

2.3 Diffeomorphic Group Action on Generalized Square-Root ODF

We have shown in §2.2 how to reorient a $\sqrt{\text{gODF}}$ ψ located at a fixed spatial position x in the image volume $\Omega \subset \mathbb{R}^3$ when it undergoes an affine transformation. In this section, we define an action of diffeomorphisms $\phi : \Omega \rightarrow \Omega$ on the $\sqrt{\text{gODF}}$, which takes into consideration the reorientation of ψ as well as the transformation of the spatial volume in $\Omega \subset \mathbb{R}^3$. We denote $\psi(\mathbf{s}, x)$ as the $\sqrt{\text{gODF}}$ with the gradient direction $\mathbf{s} \in \mathbb{R}^3$ located at $x \in \Omega$. We define the action of diffeomorphisms on $\psi(\mathbf{s}, x)$ in the form of

$$\phi \cdot \psi(\mathbf{s}, x) = A_{\phi^{-1}(x)} \psi(\mathbf{s}, \phi^{-1}(x)),$$

where the local affine transformation A_x at spatial coordinates x is defined as the Jacobian matrix of ϕ evaluated at x , i.e., $A_x = D_x \phi$. According to Eq. (10), the action of diffeomorphisms on $\psi(\mathbf{s}, x)$ can be computed as

$$\phi \cdot \psi(\mathbf{s}, x) = \sqrt{\frac{\det (D_{\phi^{-1}\phi})^{-1} \|\mathbf{s}\|}{\| (D_{\phi^{-1}\phi})^{-1} \mathbf{s} \|}} \psi((D_{\phi^{-1}\phi})^{-1} \mathbf{s}, \phi^{-1}(x)).$$

For the remainder of this paper, for the sake of simplicity, we denote $\phi \cdot \psi(\mathbf{s}, x)$ as

$$\phi \cdot \psi(\mathbf{s}, x) = A\psi \circ \phi^{-1}(x). \tag{11}$$

For the purpose of HARDI registration, in order to quantify of the similarity of $\sqrt{\text{gODF}} \phi \cdot \psi(\mathbf{s}, x)$ to other $\sqrt{\text{gODF}}$ s, we use the Riemannian distance given in §2.1

2.4 Large Deformation Diffeomorphic Metric Mapping for ODFs

The previous sections equip us with an appropriate representation of the ODF and its diffeomorphic action. Now, we state a variational problem for mapping ODFs from one subject volume to another. We define this problem in the “large deformation” setting of Grenander’s group action approach for modeling shapes, that is, ODF volumes are modeled by assuming that they can be generated from one to another via flows of diffeomorphisms ϕ_t , which are solutions of ordinary differential equations $\dot{\phi}_t = v_t(\phi_t), t \in [0, 1]$, starting from the identity map $\phi_0 = \text{Id}$. They are therefore characterized by time-dependent velocity vector fields $v_t, t \in [0, 1]$. We define a metric distance between a target volume ψ_{targ} and a template volume ψ_{temp} as the minimal length of curves $\phi_t \cdot \psi_{\text{temp}}, t \in [0, 1]$, in a shape space such that, at time $t = 1, \phi_1 \cdot \psi_{\text{temp}} = \psi_{\text{targ}}$. Lengths of such curves are computed as the integrated norm $\|v_t\|_V$ of the vector field generating the transformation, where $v_t \in V$, where V is a reproducing kernel Hilbert space with kernel k_V and norm $\|\cdot\|_V$. To ensure solutions are diffeomorphisms, V must be a space of smooth vector fields [12]. Using the duality isometry in Hilbert spaces, one can equivalently express the lengths in terms of m_t , interpreted as momentum such that for each $u \in V, \langle m_t, u \circ \phi_t \rangle_2 = \langle k_V^{-1}v_t, u \rangle_2$, where we let $\langle m, u \rangle_2$ denote the \mathbb{L}^2 inner product between m and u , but also, with a slight abuse, the result of the natural pairing between m and v in cases where m is singular (e.g., a measure). This identity is classically written as $\phi_t^*m_t = k_V^{-1}v_t$, where ϕ_t^* is referred to as the pullback operation on a vector measure, m_t . Using the identity $\|v_t\|_V^2 = \langle k_V^{-1}v_t, v_t \rangle_2 = \langle m_t, k_V m_t \rangle_2$ and the standard fact that energy-minimizing curves coincide with constant-speed length-minimizing curves, one can obtain the metric distance between the template and target $\sqrt{\text{gODF}}$ volumes, $\rho(\psi_{\text{temp}}, \psi_{\text{targ}})$, by minimizing $\int_0^1 \langle m_t, k_V m_t \rangle_2 dt$ such that $\phi_1 \cdot \psi_{\text{temp}} = \psi_{\text{targ}}$ at time $t = 1$. We associate this with the variational problem in the form of

$$J(m_t) = \inf_{m_t: \dot{\phi}_t = k_V m_t(\phi_t), \phi_0 = \text{Id}} \rho(\psi_{\text{temp}}, \psi_{\text{targ}})^2 + \lambda \int_{x \in \Omega} E_x(\phi_1 \cdot \psi_{\text{temp}}(\mathbf{s}, x), \psi_{\text{targ}}(\mathbf{s}, x)) dx \tag{12}$$

with E_x as the metric distance between the deformed $\sqrt{\text{gODF}}$ template, $\phi_1 \cdot \psi_{\text{temp}}(\mathbf{s}, x)$, and the target, $\psi_{\text{targ}}(\mathbf{s}, x)$. We use the Riemannian metric given in §2.1 and rewrite Eq. (12) as

$$J(m_t) = \inf_{m_t: \dot{\phi}_t = k_V m_t(\phi_t), \phi_0 = \text{Id}} \int_0^1 \langle m_t, k_V m_t \rangle_2 dt + \lambda \int_{x \in \Omega} \|\log_{A\psi_{\text{temp}} \circ \phi_1^{-1}(x)}(\psi_{\text{targ}}(x))\|_{A\psi_{\text{temp}} \circ \phi_1^{-1}(x)}^2 dx, \tag{13}$$

where $A = D\phi_1$, the Jacobian of ϕ_1 . For the sake of simplicity, we denote $\psi_{\text{targ}}(\mathbf{s}, x)$ as $\psi_{\text{targ}}(x)$. Note that since we are dealing with vector fields in \mathbb{R}^3 , the kernel of V is a matrix kernel operator in order to get a proper definition. We define this kernel as $k_V \text{Id}_{3 \times 3}$, where $\text{Id}_{3 \times 3}$ is an identity matrix, such that k_V can be a scalar kernel.

2.5 Gradient of J with Respect to m_t

The gradient of J with respect to m_t can be computed via studying a variation $m_t^\epsilon = m_t + \epsilon \tilde{m}_t$ on J such that the derivative of J with respect to ϵ is expressed in function of \tilde{m}_t . For the sake of simplicity, we directly give the expression of **the gradient of J with respect to m_t** as

$$\nabla J(m_t) = 2m_t + \lambda \eta_t, \tag{14}$$

where

$$\eta_t = \nabla_{\phi_1} E + \int_t^1 [\partial_{\phi_s}(k_V m_s)]^\top (\eta_s + m_s) ds, \tag{15}$$

where $E = \int_{x \in \Omega} E_x dx$. Eq. (15) can be solved backward given $\eta_1 = \nabla_{\phi_1} E$. $\partial_{\phi_s}(k_V m_s)$ is the partial derivative of $k_V m_s$ with respect to ϕ_s . $\nabla J(m_t)$ can be directly obtained by following the derivation in [16]. However, the term $\nabla_{\phi_1} E$ in $\nabla J(m_t)$ is not easy to compute and we will now discuss this in detail.

Gradient of E with respect to ϕ_1 : We now need to compute $\nabla_{\phi_1} E$, which is not straightforward and has to be computed by incorporating the Riemannian structure of ODFs. Let's first compute $\nabla_{\phi_1} E_x$ at a fixed location, x . Taking a variation $\phi_1^\epsilon = \phi_1 + \epsilon h$ of ϕ_1 , by chain rule, we get

$$\begin{aligned} \partial_\epsilon E_x|_{\epsilon=0} = & 2 \left\langle \log_{A\psi_{\text{temp}} \circ \phi_1^{-1}}(x) \psi_{\text{targ}}(x), \nabla_x (A\psi_{\text{temp}}) \circ \phi_1^{-1} \right\rangle_{A\psi_{\text{temp}}(x)}, \\ & \left. \frac{\partial(\phi_1 + \epsilon h)^{-1}(x)}{\partial \epsilon} \right|_{\epsilon=0} \Bigg\rangle_2 \end{aligned}$$

where

$$\frac{\partial(\phi_1 + \epsilon h)^{-1}(x)}{\partial \epsilon} \Big|_{\epsilon=0} = - [(D\phi_1)^{-1}h] \circ \phi_1^{-1}(x)$$

. Since $A\psi_{\text{temp}}$ also lies in the Riemannian manifold of $\sqrt{g\text{ODFs}}$, $\nabla_x (A\psi_{\text{temp}})$ are logarithm maps of $A\psi_{\text{temp}}$ defined as

$$\nabla_x [A\psi_{\text{temp}}(x)] = \begin{bmatrix} \frac{1}{|\Delta e_1|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_1) \\ \frac{1}{|\Delta e_2|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_2) \\ \frac{1}{|\Delta e_3|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_3) \end{bmatrix},$$

where $\Delta e_1, \Delta e_2$ and Δe_3 indicate small variations in three orthonormal directions of \mathbb{R}^3 , respectively. $\langle \cdot, \cdot \rangle_{A\psi_{\text{temp}}(x)}$ is the Fisher-Rao metric defined in Eq. (4). Notice that we do not explicitly derive the gradient of A with respect to x , which involves the derivative of $D\phi_1$.

In sum, we integrate $\nabla_{\phi_1} E_x$ over the image space and with a change of variable from x to $\phi_1^{-1}(x)$, this yields $\nabla_{\phi_1} E$ in the form of

$$\nabla_{\phi_1} E = -2 \int_{x \in \Omega} \det(\phi_1(x)) (D_x \phi_1)^{-\top} \tag{16}$$

$$\begin{bmatrix} \langle \log_{A\psi_{\text{temp}}(x)} \psi_{\text{targ}}(\phi_1(x)), \frac{1}{|\Delta e_1|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_1) \rangle_{A\psi_{\text{temp}}(x)} \\ \langle \log_{A\psi_{\text{temp}}(x)} \psi_{\text{targ}}(\phi_1(x)), \frac{1}{|\Delta e_2|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_2) \rangle_{A\psi_{\text{temp}}(x)} \\ \langle \log_{A\psi_{\text{temp}}(x)} \psi_{\text{targ}}(\phi_1(x)), \frac{1}{|\Delta e_3|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_3) \rangle_{A\psi_{\text{temp}}(x)} \end{bmatrix} dx .$$

2.6 Numerical Implementation

We so far derive J and its gradient $\nabla J(m_t)$ in the continuous setting. In this section, we elaborate the numerical implementation of our algorithm under the discrete setting, in particular, the numerical computation of $\nabla_{\phi_1} E$.

In discretization of the spatial domain, we first represent the ambient space, Ω , using a finite number of points on the image grid, $\Omega \cong \{(x_i)_{i=1}^N\}$. In this setting, we can assume m_t to be the sum of Dirac measures such that $m_t = \sum_{i=1}^N \alpha_i(t) \otimes \delta_{\phi_t(x_i)}$ such that

$$\rho(\psi_{\text{temp}}, \psi_{\text{targ}})^2 = \int_0^1 \sum_{i=1}^n \sum_{j=1}^n \alpha_i(t)^\top [k_V(\phi_t(x_i), \phi_t(x_j)) \alpha_j(t)],$$

where $\alpha_i(t)$ is the momentum vector at x_i and time t . In discretization of the spherical domain \mathbb{S}^2 , we discretize it into N_S equally distributed gradient directions on the sphere. For each gradient direction k , it can be represented as 3D vector with unit length \mathbf{s}_k in Cartesian coordinate and $(r_k, \theta_k, \varphi_k)$ in the spherical coordinate. We use a conjugate gradient routine to perform the minimization of J with respect to $\alpha_i(t)$. We summarize steps required in each iteration during the minimization process below:

1. Use the forward Euler method to compute the trajectory based on the flow equation:

$$\frac{d\phi_t(x_i)}{dt} = \sum_{j=1}^N k_V(\phi_t(x_i), \phi_t(x_j)) \alpha_j(t) .$$

2. Compute $\nabla_{\phi_1(x_i)} E$ in Eq. (16), which is described in details below.
3. Solve $\eta_t = [\eta_i(t)]_{i=1}^N$ in Eq. (15) using the backward Euler integration, where i indices x_i .
4. Compute the gradient $\nabla J(\alpha_i(t)) = 2\alpha_i(t) + \eta_i(t)$.
5. Evaluate J when $\alpha_i(t) = \alpha_i^{\text{old}}(t) - \epsilon \nabla J(\alpha_i(t))$, where ϵ is the adaptive step size determined by a golden section search.

Since steps 1, 3 – 5 only involve the spatial information, we follow the numerical computation proposed in the previous LDDMM algorithm [16]. We now discuss how to compute $\nabla_{\phi_1(x_i)} E$ in Eq. (16), which involves the $\sqrt{\text{gODF}}$ interpolation in the spherical coordinate for $A\psi_{\text{temp}}(x_i)$ at a fixed x_i and the $\sqrt{\text{gODF}}$

interpolation in the image spatial domain for $\psi_{\text{targ}}(\phi_1(x))$. To do so, we rewrite $A\psi_{\text{temp}}(x_i)$ as $A\psi_{\text{temp}}(\mathbf{s}_k, x_i)$ and $\psi_{\text{targ}}(\phi_1(x_i))$ as $\psi_{\text{targ}}(\mathbf{s}_k, \phi_1(x_i))$, where \mathbf{s}_k is in the Cartesian coordinate and corresponds to $(r_k, \theta_k, \varphi_k)$ in the spherical coordinate. Denote $\widehat{\mathbf{s}}_k = A^{-1}\mathbf{s} = (\widehat{r}_k, \widehat{\theta}_k, \widehat{\varphi}_k)$. We first compute $A\psi_{\text{temp}}(\mathbf{s}_k, x_i)$ according to Eq. (10) as

$$A\psi(\mathbf{s}_k, x_i) = \sqrt{\frac{\det(D_{x_i}\phi_1)^{-1}\|\mathbf{s}\|}{\|(D_{x_i}\phi_1)^{-1}\mathbf{s}\|}}\psi((D_{x_i}\phi_1)^{-1}\mathbf{s}, x_i).$$

It is more convenient to compute $A\psi(\mathbf{s}_k, x_i)$ using the form of the square-root ODF

$$A\psi(\mathbf{s}_k, x_i) = \sqrt{\frac{\det(D_{x_i}\phi_1)^{-1}r_k}{\widehat{r}_k^3}}\sqrt{\mathbf{p}(\widehat{\theta}_k, \widehat{\varphi}_k, x_i)}\delta(r_k - 1)$$

where $\mathbf{p}(\widehat{\theta}_k, \widehat{\varphi}_k, x_i)$ is computed analytically using angular interpolation on \mathbb{S}^2 based on spherical harmonics.

We compute $\psi_{\text{targ}}(\mathbf{s}_k, \phi_1(x_i))$ under the Riemannian framework in §2.1 as

$$\psi_{\text{targ}}(\mathbf{s}_k, \phi_1(x_i)) = \exp_{\psi_{\text{targ}}(\mathbf{s}_k, \phi_1(x_i))} \sum_{j \in \mathcal{N}_i} w_j \log_{\psi_{\text{targ}}(\mathbf{s}_k, \phi_1(x_i))}(\psi_{\text{targ}}(\mathbf{s}_k, x_j)),$$

where \mathcal{N}_i is the neighborhood of x_i , and w_j is the weight of x_j based on the distance between $\phi_1(x_i)$ and x_j . The exponential maps and logarithm maps can be computed via Eq. (6) and Eq. (7) respectively. Finally, the inner product in Eq. (16),

$$\langle \log_{A\psi_{\text{temp}}(x)} \psi_{\text{targ}}(\phi_1(x)), \frac{1}{|\Delta e_i|} \log_{A\psi_{\text{temp}}(x)} A\psi_{\text{temp}}(x + \Delta e_i) \rangle_{A\psi_{\text{temp}}(x)},$$

can be computed using Eq. (4), where Δe_i is the voxel size.

3 Experiments

In this section, we present experiments on synthetic and real datasets using the proposed registration algorithm. We first illustrate the affine transformations of ODFs, described in §2.2, on two types of ODFs: a single fiber and a crossing fiber. Three types of transformations are applied to the ODFs: a rotation with angle θ_z where $A = [\cos \theta_z \ -\sin \theta_z \ 0; \sin \theta_z \ \cos \theta_z \ 0; 0 \ 0 \ 1]$; a vertical scaling with factor ρ_y where $A = [1 \ 0 \ 0; 0 \ \rho_y \ 0; 0 \ 0 \ 1]$; and a vertical shearing with factor ζ_y where $A = [1 \ 0 \ 0; -\zeta_y \ 1 \ 0; 0 \ 0 \ 1]$. From Fig. 1, we see that the resulting ODFs are reasonable estimations if the fibers undergo the described transformations. In addition, we notice that the resulting ODFs are non-negative and sum to 1, thus are elements of \mathcal{Q} .

We next evaluate our method’s performance on synthetic data generated using the multi-tensor method [11]. The HARDI signal $S(\mathbf{g})$ in the direction \mathbf{g} , where

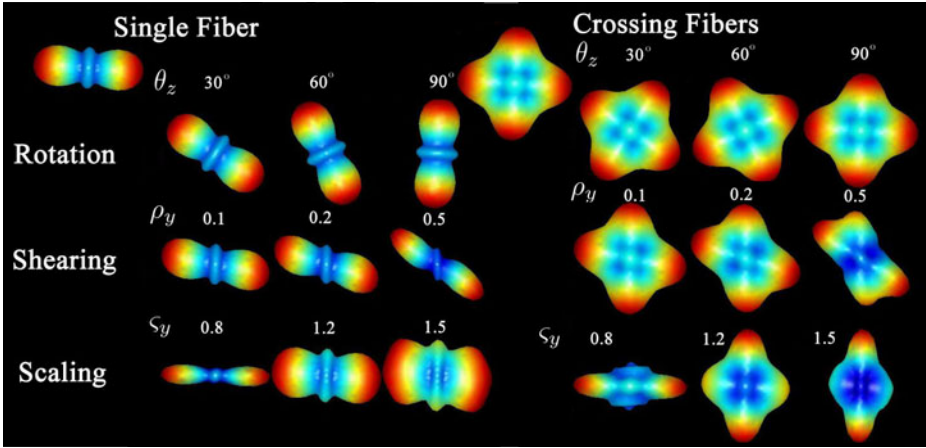


Fig. 1. Transformed ODFs after rotation, scaling and shearing

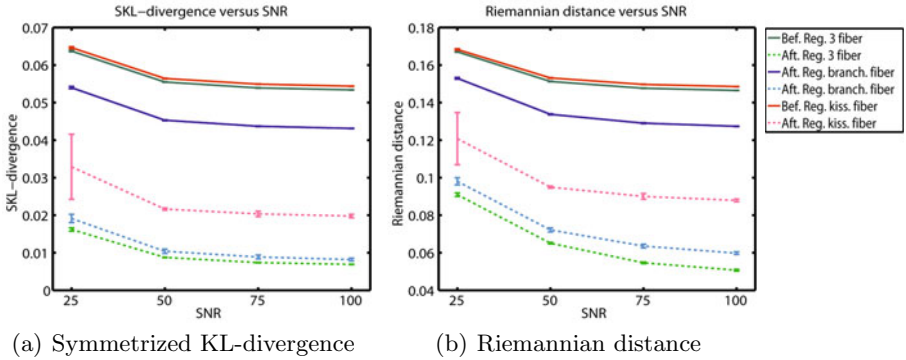


Fig. 2. Registration error vs varying SNR for synthetic data

\mathbf{g} is a unit vector, is generated by $S(\mathbf{g}) = \sum_{k=1}^N p_k \exp(-b\mathbf{g}^\top \mathbf{D}_k \mathbf{g})$, where the b -value is given, n is the number of fibers, p_k is the volume fraction of the k -th fiber and \mathbf{D}_k is the k -th diffusion tensor profile. We generate three ODF fields: three fibers crossing, branching fiber, and kissing fiber. Registration results are shown in Fig. 3. Results of three fibers crossing are shown in Figs. 3(a)-3(g) with the zoomed-in region in green shown in Figs. 3(e)-3(g), branching fibers in Figs. 3(h)-3(n) with the zoomed-in region in red shown in Figs. 3(l)-3(n), kissing fibers in Figs. 3(o)-3(u) with the zoomed-in region in blue shown in Figs. 3(s)-3(u). Visually, we see that the registration for the three fibers is the best, followed by branching fiber and, lastly, kissing fibers. Next, we examine the effects of noise on registration accuracy. Noisy versions of $S(\mathbf{g})$ are generated by adding complex Gaussian noise with zero mean and standard deviation $\sigma = \frac{1}{\zeta}$, where

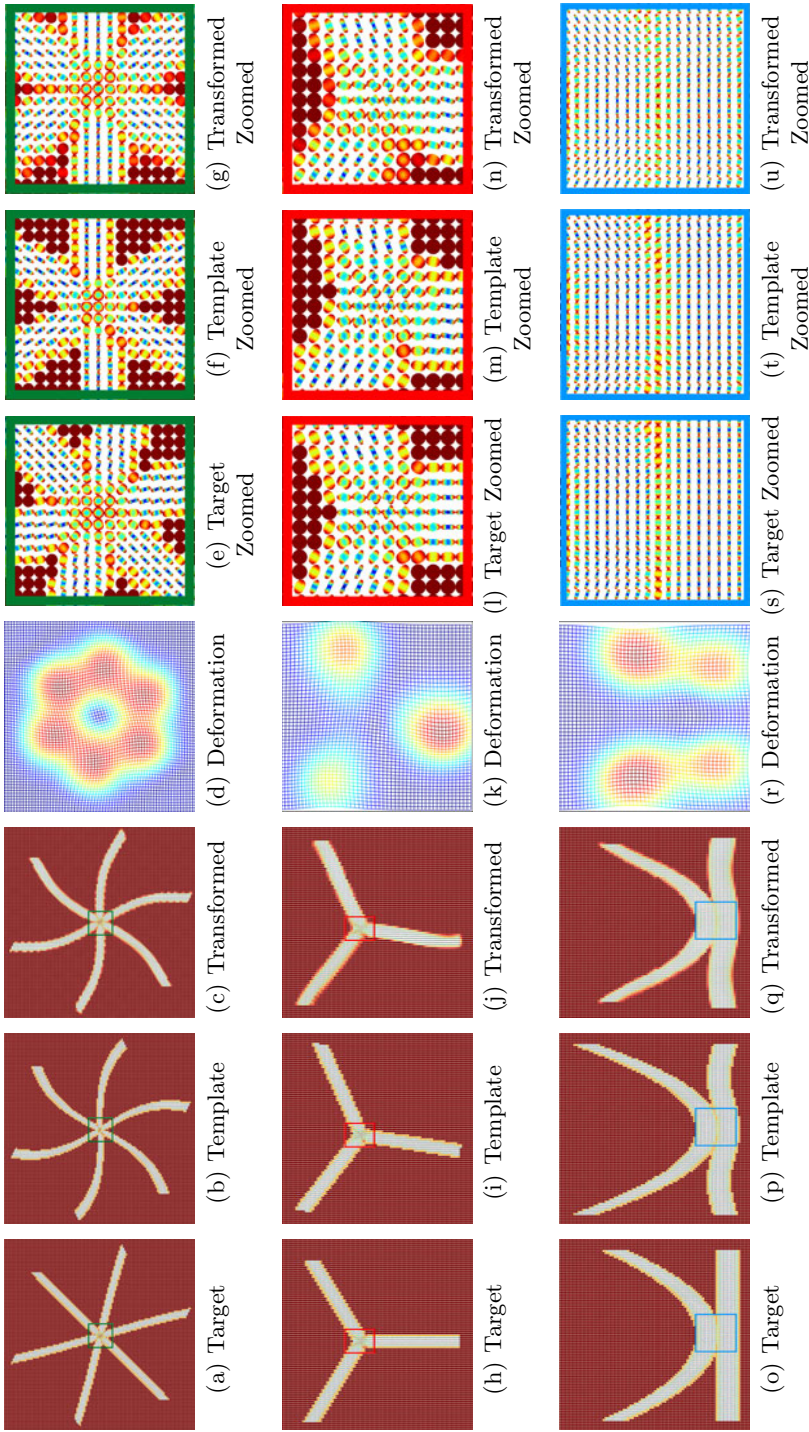


Fig. 3. Registration of synthetic ODF fields. Figs. 3(a)-3(g) shows three fibers, Figs. 3(h)-3(n) branching fibers, and Figs. 3(o)-3(u) kissing fibers.

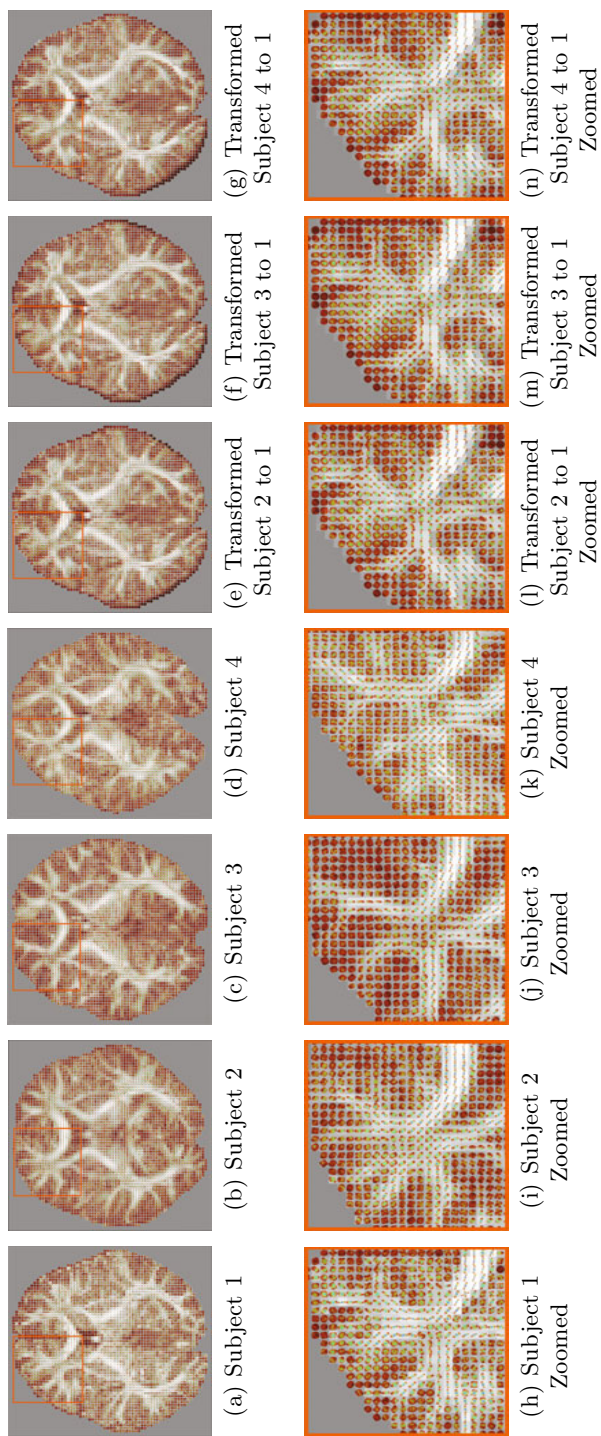


Fig. 4. Registration of real brain data. Subjects 2 – 4 are registered to Subject 1.

ζ is the signal-to-noise ratio (SNR) [11]. We use the two distances to measure the similarity between the ODFs field: the Riemannian distance used in our cost function and the symmetrized Kullback–Leibler (sKL) divergence, normalized by the number of voxels. We vary the SNR from 25 to 100 and perform 50 trials for each SNR. Fig. 2 shows the mean and standard deviation of the sKL in Fig. 2(a) and the Riemannian distance in Fig. 2(b). From the results, we see that both distances of the ODFs field are reduced after registration. In addition, when SNR is lower, there is an increase in the errors. Finally, both errors suggest the registration for the three fibers is the best, followed by branching and kissing fibers in this order. This result is observed in Fig. 3.

Finally, we apply our registration algorithm framework to a subset of a HARDI human brain database of 4 children and the results are shown in Fig. 4. Diffusion weighted MR images were obtained using the following imaging parameters: 55 axial slices (2.3 mm thick), TR/TE=6800/89 ms, with a 96×96 acquisition matrix (2.3 mm in-plane resolution). 7 with no diffusion sensitization and $n = 61$ diffusion weighted images at $b = 900$ s/mm² were acquired. Gradient directions were evenly distributed on the hemisphere. The ODF fields of the four subjects are shown in Figs. 4(a)–4(d), with the zoomed-in region in red shown in Figs. 4(h)–4(k). We map the ODF fields of subjects 2 – 4 to subject 1 are shown in Figs. 4(e)–4(g), with the zoomed-in region in red shown in Figs. 4(l)–4(n). With respect to subject 1, the Riemannian distance (normalized by number of voxels) of subject 2 before and after registration is 0.187 and 0.148, subject 3 is 0.184 and 0.156, and subject 4 is 0.187 and 0.146. With respect to subject 1, the sKL divergence (normalized by number of voxels) of subject 2 before and after registration is 0.160 and 0.125, subject 3 is 0.161 and 0.136, and subject 4 is 0.154 and 0.115.

4 Conclusion

We present a novel diffeomorphic metric mapping algorithm for aligning HARDI data in the setting of large deformations. Our mapping algorithm seeks an optimal diffeomorphic flow connecting one HARDI to another in a diffeomorphic metric space and locally reorients ODFs due to the diffeomorphic transformation at each location of the 3D HARDI volume. This is achieved through a new generalization of the traditional ODF from S^2 to \mathbb{R}^3 , used in defining affine transformation actions. Additionally, we impose a Riemannian space on this generalized ODF for easy metric computation. The diffeomorphic metric space combined with the ODF Riemannian metric space provides a natural framework for computing the gradient of our mapping functional. We demonstrate that our algorithm is robust against noise in the case of synthetic data and performs well on real brain HARDI data. This registration approach will facilitate atlas generation and group analysis of HARDI for a variety of clinical studies. We are currently investigating the effects of our registration algorithm on fiber tractography.

Acknowledgments. Work supported by grants A*STAR SERC 082-101-0025, A*STAR SICS-09/1/1/001, a center grant from the National Medical Research Council (NMRC/CG/NUHS/2010), the Young Investigator Award at National University of Singapore (NUSYIA FY10 P07), and National University of Singapore MOE AcRF Tier 1.

References

1. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. *MRM* 64, 554–566 (2010)
2. Amari, S.: *Differential-Geometrical Methods in Statistics*. Springer, Heidelberg (1985)
3. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth order tensor field estimation from DW-MRI. *NeuroImage* 45(1, suppl. 1), S153–S162 (2009)
4. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Registration of high angular resolution diffusion MRI images using 4th order tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 908–915. Springer, Heidelberg (2007)
5. Basser, P.J., Mattiello, J., Lebihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B* 103, 247–254 (1994)
6. Behrens, T.E.J., Berg, H.J., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage* 34(1), 144–155 (2007)
7. Bloy, L., Verma, R.: Demons registration of high angular resolution diffusion images. In: *ISBI* (2010)
8. Cencov, N.N.: Statistical decision rules and optimal inference. In: *Translations of Mathematical Monographs*, vol. 53. AMS, Providence (1982)
9. Cheng, G., Vemuri, B.C., Carney, P.R., Mareci, T.H.: Non-rigid registration of high angular resolution diffusion images represented by gaussian mixture fields. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5761, pp. 190–197. Springer, Heidelberg (2009)
10. Cheng, J., Ghosh, A., Jiang, T., Deriche, R.: A riemannian framework for orientation distribution function computing. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009. LNCS*, vol. 5761, pp. 911–918. Springer, Heidelberg (2009)
11. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical Q-ball imaging. *MRM* 58, 497–510 (2007)
12. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. *Quart. App. Math.* 56, 587–600 (1998)
13. Frank, L.R.: Characterization of anisotropy in high angular resolution diffusion-weighted MRI. *MRM* 47(6), 1083–1099 (2002)
14. Geng, X., Ross, T.J., Gu, H., Shin, W., Zhan, W., Chao, Y.-P., Lin, C.-P., Schuff, N., Yang, Y.: Diffeomorphic image registration of diffusion MRI using spherical harmonics. *IEEE TMI* 30(3), 747 (2011)
15. Ghosh, A., Descoteaux, M., Deriche, R.: Riemannian framework for estimating symmetric positive definite 4th order diffusion tensors. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 858–865. Springer, Heidelberg (2008)

16. Glaunès, J., Qiu, A., Miller, M., Younes, L.: Large deformation diffeomorphic metric curve mapping. *IJCV* 80(3), 317–336 (2008)
17. Goh, A., Lenglet, C., Thompson, P.M., Vidal, R.: A nonparametric Riemannian framework for processing High Angular Resolution Diffusion Images and its applications to ODF-based morphometry. *NeuroImage* 56(3), 1181–1201 (2011)
18. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *MRM* 56(1), 104–117 (2006)
19. Hong, X., Arlinghaus, L.R., Anderson, A.W.: Spatial normalization of the fiber orientation distribution based on high angular resolution diffusion imaging data. *MRM* 61, 1520–1527 (2009)
20. Leergaard, T.B., White, N.S., de Crespigny, A., Bolstad, I., D’Arceuil, H., Bjaalie, J.G., Dale, A.M.: Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain. *PLoS One* 5, e8595 (2010)
21. Miller, M.I., Beg, M.F., Ceritoglu, C., Stark, C.: Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *PNAS* 102, 9685–9690 (2005)
22. Özarslan, E., Mareci, T.H.: Generalized DTI and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *MRM* 50, 955–965 (2003)
23. Rao, C.R.: Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* 37, 81–89 (1945)
24. Srivastava, A., Jermyn, I., Joshi, S.H.: Riemannian analysis of probability density functions with applications in vision. In: *IEEE CVPR* (2007)
25. Tuch, D.S.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *MRM* 48, 577–582 (2002)
26. Yap, P.-T., Chen, Y., An, H., Yang, Y., Gilmore, J.H., Lin, W., Shen, D.: SPHERE: SPHERical Harmonic Elastic REGistration of HARDI data. *NeuroImage* 55(2), 545–556 (2011)

Schild's Ladder for the Parallel Transport of Deformations in Time Series of Images

Marco Lorenzi^{1,2}, Nicholas Ayache¹, Xavier Pennec¹,
and the Alzheimer's Disease Neuroimaging Initiative*

¹ Project Team Asclepios, INRIA Sophia Antipolis, France

² LENITEM, IRCCS San Giovanni di Dio, Fatebenefratelli, Italy

Abstract. Follow-up imaging studies require the evaluation of the anatomical changes over time for specific clinical groups. The longitudinal changes for a specific subject can be evaluated through the non-rigid registration of successive anatomical images. However, to perform a longitudinal group-wise analysis, the subject-specific longitudinal trajectories of anatomical points need to be transported in a common reference frame. In this work, we propose the Schild's Ladder framework as an effective method to transport longitudinal deformations in time series of images in a common space using diffeomorphic registration. We illustrate the computational advantages and demonstrate the numerical accuracy of this very simple method by comparing with standard methods of transport on simulated images with progressing brain atrophy. Finally, its application to the clinical problem of the measurement of the longitudinal progression in the Alzheimer's disease suggests that an important gain in sensitivity could be expected on group-wise comparisons.

1 Introduction

One of the main objectives in the field of computational anatomy concerns the modeling of the dynamics occurring in specific clinical groups. This is motivated by the need to generalize the characteristics observed in the single patients, by performing group-wise statistics. The aim of the Template-based analysis is therefore to describe the subject-specific characteristics in a normalized reference by transporting the individual geometries in a common space. Different normalization methods can be used for static observations, depending on the complexity of the feature of interest: resampling for scalar values, reorientation for vectors, and more advanced methods for tensors, like Preservation of the Principal Directions (PPD) or the Finite Strain (FS) [1].

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

Transport of longitudinal trajectories. The problem of modeling dynamic quantities, such as the changes measured in time series of images, is a more complex issue. This problem concerns the consistent transport of the trajectories of changes in a common reference, and we are interested in the behaviour of the transported dynamics, especially regarding the ability to preserve the original amount of expansion/contraction in homologous regions.

A key issue is the different nature of the changes occurring at the intra-subject level, which reflects the biological phenomena of interest, and the changes across different subjects. In fact, the inter-subject variability is a scale of magnitude higher than the more subtle subject-specific variations.

Different approaches for the transport have been proposed depending on the type of measure of change that is considered (time series of jacobian images, the time-varying initial momentum, displacement fields as a function of time, etc.). A simple method of transport consists in *reorienting* the longitudinal inter-subject displacement *vector field* by the Jacobian matrix of the subject-to-template mapping. A drawback of this method is that the longitudinal deformation is fully combined to the inter-subject one. The method proposed by [10] uses the *transformation conjugation* (change of coordinate system) from the group theory in order to compose the longitudinal inter-subject deformation with the subject-to-template one. As pointed in [3], this practice could potentially introduce variations in the transported deformation and relies on the inverse consistency of the estimated deformations, which can raise problems for large deformations. The *parallel transport* of relational measures was introduced in [12] in the context of the Large Deformation Diffeomorphic Metric Mapping (LDDMM) [8]. The notion of parallel transport proposed here consists in translating a vector *along a geodesic* while preserving the parallelism according to the space geometry. This framework allows to transport the geodesic diffeomorphic registration for both point supported data and images, and it was applied to study the hippocampal shape changes in Alzheimer's disease [9]. Although it represents a rigorous implementation of the parallel transport, it is limited to the transport along geodesics and it comes to the price of a computationally intense scheme. This is a limitation which could prevent the application to large datasets with multiple time series of images, which are now becoming more easily available to the imaging community.

In tensor based morphometry, the transport is used to normalize in a common reference frame the degree of changes in specific regions, such as the hippocampal volume loss in the brain. For this purpose, it is also possible to directly normalize the measures of change by interpolating the Jacobian determinant scalar map of the intra-subject longitudinal change into the template reference [3]. However, the Jacobian determinant represents only one of the several features of interest in morphometric studies (like full jacobian matrices, tensors or the flux [4, 7]). Moreover, transporting the original deformation trajectory allows multivariate group analysis like evaluation of mean deformations, PCA, etc.

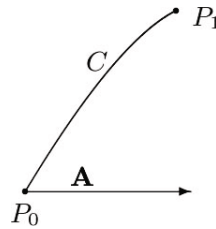
The aim of the present paper is to propose an accurate computational approach for the parallel transport of longitudinal trajectories of images, and

consequently of vectors and related measures of changes. For this purpose we introduce in Section 2 the ‘‘Schild’s Ladder’’, which proves to provide a fast and well posed diffeomorphic solution for the transport of longitudinal trajectories of images along any curve. To demonstrate the effectiveness of the scheme, we then apply it in Section 3 to the specific clinical problem of the measurements of the longitudinal progression of atrophy in Alzheimer’s disease.

2 The Schild’s Ladder Procedure

The Schild’s Ladder was introduced in the last century in the field of the general relativity by the physicist Alfred Schild [5]. It provides a straightforward method to compute a first order approximation of the parallel transport of a vector along a curve without requiring the knowledge of the tangent structure of the space.

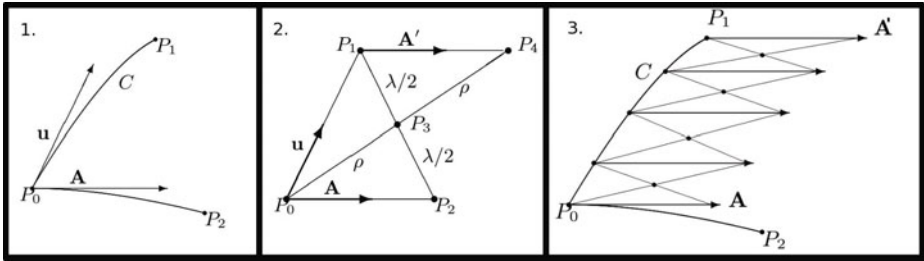
Let M a manifold and C a curve parametrized by the parameter τ with $\frac{\partial C}{\partial \tau}|_{T_0} = \mathbf{u}$, and $\mathbf{A} \in T_{P_0}M$, a tangent vector on the curve at the point $P_0 = C(0)$. Let P_1 be a point on the curve relatively close to P_0 , i.e. separated by a sufficiently small parameter value τ .



The Schild’s Ladder computes the parallel transport of \mathbf{A} along C :

1. Define a *curve* on the manifold parametrized by a parameter σ passing through the point P_0 with tangent vector $\frac{\partial}{\partial \sigma}|_{P_0} = \mathbf{A}$. Chose a point P_2 on the curve separated by P_0 by the value of the parameters σ . The values of the parameters σ and τ should be chosen in order to construct the Schild’s Ladder within a single coordinate neighborhood.
2. Let l be the *geodesic* connecting $P_2 = l(0)$ and $P_1 = l(\lambda)$, we chose the ‘‘middle point’’ $P_3 = l(\lambda/2)$. Now, let us define the *geodesic* r connecting the starting point P_0 and P_3 parametrized by ρ such that $P_3 = r(\rho)$. Extending the geodesic at the parameter 2ρ we reach the point P_4 . We can now pick a *curve* connecting P_1 and P_4 . The vector \mathbf{A}' tangent to the curve at the point P_1 is the parallel translation of \mathbf{A} along C .
3. If the distance between the points P_0 and P_1 is large, the above construction can be iterated for a sufficient number of small steps.

The Schild’s ladder geometrical approximation resides in the assumption that all the geometrical information of the space are encoded by the geodesics. Although the geodesics on the manifold are not sufficient to recover all the information about the space properties, such as the connection, it has been shown that the Schild’s Ladder describes the parallel transport with respect to the symmetric part of the connection of the space [6]. An intuitive proof is that the



construction of the above diagram is commutative and can be symmetrized with respect to the points P_1 and P_2 . If the original connection is symmetric, then this procedure provides a correct linear approximation of the parallel transport of vectors.

2.1 Application to Images

Schild’s Ladder for images. Let I_i ($i = 1 \dots n$) be a time series of images with the baseline I_0 as reference. Consider a template image T_0 , the aim of the procedure will be to compute the image T_i in order to define the transport of the sequence $I_0 - I_i$ in the reference of T_0 . In the sequel, we focus on the transport of a single image I_1 .

To apply the Schild’s Ladder in the context of the images, we define the paths in the space of images by action from the space of diffeomorphism. Let $\mathbb{I} = \{f : \mathbb{R}^3 \rightarrow \mathbb{R}\}$ the image space and let us define the action $*$: $M \times \mathbb{I} \rightarrow \mathbb{I}$ given by $(\varphi, I) \mapsto \varphi * I = I \circ \varphi^{-1}$, where M is the space of the diffeomorphisms. If the distance between two images in the image space is defined in terms of diffeomorphisms [12], then the geodesics in the image space are defined by the action of the geodesic paths in the space of the diffeomorphisms.

The Schild’s Ladder procedure can therefore be naturally translated in the following way:

1. The geodesic $l(\lambda)$ in the space \mathbb{I} connecting I_1 and T_0 is such that $l(0) = g(0) * I_1 = I_1$ and $l(1) = g(1) * I_1 = T_0$.
2. Define the half-space image $I_{\frac{1}{2}}$ as $l(1/2) = g(1/2) * I_1$.
3. Compute the geodesic $r(\rho) = h(\rho) * I_0$ connecting I_0 and $I_{\frac{1}{2}}$ such that $r(0) = I_0$ and $r(1) = h(1) * I_0 = I_{\frac{1}{2}}$.
4. Define the transported follow-up image as $T_1 = r(2) = h(2) * I_0$.

Finally one can evaluate the transported deformation, and the related measures of changes, in the new reference by registering the images T_0 and T_1 . Despite its straightforward formulation, the application of the Schild’s Ladder to the image space requires multiple evaluations of geodesics in the space of diffeomorphisms and a consequent high cost in terms of computation time and resources. Moreover, it assumes an exact matching, which is bound to lead to important numerical problems. For instance, the definition of $I_{\frac{1}{2}}$ using the forward deformation on I_1 or the backward from T_0 lead to very different results.

We propose to reformulate the above scheme in a computationally efficient and numerically stable framework using only transformations.

An effective Schild's Ladder using one parameter subgroups. We use the setting of the stationary velocity fields (SVF) diffeomorphic registration as provided by the Symmetric Log Demons algorithm [11]. In particular we base the Schild's Ladder construction on the path defined by the Lie group exponential of vectors. We note that the SVF is a valid approximation of a small step of a time-varying velocity field diffeomorphism. Given a pairs of images I_i , $i \in \{0, 1\}$, the SVF framework parametrizes the diffeomorphism φ required to match the reference I_0 to the moving image I_1 by a stationary velocity field u . The velocity field u is an element of the Lie Algebra \mathbf{G} of the Lie group of diffeomorphisms M , i.e. an element of the tangent space at the identity $T_{id}M$. The diffeomorphism φ belongs to the one parameter subgroup generated by u and is parametrized by the Lie group exponential operator $\varphi = \text{Exp}(u)$. We can therefore define the paths in the space of the diffeomorphisms from the one parameter subgroup parametrization $l(\lambda) = \text{Exp}(\lambda \cdot u)$ and by consequence the paths in the image space by the action $*$.

However the Schild's Ladder in the image space requires a number of interpolations and exponentiations, which could introduce biases due to the numerical approximations. Moreover the registration is constrained to be smooth and it is therefore impossible to reach a perfect match of corresponding intensities in the registered images. We however take advantage of the symmetry of the construction in order to be robust to the changes introduced by the registration and lead to a very simple scheme.

1. Let $I_1 = \text{Exp}(u) * I_0$.
2. Compute $v = \text{argmin}_{v \in \mathbb{G}} (\|T_0 \circ \text{Exp}(-v/2) - I_1 \circ \text{Exp}(v/2)\|^2 + \|v^2\|)$.
The half space image $I_{\frac{1}{2}}$ can be defined in terms of $v/2$ as $\text{Exp}(-v/2) * T_0$ or $\text{Exp}(v/2) * I_1$. While from the theoretical point of view the two images are identical, the choice of one of them, or even their mean, introduces a bias in the construction. The definition of the half step image can be bypassed by relying on the symmetric construction of the parallelogram (Figure 1).
3. The transformation from I_0 to $I_{\frac{1}{2}}$ is $\rho = \text{Exp}(v/2) \circ \text{Exp}(u)$ and the symmetry leads to

$$\text{Exp}(II(u)) = \rho \circ \text{Exp}(-v/2) = \text{Exp}(v/2) \circ \text{Exp}(u) \circ \text{Exp}(-v/2)$$

The transport of the deformation $\varphi = \text{Exp}(u)$ can be therefore obtained through the conjugate action operated by the deformation parametrized by $v/2$. In opposition to the standard conjugate method which operates "vertically", from the image I_0 to T_0 , here the transport operates "diagonally".

Since the direct computation of the conjugate by composition is potentially biased by the discrete approximation, we propose two different schemes to correctly evaluate the transport directly in the Lie Algebra.

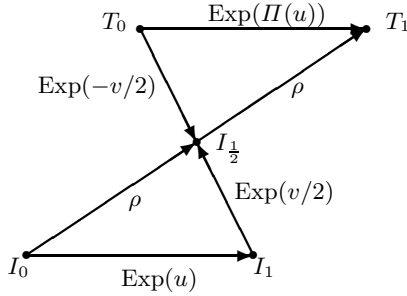


Fig. 1. Schild’s Ladder parallelogram with the one parameter subgroups. The transport $\text{Exp}(\Pi(u))$ is the deformation $\text{Exp}(v/2) \circ \text{Exp}(u) \circ \text{Exp}(-v/2)$.

BCH formula for the conjugate action. The Baker-Campbell-Haudorff (BCH) formula was introduced in the SVF diffeomorphic registration in [2] and provides an explicit way to compose diffeomorphisms by operating in the associated Lie Algebra. More specifically, if v, u are static velocity fields, then $\text{Exp}(v) \circ \text{Exp}(u) = \text{Exp}(w)$ with $w = v + u + \frac{1}{2}[v, u] + \frac{1}{12}[v, [v, u]] - \frac{1}{12}[u, [u, v]] + \dots$. In particular, for small u , the computation can be truncated to any order to obtain a valid approximation for the composition of diffeomorphisms. For example, the use of a truncated expansion $v + u + \frac{1}{2}[v, u]$ is used as updating rule for the Symmetric Log Demons algorithm. Applying the truncate BCH to the conjugate action leads to

$$\Pi_{BCH}(u) \simeq u + [v, u] + \frac{1}{2}[v, [v, u]]$$

To provide a sufficiently small vector for the computation of the conjugate we can take advantage of the properties of the one-parameter subgroups to observe that

$$\begin{aligned} & \text{Exp}(v) \circ \text{Exp}(u) \circ \text{Exp}(-v) = \\ & = \text{Exp}\left(\frac{v}{n}\right) \circ \dots \circ \text{Exp}\left(\frac{v}{n}\right) \circ \text{Exp}(u) \circ \text{Exp}\left(-\frac{v}{n}\right) \circ \dots \circ \text{Exp}\left(-\frac{v}{n}\right) \end{aligned}$$

The conjugation can then be iteratively computed in the following way:

- find n such that v/n is small.
- compute $w = u + [\frac{v}{n}, u] + \frac{1}{2}[\frac{v}{n}, [\frac{v}{n}, u]]$
- Let $u = w$
- Iterate the above construction (n steps).

Using the BCH formula allows to perform the transport directly in the Lie algebra and avoids exponentiation and the interpolations, thus reducing the bias introduced by the numerical approximations. Moreover, this methods preserves the original “Ladder” formulation, operated along the path described by $\text{Exp}(tv)$. However, it requires a number of iterations to be computed.

Conjugate action from the exponential map. We can provide an alternative and direct formula to compute the transport by conjugate action from the definition of the exponential:

$$\text{Exp}(u) = \lim_{n \rightarrow \infty} \left(\text{Id} + \frac{u}{n} \right)^n .$$

We can then write:

$$\text{Exp}(\Pi_{\text{conj}}(u)) = \lim_{n \rightarrow \infty} \left(\text{Exp}(v) \circ \left(\text{Id} + \frac{u}{n} \right) \circ \text{Exp}(-v) \right)^n$$

Let $y = \text{Exp}(-v)(x)$ and $\phi(x) = \text{Exp}(v)(x)$, then

$$\begin{aligned} \text{Exp}(\Pi_{\text{conj}}(u)) &= \lim_{n \rightarrow \infty} \left(\phi \left(y + \frac{u(y)}{n} \right) \right)^n = \\ &= \lim_{n \rightarrow \infty} \left(\text{Id} + \frac{1}{n} \left(\text{D}(\phi(y))|_{\phi^{-1}(x)} \cdot u \circ \phi^{-1}(x) + o(\|u\|^2) \right) \right)^n \end{aligned}$$

By the definition of the exponential map, we obtain then a first order approximation for the transported vector given by

$$\Pi_{\text{conj}}(u) = \text{D}(\text{Exp}(v))|_{\text{Exp}(-v)} \cdot u \circ \text{Exp}(-v)$$

We note that $\text{D}(\text{Exp}(v))|_{\text{Exp}(-v)} = \text{D}(\text{Exp}(-v))^{-1}$. This method provides a closed form formula which enables to compute the transport by reorienting the field $u \circ \text{Exp}(-v)$ by the matrix field $\text{D}(\text{Exp}(v))$ resampled by $\text{Exp}(-v)$, or equivalently, by the matrix field $\text{D}(\text{Exp}(-v))^{-1}$. The second formula requires however the inversion of a matrix which is an operation potentially unstable. In the following, the transport $\Pi_{\text{conj}}(u)$ will be evaluated through the resampling by linear interpolation of the matrix $\text{D}(\text{Exp}(v))$.

From a theoretical point of view the results obtained from the two methods are equivalent in the continuous domain

$$\Pi_{\text{conj}}(u) = u + \text{D}v \cdot u - \text{D}u \cdot v + O(\|v\|^2) \simeq \Pi_{\text{BCH}}(u)$$

3 Experiments on Synthetic and Real Data

3.1 Comparison of Different Methods of Transport

We created the realistic simulated deformations based on the deformation field that matches the baseline scan (I_0) of a patient from the ADNI dataset to the 1-year follow-up of the same patient, computed using the Symmetric Log-Demons Algorithm. The ventricular expansion was extracted by masking the corresponding SVF v for a mask including the ventricles. The deformations in the remaining areas of the brain were imposed to be negligible, by setting the velocity field to 0 and adding gaussian noise. The SVF v was then increasingly

scaled ($v_i = f_i v$, with $f_i = 0.5, 1, 2, 3$) and the resulting deformations fields $\varphi_i = \text{Exp}(v_i)$ were used to warp the baseline scan I_0 to generate a longitudinal progression of serial images I_i with increasing ventricular expansions.

The longitudinal progression was then transported in a new reference given by the image of another patient (target space T_0) along the deformation ψ_T using different methods:

- Schild’s Ladder (Π_{BCH} and Π_{conj}),
- conjugate method: $Ad_{\psi_T}(\varphi_i) = \psi_T^{-1} \circ \varphi_i \circ \psi_T$
- interpolation of the scalar maps by the mapping ψ_T ,
- reorientation of the static velocity field v_i by the Jacobian Matrix of the deformation ψ_T : $J_{\psi_T} v_i$.

As summarized in Table 1, not all the methods operate on the same features and a direct comparison is not always possible. To test the accuracy of the transport, the different methods were quantitatively compared on the scalar measures representing the amount of change induced in the ventricles mask. The analyzed features were: Jacobian determinant of the transported deformation $J = \det \nabla \varphi_i$, Log-Jacobian determinant of the transported deformation $\log(J)$, Elastic Energy $\|v_i\|_{El}^2 = \text{tr}((\text{Id} + \nabla \varphi_i)(\text{Id} + \nabla \varphi_i)^T)$, and L^2 norm of the SVF $\|v_i\|_{L^2}$.

Results. The synthetic time series of images transported by the Schild’s Ladder can be appreciated in Figure 2. The series is consistent with the original trajectory of ventricular expansion while adapting to the new reference. Concerning the quantitative comparison of the different transport methods, Figure 3 shows the Log-Jacobian scalar image derived from the different methods. As expected, either the Conjugate method and the Reorientation lead to noisy maps. We can see that for both the Schild’s Ladder (BCH scheme) and the Interpolation method, the resulting Jacobian map seems to adapt to the new reference space while remaining sufficiently smooth, coherently with the original map. The map resulting from the conjugate scheme of the Schild’s Ladder (not shown) was more noisy than the BCH one.

Table 2 shows the amount of changes measured with the different methods. For all of them, the longitudinal trend was in agreement with the original one.

Table 1. Different methods of transport and transported features. From the SVF we can infer transformations from which we can extract scalar measures, while the reverse is not possible.

	SVF (u)	Transformation ($\varphi = \text{Exp}(u)$)	Scalar Measure ($J, \log J, \ u\ , \ u\ _{El}$)
Interpolation of Scalar	No	No	Yes
Conjugate action	No	Yes	Yes
Schild’s Ladder	Yes	Yes	Yes
Reorientation	Yes	Yes	Yes

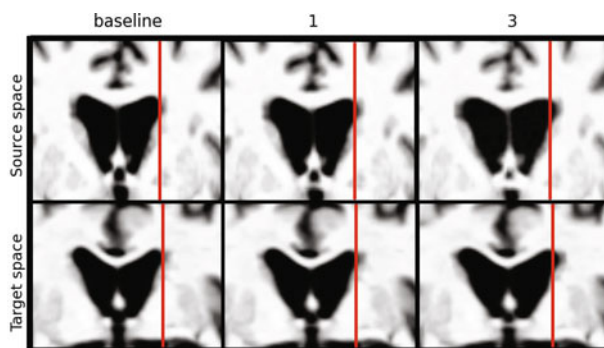


Fig. 2. Transport of time series of images. Top row: original longitudinal trajectory for the ventricular expansion at the different scaling factors. Bottom row: transported longitudinal sequence in the target space.

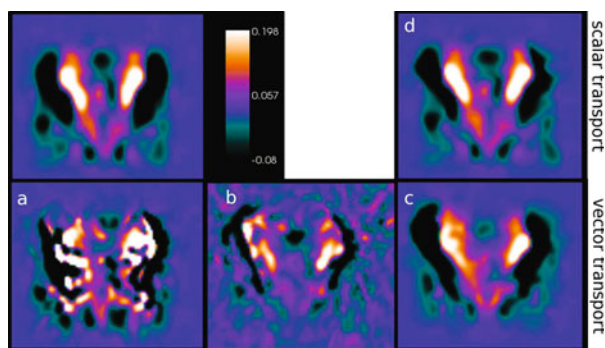


Fig. 3. Log-Jacobian maps of the transported deformation field in the target space T . Top left: Log-Jacobian map corresponding to the intra-subject deformation $\varphi_1 : I_0 \rightarrow I_1$ in the baseline space. From a) to d), Log-Jacobian maps in the target space T corresponding to the transported deformation with the different methods: a) Reorientation, b) Conjugate, c) Schild's Ladder (BCH version), d) Interpolation of the scalar Log-Jacobian original image.

The methods performed differently depending on the measure of interest, but we can note that the Schild's Ladder provides, in most of the cases, the closest measure to the one in the original reference. Moreover, the BCH version of the Schild's Ladder performs on the average better as compared to the conjugate one.

3.2 One Year Follow-Up Changes on Alzheimer's Disease

Images corresponding to the baseline I_0 and the one-year follow-up I_1 scans were selected for 70 subjects affected by Alzheimer's disease from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). For

Table 2. Average features at the different scaling factors (columns) evaluated in the ventricles mask through the different methods of transport (rows). The L^2 norm of the SVF could not be evaluated with the adjoint method. Bold: closest measure to the original one. The values were multiplied by 100.

	L^2 norm				Elastic energy				Jacobian (avg-1)				Log-Jacobian			
	0.5	1	2	3	0.5	1	2	3	0.5	1	2	3	0.5	1	2	3
Original	1	4.9	18.1	38.9	1.53	2.78	6.74	11.1	1.49	2.61	6.25	10.24	11.66	19.96	43.88	67.89
Schild's Ladder (BCH)	1	4.85	18.8	39.5	1.26	2.61	6.9	12	1.29	2.43	5.71	9.35	8.71	16.47	38.22	59.6
Schild's Ladder (conj)	0.7	2.93	13	29.1	1.3	2.59	6.12	10	1.21	2.41	6.46	11.6	9.6	17.79	37.17	57.72
Conjugate action	0.2	1.62	6.11	11.36	1.85	2.54	4.36	6.57	13.22	18.25	29.89	43.19
Reorientation	1	4.73	17.2	36.5	1.26	2.56	6.22	10.7	1.25	2.4	6.31	11.6	8.78	16.23	33.9	54.1
Interpolation of scalar maps	0.8	4	14.7	31.4	1.23	2.41	5.94	10	11.9	2.26	5.42	9.1	8.94	16.6	36.2	56.92

each subject i , the pairs of scans were bias corrected and symmetrically rigidly aligned by iterative registration to the mean intensity image. The baseline was linearly registered to the MNI standardized template and the parameters of the transformation were applied to I_1^i . Finally, for each subject, the longitudinal changes were measured by non-rigid registration using the Symmetric Log Demons algorithm with regularization parameters $\sigma_{elastic} = 2$ and $\sigma_{fluid} = 1$.

The resulting deformation fields $\varphi_i = \text{Exp}(v_i)$ were transported in a common reference by using the Schild's Ladder (BCH scheme). The group-wise longitudinal progression was modeled as the median of the transported stationary velocity fields. Finally, the areas of significant longitudinal expansion/contraction were investigated by one-sample t-test on the group of Log-Jacobian maps corresponding

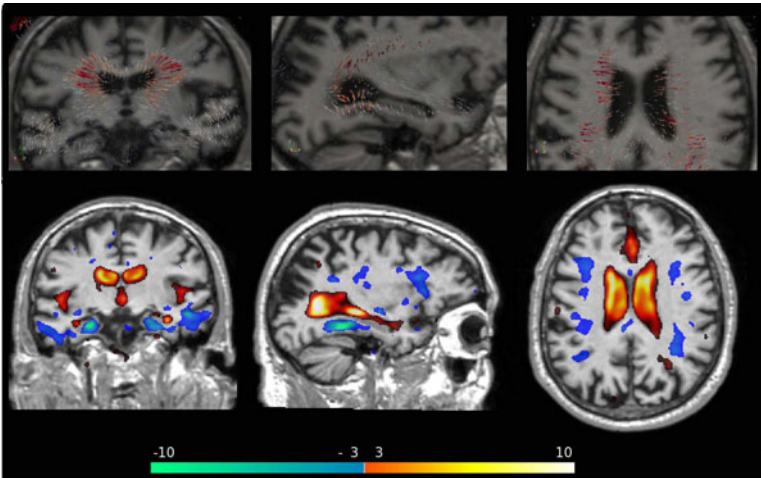


Fig. 4. One year structural changes for 70 Alzheimer's patients. Top row: Median static velocity field. We notice the lateral expansion of the ventricles as well as the contraction in the temporal lobes. Bottom: t-statistic for the areas of Log-Jacobian maps significantly different from 0 ($p < 0.001$ FDR corrected). Blue color: significant contraction, Red color: significant expansions.

to the transported deformations, with the multiple comparisons controlled by a statistical threshold of $p < 0.01$ corrected for false discovery rate (FDR).

Results. Figure 4 shows a detail from the median static velocity field from the transported one-year longitudinal trajectories. The field flows outward from the ventricular surface to indicate a pronounced ventricular lateral enlargement. Moreover, we can notice an expansion in the temporal horns of the ventricles as well as a consistent contracting flow in the temporal lobes. The same effect can be statistically quantified by evaluating the areas where the Log-Jacobian maps were statistically different from zero. The areas of significant expansion are located around the ventricles and spread in the CSF areas, while a significant contraction is appreciable in the temporal lobes, hippocampi, parahippocampal gyrus and in the posterior cingulate. We argue that the highly detailed spatial localization of the results underlines the accuracy of the transport.

4 Discussion and Conclusions

In this study we proposed a novel framework for the transport of longitudinal deformations in a reference space from time series of images. The mathematical formulation was combined with an effective computational scheme in order to provide a reliable and feasible solution for the transport of vector fields. Although designed for transporting vector quantities, the method showed also high accuracy in transporting scalar measures, by preserving smoothness of the corresponding spatial maps and providing high accuracy on numerical evaluations. This is an interesting feature which could increase the power in TBM-like group-wise statistical analysis as well as opening the way to reliable multivariate groupwise analysis. As a perspective, the association of the Schild's Ladder with specific frameworks for the estimation of longitudinal trajectories [7], will allow to model the progression of changes for specific clinical populations consistently along the temporal dimensions by including multiple time points.

Acknowledgments. This work was partially supported by the French ANR (Agence Nationale de la Recherche) "programme blanc" number ANR-09-BLAN-0332.

References

1. Alexander, D., Pierpaoli, C., Basser, P., Gee, J.: Spatial transformations of diffusion tensor magnetic resonance images. *IEEE Transactions on Medical Imaging* 20(11) (2001)
2. Bossa, M., Hernandez, M., Olmos, S.: Contributions to 3D diffeomorphic atlas estimation: Application to brain images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 667–674. Springer, Heidelberg (2007)
3. Bossa, M., Zacur, E., Olmos, S.: On changing coordinate systems for longitudinal tensor-based morphometry. In: *Spatio Temporal Image Analysis Workshop (STIA), MICCAI 2010* (2010)

4. Chung, M., Worsley, K., Paus, T., Cherif, C., Collins, D., Giedd, J., Rapoport, J., Evans, A.: A unified statistical approach to deformation-based morphometry. *NeuroImage* 14(3), 595–606 (2001)
5. Ehlers, J., Pirani, F., Schild, A.: The geometry of free fall and light propagation, in O’Raifeartaigh. *General Relativity, Papers in Honor of J. L. Synge*. Oxford University Press, Oxford (1972)
6. Kheifets, A., Miller, W., Newton, G.: Schild’s ladder parallel transport for an arbitrary connection. *International Journal of Theoretical Physics* 39(12), 41–56 (2000)
7. Lorenzi, M., Ayache, N., Frisoni, G., Pennec, X.: 4D registration of serials brain’s MR images: A robust measure of changes applied to Alzheimer’s disease. In: *Spatio Temporal Image Analysis Workshop (STIA), MICCAI* (2010)
8. Miller, M., Trounevé, A., Younes, L.: On the metrics and euler-lagrange equations of computational anatomy. *Annu. Rev. Biomed. Rev.* 4, 375–405 (2002)
9. Qiu, A., Younes, L., Miller, M., Csernansky, J.: Parallel transport in diffeomorphisms distinguish the time-dependent pattern of hippocampal surface deformation due to healthy aging and dementia of the alzheimer’s type. *Neuroimage* 40(1) (2008)
10. Rao, A., Chandrashekhara, R., Sanchez-Hortiz, G., Mohiaddin, R., Aljabar, P., Hajnal, J., Puri, B., Rueckert, D.: Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE Transactions on Medical Imaging* 23(9) (2004)
11. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: A demons-based approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)
12. Younes, L.: Jacobi fields in groups of diffeomorphisms and applications. *Q. Appl. Math.* (2007)

Dissimilarity-Based Classification of Anatomical Tree Structures

Lauge Sørensen¹, Pechin Lo¹, Asger Dirksen²,
Jens Petersen¹, and Marleen de Bruijne^{1,3}

¹ The Image Group, Department of Computer Science,
University of Copenhagen, Denmark

lauges@diku.dk

² Department of Respiratory Medicine, Gentofte University Hospital, Denmark

³ Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, The Netherlands

Abstract. A novel method for classification of abnormality in anatomical tree structures is presented. A tree is classified based on direct comparisons with other trees in a dissimilarity-based classification scheme. The pair-wise dissimilarity measure between two trees is based on a linear assignment between the branch feature vectors representing those trees. Hereby, localized information in the branches is collectively used in classification and variations in feature values across the tree are taken into account. An approximate anatomical correspondence between matched branches can be achieved by including anatomical features in the branch feature vectors. The proposed approach is applied to classify airway trees in computed tomography images of subjects with and without chronic obstructive pulmonary disease (COPD). Using the wall area percentage (WA%), a common measure of airway abnormality in COPD, as well as anatomical features to characterize each branch, an area under the receiver operating characteristic curve of 0.912 is achieved. This is significantly better than computing the average WA%.

Keywords: dissimilarity representation, linear assignment, tree classification, tree dissimilarity measure.

1 Introduction

We consider the task of classifying abnormality in anatomical tree structures in medical images. Examples of such structures and applications are: quantification of airway tree abnormalities in, e.g., chronic obstructive pulmonary disease (COPD) [7, 11, 21] or asthma [9], quantification of arterial stenosis in the circle of Willis in the brain [18], and quantification of retinopathy of prematurity in retinal vessels [8]. The goal is to relate the global state of the anatomical tree to the disease status based on local information in the branches of the tree. This goal is often achieved by considering only parts of the tree and computing an average feature across the branches in those parts [9, 8, 18, 7, 11, 21].

Abnormalities may be localized in different parts of the tree, and the underlying normal structure of the branches often varies across the tree. The value range of the feature used to capture the local information will therefore often be different for different parts of the tree, and the discriminative ability of a feature averaged across several branches may be limited. There may also be a functional difference between a tree containing one severely abnormal branch compared to a tree with several mildly abnormal branches, which can, however, give rise to the same average feature value.

In this paper, we propose to learn from data the relationship between the collective set of branches representing a tree and the global state of that tree. An anatomical tree is considered as a set of branches each characterized by one or more features, and a global label, the disease status of the subject, is associated with the entire tree. Classification is based on direct comparisons between trees, or sets of branch feature vectors, using a dissimilarity-based classification scheme [15]. The feature values extracted in individual branches are hereby collectively taken into account and related to the global state of the tree. A globally optimal one-to-one matching, according to a linear assignment [22], of the branches in two trees is used as anatomical tree dissimilarity measure on which the classification is based. More information is used compared to using an average feature since individual branches are compared, taking localized information into account which is combined into an overall tree dissimilarity measure.

Dissimilarity-based classification has previously proven useful in medical imaging, e.g., for detection of tuberculosis in chest radiographs [2] and for classification of emphysematous tissue in computed tomography (CT) [19]. A similar approach to the one proposed in this paper has previously been successfully applied to classify subjects using texture information in patches of lung tissue from CT images [20]. In the current paper, we consider tree structures. We experiment with different levels of “rigidness” by matching the branches solely based on an abnormality feature and by matching the branches using both abnormality and anatomical features. The latter approach ensures an approximate anatomical correspondence between matched branches. This is in contrast to tree proximity measures based on tree matching [3,5], which are more computationally demanding.

The dissimilarity-based classification scheme is applied to detect COPD in volumetric pulmonary CT images using automatically segmented airway tree branches. COPD is a major public health problem that is projected to become the third leading cause of death worldwide by 2020 [17]. Pulmonary function tests (PFTs), which are the current gold standard for diagnosing and monitoring COPD [17], are insensitive to early stages of the disease and lack reproducibility, and there is therefore a general interest in finding good image-based markers for COPD. In this paper, the focus is on airway disease, one of the main components of COPD, that is mainly characterized by a narrowing of the smaller airways leading to limitations of the airflow to and from the alveoli. This narrowing is caused by inflammation and fibrosis, resulting in thicker airway walls, as well as mucus, blocking the passage of air in the lumen.

The proposed approach is evaluated and compared to computing the average wall area percentage (WA%) across the branches in the tree. This is a commonly used CT-based marker of airway disease in the clinical literature that, despite its simplicity, have shown a significant correlation with PFTs [7, 11, 21].

2 Classification in Tree Dissimilarity Space

An anatomical tree $X = \{\mathbf{x}_i\}_m$ can be represented by a set of m branch feature vectors \mathbf{x}_i , each feature describing the branch shape or appearance and/or the anatomical position of a branch. A label Y is associated with the entire tree. The task is to classify an unseen tree X based on a labeled training set of n trees $\{(X_i, Y_i)\}_n$. Probabilistic classification outputs are considered as these can be directly interpreted as a probability of disease.

2.1 Dissimilarity Representation

A dissimilarity representation is used to classify the trees [15]. From the matrix of pair-wise tree dissimilarities $D = [d(X_i, X_j)]_{n \times n}$ computed from the training set $T = \{X_i\}_n$, there are different ways to derive a feature vector space where traditional vector space methods can be applied. In this work, we consider tree dissimilarity measures $d(\cdot, \cdot)$ that do not produce positive definite matrices, and the dissimilarity space approach, which does not require fulfillment of this condition, is therefore used [15]. An anatomical tree dissimilarity space is constructed of dimension equal to the size of the training set. Each feature vector describing a tree X consists of the dissimilarities $d(X, X_i)$ to each of the trees in the training set, i.e., $D(X, T) = [d(X, X_1), d(X, X_2), \dots, d(X, X_n)]$.

2.2 Dissimilarity Measure

The anatomical tree dissimilarity measure $d(\cdot, \cdot)$ is the crucial component in this construction, and it is in this part of the proposed approach that a set of branch feature vectors are collectively considered as representing a tree. The dissimilarity measure based on a linear assignment between sets of sub-objects proposed in [20] is used for this purpose. The dissimilarity between two anatomical trees, or sets of branches, $X_1 = \{\mathbf{x}_{1i}\}_m$ and $X_2 = \{\mathbf{x}_{2j}\}_n$ where \mathbf{x}_{1i} is the i th branch feature vector in X_1 , is expressed as the minimum linear sum assignment between the two sets where the cost of assigning \mathbf{x}_{1i} to \mathbf{x}_{2j} is the dissimilarity between those two branch feature vectors according to a branch dissimilarity measure $\Delta(\mathbf{x}_{1i}, \mathbf{x}_{2j})$. This can be seen as assigning all the branches in the smallest tree to the branches in the other tree in a way such that the two trees are as similar as possible while only allowing one-to-one matchings. The linear assignment problem can be formulated in terms of a bipartite graph. However, it is important to note that this construction is only used in order to compute the linear assignment between the sets of branches and that no structure is imposed on the trees.

Let $G = (X_1 \cup X_2, E)$ be a weighted undirected bipartite graph with node sets X_1 and X_2 where $|X_1| = |X_2| = n$, edge set $E = \{\{\mathbf{x}_{1i}, \mathbf{x}_{2j}\} : i, j = 1, \dots, n\}$, and with a weight $\Delta(\mathbf{x}_{1i}, \mathbf{x}_{2j})$ associated with each edge $\{\mathbf{x}_{1i}, \mathbf{x}_{2j}\} \in E$. A subset M of E is called a perfect matching, or assignment, if every node of G is incident with exactly one edge in M . The perfect matching with minimum weight M^* is given by

$$M^* = \operatorname{argmin}_M \sum_{\{\mathbf{x}_{1i}, \mathbf{x}_{2j}\} \in M} \Delta(\mathbf{x}_{1i}, \mathbf{x}_{2j}) : M \text{ is a perfect matching.} \quad (1)$$

This problem can be solved efficiently using the Hungarian algorithm [22]. The resulting anatomical tree dissimilarity measure is thus

$$d_{la}(X_1, X_2) = \sum_{\{\mathbf{x}_{1i}, \mathbf{x}_{2j}\} \in M^*} \Delta(\mathbf{x}_{1i}, \mathbf{x}_{2j}) \quad (2)$$

where M^* is obtained via (1).

Segmented anatomical trees are likely to contain differing number of branches, i.e., $|X_1| \neq |X_2|$, and the above formulation does not account for this. However, a problem with different number of branches in the two trees can be turned into a problem with the same number of branches by adding “dummy nodes” to the graph [22]. The resulting problem is now in the form of (1) and can again be solved using the Hungarian algorithm. Differences in tree sizes, in terms of number of branches, are accounted for by dividing by the number of branches in the smallest tree, in order not to favor smaller trees

$$d_{lan}(X_1, X_2) = \frac{d_{la}(X_1, X_2)}{\min(|X_1|, |X_2|)}. \quad (3)$$

A branch dissimilarity measure $\Delta(\cdot, \cdot)$ is needed in order to compute the anatomical tree dissimilarity measure $d_{lan}(\cdot, \cdot)$. In this work, branches are represented by a d -dimensional feature vector $\mathbf{x}_i = [x_{i1}, \dots, x_{id}]^T$, and the branch dissimilarity is computed as the Euclidean distance in the branch feature space $\Delta(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_2$.

2.3 Classification

Trees are classified by mapping into the anatomical tree dissimilarity space using $D(\cdot, T)$ followed by classification using the k nearest neighbor (k NN) classifier posterior probability estimate with the training set trees as prototypes

$$p(Y|D(X, T)) = \frac{k_Y(D(X, T))}{k} \quad (4)$$

where $k_Y(D(X, T))$ is the number of nearest neighbors of X , in the dissimilarity space, belonging to class Y out of a total of k nearest neighbors. The Euclidean distance between the feature vector representation of the trees of X and X_i in the dissimilarity space, $\|D(X, T) - D(X_i, T)\|_2$, is used as distance in the k NN classifier. k is fixed according to $k = \sqrt{n}$ where n is the number of prototypes in the k NN classifier [10].

3 Extraction of Airway Tree Branch Feature Vectors

In order to extract the branch feature vectors representing each airway tree, the whole airway tree is segmented in the CT image into lumen and outer wall using the algorithm described in [16]. Figure 1(a) shows an example of a segmented airway tree using this algorithm. The segmentation is subdivided into branches and branch centerlines are extracted using the algorithm described in [11]. Figure 1(b) shows a coloring of the identified branches using this algorithm. Branch generations are obtained by assigning generation 0 to the trachea and incrementing generation number by one when propagating the generation number from a parent centerline to its child centerlines. All the steps taken are fully automatic, and the reader is referred to [11,16] for further details.

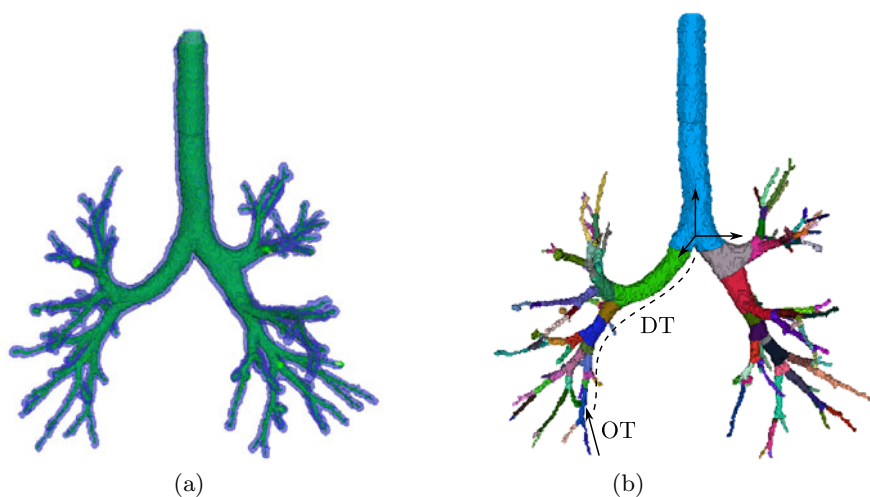


Fig. 1. (a) Airway segmentation with the inner airway wall shown in green and the outer airway wall shown in transparent blue. (b) Branches found in the airway segmentation shown in (a) overlaid with anatomical features.

Each branch is represented by a 5-dimensional feature vector \mathbf{x}_i comprising one measure known to be related to COPD as well as four anatomical features roughly capturing the location and orientation of the branch in the airway tree. The features are the following:

COPD feature

- wall area percentage (WA%). The number of voxels in the branch wall relative to the total amount of voxels in a branch, i.e., the number of voxels in both the wall and the lumen. WA% is computed from the segmented airway tree (lumen and outer wall). This is the most commonly used measure of airway disease in CT [12,17,11,21]. It is also used in asthma [9]. WA% is

computed from cross-sectional slices in the mentioned references whereas it is computed from volumetric segmentations of the branches in this paper. The reader is referred to [16] for details.

Anatomical features. The anatomical features are computed from the centerline representation of the tree.

- distance to the trachea (DT). The distance from the root of the branch to the carina along the centerlines connecting that branch to the trachea. DT is scaled w.r.t. the subject size by dividing by the height of the subject. This feature can be seen as a continuous version of the branch generation, and is less affected by errors in the segmentation and branch detection.
- orientation w.r.t. the trachea (OT). The orientation of the branch in a global coordinate system based on the trachea, carina, and main bronchi. This is expressed as a three-dimensional vector.

The global coordinate system as well as the anatomical features for a specific branch are illustrated in Fig. 1(b). Each branch feature is standardized to zero mean and unit variance.

4 Experiments

4.1 Data

The data consists of 296 low-dose volumetric CT images from 296 participants of the Danish Lung Cancer Screening Trial [14] with the following scan parameters: tube voltage 120 kV, exposure 40 mAs, slice thickness 1 mm, and in-plane resolution ranging from 0.72 to 0.78 mm. 144 images are from subjects without COPD and 152 images are from subjects diagnosed with mild to very severe COPD according to spirometry [17].

4.2 Evaluation

The proposed anatomical tree classification approach is evaluated using leave-one-out estimation on the CT data, and the area under the receiver operating characteristic (ROC) curve (AUC) is computed on the soft outputs of the classifier of Eqn. (4). The anatomical tree dissimilarity space considered in each leave-out trial is of dimension equal to the size of the training set, i.e., 295-dimensional. The proposed approach is compared to average WA%, and the analysis is performed for different individual generations in the airway tree in the range 3 to 7 and for several ranges from 3-4 to 3-7. Similar generations were considered in [1,7,6].

Different feature representations of the branches are considered, i.e., using WA% only as well as using WA% jointly with anatomical features in order to enforce an approximate anatomical correspondence in the matching.

Table 1. AUCs for COPD diagnosis for the proposed approach (top part) and for average WA% (bottom part). The best performing measure in each column is marked in bold-face, the best performing measure for each approach both for using an individual generation and a range of generations is underlined, and the overall best performing measure is marked in italics.

approach	generation					generations range			
	3	4	5	6	7	3-4	3-5	3-6	3-7
<i>d_{tan}</i>									
WA%	0.792	0.856	0.813	0.883	0.884	0.840	0.865	0.878	0.892
WA%+DT+OT	0.828	0.867	0.864	0.897	0.880	0.868	0.881	0.904	<i>0.912</i>
DT+OT	0.678	0.659	0.739	0.891	0.871	0.706	0.857	0.893	0.908
average feature									
WA%	0.811	<u>0.858</u>	0.838	0.762	0.704	<u>0.849</u>	0.845	0.818	0.788

Table 2. Distribution of how often branches of different generations in the range 3 to 7 are matched in the whole data set when considering WA% and when considering WA% together with anatomy, respectively. The percentages of a total of 43660 matches are reported, and the total number of branches in generations 3 - 7 are as follows: 2467, 4927, 8633, 10004, 7366. Note that these matrices are symmetric.

	WA%					WA%+DT+OT				
	3	4	5	6	7	3	4	5	6	7
3	1.6					5.1				
4	1.8	3.1				1.8	6.6			
5	1.7	3.9	6.9			0.8	4.4	10.2		
6	1.3	3.5	7.8	10.0		0.3	2.1	7.8	11.7	
7	0.9	2.5	5.7	7.4	5.4	0.1	0.8	3.6	7.5	8.1

4.3 Results

The results are reported in Table 1. The best overall performing measure is based on matching branches in generations 3 to 7 using WA% as well as anatomical features. This is significantly better than using the average WA% in all considered generation ranges ($p < 10^{-4}$) according to DeLong, DeLong, and Clarke-Pearson’s test [4]. It is also significantly better than matching branches in generations 3 to 7 using WA% without implicit correspondence by anatomical features ($p = 0.001$), but not significantly better than using anatomical features only ($p = 0.613$). ROC-curves for the underlined AUCs in Table 1 are shown in Fig. 2(a).

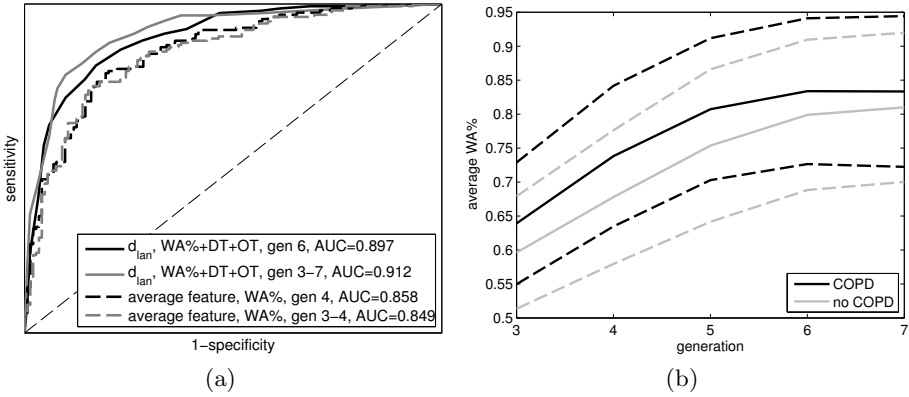


Fig. 2. (a) ROC-curves using the best performing generation and generations range for each approach corresponding to the underlined AUCs in Table 1 (b) WA% as a function of generation for the two groups. Solid line is the mean and dashed lines are the mean \pm the standard deviation.

Table 3. AUCs for COPD diagnosis on the pruned trees for the proposed approach (top part) and for using an average COPD feature (bottom part). The best performing measure in each column is marked in bold-face, the best performing measure for each approach both for using an individual generation and a range of generations is underlined, and the overall best performing measure is marked in italics. Results are not reported for generation 6 and 7 since some pruned trees did not contain branches in these generations.

approach	generation					generations range			
	3	4	5	6	7	3-4	3-5	3-6	3-7
<i>d_{lan}</i>									
WA%	0.792	0.856	0.813	—	—	0.840	0.848	0.861	0.859
WA%+DT+OT	0.828	<u>0.867</u>	0.864	—	—	0.868	0.885	0.886	<i>0.887</i>
DT+OT	0.678	0.659	0.739	—	—	0.705	0.759	0.750	0.747
average feature									
WA%	0.811	<u>0.858</u>	0.814	—	—	0.849	0.861	<u>0.862</u>	<u>0.862</u>

The effect on the matching in Eqn.(1) of including anatomical branch features, in terms of how often a branch from one generation is matched to other generations, is inspected in Table 2. More branches from the same generations are matched when including anatomical features compared to using WA% only, i.e., a larger percentage of the matches is concentrated in the diagonal of the matrix for WA%+DT+OT.

5 Discussion

Average WA% performed best at generation 4. Considering more generations, in isolation or as a range, deteriorated the performance, see Table 1. WA% on average grows as a function of generation, as seen in Fig. 2(b). This is due to anatomy, but also likely due to the resolution limit of the CT image being reached leading to size overestimation because of partial volume effects. This variation in feature value across the tree may explain why average WA% deteriorates for larger generations, i.e., the discriminative information is lost. The proposed approach naturally deals with this phenomenon by matching individual branches and can therefore incorporate branches from larger generations while preserving discriminative information. This may explain the significantly better performance.

The segmentation algorithm generally finds more branches in airway trees from the no COPD group compared to the COPD group in the data set. This skew in the number of branches may bias the results since partial matchings are allowed in Eqn. (3). This could explain why using anatomical branch features alone performs well, see Table 1. We inspect this phenomenon by pruning the smallest branches in all the trees in the data set to the size of the smallest tree. The leave-one-out estimation is then repeated for the pruned trees, and the results are shown in Table 3. As seen, it is still possible to achieve high AUCs, and the proposed approach achieves the best AUC of 0.887 which is still significantly better than average WA% for all the considered generation ranges ($p < 10^{-4}$). Moreover, the performance of using only anatomical features has become significantly worse than average WA% in all cases ($p < 10^{-4}$).

The computational complexity of the proposed anatomical tree dissimilarity measure $d_{lan}(X_1, X_2)$ is bounded by the complexity of the algorithm used to compute the linear assignment. Using the Hungarian algorithm this is $\mathcal{O}(m^4)$ [22] where m is the number of branches in the larger tree, i.e., $m = \max(|X_1|, |X_2|)$. This is in contrast to NP-complete tree proximity measures that take the tree topology into account, such as [3, 5]. A test tree X is classified by mapping into the anatomical tree dissimilarity space which has complexity $\mathcal{O}(nm^4)$ where n is the size of the training set, i.e., $n = |T|$ and m in this case is the number of branches in the largest three among all training set trees as well as the test tree, i.e., $m = \max(|X|, |X_1|, |X_2|, \dots, |X_n|)$, $X_i \in T$. This is followed by classification in the dissimilarity space using a trained classifier.

The relative contribution of airway disease and emphysema, the other main component of COPD, to COPD varies [7] and may in fact be independent [13]. It would therefore be interesting to combine the current approach with measures of emphysema.

6 Conclusions

We propose a novel method for classification of abnormality in anatomical tree structures that is based on dissimilarities computed directly between anatomical trees as represented by sets of branch feature vectors. This is an alternative

to a common approach in the clinical literature of averaging one branch feature or measuring a small number of specific locations. A k NN classifier in the dissimilarity space obtained using a tree dissimilarity measure based on a linear assignment between the branches in the trees using WA% together with anatomical features, achieved an AUC of 0.912 on a COPD classification task in volumetric pulmonary CT. This was significantly better than the AUC of using the average WA%.

Acknowledgements. This work is partly funded by the Netherlands Organisation for Scientific Research (NWO), and AstraZeneca, Sweden.

References

1. Achenbach, T., Weinheimer, O., Biedermann, A., Schmitt, S., Freudenstein, D., Goutham, E., Kunz, R.P., Buhl, R., Dueber, C., Heussel, C.P.: MDCT assessment of airway wall thickness in COPD patients using a new method: correlations with pulmonary function tests. *Eur. Radiol.* 18(12), 2731–2738 (2008), <http://dx.doi.org/10.1007/s00330-008-1089-4>
2. Arzhaeva, Y., Hogeweg, L., de Jong, P.A., Viergever, M.A., de Ginneken, B.v.: Global and local multi-valued dissimilarity-based classification: Application to computer-aided detection of tuberculosis. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 724–731. Springer, Heidelberg (2009)
3. Bille, P.: A survey on tree edit distance and related problems. *Theor. Comput. Sci.* 337, 217–239 (2005), <http://dx.doi.org/10.1016/j.tcs.2004.12.030>
4. DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L.: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3), 837–845 (1988)
5. Feragen, A., Lauze, F., Lo, P., de Bruijne, M., Nielsen, M.: Geometries on spaces of treelike shapes. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part II. LNCS, vol. 6493, pp. 160–173. Springer, Heidelberg (2011)
6. Hasegawa, M., Makita, H., Nasuhara, Y., Odajima, N., Nagai, K., Ito, Y., Betsuyaku, T., Nishimura, M.: Relationship between improved airflow limitation and changes in airway calibre induced by inhaled anticholinergic agents in COPD. *Thorax* 64(4), 332–338 (2009), <http://dx.doi.org/10.1136/thx.2008.103671>
7. Hasegawa, M., Nasuhara, Y., Onodera, Y., Makita, H., Nagai, K., Fuke, S., Ito, Y., Betsuyaku, T., Nishimura, M.: Airflow limitation and airway dimensions in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 173(12), 1309–1315 (2006), <http://dx.doi.org/10.1164/rccm.200601-0370C>
8. Heneghan, C., Flynn, J., O’Keefe, M., Cahill, M.: Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Med. Image Anal.* 6(4), 407–429 (2002)
9. Kasahara, K., Shiba, K., Ozawa, T., Okuda, K., Adachi, M.: Correlation between the bronchial subepithelial layer and whole airway wall thickness in patients with asthma. *Thorax* 57(3), 242–246 (2002)
10. Kittler, J., Alkoot, F.M.: Moderating k -NN classifiers. *Pattern Analysis and Applications* 5(3), 326–332 (2002)

11. Lo, P., van Ginneken, B., Reinhardt, J., de Bruijne, M.: Extraction of airways from CT (EXACT 2009). In: Brown, M., de Bruijne, M., van Ginneken, B., Kiraly, A., Kuhnigk, J.M., Lorenz, C., McClelland, J., Mori, K., Reeves, A., Reinhardt, J. (eds.) Proc. of The Second International Workshop on Pulmonary Image Analysis (2009)
12. Nakano, Y., Wong, J.C., de Jong, P.A., Buzatu, L., Nagao, T., Coxson, H.O., Elliott, W.M., Hogg, J.C.: The prediction of small airway dimensions using computed tomography. *Am. J. Respir. Crit. Care Med.* 171(2), 142–146 (2005), <http://dx.doi.org/10.1164/rccm.200407-8740C>
13. Patel, B.D., Coxson, H.O., Pillai, S.G., Agustí, A.G.N., Calverley, P.M.A., Donner, C.F., Make, B.J., Müller, N.L., Rennard, S.I., Vestbo, J., Wouters, E.F.M., Hiorns, M.P., Nakano, Y., Camp, P.G., Fauerbach, P.V.N., Screaton, N.J., Campbell, E.J., Anderson, W.H., Paré, P.D., Levy, R.D., Lake, S.L., Silverman, E.K., Lomas, D.A., Network, I.C.G.: Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 178(5), 500–505 (2008)
14. Pedersen, J.H., Ashraf, H., Dirksen, A., Bach, K., Hansen, H., Toennesen, P., Thorsen, H., Brodersen, J., Skov, B.G.: The Danish randomized lung cancer CT screening trial—overall design and results of the prevalence round. *J. Thorac. Oncol.* 4(5), 608–614 (2009), <http://dx.doi.org/10.1097/JTO.0b013e3181a0d98f>
15. Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence). World Scientific Publishing Co., Inc., Singapore (2005)
16. Petersen, J., Nielsen, M., Lo, P., Saghir, Z., Dirksen, A., de Bruijne, M.: Optimal graph based segmentation using flow lines with application to airway wall segmentation. In: Székely, G., Hahn, H. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 49–60. Springer, Heidelberg (2011)
17. Rabe, K.F., Hurd, S., Anzueto, A., Barnes, P.J., Buist, S.A., Calverley, P., Fukuchi, Y., Jenkins, C., Rodriguez-Roisin, R., van Weel, C., Zielinski, J.: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am. J. Respir. Crit. Care Med.* 176(6), 532–555 (2007)
18. Roher, A.E., Esh, C., Kokjohn, T.A., Kalback, W., Luehrs, D.C., Seward, J.D., Sue, L.I., Beach, T.G.: Circle of willis atherosclerosis is a risk factor for sporadic alzheimer's disease. *Arterioscler Thromb Vasc. Biol.* 23(11), 2055–2062 (2003), <http://dx.doi.org/10.1161/01.ATV.0000095973.42032.44>
19. Sørensen, L., de Bruijne, M.: Dissimilarity representations in lung parenchyma classification. In: Karssemeijer, N., Giger, M. (eds.) Medical Imaging: Computer-Aided Diagnosis. Proceedings of SPIE, vol. 7260 (2009)
20. Sørensen, L., Loog, M., Lo, P., Ashraf, H., Dirksen, A., Duin, R.P.W., de Bruijne, M.: Image dissimilarity-based quantification of lung disease from CT. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 37–44. Springer, Heidelberg (2010)
21. Washko, G.R., Dransfield, M.T., Estépar, R.S.J., Diaz, A., Matsuoka, S., Yamashiro, T., Hatabu, H., Silverman, E.K., Bailey, W.C., Reilly, J.J.: Airway wall attenuation: a biomarker of airway disease in subjects with COPD. *J. Appl. Physiol.* 107(1), 185–191 (2009), <http://dx.doi.org/10.1152/jappphysiol.00216.2009>
22. Wolsey, L.A.: Integer Programming. Wiley Interscience, Hoboken (1998)

Automated Detection of Junction Structures and Tracking of Their Trajectories in 4D Images

Guanglei Xiong¹ and Lei Xing²

¹ Biomedical Informatics Program ² Department of Radiation Oncology,
Stanford University, Stanford, CA 94305, USA
{glxiong, lei}@stanford.edu

Abstract. Junction structures, as the natural anatomical markers, are useful to study the organ or tumor motion. However, detection and tracking of the junctions in four-dimensional (4D) images are challenging. The paper presents a novel framework to automate this task. Detection of their centers and sizes is first achieved by an analysis of local shape profiles on one segmented reference image. Junctions are then separately tracked by simultaneously using neighboring intensity features from all images. Defined by a closed B-spline space curve, the individual trajectory is assumed to be cyclic and obtained by maximizing the metric of combined correlation coefficients. Local extrema are suppressed by improving the initial conditions using random walks from pair-wise optimizations. Our approach has been applied to analyze the vessel junctions in five real 4D respiration-gated computed tomography (CT) image datasets with promising results. More than 500 junctions in the lung are detected with an average accuracy of greater than 85% and the mean error between the automated and the manual tracking is sub-voxel.

1 Introduction

Accurate physiological motion information is valuable in many clinical applications. Junction structures, when present, are naturally the anatomical markers to characterize the realistic organ or tumor motion. In the lung for example, vessel and airway bifurcations are abundant and their respiration-induced motions are visually apparent using 4D CT, developed for diagnosis and radiation therapy. While metal seeds are practically used as the fiducial markers for tumor tracking, bifurcations can ideally be used as the non-invasive markers in both the planning and treatment stages. Unfortunately, detection and tracking of junction structures in 4D data is challenging due to its large number, wide range of size and shape, inhomogeneous and hysteretic motion, which are even complicated in the presence of imaging noises and artifacts. These challenges make manual processing intractable and hence call for automation.

Most of the previous work addressed the problems of detection and tracking separately. The typical approach for detection is to first extract the centerline using some thinning operation on the segmented image and then consider centerline nodes with three or more neighbors as junctions [1]. While straightforward, this may lead to several drawbacks, such as no sizing information, false positives caused by small bulges, and unnecessary computational cost for the whole centerline. Tools to

facilitate extracting a large set of landmarks manually [2] or semi-automatically [3] has also been developed for the purpose of evaluation of image registration methods. On the other hand, tracking is simply achieved by propagating manually delineated junctions in one reference image using the deformation fields from image registration, either between consecutive or all-to-one image pairs [4], or utilizing all images simultaneously, in which temporal smoothing constraints are applied [5]. We argue that individual trajectories obtained in this manner may not be optimal since the registration metric is defined over a larger region. There are studies of particular interest to track each junction separately. Tashiro et al. [6] matched automatically identified junctions on two consecutive images using a probabilistic relaxation method. Other matching-based approaches were also proposed in [7]. However, the rate of mismatching led by ambiguities limits the accuracy and how to operate on all images together is unclear. Castillo et al. [8] proposed to extract the trajectories of a coarse grid of voxels based on a compressible optical flow algorithm. But the trajectory was described a general polynomial and modeled only during the expiration. Therefore, no cyclic constraint was imposed and the curve may oscillate due to the Runge's phenomenon if the order is too high.

In this paper, we propose a novel framework to both detect and track junction structures in 4D images. Our approach contributes in several aspects. First in Section 2.1, an insight on local shape profiles of junctions is made and utilized to develop an automated detection method that can overcome the aforementioned drawbacks of the centerline-based approach. Second in Section 2.2, each identified junction is tracked separately. The trajectory is defined by a closed B-spline space curve, with imposed cyclic constraints and guaranteed no oscillations. Third in Section 2.3, a metric that combines the correlation coefficients from all images is used to measure how well the image intensities on the trajectory match. Correlation coefficient is employed since it is bounded and straightforward to indicate the reliability of a tracked trajectory. Finally in Section 2.4, an optimization strategy based on the L-BFGS method is employed and tested to optimize the metric parameterized by the control points. Local extrema, if found, are suppressed by improving the initial conditions by random walks from pair-wise optimizations. The results and discussions are in Sections 3 and 4.

2 Methods

2.1 Junction Detection by Analysis of Local Shape Profile

As a prerequisite, we assume the reference image in the 4D data is preprocessed and an object of interest (e.g. vessel or airway trees, possibly composed of several components) is segmented [9]. Note only one image needs to be segmented in our approach. Our goal is to search for the center and size (the minimal distance from the center to the boundary) of every junction. The basic idea for junction detection comes from an insight of the unique characteristics of the local shape profile of a junction. There are three or more disconnected regions inside the spherical shell between two concentric spheres of appropriate sizes at the center of a junction, whereas two or less

at other locations (e.g. near/off the center of a tube, near/off the center of an end, and off the center of a junction). Fig. 1(a) illustrates several typical cases in 2D.

To analyze the local shape profiles, we plot the maximal distance d of a ray traveling within the object in the direction θ at different locations in Fig. 1(b). The corresponding inner r_{in} and outer r_{out} radii of the two bounding spheres in Fig. 1(a) are marked as dashed lines. It is important to determine the appropriate r_{in} and r_{out} that robustly differentiate junctions from other structures for a variety of scales. Let $d_{min} = \min_{\theta \in [0, 2\pi)} d(\theta)$ and it indicates the scale of the structure. Then, $r_{in} \geq d_{min}$, because there is always only one disconnected region otherwise. To adapt to different scales, simple linear relations of $r_{in} = \alpha_{in}d_{min}$ and $r_{out} = \alpha_{out}d_{min}$ are chosen to make the spherical shell captures the size of a local shape, where α_{in} and α_{out} are user-defined constants. Furthermore, α_{in} should be larger than 1 by an amount corresponding to the tolerance of small bulges. In addition, α_{out} should be larger than α_{in} considering the implementation of discrete images but not too large to save the computation cost. We empirically use $\alpha_{in} = 3.0$ and $\alpha_{out} = 4.0$.

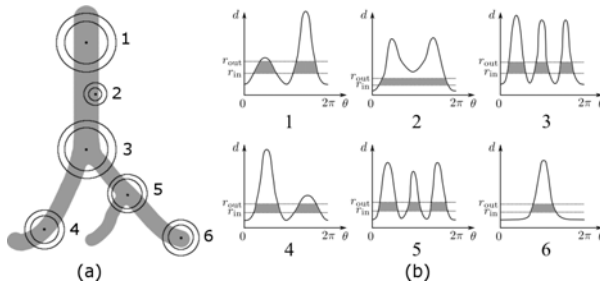


Fig. 1. (a) Typical cases of concentric spheres at different locations; (b) their corresponding local shape profiles and disconnected regions between inner and outer radii

Based on the insight of local shape profile, each candidate voxel within the object of interest is checked whether it is a junction. d_{min} for all candidates can be obtained in one pass using a fast distance transform method [10]. Unlike in Fig. 1(a), Euclidean distance can lead to a wrong number of disconnected regions when there is another non-contacting object between the bounding spheres. Instead, the minimal distance from the candidate is used. It is computed using the fast marching method seeded at the candidate [11], modified to keep track of voxels inside the spherical shell. Then, these voxels are assigned to different regions based on the connectedness. As mentioned before, the candidate with three or more disconnected region is considered as a junction point. It should be noted that multiple candidates can occur for a junction due to similar profiles among voxels in the center of a junction. Fortunately, those for a given junction are adjacent to each other and can be easily clustered using the mean shift method [12]. The candidate with the largest d_{min} in each cluster is chosen as the center of the junction and its d_{min} is the size of the junction.

To speed up the detection, it is not necessary to consider all but only the voxels close to the centerline. Thus, we narrow the set of candidates down as $\mathcal{C} = \{v \in \mathcal{S} \mid \mathcal{D}_x(v) \vee \mathcal{D}_y(v) \vee \mathcal{D}_z(v)\}$, where \mathcal{S} is the set of all voxels within the object and $\mathcal{D}_x(v) := D_{-x}(v) < D(v) \wedge D_{+x}(v) < D(v)$ with $D(v)$, $D_{-x}(v)$, and

$D_{+x}(v)$ is the distance transform at the voxel v and the adjacent voxels in both directions along x axis. $\mathcal{D}_y(v)$ and $\mathcal{D}_z(v)$ are defined similarly. Furthermore, the test of candidates in \mathcal{C} is independent with each other and is thus parallelizable.

2.2 Trajectory Modeling Using B-Spline Curve

A B-spline curve is a sequence of lower-order polynomial curve segments that joins continuously [13]. It is used here to represent the trajectory that adapts the temporal data while is smoothly regularized to suppress noises and artifacts. Thanks to the convex hull property of the B-spline curve, the oscillation effect by a single higher-order polynomial is completely avoided and the trajectory is completely bounded by the control polygon. Another advantage is that each control vertex only influences locally as a consequence of the compact support of B-spline basis functions. Without loss of generality, fourth-order B-splines with periodic basis functions and uniform knot vectors are considered. In addition, the assumption of cyclic motion from all images of the 4D data is imposed by closing the curve; i.e. the last image is temporally adjacent to the first image. Mathematically, the k -th curve segment of the B-spline curve is

$$\mathbf{C}_k(s) = [b_0(s) \quad b_1(s) \quad b_2(s) \quad b_3(s)] \begin{bmatrix} \mathbf{P}_{(k+0) \bmod m} \\ \mathbf{P}_{(k+1) \bmod m} \\ \mathbf{P}_{(k+2) \bmod m} \\ \mathbf{P}_{(k+3) \bmod m} \end{bmatrix}, \quad k = 1 \dots m \quad (1)$$

$$[b_0(s) \quad b_1(s) \quad b_2(s) \quad b_3(s)] = \frac{1}{6} [s^3 \quad s^2 \quad s \quad 1] \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \quad (2)$$

where, b_h , $h = 0 \dots 3$ are the basis functions with $\sum_h b_h = 1$, \bmod is the modulo, $s \in [0, 1)$ is the time fraction, and m is the number of curve segments, or equally, control vertices \mathbf{P}_l , $l = 1 \dots m$. It determines the flexibility of the curve. Notice that $m \in [4, n]$ is required, where n is the number of images in the 4D data. The choice of m depends on the complexity of the actual trajectory and the temporal resolution of the data. In the presence of noises and artifacts, smaller m is generally preferred when sufficient to avoid over-fitting. We choose $m = 5$ considering $n = 10$ in our case. Fig. 2 shows a B-spline curve with 5 control vertices.

Assuming the images are evenly distributed in time and the interval is unity, the i -th image is at time $t = i$. For any intermediate time $t \in [1, n + 1)$, the corresponding curve segment is $k = \lfloor \frac{m(t-1)}{n} + 1 \rfloor$ and the time fraction within the segment is $s = k - \lfloor \frac{m(t-1)}{n} + 1 \rfloor$. The derivatives of the position at s with respect to the support control vertices are the basis functions and zero for non-support control vertices, i.e.

$$\frac{\partial \mathbf{C}_k(s)}{\partial \mathbf{P}_l} = \begin{cases} b_h(s) & \text{if } l = (k + h) \bmod m, \quad h = 0 \dots 3, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

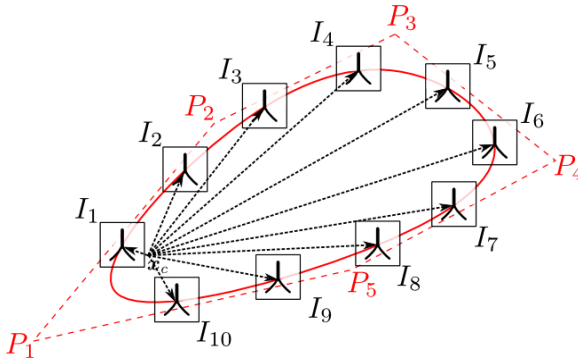


Fig. 2. An example of trajectory modeling with 5 control vertices and tracking using 10 images. The B-spline curve (solid line) and its control polygon (dashed line) are in red. A tracked trajectory passes through corresponding junction centers in all images.

2.3 Trajectory Tracking Using Correlation Coefficients

Let the center of a junction detected in the reference image ($t = r$) be \mathbf{x}_c , the size be d_c , and its motion trajectory during respiration modeled as $\mathbf{T}(\mathbf{x}_c, t) = \mathbf{x}_c + \mathbf{C}_k(s)$. An observation that we utilize to simplify the tracking problem is the shape of an individual junction and the intensity values over its neighborhood region only minimally change during respiration. Therefore, the trajectory of any voxel \mathbf{x}_j in the neighborhood region $\mathcal{N}(\mathbf{x}_c)$ is assumed to be synchronized as that at \mathbf{x}_c and the deviation between them stays constant as in the reference image

$$\mathbf{T}(\mathbf{x}_j, t) - \mathbf{T}(\mathbf{x}_c, t) = \mathbf{x}_j - \mathbf{x}_c, \mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c) \tag{4}$$

The size of the region $\mathcal{N}(\mathbf{x}_c)$ is junction-specific and is chosen to be $\alpha_{\text{neighbor}}d_c$ where $\alpha_{\text{neighbor}} = 4.0$ empirically. As shown in Fig. 2, let $I_i(\mathbf{x}), i = 1 \dots n$ denote the images in the 4D data. Evaluating the intensity at $\mathbf{T}(\mathbf{x}_j, i)$, usually a non-grid position, requires a continuous representation of the image. It is obtained by interpolation using fourth-order B-splines.

We need to define a metric that measures the similarity among the intensity values $I_i(\mathbf{T}(\mathbf{x}_j, i))$ on the trajectory. Unlike the popular choice of mean squared intensity differences, we propose to use the metric as the sum of normalized correlation coefficients with subtracting means. There are two advantages compared to the former. One is the metric is bounded and straightforward to indicate the reliability of a tracked trajectory. The other is that it is insensitive to constant additive intensity fluctuations among the images, which are not uncommon in 4D data, especially considering the neighborhood region at the junction center in our case. Mathematically, the metric with respect to the reference image I_r is

$$M(\mathbf{p}) = \sum_{i=1}^n M_i(\mathbf{p}) \tag{5}$$

$$M_i(\mathbf{p}) = \frac{\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_r(\mathbf{x}_j) - \bar{I}_r) (I_i(\mathbf{T}(\mathbf{x}_j, i)) - \bar{I}_i)}{\sqrt{\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_r(\mathbf{x}_j) - \bar{I}_r)^2 \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_i(\mathbf{T}(\mathbf{x}_j, i)) - \bar{I}_i)^2}} \quad (6)$$

where, \mathbf{p} is a vector of the coordinates of all control vertices $\mathbf{P}_l, l = 1 \dots m$. The intensity means \bar{I}_r and \bar{I}_i are computed in the reference and the i -th image as $\bar{I}_r = \frac{1}{|\mathcal{N}(\mathbf{x}_c)|} \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} I_r(\mathbf{x}_j)$ and $\bar{I}_i = \frac{1}{|\mathcal{N}(\mathbf{x}_c)|} \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} I_i(\mathbf{T}(\mathbf{x}_j, i))$. It is straightforward to show that $-n \leq M(\mathbf{p}) \leq n$. Ideally, the relation $M(\mathbf{p}) = n$ holds if the intensities on the trajectory perfectly correlate. In practice, serious violation of the original observation or the convergence to the local extrema can cause the value of $M(\mathbf{p})$ far away from n . Each $M_i(\mathbf{p})$ also indicates the reliability of a tracked trajectory passing through the i -th image.

Solving the trajectory tracking problem becomes the estimation of the optimal parameters \mathbf{p}^* that maximizes the metric $M(\mathbf{p})$

$$\mathbf{p}^* = \arg \max_{\mathbf{p}} M(\mathbf{p}) \quad (7)$$

Notice that in Eq.(5), the summation does include the contribution from the reference image $M_r(\mathbf{p})$. In other words, the position of the junction at the reference image is not constrained as $\mathbf{T}(\mathbf{x}_c, r) = \mathbf{x}_c$ (as shown in Fig. 2 with $r = 1$), but leaves this particular degrees of freedom to the optimization. It is deliberately used to account for the fact that the trajectory may need to deviate slightly from \mathbf{x}_c due to the smoothing regularization in order to achieve a better overall metric.

2.4 Optimization Strategy

Optimization of Eq. (7) is handled by a Quasi-Newton approach in the form of the limited memory BFGS (L-BFGS) method [14]. The main advantage is due to its high precision and improved convergence rate compared to simple gradient descent algorithms. Another feature of L-BFGS is that simple bounds can be placed to constrain the range of parameters [15]. Thanks to the convex hull property of B-spline curves, the feature can be used to explicitly incorporate a prior knowledge of the scope of the trajectory (e.g. in each of the three dimensions).

We start the optimization with zero initial conditions, i.e. the whole trajectory is simply a single point at \mathbf{x}_c . In most cases, this trivial choice leads to converge to the actual trajectory. However, in some cases, especially for the large movement or poor image quality, the optimization converges to a local extreme, which can be easily detected if the value of the metric $M(\mathbf{p})$ is far away from n . It is necessary to resort to a more heuristic choice, which is closer to the actual trajectory.

As illustrated in Fig. 3, we propose a technique using random walks from pair-wise optimizations. The iterative procedure is as follows: Starting from \mathbf{x}_c at $t = r$, suppose a good match ($M_i(\mathbf{p})$ is close to 1) is found at $t = i$ and the corresponding position is $\mathbf{T}'(\mathbf{x}_c, i)$. To search for the position at the next time step $t = i + 1$, a pair-wise optimization based on L-BFGS can be performed to maximize $M_{i+1}(\mathbf{p})$ in Eq. (6), which involves only $I_r(\mathbf{x}_j)$ and $I_{i+1}(\mathbf{T}(\mathbf{x}_j, i + 1))$, $\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)$. It is simpler to optimize in the pair-wise fashion than globally since the movement of the junction between the consecutive steps is smaller. To further improve the coverage, a number

of random positions in a neighborhood centered at $T'(x_c, i)$ are sampled and used to initialize a series of optimizations. The position with the largest $M_{i+1}(\mathbf{p})$ is then chosen to be $T'(x_c, i + 1)$. The iteration continues until the step returns to r . However, uncertainties in each step may accumulate in this manner. We instead walk halfway in both the forward and backward directions. Finally, the control vertices $P'_l, l = 1 \dots m$ of the B-spline curve that best fit $T'(x_c, i), i = 1 \dots n$ are computed by solving a least squares problem based on Eq. (1). The parameters \mathbf{p}' from this set of control vertices serves as the new initial conditions to repeat the optimization.

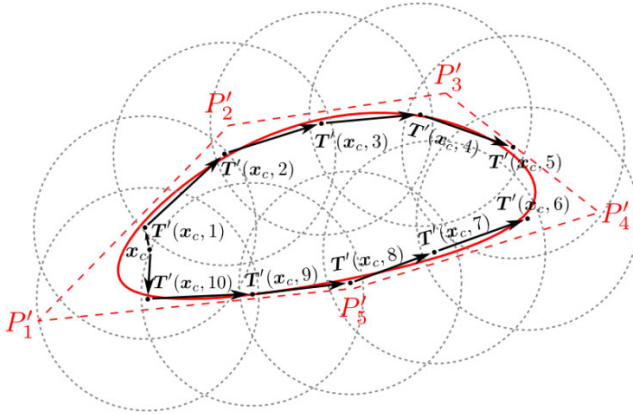


Fig. 3. An example of searching for a more heuristic initial condition by random walks. Random positions in the neighborhood (gray dashed circle) of the previous step are used to initialize pair-wise optimizations. Both forward and backward directions are searched.

L-BFGS requires the derivatives of the metric $M(\mathbf{p})$ with respect to \mathbf{p} is known explicitly, which can be obtained by differentiating Eq. (5) as

$$\frac{\partial M(\mathbf{p})}{\partial \mathbf{p}} = \sum_{i=1}^n \frac{1}{\sqrt{\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_r(\mathbf{x}_j) - \bar{I}_r)^2 \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_i(T(\mathbf{x}_j, i)) - \bar{I}_i)^2}} \left[\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_r(\mathbf{x}_j) - \bar{I}_r) \frac{\partial I_i(T(\mathbf{x}_j, i))}{\partial \mathbf{p}} - \frac{\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_r(\mathbf{x}_j) - \bar{I}_r) (I_i(T(\mathbf{x}_j, i)) - \bar{I}_i)}{\sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_i(T(\mathbf{x}_j, i)) - \bar{I}_i)^2} \sum_{\mathbf{x}_j \in \mathcal{N}(\mathbf{x}_c)} (I_i(T(\mathbf{x}_j, i)) - \bar{I}_i) \frac{\partial I_i(T(\mathbf{x}_j, i))}{\partial \mathbf{p}} \right] \quad (8)$$

where, \bar{I}_r and \bar{I}_i are defined as before. The derivatives $\frac{\partial I_i(T(\mathbf{x}_j, i))}{\partial \mathbf{p}}$ represent the change of the intensity at position $T(\mathbf{x}_j, i)$ on the trajectory with respect to the change of the parameters \mathbf{p} , which is computed using the chain rule as

$$\frac{\partial I_i(\mathbf{T}(\mathbf{x}_j, i))}{\partial \mathbf{p}} = \frac{\partial I_i(\mathbf{T}(\mathbf{x}_j, i))}{\partial \mathbf{T}(\mathbf{x}_j, i)} \frac{\partial \mathbf{T}(\mathbf{x}_j, i)}{\partial \mathbf{p}} \quad (9)$$

The first term is the intensity gradient, which can be evaluated using fourth-order B-splines. The second term is readily available in Eq. (3).

3 Results

We now present the experimental results on the application of our approach to five 4D thoracic CT datasets (DIR-lab) publicly available from University of Texas M. D. Anderson Cancer Center [2]. For each case, there are 10 images with the 1st at the maximum inhale phase and the 5th at the maximum exhale phase. All images are preprocessed (cropped and linearly re-sampled to be isotropic). The 1st image is selected as the reference, in which a lung mask is extracted using seeded thresholding, connected component analysis, hole filling, and morphological operations, which is similar to [16]. Finally, blood vessels inside the lung are segmented using a simple thresholding and regions with too few voxels are discarded.

3.1 Junction Detection

We first detect the junctions on the reference images for the five cases. The same set of parameters (mentioned in text) is used in all cases. Fig. 4 shows the results for case 1 with the location of the junctions marked by red spheres.

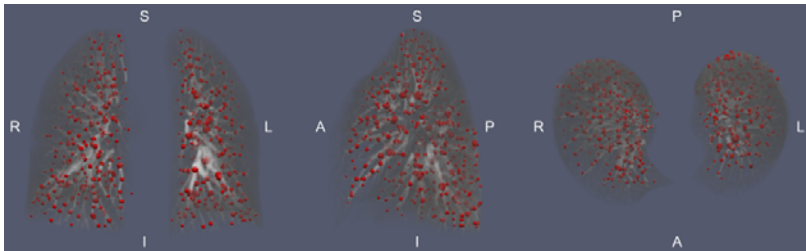


Fig. 4. Junctions detected in case 1. The reference image masked by the lung region is volume-rendered in coronal, sagittal, and transverse views. Junctions are marked as red spheres of radii corresponding to junction sizes. Letters indicate anatomical directions.

Table 1 lists the statistics of the results for all cases. To evaluate the performance, we have visually examined every junction detected. This was done by overlaying the junctions (spheres in different colors) on the original image. Upon selection, the particular junction is focused in three orthogonal views. The accuracy is defined by the ratio of true and all detected junctions. We have detected more than 500 junctions in each case with an average accuracy of 85.6%. The computation was performed on a computer with a dual-core 3.0 GHz processor and 8 GB memory. In practice, a sufficient of junctions covering the lung region is desired, especially in applications to quantify lung motion due to respiration. Fig. 5 shows the spatial distribution of detected junctions in each of the eight ranges in three main directions. The variability

of the quantity (esp. in A-P direction) may attribute to the distribution of lung volume and the actual density of junctions. It suggests that a post-processing step (e.g. sampling) may be necessary if the non-uniformity of junctions is a concern. In regard to the junction sizes, we plot the histograms using seven bins in Fig. 6. A clear exponential trend exhibits in all cases, which agrees with the exponential branching pattern of the pulmonary vessels.

Table 1. The results of junction detection for all cases

	Image size	Resolution (mm)	Junctions	Accuracy	Segmentation time (s)	Detection time (s)
Case 1	256×256×94	0.97×0.97×2.5	609	84.9%	16	35
Case 2	256×256×112	1.16×1.16×2.5	598	87.4%	18	33
Case 3	256×256×104	1.15×1.15×2.5	514	83.9%	18	39
Case 4	256×256×99	1.13×1.13×2.5	556	84.5%	17	45
Case 5	256×256×106	1.10×1.10×2.5	501	87.2%	19	33

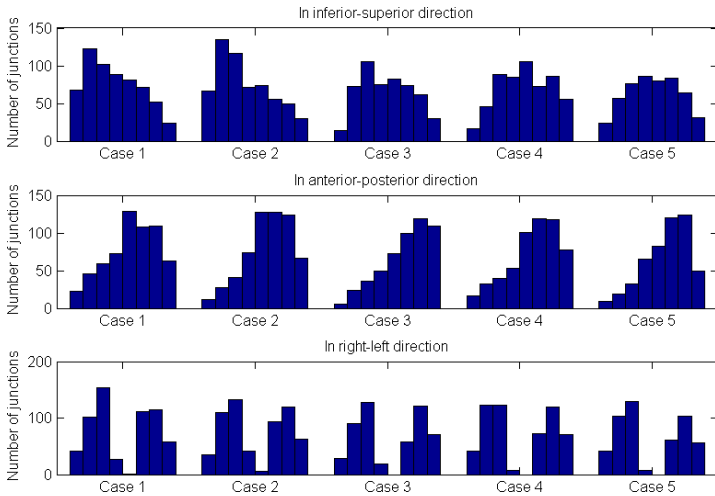


Fig. 5. Spatial distributions of the detected junctions in three main directions

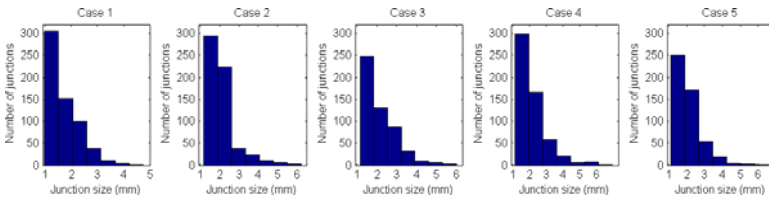


Fig. 6. Histograms of the detected junction sizes

3.2 Tracking Using Automated Detected Junctions

Next, we track the junctions detected in the previous section using all the images. Fig. 7 shows some sample trajectories tracked in case 1. Note the trajectories are actually curves in 3D and can be quite complex. Starting from the maximum inhale (marked by a sphere), the trajectory passes forward through the successive phases (marked by cones, tip pointing to the direction) towards the maximum exhale. The trajectory clearly has a different return path, i.e. hysteresis. To characterize the hysteresis, we define two axes (marked by arrows). The major axis is from the maximum inhale to maximum exhale. The minor axis is from the point to its opposite counterpart which has the shortest distance. The ratio of the lengths between the minor and major axes is used to indicate the degree of hysteresis (1 for full hysteresis, 0 for no hysteresis). Fig. 8 displays all trajectories with $M(p^*) > 9.0$ colored by these values. It suggests that the inferior portion of the lung has larger motion but lower degree of hysteresis. Fig. 9 shows the histograms of $M(p^*)$, where a majority is larger than 8.0 for all cases.

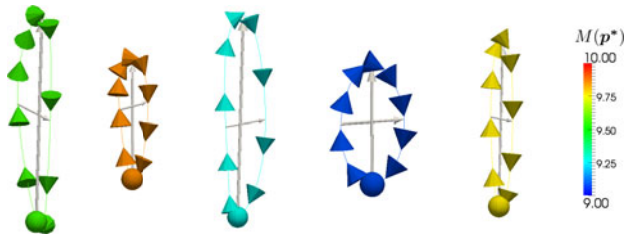


Fig. 7. Sample trajectories in case 1. The color indicates the correlation coefficients $M(p^*)$.

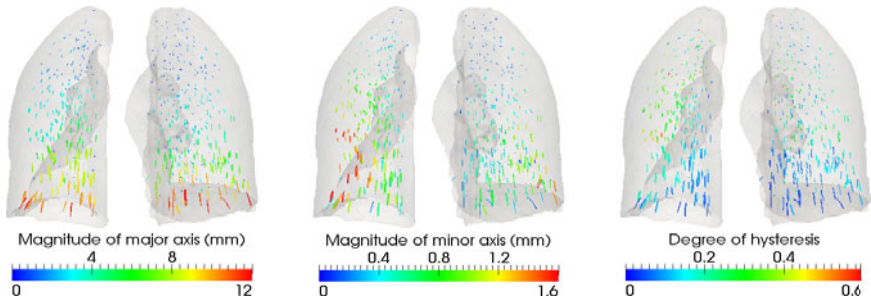


Fig. 8. Magnitudes of major and minor axes, degree of hysteresis in case 1 with $M(p^*) > 9.0$

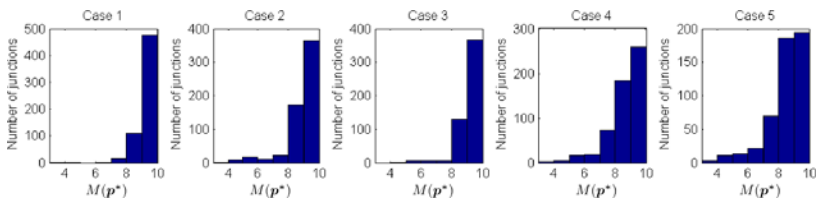


Fig. 9. Histograms of $M(p^*)$ for all cases

3.3 Tracking Using Manually Delineated Landmarks

It is difficult to generate the ground truth data to evaluate tracking methods, especially in 4D. Fortunately, DIR-lab data have 75 landmarks manually delineated on the images $I_1 \dots I_6$, i.e. during expiratory phases. Therefore, we assess the tracking accuracy by computing the tracking error as the mean Euclidean distance between the tracked and the manual landmark locations at any time. It should be noted that these manual landmarks are not necessarily all junctions and their locations are delineated only in the voxel not sub-voxel accuracy. Furthermore, we have assumed the size of all landmarks is 2.0 mm since no sizing information is available. In addition, tracking is done in all images but accuracy is assessed only on $I_1 \dots I_6$. Fig. 10 shows the box plots of the tracking error. A small error at $t = 1$ exists since it is not explicitly constrained as mentioned before. Errors at other times do not increase with the magnitude of the motion. For all cases, we achieve mean sub-voxel accuracy at all phases. Although better initialization suppresses local extrema, they can happen in some rare cases, leading to larger errors.

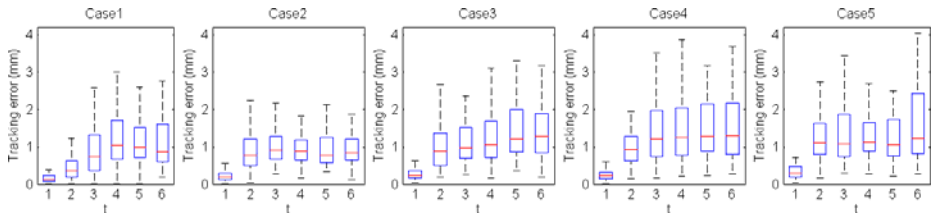


Fig. 10. Box plots of the tracking error at $t = 1 \dots 6$. Outliers are not shown for clarity.

4 Discussions

In summary, we have proposed a framework to both detect junction structures and track their trajectories in 4D images. In the problem of detection, we identified more than 500 junctions with an average accuracy of greater than 85%. We believe the automation of landmark generation should be very useful in various applications, where manual annotation is a common practice. More sophisticated lung vessel segmentation is under investigation to substitute simple thresholding which often misses small vessels. In the problem of tracking, we do not pursue the common approach which considers the whole image space which often assumes the spatial smoothness of transformation and compromises the accuracy of local deformation. Furthermore, only local deformation is of interest in some applications, e.g. tumor tracking. Instead, trajectories of the landmarks are separately tracked by optimizing a combination of correlation coefficients of intensities on the trajectory from all images. For a majority of landmarks, sub-voxel accuracy is achieved. Our approach to generate better initialization suppresses local extrema. Although there are rare worse cases, they can be detected by extremely low correlation coefficients and disregarded for further consideration.

Acknowledgement

Guanglei Xiong was supported by Stanford Graduate Fellowship. This project was supported in part by grants from NCI 1R01 CA133474 and 1R21 CA153587. We acknowledge the publicly available datasets from <http://www.dir-lab.com/>

References

1. Palagyi, K., Tschirren, J., Hoffman, E.A., Sonka, M.: Quantitative Analysis of Pulmonary Airway Tree Structures. *Comput. Biol. Med.* 36(9), 974–996 (2006)
2. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A Framework for Evaluation of Deformable Image Registration Spatial Accuracy Using Large Landmark Point Sets. *Phys. Med. Biol.* 54(7), 1849–1870 (2009)
3. Murphy, K., van Ginneken, B., Klein, S., Staring, M., de Hoop, B.J., Viergever, M.A., Pluim, J.P.: Semi-Automatic Construction of Reference Standards for Evaluation of Image Registration. *Med. Image Anal.* 15(1), 71–84 (2011)
4. Boldea, V., Sharp, G.C., Jiang, S.B., Sarrut, D.: 4D-CT Lung Motion Estimation with Deformable Registration: Quantification of Motion Nonlinearity and Hysteresis. *Med. Phys.* 35(3), 1008–1018 (2008)
5. Metz, C.T., Klein, S., Schaap, M., van Walsum, T., Niessen, W.J.: Nonrigid Registration of Dynamic Medical Imaging Data Using Nd Plus T B-Splines and a Groupwise Optimization Approach. *Med. Image Anal.* 15(2), 238–249 (2011)
6. Tashiro, M., Minohara, S., Kanai, T., Yusa, K., Sakurai, H., Nakano, T.: Three-Dimensional Velocity Mapping of Lung Motion Using Vessel Bifurcation Pattern Matching. *Med. Phys.* 33(6), 1747–1757 (2006)
7. Smeets, D., Bruyninckx, P., Keustermans, J., Vandermeulen, D., Suetens, P.: Robust Matching of 3D Lung Vessel Trees. In: *The Third International Workshop on Pulmonary Image Analysis*, Beijing, China, pp. 61–70 (2010)
8. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T.: Four-Dimensional Deformable Image Registration Using Trajectory Modeling. *Phys. Med. Biol.* 55(1), 305–327 (2010)
9. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer Analysis of Computed Tomography Scans of the Lung: A Survey. *IEEE T. Med. Imaging* 25(4), 385–405 (2006)
10. Maurer, C.R., Qi, R.S., Raghavan, V.: A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions. *IEEE T. Pattern Anal.* 25(2), 265–270 (2003)
11. Sethian, J.A.: *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge (1999)
12. Comaniciu, D., Meer, P.: Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE T. Pattern Anal.* 24(5), 603–619 (2002)
13. De Boor, C.: *A Practical Guide to Splines*. Springer, New York (2000)
14. Nocedal, J.: Updating Quasi-Newton Matrices with Limited Storage. *Math. Comput.* 35(151), 773–782 (1980)
15. Byrd, R.H., Lu, P.H., Nocedal, J., Zhu, C.Y.: A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.* 16(5), 1190–1208 (1995)
16. Hu, S.Y., Hoffman, E.A., Reinhardt, J.M.: Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images. *IEEE T. Med. Imaging* 20(6), 490–498 (2001)

Systematic Assessment of Performance Prediction Techniques in Medical Image Classification

A Case Study on Celiac Disease

Sebastian Hegenbart¹, Andreas Uhl¹, and Andreas Vécsei²

¹ Department of Computer Sciences, University of Salzburg

² St. Anna Children's Hospital, Vienna

Abstract. In the context of automated classification of medical images, many authors report a lack of available test data. Therefore techniques such as the leave-one-out cross validation or k-fold validation are used to assess how well methods will perform in practice. In case of methods based on feature subset selection, cross validation might provide bad estimations of how well the optimized technique generalizes on an independent data set. In this work, we assess how well cross validation techniques are suited to predict the outcome of a preferred setup of distinct test- and training data sets. This is accomplished by creating two distinct sets of images, used separately as training- and test-data. The experiments are conducted using a set of Local Binary Pattern based operators for feature extraction which are using histogram subset selection to improve the feature discrimination. Common problems such as the effects of over fitting data during cross validation as well as using biased image sets due to multiple images from a single patient are considered.

Keywords: celiac disease, classification, cross validation, over fitting, LOPO.

1 Introduction

A desirable data setup for experimentation within the field of medical image classification consists of two distinct sets of image samples with a balanced number of images and patients among the specific classes. In this case one set is used for training a classifier as well as performing feature selection and parameter optimization. A method's classification accuracy is then evaluated by using the trained classification method with its specific parameters on the other set of data samples. In the context of automated classification of medical images however, the available amount of test data is often very limited. Often it is not possible to build distinct data sets for training and evaluation. This can be due to a limited number of patients (e.g. a low prevalence of the specific disease), a limited number of usable images caused by qualitative problems or a high number of classes used to categorize the pathological changes in relation to the available images. In

this case, the evaluation and development of methods, is usually based on cross validation techniques such as the leave-one-out cross validation or k-fold cross validation. By applying these techniques, a prediction of how well developed methods for classification and feature extraction will generalize on an independent data set, is made. Especially in the context of medical image classification, care has to be taken when using cross validation techniques. Depending on how the used sets of image data were created, the leave-one-out or k-fold cross validation techniques might not be sufficient to assess how well developed methods will perform in a realistic scenario. In this work we will study how well different approaches to cross validation perform in the context of classifying celiac disease. We construct two distinct sets for training and evaluation to validate how well different cross validation techniques predict this “optimal” case. By using feature subset selection in combination with Local Binary Pattern (LBP)-based feature extraction we are able to study the effects of over-fitting and discuss adapted techniques for their use in the context of medical image classification such as the leave-one-patient-out cross validation. In particular we will assess how accurate the predictions of the leave-one-patient-out, leave-one-out and k-fold cross validation techniques are compared to a preferred setup using two distinct image sets. We will also study two approaches towards feature subset selection and parameter optimization in combination with cross validation techniques (the so called inner- and outer-approaches).

In Section 2 we identify common problems of constructing image sets for experimentation and explain how the image sets used during this work were constructed. In Section 3 the methods used for feature extraction and classification are presented. We also discuss the methods used for feature (histogram) subset selection. Section 4 deals with methods for cross validation and possible problems in the context of medical image classification. Also two approaches for feature subset selection and parameter optimization during cross validation are discussed. Section 5 presents the results of the conducted experiments. Finally the results are discussed in Section 6.

2 Image Set Construction

The creation of image data sets for experimentation requires the consideration of several possible problems:

- An unbalanced number of samples per class can lead to a bias towards the class with the largest number of samples when using the overall classification rate as criterion for feature selection and parameter optimization. As a consequence the overall classification rate might not be a significant measure for the performance of developed methods. It is desirable to have a balanced number of samples among each class.
- Images from a single patient usually have a higher similarity among each other as compared to images among different patients from a single class (or at least this might be conjectured). Depending on the classification method, this could have an impact on the classification outcome.

Table 1. Distribution of Image Data

	Class ₀	Class ₁	Total	Class ₀	Class ₁	Total
	Images			Patients		
Image-Set 1	155	157	312	66	21	87
Image-Set 2	151	149	300	65	19	84

- In some cases, the low number of original images from a specific class requires the extraction of multiple sub-images from a single parent image. Due to the common camera perspective and illumination these sub-images usually have the highest similarity among each other. This also might influence the classification method.

2.1 Image Data

We construct our image test sets based on images taken during duodenoscopies at the St. Anna Children’s Hospital using pediatric gastroscopes without magnification (GIF-Q165 and GIF-N180, Olympus, Hamburg). The main indications for endoscopy were the diagnostic evaluation of dyspeptic symptoms, positive celiac serology, anemia, malabsorption syndromes, inflammatory bowel disease, and gastrointestinal bleeding. Images were recorded by using the modified immersion technique, which is based on the instillation of water into the duodenal lumen for better visibility of the villi. The tip of the gastroscope is inserted into the water and images of interesting areas are taken. Gasbarrini et al. [2] showed that the visualization of villi with the immersion technique has a higher positive predictive value. Hegenbart et al. [3] state that the modified immersion technique is more suitable for automated classification purposes as compared to the classical image capturing technique. Images from a single patient were recorded during a single endoscopic session.

To study the prediction accuracy of cross validation techniques we manually created an “idealistic” set of textured image patches with optimal quality. The texture patches have a fixed size of 128×128 pixels, a size which turned out to be optimal as reported by Hegenbart et al. [3]. In a fully automated system the process of frame identification as well as segmentation would be automated as well. These techniques are beyond the scope of this paper though.

In order to generate the ground truth for the texture patches used in experimentation, the condition of the mucosal areas covered by the images was determined by histological examination of biopsies from the corresponding regions. Severity of villous atrophy was classified according to the modified Marsh classification in Oberhuber et al. [8]. This histological classification scheme identifies six classes of severity of celiac disease, ranging from class Marsh-0 (no visible change of villi structure) up to class Marsh-3C (absent villi). In this work a reduced scheme is considered using Marsh-0 (no celiac disease) and the joint set of the classes Marsh-3A, Marsh-3B and Marsh-3C (indicating celiac disease). We will refer to the non-celiac images as Class₀ and to the celiac images as Class₁ from here on. Figure 1 shows an example of the four interesting Marsh classes.

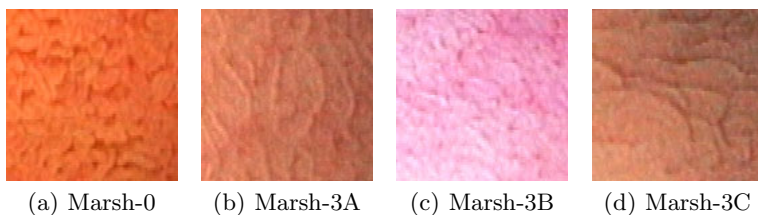


Fig. 1. Examples of Duodenal Image-Patches used for Experimentation

2.2 Construction of Distinct Data Sets

The constructed image sets originate from 171 patients (131 control patients and 40 patients with diagnosed celiac disease). In order to guarantee an image set of reasonable size, more than a single texture patch was extracted for each patient from the original images. In total 753 texture patches met the required qualitative properties. Based on this set of texture patches two distinct sets for training and evaluation were created. The construction was done in an automated way such that the number of images is balanced between the non-celiac class Marsh-0 and the celiac classes Marsh-3A to Marsh-3C. While creating the two distinct sets, care was taken that the number of patches per patient is as evenly balanced as possible. Also, no images from a single patient are within both image sets. The actual construction was done using a pseudo random number generator based on a Gaussian distribution to avoid any bias within the data sets. Table 1 shows the distribution of images and patients per class.

3 Feature Extraction and Classification

The basic LBP operator was introduced to the community by Ojala et al. [9]. We use three operators that are based on LBP to conduct our experiments. The operators are LBP (Local Binary Patterns, [11]), ELBP (extended Local Binary Patterns, [4]), and a modified version of the ELBP operator that is introduced in this work, the ELTP (extended Local Ternary Patterns) operator. The entire family of operators is used to model a pixel neighborhood in terms of pixel intensity differences. The operators assign a binary label to each possible pixel neighborhood. The distributions of these labels are then used as features, which are represented by histograms. We compute the pattern distributions for each color channel (RGB), each LBP-Scale (1-3) (see Section 3.1) as well as filter orientation (in case of the extended LBP based operators: horizontal, vertical and diagonal). In total we result in 9-histograms for LBP and 27-histograms for ELBP and ELTP. For each histogram, only a subset of dominant patterns known as the uniform patterns [7], which make up the majority of discriminative patterns, is used. In case of the LBP and ELBP operator this subset consists of 58-patterns for 8 considered neighbors. In case of the ELTP operator two histograms with 58-bins are concatenated, therefore the dimensionality of the ELTP histograms is 116 bins.

3.1 Local Binary Patterns

For the radius r and the number of considered neighbors p , the LBP operator is defined as

$$LBP_{r,p}(x, y) = \sum_{k=0}^{p-1} 2^k s(I_k - I_c), \quad (1)$$

with I_k being the value of neighbor number k and I_c being the value of the corresponding center pixel. The s function acts as sign function, mapping to 1 if the difference is smaller or equal to 0 and mapping to 0 else. The basic operator uses an eight-neighborhood with a 1-pixel radius. To overcome this limitation, the notion of scale is used as discussed by Ojala et al. in [10] by applying averaging filters to the image data before the operators are applied. Thus, information about neighboring pixels is implicitly encoded by the operator. The appropriate filter sizes for a certain scale is calculated as described in [6].

3.2 Extended Local Binary Patterns and Extended Local Ternary Patterns with Adaptive Threshold

Information extracted by the LBP-based operators from the intensity function of a digital image can only reflect first derivative information. This might not be optimal, therefore Huang et al. [4] suggest using a gradient filtering before feature extraction and call this operator ELBP or extended LBP. By doing this the velocity of local variation is described by the pixel neighborhoods.

We introduce the extended LTP (ELTP) operator consequently in perfect analogy to the ELBP operator. ELTP is based on the LTP operator instead of the LBP operator to suppress unwanted noise in the gradient filtered data. The Local Ternary Pattern operator (LTP) was introduced by Tan and Triggs [11]. The modification is based on a thresholding mechanism which implicitly improves the robustness against noise. In our scenario endoscopic images are used which usually are noisy as a result of the endoscopic procedure. The LTP operator is used to ensure that pixel regions that are influenced by these kind of distortions do not contribute to the computed histograms. The LTP is based on a thresholded sign function:

$$s(x) = \begin{cases} 1, & \text{if } x \geq T_h \\ 0, & \text{if } |x| < T_h \\ -1, & \text{if } x \leq -T_h. \end{cases} \quad (2)$$

The ternary decision leads to two separate histograms, one representing the distribution of the patterns resulting in a -1 , the other representing the distribution of the patterns resulting in a 1 .

$$H_{I,lower}(i) = \sum_{x,y} (LBP_{r,p}(x, y) = -i) \quad i = 0, \dots, 2^p - 1 \quad (3)$$

$$H_{I,upper}(i) = \sum_{x,y} (LBP_{r,p}(x, y) = i) \quad i = 0, \dots, 2^p - 1 \quad (4)$$

The two computed histograms are concatenated and then treated like a single histogram. Please note that in analogy to the LBP operator, only the uniform subset of patterns was used in this case. The actual optimal values to use for thresholding are unknown a priori. We apply an adaptive threshold based on the spatial image statistics to make sure that noisy regions do not contribute to the computed histograms while information present within high quality regions are not lost due to a threshold that was chosen too high. The calculation is based on an expected value for the standard deviation of the image (β). This value was found based on the training data used during experimentation and represents the average standard deviation of pixel intensity values within all images. The value α is used as a weighting factor combined with the actual pixel standard deviation of the considered image (σ) and is used to adapt the threshold to match the considered image characteristics.

$$T_h = \begin{cases} \beta^{\frac{1}{2}} + \alpha\sigma, & \text{if } \sigma > \beta \\ \beta^{\frac{1}{2}} - \alpha\sigma, & \text{if } \sigma \leq \beta. \end{cases} \quad (5)$$

3.3 Histogram Subset Selection

Depending on the specific operator, at least 9 (LBP) and at maximum 27 (ELBP and ELTP) histograms are computed for a single image. A single LBP histogram can be interpreted as a “macro” feature. Therefore the terms histogram subset selection and feature subset selection share the same meaning. Feature subset selection techniques are usually applied for two reasons.

Result Optimization. Probably not all parameters combinations are equally well suited for describing the specific textural properties. Even more, when computing a large number of histograms, this set could contain a few “bad” histograms which reduce the discriminative power.

Reduction of Dimensionality. Depending on the chosen classification method large feature vectors might be suboptimal in terms of computational complexity and classification performance. Feature subset selection can be used to reduce the number of considered histograms and therefore the final feature vector dimensionality.

The applied algorithm for histogram subset selection was the Sequential Forward Selection algorithm (SFS, [5]). The optimization criterion for this algorithm was the overall classification rate. The upper bound set on the number of selected histograms was 10. This technique of optimizing the feature subset might be subject to over fitting. We expect the operators computing a larger number of histograms (ELBP and ELTP) to be at higher risk of being over fitted when using “outer” optimization (see section 4.2 for a comparison of approaches for optimization).

3.4 Classification

The k-nearest neighbors (kNN) classifier was used for classification. A rather weak classifier was chosen to give more emphasis on the selected histogram combinations. After the histogram subset selection the candidate histograms were combined and treated as a single histogram. To compute the distance (or similarity) of two different histograms we apply the histogram intersection metric. For two histograms (H_1, H_2) with N bins and bin number i being referenced to as $H(i)$, the similarity measure is defined as

$$H(H_1, H_2) = \sum_{i=1}^N \min(H_1(i), H_2(i)). \quad (6)$$

The k-value is subject to parameter optimization and was optimized in the corresponding cross validations based on the specific training set. By using the kNN classifier we are also able to study problems caused by multiple images from the same patient or parent frame within the training and test set.

4 Cross Validation Protocols

Cross validation is used to estimate the accuracy of the general prediction of the classification method. In 85 articles known to the authors of this work on automated diagnosis in the field of medical image classification, more than half resort to either leave-one-out (LOOCV) cross validation or k-fold cross validation.

K-fold cross validation is a generalization of the leave-one-out cross validation technique. The k-fold cross validation partitions the original set of samples into k disjoint subsets. The classification uses $k - 1$ subsets as training input and classifies samples from the left out subset. This process is repeated k times. The leave-one-out cross validation can be seen as a k-fold cross validation with k corresponding to the number of data samples. Therefore each subset consists of only a single sample. Other approaches of cross validation such as random sub-sampling are special variations of the k-fold cross validation and were not considered in this work. When using k-fold cross validation, a balanced number of samples from each class should be available within the $k - 1$ subsets used for training. Theoretically all samples from a single class could be within one subset, leading to a bad estimation of the classification rate of this class. On the other hand using a high number of folds leads to small image subsets and usually brings up the problem that images from a single patient, or even worse from a single parent image, are within both the training and test data sets.

4.1 Leave-One-Patient-Out Cross Validation

The similarity of images from a single patient can be higher than the similarity between different patients from a class. A straight forward and clean solution is to use only a single image of each patient. Unfortunately in practice this is rarely possible due to a limited number of available data. An approach to take care

of this problem is the leave-one-patient-out (LOPO) cross validation technique (also used by André et al. [11]). LOPO cross validation is based on the k -fold cross validation. The partitioning of the original set of samples however is done such that each partition contains only images from a single patient. This approach implies that patient information in some, usually unambiguously anonymized, form is available. A variation that is closely related to the LOPO cross validation method is the leave-one-parent-frame out cross validation. In this technique the partitioning is performed such that each partition consists of all sub-images from a parent image. This approach can usually be used if no patient information is available. However, the LOPO cross validation technique should be preferred over the leave-one-parent-image-out technique whenever possible.

4.2 Feature Optimization Combined with Cross Validation

We distinguish between two approaches to feature subset selection and parameter optimization in combination with cross validation.

- The **outer**-approach optimizes features or parameters based on the results of the cross validation used for predicting the classifier’s accuracy. This means that the optimization criterion of the feature subset selection method is based on the estimates of a cross validation on the entire data set. These estimations are also used as classification rates later.
- The **inner**-approach optimizes features or parameters within a separate cross validation based on the $k - 1$ partitions used for training within the cross validation used for predicting classification accuracy. This means that the optimization criterion of the feature subset selection method is based on a separate cross validation using the training set ($k - 1$ partitions) of the current “outer” cross validation. Therefore, for each partition an new feature subset is selected. The classification rate is the estimation of the “outer” cross validation.

The outer-approach is the classical and easier approach frequently found within the literature. This approach however poses the problem that test data is used for optimizing feature subsets or parameters. This can have an influence on the optimization and therefore an effect on the prediction of how well the feature subset or optimized parameters generalize (the optimization over-fits the model towards the data). By using the inner-approach, the risk of over-fitting is reduced, the major drawback is that the computational power needed for this evaluation is considerably higher as compared to the other technique. This is caused by repeated feature subset selection and parameter optimization which is usually the most time consuming element in the automated classification chain.

5 Results

This Section presents the results of the experiments. Please note, that we use subscripts combined with the method names to indicate the type of optimization.

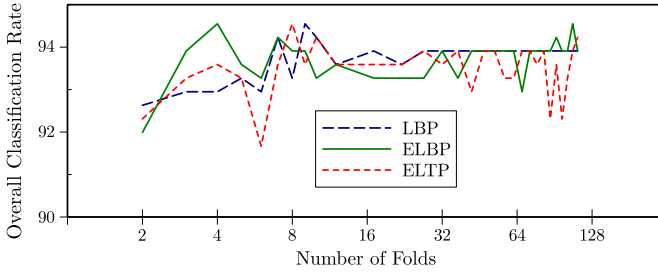


Fig. 2. Overall Classification Rate Estimates of **k-Fold_O** (outer) Cross Validations

The inner-approach is indicated by a “I”, while the outer-approach is indicated by a “O”. All results except the distinct set classification was computed using the specific cross validation technique on Image-Set1. The abbreviations “Spec.” and “Sens.” refer to the Methods’ specificity and sensitivity.

Figures 2 and 3 show the overall classification rates predicted by using k-fold cross validation. Due to computational issues, the values were computed from 2 to 10 in single steps and from 12 to 112 in steps of 5. The mean classification rates are: LBP (93.75%, $\sigma = 0.43$), ELBP (93.48%, $\sigma = 0.5$) as well as ELTP (93.48%, $\sigma = 0.64$) in case of the outer-approach and LBP (90.98%, $\sigma = 1.08$), ELBP (90.48%, $\sigma = 0.91$) as well as ELTP (89.99%, $\sigma = 1.07$) in case of the inner-approach. The columns of Table 2 labeled as Δ list the differences of the predictions of the overall classification rates between the outer- and inner-approach. The experiments based on the inner-approach used a leave-one-out cross validation as the “inner” cross validation method in all cases.

Table 3 compares the results achieved by using the “optimal” distinct set validation (Image-Set1 is used for training, Image-Set2 for evaluation) with the estimates provided by using the mentioned cross validation techniques. The columns labeled as Δ show the differences of the specific methods’ overall classification rates to the overall classification of the distinct set validation. The results with the closest proximity to the distinct set results are displayed in bold. The columns labeled as mean and max show the differences to the mean overall classification rates of the k-fold cross validation as well as the differences to the maximum

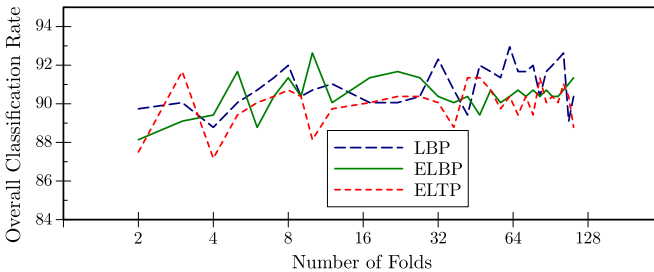


Fig. 3. Overall Classification Rate Estimates of **k-Fold_I** (inner) Cross Validations

Table 2. Cross Validation Estimates using LOOCV and LOPO

	LOOCV _O (outer)			LOOCV _I (inner)			Δ
	Spec.	Sens.	Overall	Spec.	Sens.	Overall	
LBP	93.63	94.19	93.91	90.38	90.32	90.35	3.56
ELBP	94.27	93.55	93.91	91.67	89.68	90.68	3.32
ELTP	94.27	93.55	93.91	90.32	91.03	90.68	3.32

	LOPO _O (outer)			LOPO _I (inner)			Δ
	Spec.	Sens.	Overall	Spec.	Sens.	Overall	
LBP	85.99	95.48	90.71	82.17	90.32	86.22	4.49
ELBP	91.08	94.19	92.63	81.53	90.97	86.22	6.41
ELTP	89.81	94.19	91.99	79.62	89.68	84.62	7.37

Table 3. Results of the Distinct Set Classification using Image-Set1 as Training-Data

	Distinct Sets			Δ	Δ	Δ	Δ
	Spec.	Sens.	Overall	LOPO _O	LOPO _I	LOOCV _O	LOOCV _I
LBP	79.47	87.25	83.33	7.38	2.89	10.58	7.02
ELBP	80.13	92.62	86.33	6.30	-0.11	7.58	4.35
ELTP	79.47	92.62	86.00	5.99	1.38	7.91	4.68

	Distinct Sets			Δ Mean	Δ Max	Δ Mean	Δ Max
	Spec.	Sens.	Overall	k-Fold _O	k-Fold _O	k-Fold _I	k-Fold _I
LBP	79.47	87.25	83.33	10.42	11.22	7.76	9.62
ELBP	80.13	92.62	86.33	7.39	8.22	4.15	6.30
ELTP	79.47	92.62	86.00	7.48	8.55	3.99	5.67

classification rates of the k-fold cross validations (which is also the maximum difference to all classification outcomes of the k-fold cross validation).

5.1 Performance

Beside to the actual prediction accuracy of each method, the computational complexity plays an important role of how well the method is suited for application in experimentation. A major part of the computational efforts lies within the feature subset selection. The upper bound defined on the number of histograms used to build the feature vector in this work is 10. The feature subset selection method exits if no better configuration (in terms of overall classification rate) of histograms can be found. The maximum number of performed cross validations is $\frac{n(n+1)}{2} - \frac{(n-10)(n-9)}{2}$ for n available histograms. The actual number of computations is highly dependent on the data. To be able to compare the performance among the techniques, we limit the upper bound on the histogram count to 1 for the experiments used for the performance assessment. Table 4 shows the time in seconds needed for a full cross validation of Image-Set1.

Table 4. Time in Seconds for a Full Validation

Method	Seconds	Method	Seconds
LOOCV (Outer)	2.8	LOOCV (Inner)	648.7
LOPO (Outer)	8.4	LOPO (Inner)	624.5
Distinct	2.9		

6 Discussion

The results show that there is a significant difference between the estimated rates of the cross validation methods and the distinct set evaluation. The rates of the outer-optimization indicate some degree of over-fitting during optimization. In case of the LOOCV method, the results show that the classification rates using outer-optimization are approximately 3.5 percentage points above the inner optimization. In case of the LOPO and the k-fold methods this effect can also be observed. For the LOPO method, the differences between inner- and outer-optimizations are even higher as compared to k-fold and LOOCV. We assume that this is due to a combined effect of over-fitting and image set bias of the LOOCV and k-fold methods. The mean estimates of the k-fold cross validations are comparable to the LOOCV cross validation. The prediction accuracy of methods using the outer-optimizations is further off the rates achieved by the distinct set evaluation as compared to the inner-optimization.

Table 4 shows, that the higher accuracy of the inner-optimization, comes at the cost of a considerably higher computational effort. The differences in computational complexity among the cross-validation methods is significantly smaller. Considering the results we see that the inner-approach is the best suited technique (if its complexity can be handled) for evaluating methods using features optimization.

Compared to the distinct set evaluation, the LOOCV method is off by an average of 8.7 percentage points (outer) as well as 5.35 (inner). The prediction of the LOPO method seems to be more accurate with an average difference of 6.5 percentage points (outer) as well as excellent 1.39 percentage points (inner). Considering the results of the k-fold cross validations a significant variance of the rates at low number of folds is observed. In general the standard deviation is below one percentage point for both approaches. If k-fold validation is applied we suggest using a fixed number of folds for all experiments to avoid an additional effect of over-fitting. To avoid biased image sets caused by multiple images from a patient the LOPO method should be preferred whenever possible. In general the LOPO method combined with inner-optimization seems to be the most adequate approach if no distinct sets for training and evaluation can be constructed.

References

1. André, B., Vercauteren, T., Wallace, M.B., Buchner, A.M., Ayache, N.: Endomicroscopic video retrieval using mosaicing and visual words. In: Proceedings of the 7th IEEE International Symposium on Biomedical Imaging. IEEE, Los Alamitos (2010) (to appear)

2. Gasbarrini, A., Ojetti, V., Cuoco, L., Cammarota, G., Migneco, A., Armuzzi, A., Pola, P., Gasbarrini, G.: Lack of endoscopic visualization of intestinal villi with the immersion technique in overt atrophic celiac disease. *Gastrointestinal Endoscopy* 57, 348–351 (2003)
3. Hegenbart, S., Kwitt, R., Liedlgruber, M., Uhl, A., Vécsei, A.: Impact of duodenal image capturing techniques and duodenal regions on the performance of automated diagnosis of celiac disease. In: *Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis*, Salzburg, Austria, pp. 718–723 (2009)
4. Huang, X., Li, S., Wang, Y.: Shape localization based on statistical method using extended local binary pattern. In: *Proceedings of the 3rd International Conference on Image and Graphics*, Hong Kong, China, pp. 1–4 (2004)
5. Jain, A., Zongker, D.: Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 153–158 (1997)
6. Mäenpää, T.: *The Local Binary Pattern Approach to Texture Analysis - Extensions and Applications*. Ph.D. thesis, University of Oulu (2003)
7. Mäenpää, T., Ojala, T., Pietikäinen, M., Soriano, M.: Robust texture classification by subsets of local binary patterns. In: *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 3, p. 3947. IEEE Computer Society, Los Alamitos (2000)
8. Oberhuber, G., Granditsch, G., Vogelsang, H.: The histopathology of coeliac disease: time for a standardized report scheme for pathologists. *European Journal of Gastroenterology and Hepatology* 11, 1185–1194 (1999)
9. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition* 29(1), 51–59 (1996)
10. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-Scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(7), 971–987 (2002)
11. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) *AMFG 2007*. LNCS, vol. 4778, pp. 168–182. Springer, Heidelberg (2007)

Detecting and Classifying Linear Structures in Mammograms Using Random Forests

Michael Berks, Zezhi Chen, Sue Astley, and Chris Taylor

Imaging Science and Biomedical Engineering, School of Cancer and Enabling Sciences,
University of Manchester, Oxford Road, Manchester, M13 9PT, UK
michael.berks@manchester.ac.uk

Abstract. Detecting and classifying curvilinear structure is important in many image interpretation tasks. We focus on the challenging problem of detecting such structure in mammograms and deciding whether it is normal or abnormal. We adopt a discriminative learning approach based on a Dual-Tree Complex Wavelet representation and random forest classification. We present results of a quantitative comparison of our approach with three leading methods from the literature and with learning-based variants of those methods. We show that our new approach gives significantly better results than any of the other methods, achieving an area under the ROC curve $A_z = 0.923$ for curvilinear structure detection, and $A_z = 0.761$ for distinguishing between normal and abnormal structure (spicules). A detailed analysis suggests that some of the improvement is due to discriminative learning, and some due to the DT-CWT representation, which provides local phase information and good angular resolution.

Keywords: mammography, linear structure detection, spiculated lesions, classification, regression, random forests, dual-tree complex wavelet.

1 Introduction

The detection and classification of curvilinear structure is important in many image interpretation tasks – from roads and rivers in aerial images, through cracks in manufactured components, to blood vessels and ducts in medical images. Often the detection task is difficult because the structures of interest appear at relatively low contrast against a cluttered background. Classification is of interest when it is important to distinguish between subtly different structures which may be present within the same image – for example, rivers and roads. In this paper we focus on the challenging problem of detecting and classifying curvilinear structures in x-ray mammograms, but the methods we describe are equally applicable to problems in other domains.

We adopt a discriminative learning approach, applied to a rich description of local image structure, both to distinguish between curvilinear structure and background clutter, and to classify detected structures into different types. This approach achieves high detection accuracy by taking account of the domain-specific characteristics of both the structures of interest and the background clutter, whilst capturing subtle distinctions between different types of curvilinear structure. To describe local image

structure we use the Dual-Tree Complex Wavelet Transform (DT-CWT) [1], whilst for detection and classification we use random forest (RF) classification [2].

In the remainder of the paper we provide more detailed motivation, explain the details of our approach and present the results of an experimental evaluation using both synthetic and real data. We show that, overall, our approach is significantly better than the current state-of-the-art, and that we can distinguish effectively between curvilinear structures associated with malignancy (spicules) and those associated with normal structure (vessels, ducts etc).

2 Background

2.1 Detecting Abnormalities in Mammograms

Breast screening programmes using x-ray mammography have been deployed widely to detect early signs of cancer. The use of computer-aided detection systems to support breast radiology experts is also widespread. In mammograms, the projection of a complex network of vessels, ducts and connective tissue in the breast results in images that are rich in linear structures at a variety of scales and orientations. Mammographic signs of cancer include: the presence of a radio-opaque mass, often with associated radiating curvilinear structures (spicules); the presence of microcalcifications (clusters of highly radio-opaque ‘dots’); and the presence of abnormal or abnormally organised curvilinear structures (architectural distortion – AD). Fig 3(a) shows an approximately 4 x 4 cm region extracted from a mammogram, centred on a malignant tumour, and demonstrates a central mass and radiating spicules, along with normal curvilinear structures (vessels, ducts, glandular tissue etc). The signs of abnormality often appear in conjunction, but when only AD is present, detection of the abnormality is extremely difficult. It has been estimated that approximately a third of all cancers missed during mammographic screening present as AD [3], and it has been reported that computer-aided detection systems do not yet detect AD with adequate sensitivity or specificity [4].

Previous attempts at detecting both patterns of spicules associated with malignant masses and more general distortions in which no focal mass is visible have employed a two stage approach comprising i) the detection of all curvilinear structures in a specified region, ii) an analysis of the orientation pattern of the detected structures to determine whether abnormal patterns are present [5-7]. The work we report here seeks to provide a firmer foundation for this approach by improving the detection of curvilinear structures and identifying those structures likely to be (abnormal) spicules.

2.2 Detecting and Classifying Curvilinear Structures

There is a substantial literature on the detection of curvilinear structure in mammograms which we will review briefly. There is also a more general body of work on the detection of curvilinear structure which space does not allow us to cover, though most of the relevant approaches have been applied to mammograms.

Dixon and Taylor described a curvilinear structure detector based on comparing the integrated intensity along oriented lines (generally referred to as Linop) [8], which Parr et al. applied to mammograms [6]. Cerneaz and Brady [9] described an approach

based on estimating the second derivatives of the mammographic image surface. Karssemeijer and te Brake described an approach based on multiscale directional second-order Gaussian derivatives [5]. Zwiggelaar et al. compared all these methods using simulated mammographic images and reported that Linop gave the best results in terms of area under the receiver operating characteristic (ROC) curve [10]. More recently, researchers have recognised the importance of using local phase information, particularly in distinguishing between strong edges and genuine curvilinear structure. Rangayyan and Ayres described an approach based on Gabor filters [7]. Schenk and Brady [11] and McLoughlin et al. [12] used sets of steerable complex filters [13] at multiple scales to compute local energy, orientation and phase at each pixel. Wai et al. [14] used the monogenic signal as more efficient way of calculating local phase and orientation at multiple scales, detecting curvilinear structure using amplitude-weighted local phase congruency. Overall, the conclusion that can be drawn from the literature is that the detection of curvilinear structure benefits from access to local phase and magnitude information at multiple scales.

The literature on classifying curvilinear structures in mammograms is much more limited. We are aware of the work of Zwiggelaar et al. [10], which demonstrated the feasibility of distinguishing between different types of structure using cross-sectional profiles obtained from manually annotated curvilinear structures, but did not obtain very satisfactory results when the method was fully automated. We recently reported preliminary classification (but not detection) results using our current approach [15].

2.3 Dual-Tree Complex Wavelet Transform

Wavelet transforms have been used extensively in image processing and analysis to provide a rich description of local structure. The dual-tree complex wavelet transform has particular advantages because it provides a directionally selective representation with approximately shift-invariant coefficient magnitudes and local phase information [1]. The DT-CWT combines the outputs of two discrete transforms, using real wavelets differing in phase by 90° , to form the real and imaginary parts of complex coefficients. For 2-D images, it produces six directional sub-bands, oriented at $\pm 15^\circ$, $\pm 45^\circ$ and $\pm 75^\circ$, at each of a series of scales separated by factors of two. The DT-CWT is less expensive to compute and provides a richer description (magnitude and phase at six orientations rather than one) than the monogenic signal used by Wai et al. [14].

2.4 Random Forest Classification and Regression

Random forest classifiers have become popular because of their ease of use (rapid training, no critical parameters to select), resistance to overtraining, and near state-of-the-art performance. Given a set of training data consisting of N samples, each of which is a D -dimensional feature vector labelled as belonging to one of C classes, a random forest comprises a set of tree predictors constructed from the training data [2]. Each tree in the forest is built from a bootstrap sample of the training data (that is, a set of N samples chosen randomly, with replacement, from the original data). The trees are built using a standard classification and regression tree (CART) algorithm; however, rather than assessing all D dimensions for the optimal split at each tree node, only a random subset of $d < D$ dimensions are considered. The trees are built to

full size (i.e. until a leaf is reached containing samples from only one class) and do not need to be pruned (although they can be for the sake of efficiency).

During classification, an unseen feature vector is classified independently by each tree in the forest; each tree casts a unit class vote, and the most popular class can be assigned to the input vector. Alternatively, the proportion of votes assigned to each class can be used to provide a probabilistic labelling of the input vector. Random forests are particularly suited to learning non-linear relationships in high-dimensional multi-class training data, and have been shown to perform as well as classifiers such as Adaboost or support vector machines, whilst being more efficient to compute [2].

Similarly, by building regression rather than classification trees, a random forest can be constructed to learn the relationship between patterns of high-dimensional training data and a single output variable.

3 Detecting Curvilinear Structures

3.1 Detection as Learning

We have argued for the rich description of local structure provided by the DT-CWT. We now have to decide how to use this information to compute a single measure of curvilinear structure probability. One solution would be to select the maximum of the six oriented sub-band coefficients at each scale, and combine them across scale in a measure of phase congruency, as in the monogenic signal based method of Wai et al. [14], but this would discard potentially useful information. Instead, we pose the problem as a classification task in which we attempt to learn the patterns of DT-CWT coefficients associated with pixels belonging to either linear structures or background. We construct a feature vector characterising each pixel, by sampling DT-CWT coefficients (in phase/magnitude form) from the six oriented sub-bands in each of the s finest decomposition scales from a $w \times w$ neighbourhood centred on the pixel. This involves interpolating the coefficients obtained at coarse scales to provide a response at every pixel using the band-pass interpolation method of Anderson et al. [16]. To improve the rotational symmetry of the coefficients, we apply the transformations recommended by Kingsbury [17]. That is, an additional set of filters are used to reduce the wave frequencies of the 45° and 135° sub-bands so that they lie closer to the 15° , 75° , 105° and 165° bands. The six sub-bands are then multiplied by $\{i, -i, i, -1, 1, -1\}$ respectively, so that the phase at the centre of the impulse response of each wavelet is zero. Finally, to achieve 180° rotational symmetry, we replace any coefficient with negative imaginary part with its complex conjugate.

By forming such a feature vector at each pixel in a training set of ground-truth labelled images, we can obtain a large set of data which can be used to train a random forest classifier to differentiate between linear structures and background. In constructing the best detector, this discriminative approach takes account of the characteristics of the background pixels as well as the curvilinear structure pixels.

3.2 Orientation as Regression

In addition to detecting curvilinear structure in mammograms, for many subsequent analysis tasks (e.g. detecting abnormal patterns of structure indicative of disease) it is

equally important that we can make accurate estimates of orientation. As with detection, rather than doing this prescriptively, we pose the problem as a regression task in which we learn how patterns of DT-CWT coefficients correspond to structures at varying orientations. Feature vectors of DT-CWT coefficients are formed as in section 2.1, with the difference that the image ground truth specifies the orientation of a pixel as opposed to its class membership. A random forest regressor can then be trained to predict the orientation of any pixel given its DT-CWT representation.

3.3 Training on Synthetic Data

To apply our learning algorithm to mammograms we require a precise binary labelling of curvilinear structure and background pixels. Given the complexities of structure typically found in mammograms, such data are impractical to obtain. Our solution is to train a classifier on synthetic images containing local structure that closely resembles that which we expect to find in real mammograms. We form these synthetic images by pre-processing regions of real mammograms to remove existing linear structure, whilst maintaining both the coarse scale texture properties of the region and the local image noise across all frequencies, before superimposing synthetic curvilinear structures on the regions. We have used straight lines, each defined by its orientation, width, peak intensity and cross-sectional intensity profile. For each property, we define a probability density function from which samples can be drawn, allowing arbitrarily large numbers of lines to be generated. An example of a synthetic image is shown in Fig 1(a). More details are given in section 5.1.

Given a set of synthetic images generated as described above, we can train a random forest classifier. The literature suggests (and this is borne out by our experience) that the best performance is obtained by training with a balanced dataset – in this case with the same number of foreground (curvilinear structure) and background examples. To achieve this we randomly sample feature vectors from pixels lying on the centre-lines of synthetic curvilinear structures, and randomly sample the same number of feature vectors from background pixels. Having trained a classifier using the synthetic data, we can classify feature vectors extracted about every pixel in a synthetic or real mammogram image to obtain a line probability image (using the probabilistic labelling scheme as opposed to a hard binary classification). Example results are shown in Fig 1(b)-(h) and Fig 3 (b)-(h).

To construct the orientation regression random forests, we sample points randomly from line pixels in the synthetic images (background pixels are not used because the orientation is undefined).

4 Classifying Curvilinear Structure

The learning approach described above can also be used to differentiate between different kinds of curvilinear structure. The hypothesis is that the cross-sectional intensity profiles of structures differ in systematic ways between types of mammographic structure (as suggested by Zwiggelaar et al. [10]), and that profile shape is effectively captured by the DT-CWT coefficients – particularly in the phase components. In the experiments described below we concentrated on the task of distinguish-

ing between spicules, which are a sign of malignancy, and other curvilinear structures. Although it was not realistic to create a training set of mammogram images with all spicules annotated, we were able to obtain expert annotations of a reasonably large number of spicules in a set of mammogram patches containing malignant tumours (see section 5.1 for details). These annotations were used to select pixels for the positive training set. To create a balanced training set we sampled feature vectors from the same number of pixels in a set of normal mammogram patches, such that the distribution of curvilinear structure probabilities was the same as for the spicule training set. Using this balanced training set, we built a random forest classifier to perform spicule/non-spicule classification.

5 Experimental Evaluation

We have conducted a systematic evaluation of the performance of our method for curvilinear structure detection and classification, using both synthetic and real mammogram data, comparing seven competing approaches:

- **DT-CWT/RF:** the method described in this paper.
- **Monogenic:** the monogenic-signal phase congruency approach of Wai et al. [14].
- **Linop:** the Line operator method of Dixon and Taylor [8] and Parr et al. [6].
- **Gaussian:** the directional Gaussian 2nd derivatives method employed by Karssemeijer and Te Brake [5].
- **Monogenic/RF:** the raw responses used in *Monogenic*, combined using RF classification.
- **Linop/RF:** the raw responses used in *Linop*, combined using RF classification.
- **Gaussian/RF:** the raw responses used in *Gaussian*, combined using RF classification.

Monogenic, Linop and Gaussian are representative of the current state of the art in line detection. Monogenic/RF, Linop/RF and Gaussian/RF are natural variants, in which the intermediate multiscale responses previously used to construct the detection outputs are instead combined to given feature representation at each pixel that can subsequently be classified using a random forest. These learning variants were developed for two reasons: firstly, when analyzing quantitative results for detection performance, they allow us to decouple the effect of random forest learning from the effect due to the type of representation used. Secondly, unlike their original variants, Monogenic/RF, Linop/RF and Gaussian/RF can be used in the spicule classification experiment described in 5.3.

In what follows, we present both qualitative and quantitative results for detection and classification for each method outlined above.

5.1 Data

Synthetic Data. We used two sets of synthetic images containing curvilinear structures, with known ground truth, to train and test curvilinear structure detection. We

randomly extracted 4×4 cm (512×512 pixel) mammographic backgrounds with 256 grey-levels 72 (30) normal mammograms for the training (test) set, resulting in 10460 training patches and 4903 test patches, from which naturally occurring linear structures were filtered out. Lines were added to these backgrounds, with parameters drawn from the following distributions: orientation $[0, \pi]$ uniform; width $[4, 16]$ pixels uniform; peak intensity $[1, 255]$ grey-levels (relative to images scaled 0 – 255) from an exponential distribution with half-width 4 grey-levels; profile shape β determined by the equation $\beta = \alpha + (1 - \alpha) \sin x$ for offsets $x \in (0, \pi)$, where the ‘squareness’ parameter α determines how close the shape is to a pure half-cycle sin or a rectangle and is drawn uniformly from $[0, 1]$. The range of widths, distribution of contrasts (with low contrast lines much more likely than high contrast lines) and variability in profile shape were chosen to mimic what is found in real mammograms.

During training, backgrounds were randomly sampled from the 10460 training patches and a single line was added to the centre of the patch. These images were produced ‘on-the-fly’ during each tree-building step of random forest construction as described in section 5.2 and no permanent set was maintained.

For testing, 100 backgrounds were randomly selected from the test patches. To each, multiple lines were added sequentially, with the number and position of lines varying randomly. An example of a synthetic image is shown in Fig 1(a).

Note that all synthetic lines used were straight lines. We conducted experiments explicitly using curved structures, however as there was no performance difference between training on curved or straight lines when detecting curved lines, it was decided that including curves was unnecessary.

Real Mammogram Data. We used real mammogram patches to illustrate qualitative results for curvilinear structure detection and to train and test spicule/non-spicule classification. Data were taken from a sequential set of 84 abnormal mammograms with biopsy-proven malignancy, drawn from a screening population (Nightingale Breast Centre, South Manchester University Hospitals Trust, UK), and from a set of 89 normal mammograms of the contralateral breasts of the same individuals (where disease was radiologically confirmed to be confined to one breast). All mammograms were digitised to a resolution of $90\mu\text{m}$, using a Vidar CADPRO scanner. A 4×4 cm patch was extracted around each abnormality, and a similar patch was sampled randomly from each of the normal mammograms; examples are shown in Fig 4. For each abnormal patch an expert radiologist manually annotated some (though not necessarily all) of the spicules associated with the abnormality, resulting in a total of 555 spicule annotations. The expert spicule annotations for the abnormal images were used as a basis for selecting spicule pixels, though they were not sufficiently accurate to be used directly. To refine the annotations, we initialised a snake [18] using each original annotation, and iterated it to convergence, using evidence from the linear structure probability image. We note that a similar technique has recently been published in detail by Muralidhar et al. [19]. As a result, the 555 refined spicule annotations identified a set of 36,514 spicule pixels. As outlined in section 4, we sampled the same number of pixels from the normal images, such that the detection probability distributions for the spicule and non-spicule samples were the same.

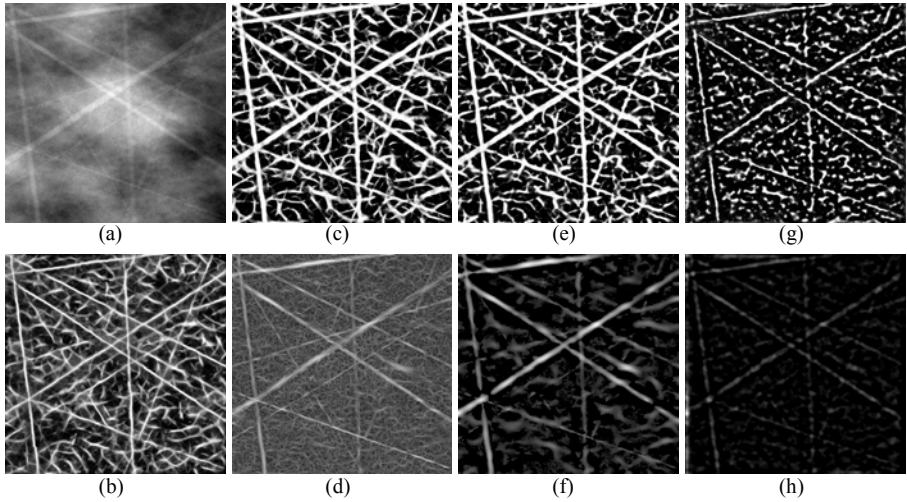


Fig. 1. **a)** Synthetic test image; **b) - h)** results of line detection algorithms (see text for details): **b)** *DT-CWT/RF*; **c)** *Linop/RF*; **d)** *Linop*; **e)** *Gaussian/RF*; **f)** *Gaussian*; **g)** *Monogenic/RF*; **h)** *Monogenic*

5.2 Curvilinear Structure Detection and Orientation Estimation

Curvilinear structure detection and orientation estimation was tested for all seven methods outlined at the beginning of section 5.

For the four learning methods (*DT-CWT/RF*, *Monogenic/RF*, *Linop/RF*, *Gaussian/RF*), we first constructed random forests to classify between structure and background pixels and to compute an estimate of orientation at each pixel.

All forests were constructed with 200 trees and $d = \sqrt{D}$, following published guidelines [2]. However, rather than using a single set of training data from which samples were drawn with replacement (i.e. bootstrapping), we used our method for randomly generating unique synthetic lines images (as described in section 5.1) to construct a new training sample at each tree-building step. For detection, each sample comprised 100k curvilinear structure pixels and 100k background pixels, whilst for orientation regression we used 200k pixels all belonging to curvilinear structure.

Thus for any given representation (*DT-CWT*, *Monogenic*, *Linop*, *Gaussian*) and forest (classification, regression) we applied the following scheme:

1. Generate a new synthetic line image with known ground truth
2. Extract feature vectors for a random set of pixels in the image
3. Repeat 1 & 2 until training sample complete
4. Construct tree using the CART algorithm
5. Repeat 1 to 4 until 200 trees constructed

Details on the composition of feature vectors for each representation type are given below. Note for all methods, the number of scales used s and the neighbourhood size w were empirically tested to select the best combination for each method. In each

case, we tried $s = 3, 4, 5, 6$ and $w = 1, 3$. For reasons of space, results are shown only for the best combination in each method.

- **DT-CWT/RF:** images were decomposed using the DT-CWT to s scales. Neighbourhoods of interpolated phase and magnitude coefficients were extracted in each of the 6 oriented subbands producing a feature vector of $12w^2s$ elements.
- **Monogenic/RF:** the monogenic signal was extracted across s scales, with the initial wavelength set at $\lambda = 4$ pixels. Neighbourhoods of phase, amplitude and orientation values were computed giving a total feature size of $3w^2s$.
- **Linop/RF:** 8 oriented line filters were computed at each scale. Collecting neighbourhoods of the oriented filter responses produced $8w^2s$ elements in each feature vectors.
- **Gaussian/RF:** for the Gaussian 2nd derivative method, the three directional derivatives were applied to an image at each scale. The standard deviation of the smallest kernel was 1 pixel, subsequent kernels increased by a factor of 2 at each scale. As with *Monogenic/RF* this resulted in feature vectors with $3w^2s$ elements.

For testing, feature vectors for each representation were extracted at all pixels in the 100 synthetic test images. The classification and regression forests were then used to compute a line detection score (the proportion of votes for the curvilinear structure class) and orientation (the mean output of each regression tree, with appropriate angular wrapping at 0° and 180°) at each pixel. Example results of line detection are shown in Fig 1 (b)-(e).

In addition to the four learning methods, the prescriptive variants of the Monogenic, Linop and Gaussian approaches were applied to the test images, example results of which are depicted in Fig 1 (f)-(h).

ROC curves for the seven methods tested are shown in Fig 2, using the known ground-truth for the test images to define pixels on the centrelines of curvilinear structures as foreground, and pixels lying outside the structures as background. Areas under the ROC curves and detection sensitivities at a fixed specificity of 90% are shown in table 1. For orientation, only pixels belonging to curvilinear structures (although not necessarily at the centerline) were included in the results. The absolute differences between the predicted and known orientations (with appropriate angle wrapping) were taken, and used to generate cumulative distribution functions of orientation errors for each method, as shown in Fig 2. The mean absolute errors of the estimated orientations are also included in Table 1.

These results show that the four learning methods perform significantly better than the three prescriptive methods (with the exception of orientation computation in *Monogenic/RF*). *DT-CWT/RF* is significantly the strongest performing for both line detection and orientation estimation, followed by *Linop/RF* then *Gaussian/RF*.

As expected, because *Linop*, of the three prescriptive methods, discards the highest proportion of filter responses, *Linop/RF* gains the most from training. This highlights the ability of the random forests to extract useful information from a rich local description of image content, and whilst we do not have a prescriptive variant to compare it to, we believe this shows the importance of training in maximizing the benefit of using the DT-CWT. We also note that the added information that can be gained from the DT-CWT representation results from an increase in the richness of

the local description of texture and is not due to increasing the support of the descriptor. Indeed, as described above we tested all representations over an equivalent range of filter scales so that the same image support was available to each method.

The orientation results for both *Monogenic* and *Monogenic/RF* were surprisingly poor. Counter-intuitively, visual analysis of the test outputs showed that the *Monogenic* methods performed particularly badly at the exact centerline of curvilinear structures, where an essentially random orientation appeared to be returned. This is in contrast to the other tested methods that, as expected, performed strongest along the centerlines of structures. Computing estimation errors at pixels belonging to structures but not lying on the centerline produces a small improvement in the results (mean absolute errors of 32.55° and 29.39° for the *RF* and prescriptive variant respectively), though even then performance lags behind the other tested methods and of course in a real image we do not know the precise location of structure centerlines. Determining why orientations are estimated so poorly by the monogenic signal at the centre of structures, where in theory the signal is strongest, may warrant further investigation.

Table 1. Line detection and orientation computation results

Line detection algorithm	ROC A_z	Sensitivity for fixed 90% specificity	Mean Absolute Error of orientation
DT-CWT/RF, $w = 1, s = 5$	0.923 ± 0.00036	0.792	15.88
Linop/RF, $w = 1, s = 5, 8$ orientations	0.911 ± 0.00048	0.757	19.35
Gaussian/RF, $w = 1, s = 4, \sigma_{min} = 1$	0.901 ± 0.00048	0.731	21.37
Monogenic/RF, $w = 1, s = 4, \lambda = 4$	0.868 ± 0.00045	0.643	33.73
Monogenic, $s = 3, \lambda = 4$	0.818 ± 0.00055	0.547	30.86
Gaussian, $s = 4, \sigma_{min} = 1$	0.813 ± 0.00053	0.533	24.27
Linop, $s = 5, 8$ orientations	0.737 ± 0.00060	0.413	29.32

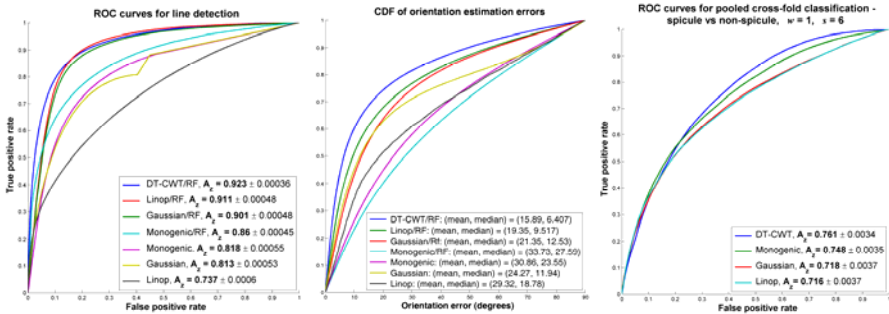


Fig. 2. *left*: Receiver operating characteristic (ROC) curves for different curvilinear structure detection methods; *centre*: Cumulative distribution functions (CDF) of errors in orientation estimation for the different methods; *right*: ROC curves for spicule classification.

To show qualitative results for real mammograms, the seven methods were applied to detect curvilinear structures and estimate orientation for the malignant regions described in section 5.1. Example detection results are shown in Fig 3. Assessing the results visually, it would appear that the outputs of the learning based methods (and particularly *DT-CWT/RF*, *Linop/RF* and *Gaussian/RF*) are similar to the output of synthetic data whilst capturing the key curvilinear structures in the real regions. This would suggest our model for producing synthetic lines generates good data from which to train random forests for real images. Of course validating this hypothesis is important and we are currently working on a quantitative evaluation of the methods in real data when used, for example, as part of a larger lesion detection scheme.

In terms of applying the learning methods to real images, it is worth noting how the methods scale with increasing image size – particularly above the point at which the set of feature vectors for all image pixels can be stored in working memory. For the DT-CWT, provided the initial decomposition can be stored in memory (which due to its low-redundant decimating construction is possible even for full size mammograms of the order 3000x2400 pixels) then interpolated coefficients can be efficiently sampled to generate feature vectors for block-wise regions of the image. Each block of feature vectors can be classified by the forest and cleared from working from memory storing only the output of the forest. In this way only a small overhead is introduced for block-wise classifying the image. However, for the other methods it becomes necessary to interleave the decomposition of the image with the sampling of feature vectors. For example, it may be necessary to apply the filters at a single scale, extract features for that scale for a particular block of the image, filter at the next scale and extract those features, and so on. Of course, when the complete feature vectors for a single block have been classified, the process repeats. Thus a large image may in

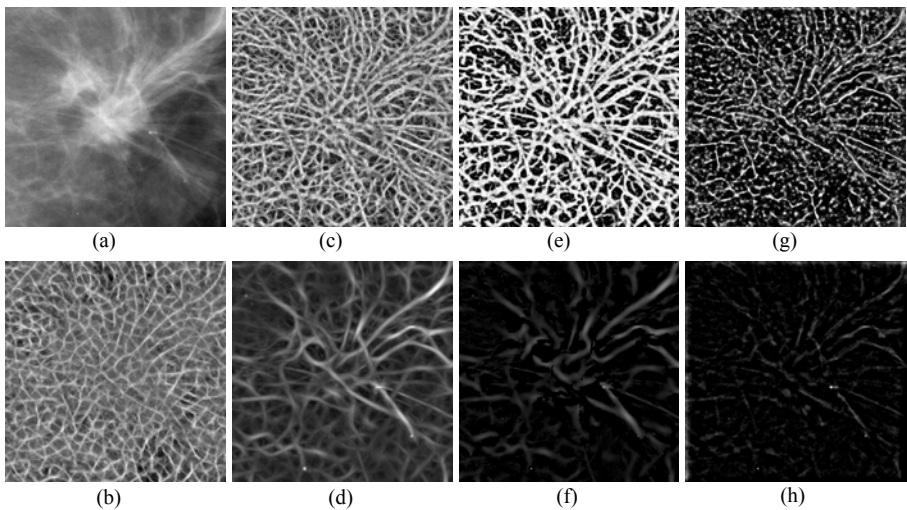


Fig. 3. a) Mammogram region containing malignant spiculated mass; b) - h) results of line detection algorithms (see text for details): b) *DT-CWT/RF*; c) *Linop/RF*; d) *Linop*; e) *Gaussian/RF*; f) *Gaussian*; g) *Monogenic/RF*; h) *Monogenic*

fact end up by decomposed many times over introducing a large computational overhead for block-wise processing. The point at which this cost occurs will depend on the size of the image and the type of representation used. Obviously the cost is worst for *Linop* which requires storing 8 (i.e. the number orientations) full copies of the image at each scale, compared to just 3 for the Gaussian and Monogenic methods.

5.3 Spicule Classification

The four learning-based methods were also applied to the problem of spicule/non-spicule classification. Feature vectors were formed as above, and random forest classifiers were trained using balanced spicule/non-spicule training data, as outlined in section 4. To make effective use of the available data, we used a 10-fold cross-validation design. The set of normal and abnormal regions were divided into 10 groups so that the total number of normal and spicule pixels in each group were as close as possible to a 10^{th} of the total and no two views from the same case were included in different groups. The samples in each group were then classified using a random forest trained on the samples from the remaining 9 groups. The classification results from each group were pooled to generate an unbiased class probability for each sampled pixel. These probabilities were used to compute an ROC curve for each training regime, and the area under the curve (A_z) was computed and used as a measure of classification performance. The ROC curves and A_z values for the three methods are shown in Fig 2 and Table 2. These results demonstrate a clear advantage for *DT-CWT/RF*. As might be expected, because the *Linop* and *Gaussian* representations do not really capture profile shape, they perform significantly worse than the two representations that include phase.

Table 2. Results for spicule classification

Feature representation for spicule classification	ROC A_z
<i>DT-CWT/RF</i> , $w = 1$, $s = 6$, all orientations	0.761±0.0034
Monogenic/RF, $w = 1$, $s = 5$, 8 orientations	0.748±0.0035
Gaussian/RF, $w = 1$, $s = 5$, $\sigma_{\min} = 1$	0.718±0.0037
<i>Linop/RF</i> , $w = 1$, $s = 5$, $\lambda = 4$	0.716±0.0037

In addition to computing a class vote for spicule membership at only those pixels in the selected training sets, the forests we have constructed can be used to compute results for whole region in each cross-fold group. Typical qualitative *DT-CWT/RF* results for a normal and abnormal region are shown in Fig 4. In the left column, the original regions are shown. The spiculations of the mass are clear and well defined, particularly to the south-east of the central mass. In the normal region, there are a set of structures that intersect in an approximate radial pattern that may trigger a feature detector erroneously. In the right column, the predicted spicule class membership is shown as hue varying from cyan (normal) to pink (spicule), modulated by the output of the *DT-CWT/RF* detection method. Note how the linear structures in the region of the mass are deemed highly likely to be spicules, whilst those in the normal region are not. This shows excellent promise as a means of providing a relevance measure to methods for abnormality detection.

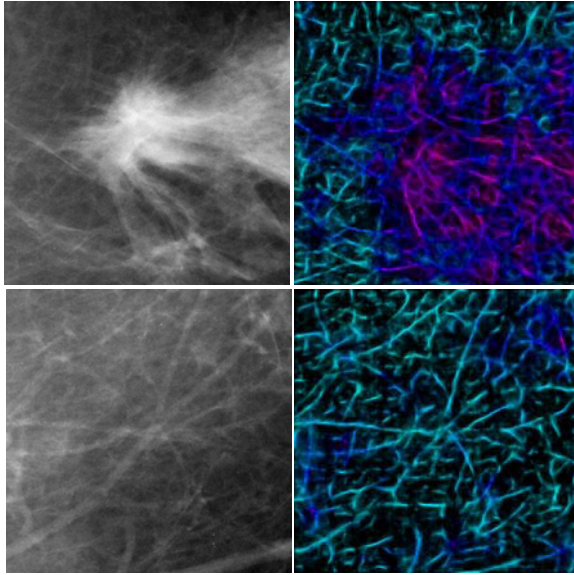


Fig. 4. Left column: an malignant (top) and normal (bottom) region; right column: spicule probability depicted as hue from cyan (normal) to pink (spicule), modulated by strength line detection output from the *DT-CWT* method

6 Discussion and Conclusions

We have presented a discriminative learning-based approach to the detection and classification of curvilinear structures, based on a combination of *DT-CWT* representation of local structure and random forest classification. We have applied the method to the challenging problem of detecting and estimating the orientation of curvilinear structures in mammograms and distinguishing between normal and abnormal structures. The results of our experimental evaluation are extremely encouraging, and represent a significant improvement over the current state of the art.

We have also introduced learning-based variants of three existing methods, demonstrating that whilst learning accounts for a significant part of this improvement, the choice of representation is also important and will have a different effect on performance depending on the task in hand. For example, constructing a representation based on the raw responses to Linop filters produces features that are excellent for estimating structure orientation but provide less information for determining structure shape and thus type. Conversely, features formed from the monogenic signal are good at determining structure type – most likely because of the inclusion of the phase measure – whilst they perform relatively poorly at detection and orientation estimation. For these reasons, it seems fair to conclude that the *DT-CWT* provides the best all round representation. It produced the strongest performance for all three tasks (curvilinear structure detection, orientation estimation and spicule classification). Moreover, as discussed in section 5.2, of all the methods, the *DT-CWT* incurs the least overhead

when working with full-size real images that require block-wise classification/regression. For example, initial tests show that the structure detection and orientation regression can be performed on a full-size (~3000 x 2400 pixels) mammogram in ~1hr 30mins.

Our next goal is to show that improving the individual steps of curvilinear structure and orientation estimation result in a subsequent improvement for a high level task such as detecting patterns of spiculations indicative of disease. Moreover we hope to show that classification into structure type can further aid such tasks by focusing only (or at least more) on those structures most likely to be associated with disease.

Acknowledgements

We thank Nick Kingsbury for the DT-CWT Matlab toolbox. Mammograms were provided by the Nightingale Breast Centre, South Manchester University Hospitals Trust, UK and were annotated by Dr Caroline Boggis and Dr Rumana Rahim. This work was funded by EPSRC grant EP/E031307/1.

References

1. Kingsbury, N.: Complex wavelets for shift invariant analysis and filtering of signals. *Applied and Computational Harmonic Analysis* 10(3), 234–253 (2001)
2. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
3. Burrell, H.C., Evans, A.J., Robin, A., Wilson, M., et al.: False-negative breast screening assessment: what lessons can we learn? *Clin. Radiol.* 56(5), 385–388 (2001)
4. Baker, J.A., Rosen, E.L., Lo, J.Y., Gimenez, E.I., et al.: Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion. *Am. J. Roentgenol.* 181(4), 1083–1088 (2003)
5. Karssemeijer, N., te Brake, G.M.: Detection of stellate distortions in mammograms. *IEEE Trans. Med. Im.* 15(5), 611–619 (1996)
6. Parr, T.C., Taylor, C.J., Astley, S.M., Boggis, C.R.M.: Statistical modeling of oriented line patterns in mammograms. In: *Proc. SPIE Medical Imaging*, vol. 3034, pp. 44–55 (1997)
7. Rangayyan, R., Ayres, F.: Gabor filters and phase portraits for the detection of architectural distortion in mammograms. *Medical and Biological Engineering and Computing* 44(10), 883–894 (2006)
8. Dixon, R.N., Taylor, C.J.: Automated asbestos fibre counting. In: *Institute of Physics Conference Series*, vol. 44, pp. 178–185 (1979)
9. Cerneaz, N., Brady, M.: Finding curvilinear structures in mammograms. In: *Computer Vision, Virtual Reality and Robotics in Medicine*, pp. 372–382 (1995)
10. Zwiggelaar, R., Astley, S.M., Boggis, C.R.M., Taylor, C.J.: Linear structures in mammographic images: detection and classification. *IEEE Trans. Med. Im.* 23(9), 1077–1086 (2004)
11. Schenk, V., Brady, M.: Finding CLS Using Multiresolution Orientated Local Energy Feature Detection. In: *Proc. 6th International Workshop on Digital Mammography*, Bremen, Germany (2002)
12. McLoughlin, K.J., Bones, P.J., Kovesi, P.: Connective tissue representation for detection of microcalcifications in digital mammograms. In: *Proc. SPIE Medical Imaging*, vol. 4684, pp. 1246–1256 (2002)

13. Freeman, W.T., Adelson, E.H.: The Design and Use of Steerable Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 891–906 (1991)
14. Wai, L.C.C., Mellor, M., Brady, J.M.: A Multi-resolution CLS Detection Algorithm for Mammographic Image Analysis. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004*. LNCS, vol. 3217, pp. 865–872. Springer, Heidelberg (2004)
15. Chen, Z., Berks, M., Astley, S.M., Taylor, C.J.: Classification of Linear Structures in Mammograms Using Random Forests. In: Martí, J., Oliver, A., Freixenet, J., Martí, R. (eds.) *IWDM 2010*. LNCS, vol. 6136, pp. 153–160. Springer, Heidelberg (2010)
16. Anderson, R., Kingsbury, N., Fauqueur, J.: Coarse-level object recognition using interlevel products of complex wavelets. In: *Proc. IEEE International Conference on Image Processing*, pp. 745–748 (2005)
17. Kingsbury, N.: Rotation-Invariant Local Feature Matching with Complex Wavelets. In: *Proc. European Conference on Signal Processing, EUSIPCO* (2006)
18. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
19. Muralidhar, G.S., Bovik, A.C., Giese, J.D., Sampat, M.P., et al.: Snakules: A Model-Based Active Contour Algorithm for the Annotation of Spicules on Mammography. *IEEE Trans. Med. Im.* 29(10), 1768–1780 (2010)

A Unified Framework for Joint Segmentation, Nonrigid Registration and Tumor Detection: Application to MR-Guided Radiotherapy*

Chao Lu¹, Sudhakar Chelikani², and James S. Duncan^{1,2}

¹ Department of Electrical Engineering, School of Engineering & Applied Science

² Department of Diagnostic Radiology, School of Medicine

Yale University, New Haven, CT, USA

{chao.lu,sudhakar.chelikani,james.duncan}@yale.edu

Abstract. Image guided external beam radiation therapy (EBRT) for the treatment of cancer enables accurate placement of radiation dose to the cancerous region. However, the deformation of soft tissue during the course of treatment, such as in cervical cancer, presents significant challenges. Furthermore, the presence of pathologies such as tumors may violate registration constraints and cause registration errors. In this paper, we present a unified MAP framework that performs automatic segmentation, nonrigid registration and tumor detection simultaneously. It can generate a tumor probability map while progressively identifying the boundary of an organ of interest based on the achieved transformation. We demonstrate the approach on a set of 30 T2-weighted MR images, and the results show that the approach performs better than similar methods which separate the registration and segmentation problems. In addition, the detection result generated by the proposed method has a high agreement with the manual delineation by a qualified clinician.

1 Introduction

External beam radiation therapy(EBRT) is the primary modality of treatment for cervical cancer [12]. Imaging systems capable of visualizing patient soft-tissue structure in the treatment position have become the dominant method of positioning patients for both conformal and stereotactic radiation therapy. Traditional CT images suffer from low resolution and low soft tissue contrast, while Magnetic Resonance (MR) imaging can characterize deformable structure with superior visualization and differentiation of normal soft tissue and tumor-infiltrated soft tissue. In addition, advanced MR imaging with modalities such as diffusion, perfusion and spectroscopic imaging has the potential to better localize and understand the disease and its response to treatment. Therefore, magnetic resonance guided radiation therapy (MRgRT) systems with integrated MR imaging in the treatment room are now being developed as an advanced system in radiation therapy [7].

* This work is supported by NIH/NIBIB Grant R01EB002164.

The goal in radiotherapy cancer treatment is to deliver as much dose as possible to the clinical target volume (CTV), while trying to deliver as little dose as possible to surrounding organs at risk [4,10]. When higher doses are to be delivered, precise and accurate targeting is essential because of unpredictable inter- and intra-fractional organ motions over the process of the weekly treatments, which require accurate nonrigid mapping between treatment and planning day. However, the presence of pathologies such as tumors may violate registration constraints and cause registration errors, because the abnormalities often invalidate gray-level dependency assumptions that are usually made in the intensity-based registration. In addition, the registration problem is challenging due to the missing correspondences caused by tumor regression. Incorporating some knowledge about abnormalities can improve the registration. Meanwhile, segmentation of the soft tissue structure over the course of EBRT is challenging due to structure adjacency and deformation over treatment.

Some initial work has been performed in simultaneously registration and segmentation or incorporating abnormality detection. Chelikani *et.al.* integrated rigid 2D portal to 3D CT registration and pixel classification in an entropy-based formulation [3]. Yezzi *et.al.* integrated segmentation using deformable model with rigid and affine registration [17]. Pohl *et.al.* performed voxel-wise classification and registration which could align an atlas to MRI [14]. The automatic mass detection problem mainly focuses on the application on mammography [13]. Hachama *et.al.* proposed a registration technique which neglects the deformation generated by abnormalities [5], which is not always the case, and raises questions regarding the applicability of the method.

In this paper, we present a probability based technique as an extension to the work in [11]. Our model is based on a MAP framework which can achieve segmentation and nonrigid registration while simultaneously estimating a probability tumor map, which estimates each voxel's probability of belonging to a tumor. In this manner, we can interleave these processes and take proper advantage of the dependence among them. Using the proposed approach, it is easy to calculate the location changes of the lesion for diagnosis and assessment, thus we can precisely guide the interventional devices toward the tumor during radiation therapy. Escalated dosages can then be administered while maintaining or lowering normal tissue irradiation. There are a number of clinical treatment sites that will benefit from our MR-guided radiotherapy technology, especially when tumor regression and its effect on surrounding tissue can be significant. Here we focus on the treatment of cervical cancer as the key example application.

2 Description of the Model

2.1 A Unified Framework: Bayesian Formulation

Let I_0 and I_d be the planning day (day 0) and treatment day (day d) 3D MR images respectively, and S_0 be the set of segmented planning day organs. A unified framework is developed using Bayesian analysis to calculate the most likely segmentation in treatment day fractions S_d , the deformation field between the

planning day and treatment day data T_{0d} , and the tumor map M at treatment day which associates with each pixel of the image its probability of belonging to a tumor.

$$\widehat{S}_d, \widehat{T}_{0d}, \widehat{M} = \arg \max_{S_d, T_{0d}, M} [p(S_d, T_{0d}, M | I_d, S_0)] \quad (1)$$

This multiparameter MAP estimation problem in general is difficult to solve, however, we reformulate this problem such that it can be solved in two basic iterative computational stages using an iterative conditional mode (ICM) strategy [11]. With k indexing each iterative step, we have:

$$\widehat{S}_d^k = \arg \max_{S_d^k} [p(S_d^k | T_{0d}^k(S_0), I_d, M^k)] \quad (2)$$

$$\widehat{T}_{0d}^{k+1}, \widehat{M}^{k+1} = \arg \max_{T_{0d}^{k+1}, M} [p(T_{0d}^{k+1}, M^{k+1} | S_d^k, S_0, I_d, I_0)] \quad (3)$$

These two equations represent the key problems we are addressing. Equation (2) estimates the segmentation of the important day d structure (S_d^k). Equation (3) estimates the next iterative step of the mapping T_{0d}^{k+1} between the day 0 and day d spaces, as well as the probability tumor detection map M^{k+1} .

2.2 Segmentation

In this paper, we use Eq. (2) to segment the normal or non-cancerous organs, hence we assume the segmentations S_d^k here are independent of the tumor detection map M . Bayes rule is applied to Eq. (2):

$$\begin{aligned} \widehat{S}_d^k &= \arg \max_{S_d^k} [p(S_d^k | T_{0d}^k(S_0), I_d, M^k)] = \arg \max_{S_d^k} [p(S_d^k | T_{0d}^k(S_0), I_d)] \\ &= \arg \max_{S_d^k} [p(I_d | S_d^k) \cdot p(T_{0d}^k(S_0) | S_d^k) \cdot p(S_d)] \end{aligned} \quad (4)$$

Shape Prior Information. The segmentation module here is similar to the previous work in [11]. Here we assume that the priors are stationary over the iterations, so we can drop the k index for that term only, i.e. $p(S_d^k) = p(S_d)$. Instead of using a point model to represent the object, we choose level set as the representation of the object to build a model for the shape prior [9], and then define the joint probability density function in Eq.(4).

Consider a training set of n rigidly aligned images, with M objects or structures in each image. The training data were generated from manual segmentation by a qualified clinician. Each object in the training set is embedded as the zero level set of a higher dimensional level set Ψ with negative distances inside and positive distances outside the object. Using the technique developed by Leventon *et al.* in [9], the mean ($\bar{\Psi}$) and variance of the boundary of each object can be computed using principal component analysis (PCA). An estimate of the object

shape Ψ_i can be represented by k principal components and a k -dimensional vector of coefficients (where $k < n$) α_i

$$\tilde{\Psi}_i = U_k \alpha_i + \bar{\Psi} \tag{5}$$

where U_k is a matrix generated by PCA. Under the assumption of a Gaussian distribution of object represented by α_i , we can compute the probability of a certain shape in Eq.(6), where Σ is the diagonal matrix of corresponding singular values.

$$p(S_d) = p(\alpha_i) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_k|}} \exp \left[-\frac{1}{2} \alpha_i^T \Sigma_k^{-1} \alpha_i \right] \tag{6}$$

Segmentation Likelihoods. Then we imposed a key assumption: the likelihood term is separable into two independent data-related likelihoods, requiring that the estimate of the structure at day d be close to: i.) the same structure segmented at day 0, but mapped to a new estimated position at day d by the current iterative mapping estimate T_{od}^k and ii.) the intensity-based feature information derived from the day d image [11].

In the Equation (4), $p(T_{od}^k(S_0)|S_d)$ constrains the organ segmentation in day d to be adherent to the transformed day 0 organs by current mapping estimate T_{od}^k . Thus, we made the assumption about the probability density of day 0 segmentation likelihood term:

$$p(T_{od}^k(S_0)|S_d^k) = \frac{1}{Z} \prod_{(x,y,z)} \exp \left[-(\Psi_{T_{od}^k(S_0)} - \Psi_{S_d^k})^2 \right] \tag{7}$$

In equation (4), $p(I_d|S_d^k)$ indicates the probability of producing an image I_d given S_d^k . In three-dimensions, assuming gray level homogeneity within each object, we use the imaging model defined by Chan and Vese [2] in Eq.(8), where c_1 and σ_1 are the average and standard deviation of I_d inside S_d^k , c_2 and σ_2 are the average and standard deviation of I_d outside S_d^k .

$$p(I_d|S_d^k) = \prod_{inside(S_d^k)} \exp[-(I_d(\mathbf{x}) - c_1)^2/2\sigma_1^2] \cdot \prod_{outside} \exp[-(I_d(\mathbf{x}) - c_2)^2/2\sigma_2^2] \tag{8}$$

Energy Function. Combining(4),(6),(7),(8),we introduce the segmentation energy function E_{seg} defined by:

$$\begin{aligned} E_{seg} &= -\ln p(S_d^k|T_{od}^k(S_0), I_d) = -\ln p(I_d|S_d^k) \cdot p(T_{od}^k(S_0)|S_d^k) \cdot p(S_d^k) \\ &\propto \lambda_1 \int_{inside(S_d^k)} |I_d(\mathbf{x}) - c_1|^2 d\mathbf{x} + \lambda_2 \int_{outside(S_d^k)} |I_d(\mathbf{x}) - c_2|^2 d\mathbf{x} \\ &+ \gamma \int_{\mathbf{x}} \left| \Psi_{T_{od}^k(S_0)} - \Psi_{S_d^k} \right|^2 d\mathbf{x} + \omega_i \alpha_i^T \Sigma_k^{-1} \alpha_i \end{aligned} \tag{9}$$

Notice that the MAP estimation of the objects in (4), $\widehat{S_d^k}$, is also the minimizer of the above energy functional E_{seg} . This minimization problem can be formulated and solved using the level set surface evolving method.

2.3 Registration and Tumor Detection

The second stage of the proposed strategy described above in Eq. (3) can be further developed using Bayes rule as follows in Eq.(10). As indicated in Section 2.2, the segmentations S_d^k are independent of the tumor detection map M , and the priors are stationary over the iterations.

$$\begin{aligned}
 \widehat{T_{0d}^{k+1}}, \widehat{M^{k+1}} &= \arg \max_{T_{0d}^{k+1}, M^{k+1}} [p(T_{0d}^{k+1}, M^{k+1} | S_d^k, S_0, I_d, I_0)] \\
 &= \arg \max_{T_{0d}^{k+1}, M^{k+1}} [\ln p(S_d^k, S_0 | T_{0d}^{k+1}, M^{k+1}) + \ln p(I_d, I_0 | T_{0d}^{k+1}, M^{k+1}) + \ln p(T_{0d}, M)] \\
 &= \arg \max_{T_{0d}^{k+1}, M^{k+1}} [\ln p(S_d^k, S_0 | T_{0d}^{k+1}) + \ln p(I_d, I_0 | T_{0d}^{k+1}, M^{k+1}) + \ln p(T_{0d} | M) + \ln p(M)]
 \end{aligned} \tag{10}$$

The first term on the right hand side represents conditional likelihood related to mapping the segmented soft tissue structures at days 0 and d , and the second term registers the intensities of the images while simultaneously estimating the probability tumor map. The third and fourth terms represent prior assumptions and constraints on the overall nonrigid mapping and tumor map.

Segmented Organ Matching. As discussed in the segmentation section, each object is represented by the zero level set of a higher dimensional level set Ψ . Assuming the objects vary during the treatment process according to a Gaussian distribution, and given that the different organs can be mapped respectively, we further simplifies the organ matching term as:

$$\begin{aligned}
 \ln p(S_d^k, S_0 | T_{0d}^{k+1}) &= \sum_{obj=1}^N \ln p(S_{d-obj}^k, S_{0-obj} | T_{0d}^{k+1}) \\
 &= \sum_{obj=1}^N \int_{\mathbf{x}} \ln \frac{1}{\sqrt{2\pi}\sigma_{obj}} \exp \left[\frac{-(\Psi_{T_{0d}^{k+1}(S_{0-obj})} - \Psi_{S_{d-obj}^k})^2}{2\sigma_{obj}^2} \right] d\mathbf{x} \tag{11} \\
 &= \sum_{obj=1}^N -\omega_{obj} \int_{\mathbf{x}} [\Psi_{T_{0d}^{k+1}(S_{0-obj})} - \Psi_{S_{d-obj}^k}]^2 d\mathbf{x}
 \end{aligned}$$

with N represent the number of objects in each image, and ω_{obj} are used to weight different organs. When minimized w.r.t. T_{0d}^{k+1} , the organ matching term ensures the transformed day 0 organs and the segmented day d organs align over the regions [\[11\]](#).

Intensity Matching and Tumor Probability Map. In order to define the likelihood $p(I_d, I_0 | T_{0d}^{k+1}, M^{k+1})$, we assume conditional independence over the voxel locations \mathbf{x} , as discussed in [\[18\]](#).

$$p(I_d, I_0 | T_{0d}^{k+1}, M^{k+1}) = \prod_{\mathbf{x}} p(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x})) | M^{k+1}(\mathbf{x})) \tag{12}$$

Different from the work in [11] which only uses a single similarity metric, here we model the probability of the pair $p(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x})) | M^{k+1}(\mathbf{x}))$ to be dependent on the class of the pixel \mathbf{x} . Each class is characterized by a probability distribution, denoted by p_N for the normal tissue and p_T for the tumor. Let $M^{k+1}(\mathbf{x})$ be the tumor map which associates to \mathbf{x} its probability to belong to a tumor. Thus, the probability distribution $p(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x})) | M^{k+1}(\mathbf{x}))$ can be defined as a mixture of the two class distribution:

$$p(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x})) | M^{k+1}(\mathbf{x})) = (1 - M^{k+1}(\mathbf{x}))p_N(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x}))) + M^{k+1}(\mathbf{x})p_T(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x}))) \tag{13}$$

Normal Tissue Class. Across the treatment procedure, the tumor experiences a regression process. When the tumor shrinks, some part of the tumor returns to the intensity level around the normal tissue. Thereafter, the normal tissue in treatment day MR has two origins. One is the normal tissue in planning day MR, and the other comes from the tumor in planning day and returns to normal due to the tumor regression. We choose two different probability for these two types. The histograms of normal cervical tissue and tumor are shown in Fig 1. From clinical research [6] and Fig 1, we can see that the intensities of tumor are generally much higher than those of normal cervical tissue.

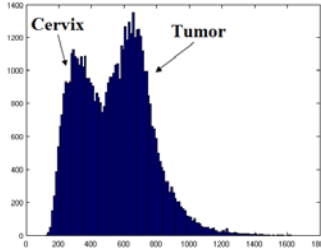


Fig. 1. The histograms of normal cervix and tumor

Therefore, for the sake of simplicity, we characterize the normal tissue from the second type (used to be tumor but returned to normal later on) as areas with much lower intensity in treatment day MR [6]. We assume a voxel labeled as this part in day d MR can match any voxel in day 0 tumor with equal probability and use a uniform distribution. The remaining voxels are labeled as type 1 normal tissue (always normal since the planning day), which are modeled assuming a discrete Gaussian distribution across the corresponding voxel locations.

$$p_N(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x}))) = \begin{cases} 1/c, & T_{0d}^{k+1}(I_0(\mathbf{x})) - I_d(\mathbf{x}) > \Delta \\ \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{|T_{0d}^{k+1}(I_0(\mathbf{x})) - I_d(\mathbf{x})|^2}{2\sigma^2}), & otherwise \end{cases} \tag{14}$$

where c is the number of voxels in the day 0 tumor, and Δ is the predetermined threshold used to differentiate the intensity of normal and tumor tissue.

Tumor Class. The definition of the tumor distribution is a difficult task. Similar to normal tissue, the tumor tissue in treatment day MR also has two origins. One is the tumor tissue in planning day MR, which represents the remaining tumor after the radiotherapy. We assume the voxel in the residual tumor can match any voxel in the initial tumor (day 0) with equal probability and use a uniform distribution. The other origin of tumor class in day d is from the normal tissue in planning MR, but the tissue grows into malignant tumor later on. This type is characterized with much higher intensity in day d image [6]. We assume that each normal voxel in the day 0 MR can turn into a tumor voxel with equal probability. Thus this type is also modeled using another uniform distribution, but with less probability, since the chance of this deterioration process is relatively small across the radiotherapy treatment.

$$p_T(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x}))) = \begin{cases} 1/(V - c), & T_{0d}^{k+1}(I_0(\mathbf{x})) - I_d(\mathbf{x}) < -\Delta \\ 1/c, & \text{otherwise} \end{cases} \quad (15)$$

where V is the total number of voxels in MR image, c is the number of voxels in the day 0 tumor, and it's the same c as in Eq.(14).

Non-negative Jacobian. We want the transformation to be non-singular hence we would like to have a large penalty on negative Jacobian for normal tissues [8]. Meanwhile, we simulate the tumor regression process by constraining the determinant of the transformation's Jacobian at $M(\mathbf{x})$ between 0 and 1.

$$p(T_{0d}(\mathbf{x})|M(\mathbf{x})) = (1 - M(\mathbf{x}))p_N(T_{0d}(\mathbf{x})) + M(\mathbf{x})p_T(T_{0d}(\mathbf{x})) \quad (16)$$

where p_N and p_T are calculated using continuous logistic functions: ($\epsilon = 0.1$)

$$p_N(T_{0d}(\mathbf{x})) = \frac{1}{V - c} \left(\frac{1}{1 + \exp(-|J(T_{0d}(\mathbf{x}))|/\epsilon)} \right)^2 \quad (17)$$

$$p_T(T_{0d}(\mathbf{x})) = \frac{1}{c} \left[\left(\frac{1}{1 + \exp(\frac{-|J(T_{0d}(\mathbf{x}))|}{\epsilon})} \right)^2 - \left(\frac{1}{1 + \exp(\frac{1-|J(T_{0d}(\mathbf{x}))|}{\epsilon})} \right)^2 \right] \quad (18)$$

V and c are defined the same as above, and they are used to normalize the probability density functions. These constraints (p_N and p_T) penalize negative Jacobians and p_T simulates the tumor regression process, thus reduce the probability of folding in the registration maps.

Tumor Map Prior. We assume that the tumor map arises from a Gibbs distribution. Many specific terms can be defined to describe the spatial configurations of different type of lesion. The previous detection approaches have a tendency to *over-detect* (to find more regions than the real ones) [13], hence in this paper, we use an energy restricting the total amount of abnormal pixels in the image.

$$p(M) = \frac{1}{Z} e^{-U(M)} = \frac{1}{Z} e^{-\sum_{\mathbf{x}} M(\mathbf{x})} \quad (19)$$

Energy Function. Combine the above equations, we introduce:

$$\begin{aligned}
 E_{reg-det}(T, M) &= -\ln p(T_{0d}^{(k+1)}, M^{(k+1)} | S_d^k, S_0, I_d, I_0) \\
 &= \sum_{obj=1}^N \omega_{obj} \int_{\mathbf{x}} \left[\Psi_{T_{0d}^{k+1}(S_{0-obj})} - \Psi_{S_d^k-obj} \right]^2 d\mathbf{x} \\
 &\quad - \int_{\mathbf{x}} \ln p(I_d(\mathbf{x}), T_{0d}^{k+1}(I_0(\mathbf{x})) | M^{k+1}(\mathbf{x})) d\mathbf{x} - \int_{\mathbf{x}} \ln p(T_{0d}(\mathbf{x}) | M(\mathbf{x})) d\mathbf{x} + \sum_{\mathbf{x}} M(\mathbf{x})
 \end{aligned} \tag{20}$$

T_{0d}^{k+1} and M^{k+1} are estimated using a conjugate gradient method to minimize Eq.(20). Eqs (9) and (20) run alternatively until convergence, then the soft tissue segmentation as well as the nonrigid registration and tumor detection in day d benefit from each other and can be estimated simultaneously.

3 Results

We tested our proposed method on 30 sets of MR data acquired from five different patients undergoing EBRT for cervical cancer at Princess Margret Hospital. Each of the patient had six weekly 3D MR images. A GE EXCITE 1.5-T magnet with a torso coil was used in all cases. T2-weighted, fast spin echo images (voxel size $0.36mm \times 0.36mm \times 5mm$, and image dimension $512 \times 512 \times 38$) were acquired. The clinician performed the bias field correction so that the intensities could be directly compared, and the Δ in Eq.(14) and (15) was chosen to be 150. The MR images were resliced to be isotropic. We adopted a "leave-one-out" test so that all the tested images were not in their training sets.

Segmentation Results. Fig 2(a) shows an example of MR data set with extracted 3D bladder (green) and uterus (purple) surfaces using the proposed method. Fig 2(b)(c) show the segmentation of bladder and uterus from axial view respectively, compared with manual segmentation.

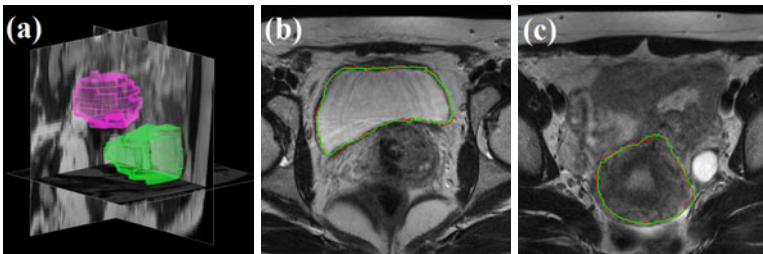


Fig. 2. Segmentation Results. (a) An example of MR data set with automatically extracted bladder (green) and uterus (purple) surfaces. (b) Comparing bladder segmentations using the proposed algorithm (green) and manual segmentation (red). (c) Uterus results using the proposed algorithm (green) and manual segmentation (red).

We compare the experimental results from our algorithm with those obtained from the segmentation using traditional level set active shape models (without priors, and with shape priors only). To quantify the accuracy of our approach, we use mean absolute distance (MAD) and Hausdorff distance (HD) to evaluate the segmentation performance. While MAD is a global measure of the match of two surfaces, HD reflects their local similarities. Table 1 and 2 quantitatively analyze the segmentation results on bladder and uterus surfaces respectively. We observe that both MAD and HD decrease with the proposed method, which implies that our method has a consistent agreement with the manual segmentation.

Table 1. Evaluation of Segmentation of Bladder

	MAD(mm)	HD(mm)
Without prior	7.61 ± 2.69	8.63 ± 3.18
With shape prior	3.28 ± 1.35	4.10 ± 1.41
Proposed method	0.82 ± 0.10	1.20 ± 0.15

Table 2. Evaluation of Segmentation of Uterus

	MAD(mm)	HD(mm)
Without prior	10.57 ± 6.19	12.13 ± 5.88
With shape prior	6.11 ± 1.34	7.28 ± 1.21
Proposed method	1.26 ± 0.38	1.92 ± 0.75

Registration Results. The mapping performance of the proposed algorithm was also evaluated. For comparison, a conventional intensity-based FFD non rigid registration (NRR) [16] and rigid registration (RR) were performed on the same sets of real patient data. The corresponding difference images of the deformed bladder are shown in Fig.3.



Fig. 3. Example of different transformations on the registration for one patient. (a) Difference image of bladder after rigid registration. (b) After nonrigid registration. (c) After the proposed method.

Organ overlaps between the ground truth in day d and the transformed organs from day 0 were used as metrics to assess the quality of the registration (Table 3). We also tracked the registration error percentage between the ground truth in day d and the transformed organs from day 0 for bladder and uterus, as shown in Table 4. The registration error is represented as percentage of false positives (PFP), which calculates the percentage of a non-match being declared to be a match. Let Ω_A and Ω_B be two regions enclosed by the surfaces A and B respectively, PFP is defined as follows:

$$PFP = \frac{Volume(\Omega_B) - Volume(\Omega_A \cap \Omega_B)}{Volume(\Omega_A)}$$

From the Table 3 and Table 4, we found that the RR performed the poorest out of all the registrations algorithms, while the proposed method significantly outperformed the NRR at aligning segmented organs.

Table 3. Evaluation of Registration: Organ Overlaps (%)

	Bladder	Uterus
Rigid Registration	65.76 ± 6.69	68.32 ± 5.18
Nonrigid Registration	80.34 ± 2.39	77.28 ± 3.84
Proposed Method	90.58 ± 1.95	87.80 ± 2.24

Table 4. Evaluation of Registration Error: PFP (%)

	Bladder	Uterus
Rigid Registration	33.57 ± 5.34	30.92 ± 5.22
Nonrigid Registration	19.21 ± 3.41	21.78 ± 4.03
Proposed Method	9.36 ± 1.40	12.06 ± 1.96

Detection Results. We compare the tumor images obtained using our method with the manual detection performed by a clinician. For the proposed method, we set a threshold for the tumor probability map M . Fig.4(b) shows the detection rely entirely on the histogram in Fig.1 alone, which can not provide a meaningful result. Fig.4(c)-(h) provide the comparison between the proposed method and the detection by an expert. The tumor binary images are obtained with the probability threshold 0.7, which presents all the voxels that have an over 70% chance to be in the tumor. From our experiments, we have found that the detection results are not sensitive to the threshold. The thresholds between 0.5 and 0.85 give almost the same detection results.

From the experiments, we find that the tumor shape has a strong influence on the performance of the detection results. As a global trend, algorithms show more accurate detection on well-defined (e.g. Fig.4(c)(d)) than on ill-defined masses (e.g. Fig.4(e)(f)). Meanwhile, the tumor size has a weak influence on the detection performance. There is not a specific size where the algorithm performs poorly, i.e. the algorithm is not sensitive to the tumor size, as shown in Fig.4.

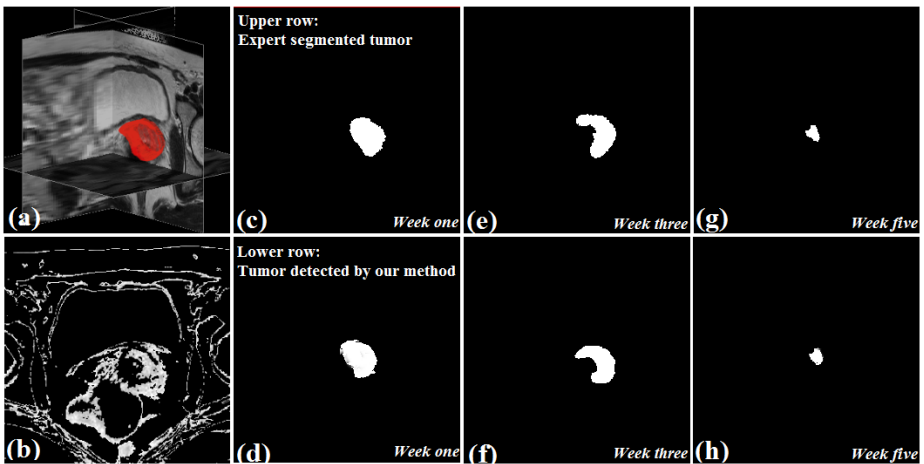


Fig. 4. Detection of cervical tumor for a patient. (a)3D surface of the detected tumor using our method. (b) Detection using the intensity information according to the histogram in Fig.1 (c)-(h): Comparison. Top: Tumor outlined manually by a clinician. Bottom: Detection results using our unified algorithm. Results from the first, third, and fifth week during the treatment are shown here.

The detections of the algorithm are considered true positives if they overlapped with a ground truth (manual detection), and false positives otherwise. Free-response receiver operating characteristic (FROC) curves [13] are produced as validation of the algorithm. For performing the FROC analysis, a connected component algorithm is run to group neighboring voxels as a single detection. The algorithm achieved 90% sensitivity at 6.0 false positives per case. In comparison, the algorithm from [15] gave only 74% sensitivity at the same false positive rate. The corresponding FROC curves are shown in Fig 5. The proposed method shows better specificity and achieves a higher overall maximum sensitivity.

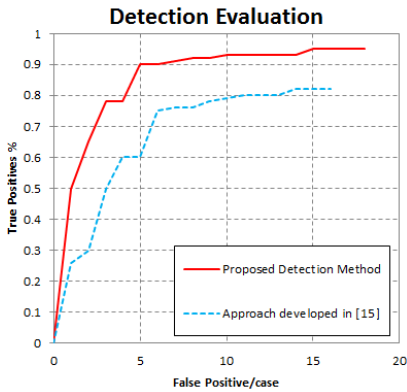


Fig. 5. FROC curves, compared with the approach in [15]

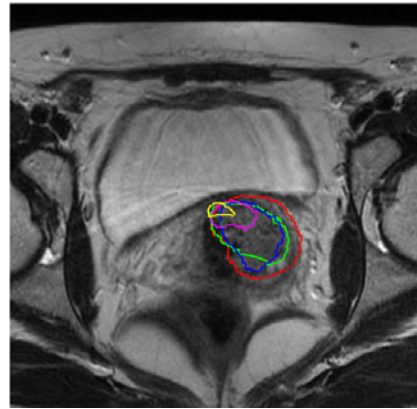


Fig. 6. Mapping the detected tumor contours to planning day MRI

Fig. 6 shows a representative slice from the planning day MRI with overlaid five treatment day tumor detection contours. It can be seen that the detected tumor appears in the same slice and the same location in the registered serial treatment images, and tumor regression process is clearly visible from the contours. Using the proposed technique, we can easily segment the organs of interest, estimate the mapping between day 0 and day d , and calculate the location changes of the tumor for diagnosis and assessment, thus we can precisely guide the interventional devices toward the lesion during image guided therapy.

4 Conclusion

In this paper, a unified framework for simultaneous segmentation, nonrigid registration and tumor detection had been presented. In the segmentation process, the surfaces evolve according to the constraints from deformed contours and image gray level information as well as prior information. The constrained nonrigid registration part matches organs and intensity information respectively while taking the tumor detection into consideration. We define the intensity matching

as a mixture of two distributions which describe statistically image gray-level variations for both pixel classes (i.e. tumor class and normal tissue class). These mixture distributions are weighted by the tumor detection map which assigns to each voxel its probability of abnormality. We also constraint the determinant of the transformation's Jacobian, which guarantees the transformation to be smooth and simulates the tumor regression process. In the future, we plan to develop a system that incorporates a physical tumor regression model.

References

1. Besag, J.: On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)* 48(3), 259–302 (1986)
2. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Transactions on Image Processing* 10(2), 266–277 (2001)
3. Chelikani, S., Purushothaman, K., Knisely, J., Chen, Z., Nath, R., Bansal, R., Duncan, J.S.: A gradient feature weighted minimax algorithm for registration of multiple portal images to 3DCT volumes in prostate radiotherapy. *Int. J. Radiation Oncology Biol. Phys.* 65(2), 535–547 (2006)
4. Greene, W.H., Chelikani, S., Purushothaman, K., Chen, Z., Papademetris, X., Staib, L.H., Duncan, J.S.: Constrained non-rigid registration for use in image-guided adaptive radiotherapy. *Medical Image Analysis* 13(5), 809–817 (2009)
5. Hachama, M., Desolneux, A., Richard, F.: Combining registration and abnormality detection in mammography. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) *WBIR 2006. LNCS*, vol. 4057, pp. 178–185. Springer, Heidelberg (2006)
6. Hamm, B., Forstner, R.: MRI and CT of the Female Pelvis, chap. 3.2.1. *General MR Appearance*, 1st edn., p. 139. Springer, Heidelberg (2007)
7. Jaffray, D.A., Carlone, M., Menard, C., Breen, S.: Image-guided radiation therapy: Emergence of MR-guided radiation treatment (MRgRT) systems. In: *Medical Imaging 2010: Physics of Medical Imaging*, vol. 7622, pp. 1–12 (2010)
8. Joshi, A., Leahy, R., Toga, A., Shattuck, D.: A framework for brain registration via simultaneous surface and volume flow. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) *IPMI 2009. LNCS*, vol. 5636, pp. 576–588. Springer, Heidelberg (2009)
9. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: *2000 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 316–323 (2000)
10. Lu, C., Chelikani, S., Papademetris, X., Staib, L., Duncan, J.: Constrained non-rigid registration using lagrange multipliers for application in prostate radiotherapy. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 133–138 (June 2010)
11. Lu, C., Chelikani, S., Chen, Z., Papademetris, X., Staib, L.H., Duncan, J.S.: Integrated segmentation and nonrigid registration for application in prostate image-guided radiotherapy. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010. LNCS*, vol. 6361, pp. 53–60. Springer, Heidelberg (2010)
12. Nag, S., Chao, C., Martinez, A., Thomadsen, B.: The american brachytherapy society recommendations for low-dose-rate brachytherapy for carcinoma of the cervix. *Int. J. Radiation Oncology Biology Physics* 52(1), 33–48 (2002)
13. Oliver, A., Freixenet, J., Marti, J., Perez, E., Pont, J., Denton, E.R.E., Zwiggelaar, R.: A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis* 14(2), 87–110 (2010)

14. Pohl, K.M., Fisher, J., Levitt, J., Shenton, M., Kikinis, R., Grimson, W., Wells, W.: A unifying approach to registration, segmentation, and intensity correction. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 310–318. Springer, Heidelberg (2005)
15. Richard, F.: A new approach for the registration of images with inconsistent differences. In: 2004 International Conference on Pattern Recognition (ICPR), vol. 4, pp. 649–652 (August 2004)
16. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging* 18(8), 712–721 (1999)
17. Yezzi, A., Zollei, L., Kapur, T.: A variational framework for integrating segmentation and registration through active contours. *Medical Image Analysis* 7(2), 171–185 (2003)
18. Zhang, J., Rangarajan, A.: Bayesian multimodality non-rigid image registration via conditional density estimation. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 499–511. Springer, Heidelberg (2003)

Detection of Crossing White Matter Fibers with High-Order Tensors and Rank- k Decompositions

Fangxiang Jiao*, Yaniv Gur*, Chris R. Johnson, and Sarang Joshi

SCI Institute, University of Utah, Salt Lake City, UT 84112, USA
{fjiao,yanivg,crj,sjoshi}@sci.utah.edu

Abstract. Fundamental to high angular resolution diffusion imaging (HARDI), is the estimation of a positive-semidefinite orientation distribution function (ODF) and extracting the diffusion properties (e.g., fiber directions). In this work we show that these two goals can be achieved efficiently by using homogeneous polynomials to represent the ODF in the spherical deconvolution approach, as was proposed in the Cartesian Tensor-ODF (CT-ODF) formulation. Based on this formulation we first suggest an estimation method for positive-semidefinite ODF by solving a linear programming problem that does not require special parameterization of the ODF. We also propose a rank- k tensor decomposition, known as CP decomposition, to extract the fibers information from the estimated ODF. We show that this decomposition is superior to the fiber direction estimation via ODF maxima detection as it enables one to reach the full fiber separation resolution of the estimation technique. We assess the accuracy of this new framework by applying it to synthetic and experimentally obtained HARDI data.

1 Introduction

The imaging technique known as Diffusion Tensor MRI (DT-MRI) measures the Brownian motion of water molecules in a tissue and enables one to reveal its diffusion properties. It is primarily used to infer the white matter connectivity of the brain. The signal attenuation model in DT-MRI is given by the Stejskal-Tanner equation

$$S(\mathbf{g}) = S_0 \exp(-bD(\mathbf{g})), \quad (1)$$

where $D(\mathbf{g})$ is the apparent diffusion coefficient (ADC) in the direction \mathbf{g} . In traditional DTI the ADC is modeled by a quadratic form $\mathbf{g}^T D \mathbf{g}$, where D is a second-order tensor known as diffusion tensor. Since $D(\mathbf{g})$ is a quadratic form, the modeled ADC is elliptic and thus cannot model complex structures such as crossing fibers. To overcome the limitations of DTI, High Angular Resolution Diffusion Imaging (HARDI) is used. Different modalities and estimation techniques associated with HARDI have been proposed over the years. These methods include the multi-compartment model [24], Q-ball imaging (QBI) [23,9], spherical deconvolution [21,2], Diffusion Orientation Transform (DOT) [16], OPDF [22]

* These authors contributed equally to this work.

and methods that describe the ADC profile using high-order tensors [4,6]. These latter methods represent $D(\mathbf{g})$ as an even-order homogeneous polynomial whose coefficients are identified with the entries of a high-order tensor. The resulting function can describe multiple maxima and can be used to model complex fiber structures. Unfortunately, in contrast to the diffusion tensor model, the maxima of the ADC profile described by a high-order homogeneous polynomial, do not correspond to the underlying fiber directions. This is solved by computing the diffusion propagator and locating the fiber directions at its local maxima. This computation involves a non-trivial Fourier transform step that adds complexity to the estimation process. To solve this problem, it was proposed in [25] to combine the high-order tensor formulation with the spherical deconvolution technique. This strategy enables one to estimate a positive-definite ODF, dubbed Cartesian Tensor-ODF (CT-ODF), whose maxima correspond to the orientations of the fibers.

Although finding all the local maxima's of high order spherical functions is not trivial, only a handful of papers have been devoted to this important issue [7,11,18]. It turns out that since each maximum has a finite width, maxima tend to interfere below a certain fiber separation angle. Therefore, using maxima finding, the maximal fiber separation resolution enabled by the data acquisition technique cannot be reached. An interesting solution to this problem was proposed in [18]. The ODFs in that case were estimated using the Q-Ball imaging technique and then were converted to high-order tensors using a linear transformation. Then, a heuristic rank- k tensor approximation was applied to the tensors to extract the fiber directions beyond the resolution limit determined by the maxima. This method was later used to initialize the ball-and-stick model [19]. Although the proposed method was applied successfully to synthetic and real data, it has some inherent limitations: To calculate the different rank-1 tensors that contribute to the rank- k approximation, rank-1 tensor subtractions were used. It is known that rank-1 tensor subtractions can potentially increase the tensor rank [20] and hence the convergence of the algorithm is not guaranteed. Furthermore, although the initial ODF is non-negative, the residuals obtained by these subtractions do not have this property. In this paper we address these problems and in addition to a new ODF estimation technique, we propose an alternative way to decompose the tensors.

The paper is organized as follows: We first develop the estimation technique for positive-semidefinite ODFs of any order. This estimation method is based on the CT-ODF formulation for high-order tensors proposed by Angelos et. al. [25]. Then, we formulate the estimation problem as a linear programming problem with linear constraints that enforce non-negativity of the ODF. For extracting the individual fiber properties we apply a rank- k tensor decomposition, known as the CP decomposition, to the ODF. In addition to providing the fiber directions, the decomposition also enables us to estimate the fiber fractions. Finally, we demonstrate our proposed technique on synthetic and real HARDI data and show that the proposed algorithm provides accurate results and can reliably resolve two crossing fibers with much higher fidelity than by maxima detection.

We confirm the accuracy of the algorithm on both synthetic, phantom and real HARDI data.

2 Spherical Deconvolution

Following the work by Tournier et al. [21], the Diffusion-Weighted MR signal can be modeled by a spherical convolution of an orientation distribution function (ODF) with an axially symmetric kernel, K :

$$S(\mathbf{g}, b) = S_0 \int_{S^2} F(\mathbf{v})K(\mathbf{g}, \mathbf{v}, b)d\mathbf{v}, \quad (2)$$

where \mathbf{g} is the gradient direction. The function F is associated with the ODF (or fiber-ODF in Tournier's original work), and it is composed of a sum of k delta functions, each is oriented along one fiber direction and weighted according to the corresponding fiber fraction. The kernel K can be chosen in various ways depending on the dataset and the region in the brain (e.g., [5,2]). A very common choice is the single fiber response which is described by the bipolar Watson function

$$K(\mathbf{g} \cdot \mathbf{v}, b) = e^{-c(\mathbf{g}^T \mathbf{v})^2}, \quad (3)$$

where the concentration parameter, c , is a function of the b value and the diffusivity. Given the measured DW-signal and a kernel, which is known a priori, the ODF is computed by performing spherical deconvolution of K from $S(\mathbf{g}, b)$. Technically, this may be solved using least-squares where the solution is given by a simple pseudo-inverse operation [21].

In [25] it was proposed to represent F as a spherical, even-order and positive-definite homogeneous polynomial induced by a high-order tensor. In that work it was suggested to use the single fiber response kernel described in Eq. (3). The concentration parameter was chosen to be large enough to describe a diffusion process which is highly restricted perpendicular to the orientation \mathbf{v} . We use the same ideas here.

2.1 Estimation of Positive-Semidefinite ODF Using Spherical Deconvolution

Any ODF estimated from the data has to be non-negative. In [25] a special parameterization was used to yield a positive-definite ODF. In this section we show that the same goal can be achieved by minimizing an objective function subject to linear constraints that enforce the positivity on the ODF. That is, given measurements in n gradient directions, we aim to solve the following problem:

$$\min_F \frac{1}{2} \sum_{i=1}^n \left\| S(\mathbf{g}_i, b) - S_0 \int_{S^2} F(\mathbf{v})K(\mathbf{g}_i, \mathbf{v}, b)d\mathbf{v} \right\|^2 \quad (4)$$

subject to

$$F(\mathbf{g}_i) \geq 0, \quad \mathbf{g}_1, \dots, \mathbf{g}_n \in S^2.$$

The solution to this problem guarantees positive-semidefiniteness in the discrete sense, that is, in the directions which were used to acquire the signal. We believe that under certain conditions the estimated ODF will be positive-semidefinite in every direction on the sphere. However, since this is an open problem, we leave the complete mathematical study as future work.

We now formulate the problem explicitly. This formulation holds for tensors of any order, however, in this paper we only consider fourth-order tensors that are also *supersymmetric*. Here we will refer to a supersymmetric tensor by using the term *symmetric*. The coefficients of a symmetric fourth-order tensor are invariant under any permutation of the indices. Thus, a symmetric fourth-order tensor has 15 unique coefficients associated with a homogeneous polynomial:

$$F(\mathbf{g}) = \sum_{a=0}^4 \sum_{b=0}^{4-a} c_{ab} g_1^a g_2^b g_3^{4-a-b}, \tag{5}$$

where c_{ab} denote the unique tensor coefficients and g_1, g_2 and g_3 are the components of the gradient direction \mathbf{g} .

Substituting F into the integral (2), we have a sum of integrals, each related to a different monomial:

$$S(\mathbf{g}, b) = \sum_{a=0}^4 \sum_{b=0}^{4-a} c_{ab} \int_{\mathbf{v} \in S^2} v_1^a v_2^b v_3^{4-a-b} K(\mathbf{g}, \mathbf{v}, b) d\mathbf{v}. \tag{6}$$

Solving these integrals analytically is intractable, hence, we approximate each one of them according to the centroid rule for integration of functions on the sphere [3]. Given a sphere triangulation with N faces, for a spherical function, $f(\mathbf{v})$, the centroid rule is given by

$$\int_{S^2} f(\mathbf{v}) d\mathbf{v} \approx \sum_{i=1}^N f(\mathbf{v}_i) A(\Delta_i) \tag{7}$$

where \mathbf{v}_i is the centroid of the i 'th face and $A(\Delta_i)$ is the area of the face. This scheme is very accurate for specific sphere triangulations. Here we choose the third-order icosahedron triangulation which results in 1280 faces (642 nodes). The evaluation of each integral according to this scheme is very fast as the centroids and the areas of the faces are computed only once.

Following these calculations we can define a matrix, C , whose entries correspond to the numerical approximation of (7) for each monomial, in each direction \mathbf{g}_i . The size of this matrix is then $n \times m$ where n is the number of gradient directions and m is the number of unique tensor coefficients.

The linear constraints that impose the positivity on F are defined by using a $n \times m$ matrix A . Each row of A corresponds to a different gradient direction, and each column corresponds to a different monomial. The multiplication $A\mathbf{x}$ results in a n -dimensional vector, each element of it corresponds to $F(\mathbf{g}_i)$, where F is defined by Eq. (5). Thus, we obtain a set of n linear constraints, each constraint is applied to a different gradient direction.

Finally, with respect to the matrices defined above, for each voxel we solve the following linear programming problem:

$$\arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{S} - \mathbf{C} \cdot \mathbf{x}\|^2 \quad \text{subject to} \quad -\mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad (8)$$

where \mathbf{S} is a vector of the n DW measurements, and \mathbf{b} is a n -dimensional vector which defines the boundary of the convex polytope on which we minimize the objective function. Setting the values of \mathbf{b} to be zero results in estimation of a positive-semidefinite ODF.

To solve this problem, the number of gradient directions has to be larger than the number of the tensor coefficients. Since, typically, in HARDI scans $n > 60$, this condition holds as a fourth-order homogeneous polynomial defined by $m = 15$ unique coefficients. This problem may be solved efficiently using MatLab optimization toolbox or through open source packages for convex optimization such as CVX [10]. Given the optimal vector of coefficients, \mathbf{x}^* , the ODF is computed by $F = \mathbf{A}\mathbf{x}^*$. The unique tensor coefficients are then arranged in a fourth-order tensor using the symmetry and the appropriate monomial factors.

Once the ODF has been estimated we proceed to extracting the fiber directions and fractions. As an ODF is associated with a finite-order expansion of spherical harmonics, its maxima has a finite width. Thus, the ODF's maxima interfere and do not correspond to the correct fiber directions below a certain separation angle. In the following section we solve this problem by using a rank- k tensor decomposition known as the CP decomposition. We show that while a rank-1 decomposition corresponds to finding the maxima of F , a decomposition with $k > 1$ corresponds to finding the different components (fibers) that contribute to F which, in turn, significantly increases the ability to separate crossing fibers.

3 High-Order Tensor (HOT) Decompositions

To discuss HOT decompositions we have to first define the notion of a *tensor-rank*. Tensor rank, denoted here as $R = \text{rank}(\mathcal{D})$, is defined as the minimal number of terms such that the following equality holds

$$\mathcal{D} = \sum_{r=1}^R \mathbf{v}_r^1 \otimes \mathbf{v}_r^2 \otimes \cdots \otimes \mathbf{v}_r^n, \quad (9)$$

where \mathbf{v} are first-order tensors (vectors). The *order* of the tensor, n , is defined by the number of its indices and it determines the number of tensor products in Eq. (9). A *cubic* tensor, is a tensor whose different modes have the same size, i.e., $\mathcal{D} \in \mathbb{R}^{d \times d \times \cdots \times d}$. The decompositions that we discuss in this section hold for a general n 'th-order tensor which is not necessarily cubic or symmetric. In our case \mathcal{D} is cubic and symmetric where $n = 4$ and $d = 3$.

Unlike the matrix case ($n = 2$), the rank of a given HOT is not known. In fact, the problem of determining the rank of a given tensor is NP-complete [11].

However, in this work we are interested in low-rank tensor approximation. For a given tensor rank $k < R$, the low-rank approximation is defined by:

$$\mathcal{D} \approx \sum_{r=1}^k \lambda_r (\mathbf{v}_r^1 \otimes \mathbf{v}_r^2 \otimes \cdots \otimes \mathbf{v}_r^n), \tag{10}$$

where $\|\mathbf{v}_r\| = 1$, and for a symmetric tensor, $\mathbf{v}_r^1 = \mathbf{v}_r^2 = \cdots = \mathbf{v}_r^n$. A low-rank tensor approximation is known as rank- k decomposition and it is applied to various branches of science and engineering. It is also known in the mathematical literature as the CANDECOMP\PARAFAC (CP) decomposition [12]. The vectors \mathbf{v}_r represents here the fiber directions, and the fiber weights are simply $w_r = \lambda_r / \sum_{i=1}^k \lambda_i$. The rank of the tensor corresponds here to the number of crossing fibers within a voxel. Since we do not expect to detect reliably more than two crossing fibers using a fourth-order tensor, we restrict ourselves to the $k = 2$ case.

The fiber model is determined in this work according to the ratio between the coefficients λ_r . That is, the weakest fiber term is rejected whenever $\lambda_{\text{strong}} / \lambda_{\text{weak}} > t$, where the threshold was set here to $t = 4$.

To compute the CP decomposition for a given tensor, \mathcal{D} , and a given rank, k , one has to solve the least-squares problem

$$\min_{\tilde{\mathcal{D}}} \|\mathcal{D} - \tilde{\mathcal{D}}\|^2 \tag{11}$$

where $\tilde{\mathcal{D}} = \sum_{r=1}^k \lambda_r (\mathbf{v}_r^1 \otimes \mathbf{v}_r^2 \otimes \cdots \otimes \mathbf{v}_r^n)$. Due to its simplicity and efficiency, the most popular technique to compute a rank- k tensor approximation is the Alternating Least Squares (ALS) [8][13]. The principal of the ALS is straightforward. In each iteration it solves a least-squares problem for the set of vectors $\{\mathbf{v}_r^i\}_{r=1}^k$, $i = m$, while keeping the vectors with $i \neq m$ fixed.

A particular case with $k = 1$ is the *rank-1* decomposition. Given a symmetric tensor \mathcal{D} , its best rank-1 approximation is computed by solving the problem (11) where $\tilde{\mathcal{D}} = \lambda \underbrace{\mathbf{v} \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{n \text{ times}}$. This problem is equivalent to the nonlinear optimization problem [14]

$$\max_{\mathbf{v}} |D(\mathbf{v})| \quad \text{subject to} \quad \|\mathbf{v}\| = 1, \tag{12}$$

where $D(\mathbf{v})$ is the homogeneous polynomial induced by the tensor and identified here with the ODF. The best rank-1 decomposition for symmetric tensors can be efficiently computed by solving the ALS for $k = 1$ or by using a high-order power method (HOPM) (e.g., [14]). This problem may have multiple non-antipodal solutions and the different solutions are found by multiple initializations. Upon converges, for each initialization the algorithm produces an eigenpair $(\mathbf{v}_i, \lambda_i)$. For each eigenpair the unit-norm vector \mathbf{v}_i specifies a global maximum location where $\lambda_i = D(\mathbf{v}_i)$. As in our case $D(\mathbf{v})$ corresponds to the ODF, *as long as the maxima are distinguished*, the resulting vectors will point in the directions of the underlying fibers.

4 Experimental Results

4.1 Synthetic Data Simulations

To assess the accuracy of our new algorithm, we applied it to synthetic, as well as measured experimental HARDI data. First, we generated synthetic data by simulating two crossing fibers according to the multi-compartment model:

$$S(\mathbf{g}, b) = S_0 \sum_{i=1}^{k=2} w_i e^{-b\mathbf{g}D_i\mathbf{g}^T}. \quad (13)$$

For both tensors we assume a prolate tensor model with eigenvalues $\lambda_1 = 1.7 \cdot 10^{-3} \text{mm}^2/\text{s}$, $\lambda_2 = \lambda_3 = 3 \cdot 10^{-4} \text{mm}^2/\text{s}$ and a b -value of $1500 \text{ s}/\text{mm}^2$. The baseline signal was set to $S_0 = 1$. One fiber direction was created randomly and the second one was obtained by rotating the corresponding tensor to get the desired separation angle between the fibers. The weights were set equally. For the convolution kernel we have used Eq. (3) with a concentration parameter of $c = 200$.

The algorithm was tested on noisy data at three levels of SNR¹: 50, 25 and 12.5, where the signal was corrupted by Rician distributed noise. For each noise level, the separation angle was varied from 30 to 90 in 5 degree steps. The signal was measured using 81 gradient directions which were computed using a second-order icosahedron sphere tessellation. For each separation angle and noise level, we performed 100 experiments where fourth-order tensors were estimated using the linear programming approach and a rank-2 CP decomposition was applied to extract the fiber directions and fractions. The mean and the standard deviation of the separation angle, the fiber direction deviation and the weights estimation were calculated for each case.

The CP decompositions were performed using the ALS algorithm [8]. Although the ALS algorithm produces non-symmetric intermediate results for symmetric tensors, we have found that eventually it converges to a symmetric tensor solution. We have implemented a symmetric version of the ALS according to [8]. Although it produces symmetric intermediate solutions, it has not obtained more accurate solutions than the non-symmetric version. For a MatLab implementation of the ALS² it takes approximately 20ms on a Linux workstation with a 2.4MHz quad core CPU and 6GB to produce a rank-2 decomposition for a given tensor.

There are only rare cases where the ALS will not converge to a stationary point. However, it may converge to local minima. While local minima solutions cannot be entirely avoided, we have found that they can be adequately treated by initializing the ALS using the singular eigenvectors of the unfolded tensor [15]. Random initialization gave less accurate results as the algorithm produced local minima solutions more often, especially in low SNR simulations and for small separation angles.

¹ Measured as the baseline signal, S_0 , divided by the noise standard deviation, σ .

² Available at: <http://csmr.ca.sandia.gov/~tgkolda/TensorToolbox/>

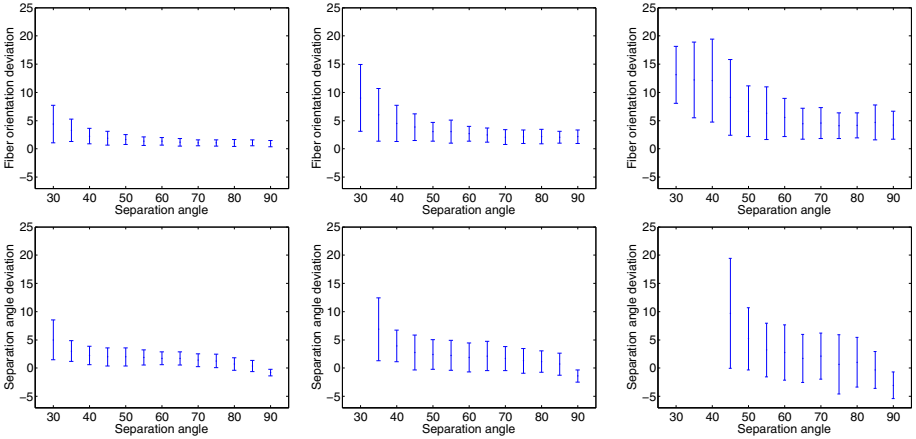


Fig. 1. Deviation from the closest actual fiber direction (top) and separation angle deviation (bottom). The SNR decreases from left to right.

In Fig. 1 we present the deviations of the estimated directions from the closest actual fiber directions, as well as the separation angle deviations. For SNR ratios of 50 and 25 it is shown that the algorithm can reliably resolve the fiber directions, especially above a separation angle of 35 degrees. When the SNR drops down to 12.5, which is a value found in real scans, below an angle of 45 degrees, we observed large biases and high standard deviations at one fiber direction. Hence, the separation angle deviation is not shown for these cases. The performance of the algorithm in this SNR level improves significantly above a separation angle of 50 degrees where both fiber directions could be resolved reliably. As shown in Fig. 2 the fiber fractions could be estimated accurately above 45 degrees for SNR levels of 50 and 25 whereas at the lowest SNR level 60 degrees is the point where the accuracy improves significantly.

In Fig. 3 we show that the rank-2 CP decomposition has an advantage over maxima finding even at large separation angles where the ODF has distinct maxima. While at a fiber separation angle of 80 degrees the rank-2 decomposition has a slight advantage only at low SNR levels (left image), at 70 degrees it outperforms the rank-1 decomposition at all noise levels (right image). Below

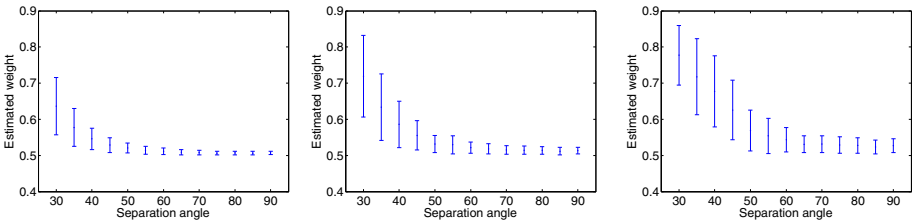


Fig. 2. Estimation results of fiber weight for one fiber using the rank-2 decomposition. The SNR ratio decreases from 50 to 12.5, from left to right.

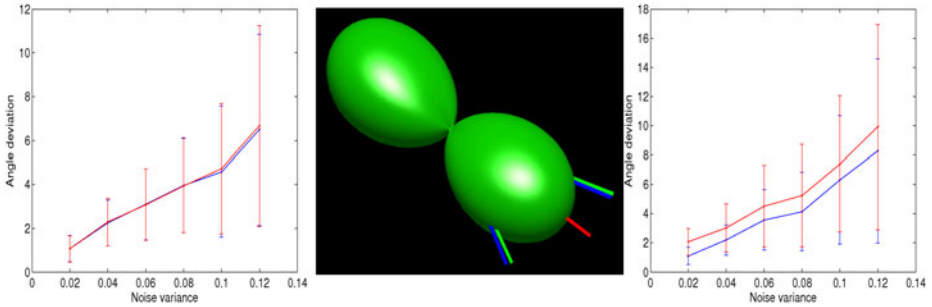


Fig. 3. Rank-2 decomposition vs. maxima finding with a rank-1 decomposition for equally weighted fibers. These results present the deviation from the closest actual fiber direction for a separation angle of 80 degrees (left) and 70 degrees (right). The middle image demonstrates the differences between the approaches when the separation angle is 45 deg. The green and the blue lines show the true and the estimated fiber directions, respectively. The red line shows the direction of the maximum obtained by a rank-1 tensor decomposition. The mean and the standard deviation were calculated from 100 runs at each noise level.

70 degrees the maxima merge and maxima finding is no more reliable. An ODF of crossing fibers at 45 degrees is shown in the middle image. In that case the correct fiber orientations can be estimated by using the rank-2 decomposition only. Maxima locations in this experiment were calculated using the SS-HOPM algorithm [14].

4.2 Phantom Data

To test our algorithm on experimentally obtained HARDI data where the ground truth fibers are known, we first apply our decomposition algorithm to the publicly available phantom data used in the MICCAI 2009 Fiber Cup contest [17]. The data was scanned at three b-values: 650, 1500 and 2650 s/mm². We used the dataset with a b-value of 2650s/mm² in this study. The top row of Fig. 4 shows the reconstructed fourth-order tensor field, the rank-1 decomposition and the rank-2 decomposition results. The two right hand side images in this row illustrate the comparison between the rank-1 and the rank-2 decomposition where the differences between the decompositions are highlighted by ellipses. The fiber directions are presented as thin cylinders at each voxel, where the length of the cylinder is determined by the fiber weight. We have integrated the ground truth fibers as a reference.

As shown, the decomposed directions clearly delineate the hidden fiber orientations. However, by using the rank-2 decomposition, our algorithm could detect more crossing fibers which are oriented along the ground truth fibers.

4.3 Cat Brain Data

To test the algorithm on real data we used a HARDI scan of a cat brain. The data was acquired using a standard 3D diffusion-weighted spin-echo sequence with

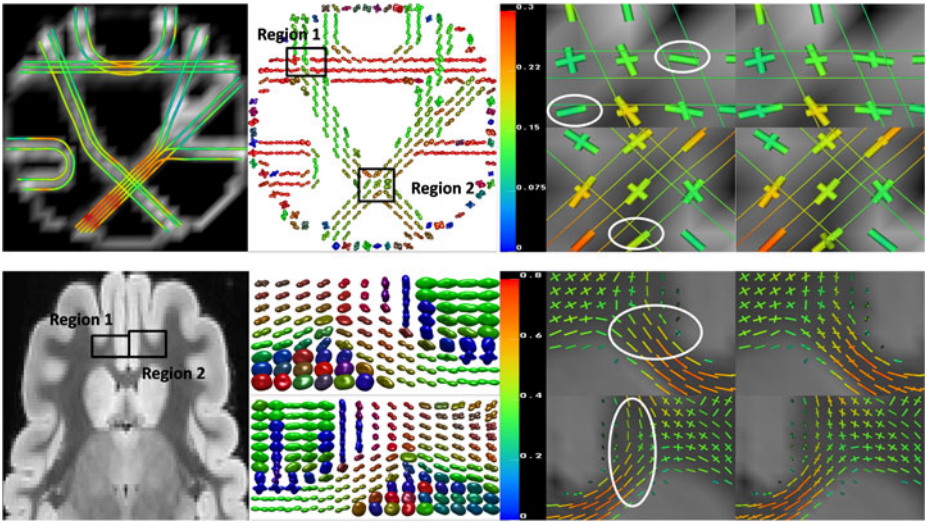


Fig. 4. The decomposition results of the ODF field of the phantom (top row) and cat brain (bottom row) show the following: b0 image (left), fourth-order tensor field of region 1 and region 2 (middle left), rank-1 decomposition (right middle) and rank-2 decomposition (right). The results of region 1 are presented on top of the results of region 2. For the phantom, the ground truth fibers are shown both in the b0 image and in the detailed views of the decomposition results. The decomposition results are color-coded according to the FA where the scale is presented in the image.

TR=500 ms, TE=39.8 ms, field of view $70 \times 40 \times 32$ mm, matrix size $175 \times 100 \times 80$, zero padded to $200 \times 100 \times 80$, yielding an isotropic resolution of 400 microns and a b-value of 6000 s/mm^2 . Four images without diffusion weighting (b0 image) and 96 diffusion weighted images were acquired. The diffusion gradients were uniformly spaced over a unit sphere. The two left images of the bottom row in Fig. 4 show the b0 image and the reconstructed fourth-order tensor field for the two specified ROIs. The two right images in this row compare between maxima finding with the rank-1 decomposition and the rank-2 decomposition. These results show that some of the crossing fibers, which are not detectable using maxima finding, could be detected using the rank-2 decomposition. Although further validation needs to be done on more datasets, promising results are already demonstrated qualitatively on this cat brain data.

5 Conclusions

In this paper we propose a novel framework that combines an ODF estimation method with a parameter extraction technique for estimation of fiber directions and fiber fractions. For the estimation method we have used a specific form of spherical deconvolution where the ODF is represented by a homogeneous polynomial induced by a high-order tensor. The ODF was constrained to be

non-negative by adding a set of linear constraints to the objective function that represents the spherical deconvolution. Then, we show that fiber directions and fiber fractions are accurately extracted by applying a rank-2 CP decomposition to the ODF. As the ODF in this case is associated with a high-order tensor we can apply the decomposition directly to the tensor without using the conversion step which was necessary in [18]. The CP decomposition optimizes the *sum* of the different rank-1 terms and no tensor subtractions are being used. Consequently, the problems of rank increasing and non-positive residuals do not exist here.

Experiments were performed on synthetic data, phantom and real data show that this method can resolve two crossing fibers reliably, even at low SNR, and at far better resolution than the maxima detection approach.

As future work we plan to make the algorithm more efficient, accurate and more robust to noise both at the spherical deconvolution and the tensor decomposition levels. Tensors of order greater than four will be considered as well.

Acknowledgments. The authors would like to thank Tamara G. Kolda of Sandia National Labs, Livermore, California, for useful discussions and comments regarding this work. This research was funded by the NIH grants: 5R01EB007688, 5R01HL092055, and by the NIH/NCRR Center for Integrative Biomedical Computing, P41-RR12553-10, Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST), and DOE SciDAC VACET.

References

1. Aganj, I., Lenglet, C., Sapiro, G.: ODF maxima extraction in spherical harmonic representation via analytical search space reduction. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 84–91. Springer, Heidelberg (2010)
2. Alexander, D.C.: Maximum entropy spherical deconvolution for diffusion MRI. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 76–87. Springer, Heidelberg (2005)
3. Atkinson, K.: Numerical integration on the sphere. *J. Austral. Math. Soc.* 23, 332–347 (1982)
4. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth-order tensor field estimation from DW-MRI. *NeuroImage* 45(1), 153–162 (2009)
5. Barmpoutis, A., Jian, B., Vemuri, B.C.: Adaptive kernels for multi-fiber reconstruction. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 338–349. Springer, Heidelberg (2009)
6. Barmpoutis, A., Vemuri, B.C.: A unified framework for estimating diffusion tensors of any order with symmetric positive-definite constraints. In: ISBI 2010, pp. 1385–1388 (2010)
7. Bloy, L., Verma, R.: On computing the underlying fiber directions from the diffusion orientation distribution function. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 1–8. Springer, Heidelberg (2008)
8. Comon, P., Luciani, X., de Almeida, A.L.F.: Tensor decompositions, alternating least squares and other tales. *J. of Chemometrics* 23, 393–405 (2009)

9. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-ball imaging. *Magn. Res. Med.* 58(3), 497–510 (2007)
10. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 1.21 (October 2010), <http://cvxr.com/cvx>
11. Håstad, J.: Tensor rank is NP-complete. *J. Algorithms* 11, 644–654 (1990)
12. Kiers, H.A.L.: Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* 14, 105–122 (2000)
13. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Review* 51, 455–500 (2009)
14. Kolda, T.G., Mayo, J.R.: Shifted power method for computing tensor eigenpairs (February 2011), <http://arxiv.org/abs/1007.1267>, arXiv:1007.1267v2 [math.NA]
15. De Lathauwer, L., De Moor, B., Vandewalle, J.: On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications* 21, 1324–1342 (2000)
16. Özarslan, E., Shepherd, T.M., Vemuri, B.C., Blackband, S.J., Mareci, T.H.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). *NeuroImage* 31(3), 1086–1103 (2006)
17. Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., Mangin, J.F.: New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging HARDI models. *Magn. Res. Med.* 60(6), 1276–1283 (2008)
18. Schultz, T., Seidel, H.-P.: Estimating crossing fibers: A tensor decomposition approach. *IEEE TVCG* 14(6), 1635–1642 (2008)
19. Schultz, T., Westin, C.-F., Kindlmann, G.: Multi-diffusion-tensor fitting via spherical deconvolution: A unifying framework. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 674–681. Springer, Heidelberg (2010)
20. Stegeman, A., Comon, P.: Subtracting a best rank-1 approximation may increase tensor rank. *Linear Algebra and its Applications* 433, 1276–1300 (2010)
21. 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)
22. Tristan-Vega, A., Westin, C.-F., Aja-Fernandez, S.: Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. *NeuroImage* 47(2), 638–650 (2009)
23. Tuch, D.S.: Q-ball imaging. *Magn. Res. Med.* 52(6), 1358–1372 (2004)
24. 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. *Magn. Res. Med.* 48(4), 577–582 (2002)
25. Weldeslassie, Y.T., Barmpoutis, A., Atkins, M.S.: Symmetric positive-definite cartesian tensor orientation distribution functions (CT-ODF). In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 582–589. Springer, Heidelberg (2010)

Nonnegative Factorization of Diffusion Tensor Images and Its Applications

Yuchen Xie, Jeffrey Ho, and Baba C. Vemuri*

Department of CISE, University of Florida, Gainesville FL, 32611, USA
{yxie, jho, vemuri}@cise.ufl.edu

Abstract. This paper proposes a novel method for computing linear basis images from tensor-valued image data. As a generalization of the nonnegative matrix factorization, the proposed method aims to approximate a collection of diffusion tensor images using nonnegative linear combinations of basis tensor images. An efficient iterative optimization algorithm is proposed to solve this factorization problem. We present two applications: the DTI segmentation problem and a novel approach to discover informative and common parts in a collection of diffusion tensor images. The proposed method has been validated using both synthetic and real data, and experimental results have shown that it offers a competitive alternative to current state-of-the-arts in terms of accuracy and efficiency.

1 Introduction

In this paper, we introduce the novel notion of nonnegative factorization of tensor fields and its application to the segmentation of diffusion tensor images (DTI). Tensor images (or tensor fields) abound in medical imaging applications, with well-known examples such as DTIs, conductivity tensor images (CTI) and elasticity tensors in Elastography. There is an increasing demand for a principled and versatile method that can automatically extract important features from single as well as multiple tensor images, and this paper takes a step forward in this direction by investigating the general factorization problem for tensor fields. Specifically, we formulate the factorization as an optimization problem, and as the main technical contribution, we present an iterative algorithm that can efficiently and reliably optimize the objective function and compute good-quality factorizations.

Matrix factorization in the form of factoring a data matrix \mathbf{V} into a product of the basis matrix \mathbf{W} and the coefficient matrix \mathbf{H} appears frequently in computer vision and image processing. Examples include the principal component analysis (PCA) and the Tomasi-Kanade factorization from multi-view 3D structure recovery. Algebraically, the factorization attempts to “discover” the linear subspace spanned by the columns of \mathbf{W} that can be used to approximate the data given as the columns of \mathbf{V} . Instead of the full subspace, the linear model can be further constrained by requiring that the data vectors belong to (or can be approximated by) the cone generated by the columns of \mathbf{W} . Since

* This research is in part supported by the NIH grant NS066340 to BCV, the NSF grant IIS 0916001 to JH and BCV and the University of Florida Alumni Fellowship to YX.

a point in a cone can be written as a nonnegative linear combination of the cone's generators, this gives rise to the nonnegative matrix factorization (NMF) first introduced in [7]. Specifically, using images as data, the matrices appeared in an NMF are all required to have nonnegative components. While the unconstrained matrix factorization admits a comparatively straightforward solution using linear algebra, nonnegative matrix factorization is generally more difficult to solve as its objective function is non-convex and there exists no known linear algebraic methods that can compute its solution in closed form. Nevertheless, the optimization is not particularly involved as the nonnegative constraint can be easily enforced, and NMF is almost always solved using an alternating sequence of nonnegative linear least squares with variables in \mathbf{W} and \mathbf{H} separately. More importantly, the nonnegativity constraint has been argued in the original paper [7] to be more capable, compared with the unconstrained factorization, of capturing features that are common among the data. In particular, the columns of the basis matrix \mathbf{W} can be considered as the discovered *common parts* among the input data, and the coefficient matrix \mathbf{H} provides the *weights* for reconstructing each input data using these parts. The power and versatility of NMF have been demonstrated to various degrees by a wide range of successful applications that include data clustering [2], parts discovery [9] and MR image analysis [6].

The proposed nonnegative factorization of tensor-valued images is an extension of the nonnegative matrix factorization method above. In our context, components in the matrices \mathbf{V} and \mathbf{W} are represented as symmetric positive semi-definite (PSD) tensors of rank two and \mathbf{H} is the coefficient matrix with nonnegative real components. In particular, the nonnegativity constraint in our more general setting has now acquired two different realizations: the nonnegative constraint on the components of \mathbf{H} as in the original NMF and the generalized nonnegative constraint on the components of \mathbf{W} that they belong to the space of $\mathbf{PSD}(n)$ of $n \times n$ symmetric positive semi-definite matrices. Geometrically, this corresponds to replacing the nonnegative real line (intensity values) with the space $\mathbf{PSD}(n)$ (diffusion tensors), and it is their respective cone structures that permit us to formulate nonnegative factorization using these spaces. While our extension is easy to understand conceptually, the resulting optimization problem in our case is considerably more difficult to solve because of the generalized nonnegative constraint on \mathbf{W} and the dimension of $\mathbf{PSD}(3)$. The former requires generalized matrix inequalities to enforce the constraint [1] and the latter introduces a large number of variables in the objective function. Therefore, a major portion of this paper is devoted to an optimization algorithm that can reliably and efficiently compute the factorization.

Having overcome this computational hurdle, we will next show that the proposed tensor images factorization method can be successfully applied to segmentation problems for single and multiple DTI images. For a single image, the data matrix \mathbf{V} is an $1 \times n$ array of PSD tensors, and a direct clustering on the coefficient matrix \mathbf{H} gives the segmentation. For multiple tensor images, the basis matrix \mathbf{W} gives as before a “part-decomposition” of the collection, and the common parts given as the columns of \mathbf{W} can usually be realized as tensor fields with local support. In the current medical imaging literature, algorithms that segment DTI images can be broadly classified into two categories, the level-set based methods (e.g., [16,18]) and the methods based on combinatorial optimization such as graph cuts [17]. In terms of its underlying motivation and

numerics, our factorization-based method offers a completely different approach to the segmentation problem. We have validated the proposed method using both synthetic and real tensor images. Preliminary segmentation results from single and multiple DTI images have indicated that the proposed factorization-based approach is a viable and competitive alternative method for segmenting tensor images.

2 Preliminaries

Given an $n \times m$ non-negative matrix \mathbf{V} , each of whose column \mathbf{v}_i represents an input image, non-negative matrix factorization (NMF) [7] attempts to factor it into two matrices with non-negative components, $\mathbf{V} \approx \mathbf{W}\mathbf{H}$, where \mathbf{W} denotes the $n \times r$ basis matrix and \mathbf{H} denotes the $r \times m$ coefficient matrix. According to the geometric interpretation elucidated in [3], NMF determines a cone $\Sigma_{\mathbf{W}} = \{\mathbf{x} | \mathbf{x} = \sum_{j=1}^r h_j \mathbf{w}_j, h_j \geq 0\}$ that approximates the input images with the basis elements (or generators of $\Sigma_{\mathbf{W}}$) \mathbf{w}_j given by the j -th column of \mathbf{W} .

For tensor-valued images, there is a $d \times d$ symmetric, positive semi-definite matrix associated to each pixel (or voxel). $d = 3$ for diffusion tensor images. The mathematics for nonnegative factorization of tensor images is straightforward as it only requires a cone structure that is provided by $\mathbf{PSD}(d)$, the space of $d \times d$ symmetric positive semi-definite matrices. A collection of m tensor images of size n can be arranged into a

block matrix $\mathbf{V} = \begin{pmatrix} V_{11} & \cdots & V_{1m} \\ \vdots & \ddots & \vdots \\ V_{n1} & \cdots & V_{nm} \end{pmatrix}$, where $V_{ki} \in \mathbf{PSD}(d)$, $k = 1, \dots, n$ and $i =$

$1, \dots, m$. Each of the m columns represents one tensor image. As before, nonnegative factorization for tensor images attempts to factor \mathbf{V} into a product of the basis matrix \mathbf{W} the coefficient matrix \mathbf{H} whose elements are non-negative real numbers:

$$\mathbf{V} \approx \mathbf{W} * \mathbf{H} = \begin{pmatrix} W_{11} & \cdots & W_{1r} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nr} \end{pmatrix} * \begin{pmatrix} h_{11} & \cdots & h_{1m} \\ \vdots & \ddots & \vdots \\ h_{r1} & \cdots & h_{rm} \end{pmatrix}, \quad (1)$$

where the blocks W_{ij} in \mathbf{W} are matrices in $\mathbf{PSD}(d)$, and the blockwise product $*$ is defined as

$$\mathbf{W} * \mathbf{H} = \begin{pmatrix} \sum_{j=1}^r W_{1j} h_{j1} & \cdots & \sum_{j=1}^r W_{1j} h_{jm} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^r W_{nj} h_{j1} & \cdots & \sum_{j=1}^r W_{nj} h_{jm} \end{pmatrix}. \quad (2)$$

r in the above equation is the number of basis elements (generators) used in the factorization, and it is clear that our nonnegative factorization reduces to the usual NMF when $d = 1$. We remark that the nonnegative factorization can be considered as a generative model for the collection of input tensor images as each input tensor image is approximated by a non-negative linear combination of r columns of \mathbf{W} , each of which can be considered as a tensor image. To determine \mathbf{W} , \mathbf{H} from the data matrix \mathbf{V} , we formulate a constrained optimization problem that minimizes the cost function

$$E(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_{i=1}^m \sum_{k=1}^n \|V_{ki} - \sum_{j=1}^r W_{kj} h_{ji}\|_F^2 \tag{3}$$

with the constraints $W_{kj} \succeq 0$ and $h_{ji} \geq 0$, $i = 1, \dots, m$, $j = 1, \dots, r$ and $k = 1, \dots, n$. \succeq denotes matrix inequality and $\|\cdot\|_F$ denotes the Frobenius norm.

3 Algorithm and Implementation Details

In this section, we present an efficient algorithm that solves the constrained optimization defined above. While the objective function $E(\mathbf{W}, \mathbf{H})$ is not convex, it is convex with respect to the two block variables \mathbf{W} and \mathbf{H} . A common approach to solve this type of constrained optimization problem is the block coordinate descent method [10], and in particular, we can alternatively fix one block variable and improve the other as shown in Algorithm 1. Grippo et al [5] have shown that every limit point of the sequence $\{\mathbf{W}^t, \mathbf{H}^t\}$ generated by Algorithm 1 is a stationary point of the optimization problem defined in (3). Since both sub-problems are convex, our algorithm can easily be shown to be provably convergent. The details for solving the two sub-problems efficiently will be discussed below.

Algorithm 1. Alternating Non-negative Factorization

- Initialize $\mathbf{H}^1 \geq 0$.
 - For $t = 1, 2, \dots$
 - $\mathbf{W}^{t+1} = \underset{\mathbf{W}}{\operatorname{argmin}} E(\mathbf{W}, \mathbf{H}^t)$, s.t. $W_{kj} \succeq 0, \forall k, j$.
 - $\mathbf{H}^{t+1} = \underset{\mathbf{H}}{\operatorname{argmin}} E(\mathbf{W}^{t+1}, \mathbf{H})$, s.t. $\mathbf{H} \geq 0$.
-

3.1 Optimization with Respect to the Basis Matrix \mathbf{W}

When \mathbf{H} is fixed, the optimization problem (3) reduces to a quadratic semi-definite programming problem with a large number of positive semi-definite matrices as the constrained variables. This is a challenging optimization problem without readily available solvers as most available semi-definite programming solvers such as SeDuMi [11] and SDPT3 [14] require the linear objective functions. Toh et al. [13] proposed an inexact primal-dual algorithm to solve a special class of convex quadratic semi-definite programming problem. However, their algorithm only deals with a single positive semi-definite matrix and cannot be directly applied to our optimization problem. Instead, we will exploit the special feature in our problem that $d = 3$ is a relatively small number and design a specific algorithm based on primal-dual path-following interior-point method to solve this subproblem.

The primal problem is given by

$$\min_{\mathbf{W}} \frac{1}{2} \sum_{i=1}^m \sum_{k=1}^n \|V_{ki} - \sum_{j=1}^r W_{kj} h_{ji}\|_F^2 \tag{4}$$

s.t. $W_{kj} \succeq 0, k = 1, \dots, n, j = 1, \dots, r.$

We introduce the $d \times d$ symmetric matrices Z_{kj} associated with the matrix inequalities $W_{kj} \succeq 0$, where $k = 1, \dots, n, j = 1, \dots, r$. The Lagrangian of problem (4) is then

$$L(\mathbf{W}, \mathbf{Z}) = \frac{1}{2} \sum_{i=1}^m \sum_{k=1}^n \|V_{ki} - \sum_{j=1}^r W_{kj} h_{ji}\|_F^2 - \sum_{k=1}^n \sum_{j=1}^r \text{Tr}(Z_{kj} W_{kj}). \tag{5}$$

If we take the derivative of $L(\mathbf{W}, \mathbf{Z})$ with respect to W_{kj} and set it to zero, we have

$$Z_{kj} = - \sum_{i=1}^m (V_{ki} - \sum_{l=1}^r W_{kl} h_{li}) h_{ji}. \tag{6}$$

Substituting this expression back into the Lagrangian gives the dual problem

$$\max_{\mathbf{Z}} \sum_{i=1}^m \sum_{k=1}^n \left(-\frac{1}{2} \|V_{ki} - \sum_{j=1}^r W_{kj} h_{ji}\|_F^2 + \langle V_{ki} - \sum_{j=1}^r W_{kj} h_{ji}, V_{ki} \rangle \right) \tag{7}$$

s.t. $\sum_{i=1}^m \left(V_{ki} h_{ji} - \sum_{l=1}^r W_{kl} h_{ji} h_{li} \right) + Z_{kj} = 0$

$Z_{kj} \succeq 0, k = 1, \dots, n, j = 1, \dots, r.$

For primal-dual interior point method, we use the perturbed Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{aligned} Z_{kj} + \sum_{i=1}^m (V_{ki} - \sum_{l=1}^r W_{kl} h_{li}) h_{ji} &= 0 \\ W_{kj} Z_{kj} &= \nu_{kj} I \\ W_{kj} \succeq 0, Z_{kj} \succeq 0 \\ k &= 1, \dots, n, j = 1, \dots, r \end{aligned} \tag{8}$$

where ν_{kj} are positive parameters. Given the current iterate (\mathbf{W}, \mathbf{Z}) , the search direction $(\Delta \mathbf{W}, \Delta \mathbf{Z})$ at each interior-point iteration is the solution of the following Newton system

$$\sum_{l=1}^r \Delta W_{kl} \left(\sum_{i=1}^m h_{li} h_{ji} \right) - \Delta Z_{kj} = R_{d_{kj}} := Z_{kj} + \sum_{i=1}^m (V_{ki} - \sum_{l=1}^r W_{kl} h_{li}) h_{ji} \tag{9}$$

$$H_{P_{kj}} (\Delta W_{kj} Z_{kj} + W_{kj} \Delta Z_{kj}) = \sigma \mu_{kj} I - H_{P_{kj}} (W_{kj} Z_{kj} + \Delta W_{kj} \Delta Z_{kj}) \tag{10}$$

where $\mu_{kj} = \langle W_{kj}, Z_{kj} \rangle / d$ and $\sigma \in (0, 1)$ is the centering parameter. The symmetrization scheme H_P defined as $H_P(M) = \frac{1}{2}[PMP^{-1} + P^{-T}M^T P^T]$ is required here to generate symmetric ΔW_{kj} and ΔZ_{kj} . Inspired by [12], we choose $P_{kj} = T_{kj}^{-1/2}$, where T_{kj} is the Nesterov-Todd scaling matrix satisfying $T_{kj}Z_{kj}T_{kj} = W_{kj}$. The linearization of equation (10) gives

$$\mathcal{E}_{kj}(\Delta W_{kj}) + \mathcal{F}_{kj}(\Delta Z_{kj}) = R_{c_{kj}} := \sigma \mu_{kj} I - H_{P_{kj}}(W_{kj}Z_{kj}) \tag{11}$$

where \mathcal{E}_{kj} and \mathcal{F}_{kj} are linear operators. By eliminating ΔZ_{kj} in equations (9) and (11), we obtain

$$\sum_{l=1}^r \left(\sum_{i=1}^m h_{li} h_{ji} \right) \Delta W_{kl} + \mathcal{F}_{kj}^{-1} \mathcal{E}_{kj}(\Delta W_{kj}) = R_{d_{kj}} + \mathcal{F}_{kj}^{-1} R_{c_{kj}}. \tag{12}$$

Therefore, we just need to solve n linear systems to get $\Delta \mathbf{W}$. Each linear system includes $\frac{d(d+1)r}{2}$ equations. Because both d and r are small for our problem, these linear systems can be solved efficiently, and $\Delta \mathbf{Z}$ can be computed easily using equation (9). The detail steps of the algorithm is shown in Algorithm 2.

3.2 Optimization with Respect to Coefficient Matrix H

With a fixed \mathbf{W} , (3) becomes a convex optimization problem. Since

$$\frac{\partial E}{\partial h_{ji}} = - \sum_{k=1}^n \text{Tr}(V_{ki} W_{kj}) + \sum_{k=1}^n \sum_{l=1}^r \text{Tr}(W_{kj} W_{kl}) h_{li}, \tag{13}$$

setting $\frac{\partial E}{\partial h_{ji}} = 0$ gives

$$\sum_{l=1}^r \left(\sum_{k=1}^n \langle W_{kj}, W_{kl} \rangle \right) h_{li} = \sum_{k=1}^n \langle V_{ki}, W_{kj} \rangle. \tag{14}$$

Thus we just need to solve a non-negative least squares problem $\mathbf{A}_{r \times r} \mathbf{H}_{r \times m} = \mathbf{B}_{r \times m}$ to get h_{ji} , where $A_{jl} = \sum_{k=1}^n \langle W_{kj}, W_{kl} \rangle$ and $B_{ji} = \sum_{k=1}^n \langle V_{ki}, W_{kj} \rangle$, for all i, j, l . In our implementation, we use the fast active set method proposed by Van Benthem and Keenan [15] to solve this large-scale nonnegative least squares problem.

Once we have obtained the basis matrix \mathbf{W} , we can easily compute the ‘‘projection’’ of a new diffusion tensor image $\mathbf{X} = (X_1, \dots, X_n)^T$ by solving the following optimization problem

$$\min_{\mathbf{y} \geq 0} \sum_{k=1}^n \|X_k - \sum_{j=1}^r W_{kj} y_j\|_F^2 \tag{15}$$

where each X_k is a diffusion tensor and $\mathbf{y} = (y_1, \dots, y_r)^T$ is the nonnegative coefficient vector. This problem, similar to the subproblem with respect to \mathbf{H} above, can also be efficiently solved using the same fast active set method.

3.3 Diffusion Tensor Image Segmentation

Because of its relations to K-means [2] and probabilistic latent semantic analysis [4], nonnegative matrix factorization has been widely used in data clustering, e.g., [19]. In this section, we formulate the diffusion tensor image segmentation problem as a special case of our more general nonnegative factorization problem with spatial constraints.

Algorithm 2. Quadratic semi-definite programming for nonnegative factorization

- 1: **Initialization** $W_{kj} \succeq 0, Z_{kj} \succeq 0, \forall k, j$ and $\tau = 0.9$.
- 2: **Convergence test** Stop the iteration if the accuracy measure ϕ is sufficiently small.

$$\phi = \frac{\sum_{k=1}^n \sum_{j=1}^r \langle W_{kj}, Z_{kj} \rangle}{1 + |\mathbf{pobj}| + |\mathbf{dobj}|}$$

where \mathbf{pobj} and \mathbf{dobj} are values of primal and dual objective functions.

- 3: **Predictor step** Compute the predictor search direction $(\delta\mathbf{W}, \delta\mathbf{Z})$ by choosing $\sigma = 0$.
- 4: **Predictor step-length** Compute $\alpha_p = \min(1, \tau\alpha)$. α is the maximum step length that can be taken so that for $k = 1, \dots, n$ and $j = 1, \dots, r$, $W_{kj} + \alpha\delta W_{kj}$ and $Z_{kj} + \alpha\delta Z_{kj}$ remain positive semidefinite.
- 5: **Centering rule** Set

$$\sigma = \frac{\sum_{k,j} \langle W_{kj} + \alpha_p \delta W_{kj}, Z_{kj} + \alpha_p \delta Z_{kj} \rangle}{\sum_{k,j} \langle W_{kj}, Z_{kj} \rangle}.$$

- 6: **Corrector step** Compute the search direction $(\Delta\mathbf{W}, \Delta\mathbf{Z})$ using

$$R_{c_{kj}} = \sigma \mu_{kj} I - H_{P_{kj}}(W_{kj} Z_{kj} + \delta W_{kj} \delta Z_{kj}).$$

- 7: **Corrector step-length** Compute α_c similar to step 4 with $(\delta\mathbf{W}, \delta\mathbf{Z})$ replaced by $(\Delta\mathbf{W}, \Delta\mathbf{Z})$.
- 8: Update (\mathbf{W}, \mathbf{Z}) to the next iterate $(\mathbf{W}^+, \mathbf{Z}^+)$.

$$W_{kj}^+ = W_{kj} + \alpha_c \Delta W_{kj}, Z_{kj}^+ = Z_{kj} + \alpha_c \Delta Z_{kj}.$$

- 9: Update the step-length parameter by $\tau^+ = 0.9 + 0.08\alpha_c$.
-

Specifically, given a diffusion tensor image of size m , we arrange the m tensors in a row to form the data matrix $\mathbf{V} = (V_1, \dots, V_m)$. The original nonnegative factorization problem is modified as

$$\min_{\mathbf{W}, \mathbf{H}} \frac{1}{2} \sum_{i=1}^m \|V_i - \sum_{j=1}^r W_j h_{ji}\|_F^2 + \frac{\lambda}{2} \sum_{(k,l) \in \Omega} \sum_{j=1}^r (h_{jk} - h_{jl})^2 \tag{16}$$

with nonnegative constraints $W_j \succeq 0$ and $h_{ji} \geq 0$, for all i, j . The first term is simply the objective function in (3) given \mathbf{V} as a row vector. The second term is the spatial smoothness (soft) constraint that requires neighboring pixels to have similar coefficients, and in the equation above, Ω denotes the edge set of the discrete image graph,

and λ the coupling parameter. The optimization problem (16) can be solved using the same alternating method discussed above as the second term is also quadratic in \mathbf{H} . Once the coefficient matrix \mathbf{H} has been determined, we cluster the columns of \mathbf{H} to produce the diffusion tensor image segmentation. In our implementation, we use K-means for this last clustering step.

4 Experiments

In this section, we present two sets of experimental results. The first experiment is on diffusion tensor image segmentation using the segmentation algorithm outlined in the previous section. For the second set of experiments, we work with multiple images, and the results demonstrated that, as for scalar-valued images, meaningful parts can be discovered or detected using nonnegative factorization.

4.1 Diffusion Tensor Image Segmentation

Synthetic Tensor Images. In this experiment, we test the accuracy of the segmentation algorithm using synthetic tensor images with various levels of added noise. We first generate the 32×32 tensor image \mathbf{T} shown in Figure 1 that contains only two tensors: diagonal matrices $diag(0.5, 0.25, 0.25)$ and $diag(0.25, 0.5, 0.25)$, and this defines the ground truth of the segmentation. Different levels of Gaussian noise $N(0, \sigma)$ are added to the tensor image \mathbf{T} to generate noisy tensor images, and we compare the segmentation accuracy of our method with the segmentation algorithm based on clustering the pixels using K-means (on the tensors). In this experiment, the segmentation accuracy is defined by the percentage of correctly segmented pixels, and the comparison across different noise levels is plotted in Figure 1. The result clearly shows the robustness of our method when compared with K-means, especially in the presence of substantial amount of noise.

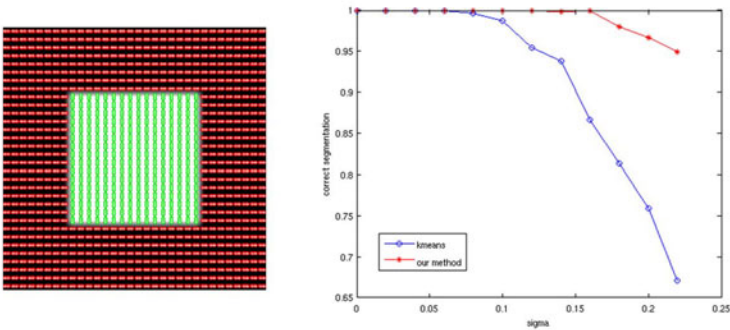


Fig. 1. Left: Input tensor field without noise. Right: Segmentation accuracy vs. different levels of added Gaussian noise $N(0, \sigma)$ with covariance σ .

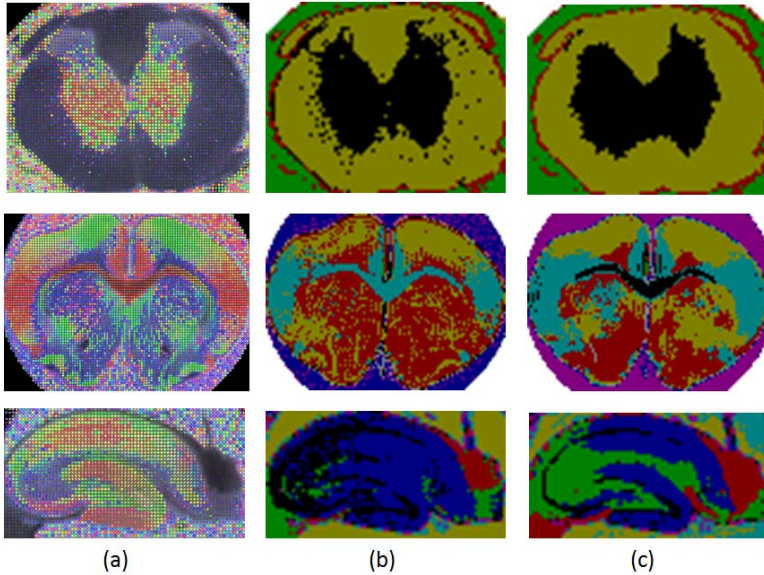


Fig. 2. First Row: Spinal cord of a rat. **Second Row:** Corpus callosum of a rat. **Third Row:** Hippocampus of a rat. Columns (a): Diffusion tensor images. (b): Segmentation results using K-means. (c): Segmentation results using our method. Segments are color-coded.

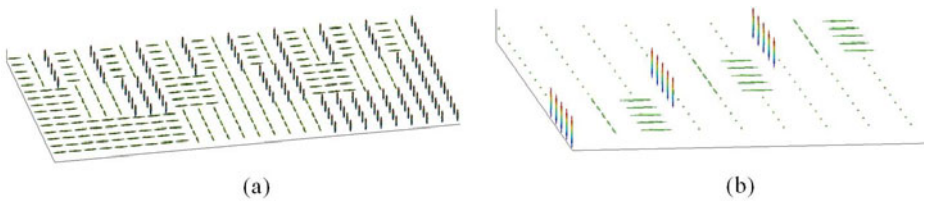


Fig. 3. (a) Visualization of the data matrix \mathbf{V} , each of whose columns represents a 5×3 tensor image. **(b)** Visualization of the basis matrix \mathbf{W} , each of whose columns represents a basis tensor image.

Diffusion Tensor Images. We next present segmentation results on three real diffusion tensor images shown in Figure 2. These are DTI images of the spinal cord, corpus callosum and an isolated hippocampus of a rat. The data were acquired using a PGSE with $TR=1.5s$, $TE=28.3ms$, bandwidth= 35Khz. Twenty-one diffusion weighted images with a b-value of $1250s/mm^2$ were collected. The sizes of the regions of interest for rat spinal cord, corpus callosum and hippocampus are 71×61 , 101×74 and 71×39 , respectively. In this experiment, the number of clusters for these images are four, eight and seven, respectively, and the number of basis elements (columns in \mathbf{W}) is set to five for all three images. Again, we compared our method with the K-means based segmentation and the results consistent demonstrate that our method can produce anatomically more accurate

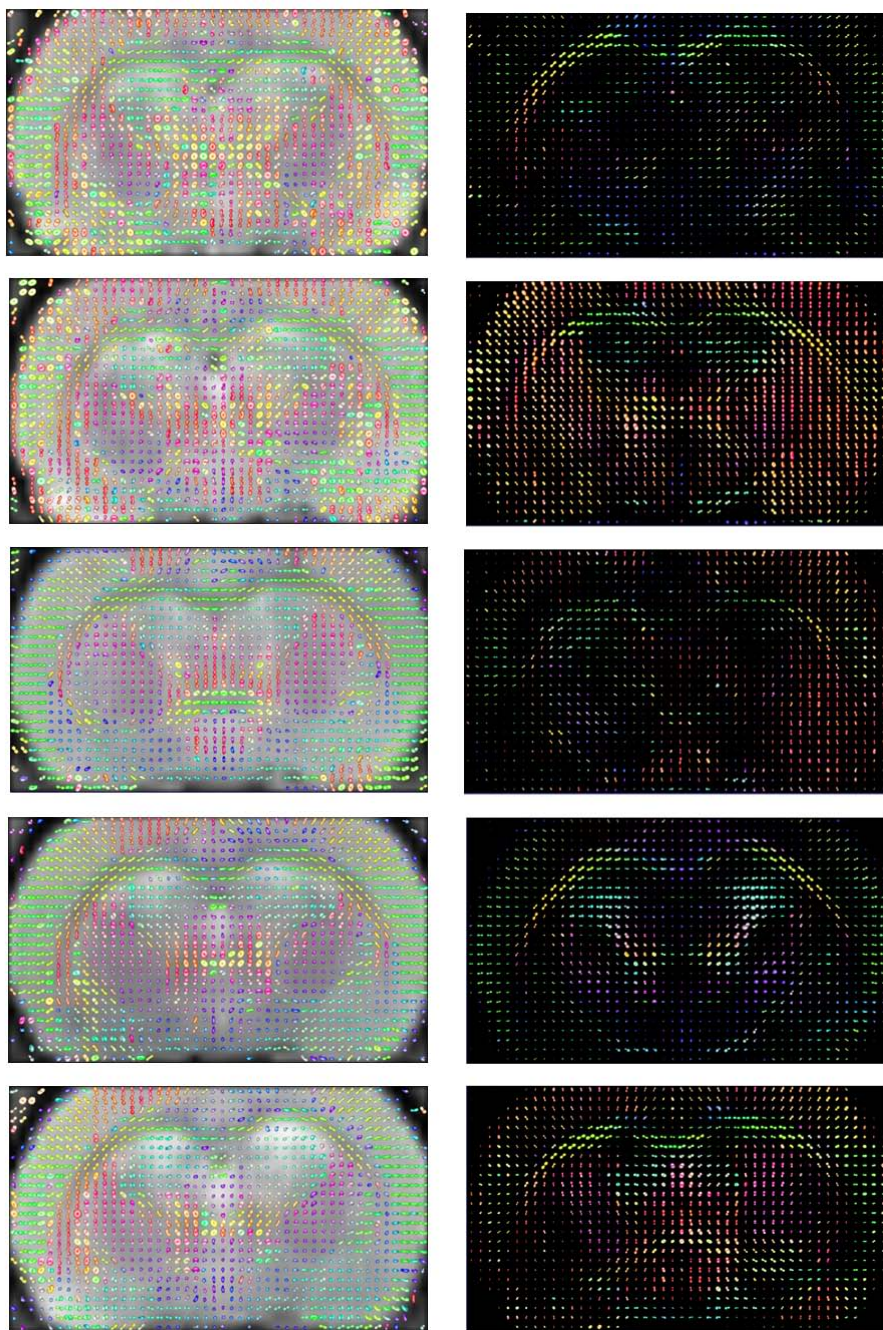


Fig. 4. **Left column:** Five sample slices from 53 input rat brain diffusion tensor images. **Right column:** Sample slices from the five basis tensor images produced by the proposed nonnegative factorization algorithm. We have enhanced the anisotropy for better visualization. The first, second and fourth basis images clearly include white matter, putamen and nucleus accumbens, respectively.

and meaningful segmentation than the straightforward K-means clustering. Finally, we note that both algorithms use K-means for clustering pixels. However, in our method, K-means is applied only to the coefficient vectors in \mathbf{H} , while in the comparison method, K-means is applied directly to the tensors. While the clustering power of the nonnegative factorization for matrices are well-known for scalar-valued data (e.g., [2]), our experimental results provide the first convincing evidence of its clustering power for tensor-valued data as the two sets of experiments have shown that there is no reason to expect that a direct clustering on tensors would produce desired segmentation results. However, direct clustering on coefficient vectors does yield satisfactory results.

4.2 Nonnegative Factorization with Multiple Tensor Fields

For a suitable collection of images, it is well-known that NMF has the ability to automatically determine similar decompositions of the images into their common parts and provide part-based representations for the images in the collection [3]. In this experiment using synthetic data, we show that nonnegative factorization also has similar capability for tensor images. Twenty-seven tensor fields with size 5×3 (15 pixels) are generated and form the columns of the data matrix \mathbf{V} as shown in Figure 3(a). We factor \mathbf{V} into the coefficient matrix \mathbf{H} and the basis matrix \mathbf{W} shown in Figure 3(b). Our factorization algorithm correctly determines the nine basis elements (columns of \mathbf{W}) required to form the data matrix \mathbf{V} , and the coefficient matrix returned by our algorithm is a sparse matrix. Furthermore, the L^2 factorization error is less than 8.57×10^{-10} .

Finally, we present a preliminary result on applying our nonnegative factorization method to automatically discover and segment anatomically important regions from a collection of 53 rat brain diffusion tensor images. In the preprocessing step, all images are aligned using similarity transforms, and in each image, a region of interest of size $46 \times 27 \times 7$ is manually selected. The left column of Figure 4 displays five sample slices from the input 3D diffusion tensor images. We apply the factorization algorithm with $r = 5$ (five basis elements) to this collection of DTIs, and one sample slice from each of the five basis images found by the algorithm is shown on the right column in Figure 4. Important anatomical regions such as white matter, putamen and nucleus accumbens are clearly represented in the five basis images.

5 Conclusions

This paper introduces the novel notion of nonnegative factorization of single and multiple tensor-valued images. The well-known method of nonnegative matrix factorization is extended to tensor-valued data, and an algorithm is proposed to efficiently and reliably solve the new factorization problem formulated as an optimization problem with a non-convex objective function. We have formulated a new approach to DTI segmentation using the proposed nonnegative factorization, and experimental results have shown that our algorithm offers a competitive alternative to currently available methods in terms of its accuracy and efficiency. Perhaps more importantly, our work has demonstrated the usefulness and versatility of the notion of nonnegative factorization, now in the more general setting of tensor-valued images. We believe that this simple yet powerful notion will find its rightful place in the analysis of tensor-valued images. For

example, it could potentially offer a new and effective approach to simultaneously analyze a large number of diffusion tensor images, an essential task in group studies and other applications, that can currently be achieved using only a limited number of available tools, such as the standard statical analysis in Euclidean space or principal geodesic analysis on Riemannian manifolds [18].

References

1. Boyd, S.P., Vandenberghe, L.: *Convex optimization*. Cambridge University Press, Cambridge (2004)
2. Ding, C., He, X., Simon, H.: On the equivalence of nonnegative matrix factorization and spectral clustering. In: *Proc. SIAM Data Mining Conf.* (2005)
3. Donoho, D., Stodden, V.: When does non-negative matrix factorization give a correct decomposition into parts? In: *NIPS* (2004)
4. Gaussier, E., Goutte, C.: Relation between PLSA and NMF and implications. In: *ACM SIGIR* (2005)
5. Grippo, L., Sciandrone, M.: On the convergence of the block nonlinear gauss-seidel method under convex constraints. *Operations Research Letters* 26(3), 127–136 (2000)
6. Joshi, S., Karthikeyan, S., Manjunath, B., Grafton, S., Kiehl, K.: Anatomical parts-based regression using non-negative matrix factorization. In: *CVPR* (2010)
7. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791 (1999)
8. Lenglet, C., Rousson, M., Deriche, R.: DTI segmentation by statistical surface evolution. *IEEE Trans. on Medical Imaging* 25(6), 685–700 (2006)
9. Li, S., Hou, X., Zhang, H., Cheng, Q.: Learning spatially localized, parts-based representation. In: *CVPR* (2001)
10. Nocedal, J., Wright, S.J.: *Numerical optimization*. Springer, Heidelberg (2000)
11. Sturm, J.: Using SeDuMi 1. 02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software* 11(1), 625–653 (1999)
12. Todd, M.J., Toh, K.C., Tutuncu, R.H.: On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization* 8(3), 769–796 (1998)
13. Toh, K.C., Tutuncu, R.H., Todd, M.J.: Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems. *Pac. J. Optim.* 3, 135–164 (2007)
14. Tutuncu, R.H., Toh, K.C., Todd, M.J.: Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming* 95(2), 189–217 (2003)
15. Benthem, M.H.V., Keenan, M.R.: Fast algorithm for the solution of large-scale non-negativity-constrained least squares problems. *Journal of Chemometrics* 18(10), 441–450 (2004)
16. Wang, Z., Vemuri, B.C.: DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Trans. on Medical Imaging* 24(10), 1267–1277 (2005)
17. Weldelessie, Y., Hamarneh, G.: DT-MRI segmentation using graph cuts. In: *Proc. of SPIE Medical Imaging: Image Processing* (2007)
18. Xie, Y., Vemuri, B.C., Ho, J.: Statistical analysis of tensor fields. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 682–689. Springer, Heidelberg (2010)
19. Xu, W., Liu, X., Gong, Y.: Document clustering based on non-negative matrix factorization. In: *ACM SIGIR* (2003)

Multi-subject Dictionary Learning to Segment an Atlas of Brain Spontaneous Activity

Gael Varoquaux^{1,2,3}, Alexandre Gramfort^{2,3}, Fabian Pedregosa^{2,3},
Vincent Michel^{2,3}, and Bertrand Thirion^{2,3}

¹ INSERM U992 Cognitive Neuroimaging unit

² INRIA, Parietal team, Saclay, France

³ LNAO/NeuroSpin, CEA Saclay, Bat. 145, 91191 Gif-sur-Yvette, cedex France

Abstract. Fluctuations in brain on-going activity can be used to reveal its intrinsic functional organization. To mine this information, we give a new hierarchical probabilistic model for brain activity patterns that does not require an experimental design to be specified. We estimate this model in the dictionary learning framework, learning simultaneously latent spatial maps and the corresponding brain activity time-series. Unlike previous dictionary learning frameworks, we introduce an explicit difference between subject-level spatial maps and their corresponding population-level maps, forming an atlas. We give a novel algorithm using convex optimization techniques to solve efficiently this problem with non-smooth penalties well-suited to image denoising. We show on simulated data that it can recover population-level maps as well as subject specificities. On resting-state fMRI data, we extract the first atlas of spontaneous brain activity and show how it defines a subject-specific functional parcellation of the brain in localized regions.

1 Introduction

The study of intrinsic brain functional organization via distant correlations in the fluctuations of brain signals measured by functional Magnetic Resonance Imaging (fMRI) is receiving increasing interest. In particular, the 1000 Functional Connectomes project aims at parceling the brain in functional regions and then at studying the correlation structure of brain function across these nodes [5]. Independent Component Analysis (ICA) is the most popular data-driven approach to analyze spontaneous activity, as it has been shown to extract interpretable spatial patterns [3] that are reproducible across subjects [9]. They form networks of functional regions that are also found in task-driven studies [22].

From a medical point of view, the development of statistically-controlled analysis of brain spontaneous is interesting as it can lead to new diagnostic or prognostic tools applicable on impaired patients. In particular, correlations in the functional signal between predefined regions have been shown to contain markers of post-stroke functional reorganization [24]. However, inferences drawn from these regions depends on the targeted regions, and on their precise delineation. In addition, subject-to-subject local variability, for instance in functional topography, may confound the changes in long-distance interactions.

We address the segmentation of functional regions directly from the fMRI signal. The challenge stems from the lack of salient features in the original signal, as well as the lack of a controlled experimental design to perform model fitting as in task-driven fMRI experiments. In particular, it is difficult to optimize the parameters (dimension and regularization) of the models, hence to obtain an arguably faithful and meaningful representation of this data. ICA tackles these difficulties by estimating a mixing matrix to minimize the mutual information between the resulting spatial components. Departing from ICA, [12] performs segmentation by clustering the time series through a mixture model. However, these approaches lack an explicit noise model and do not take into account the subject-to-subject variability nor the spatial structure of the signal. In this paper, we formulate the problem in the dictionary learning framework and reject observation noise based on the assumption that the relevant patterns are spatially sparse [10,25], and we focus on the choice of the involved parameters. The paper is organized as follows: we give in section 2 a two-level probabilistic model that involves subject-specific spatial maps as well as population-level latent maps, and in section 3 an associated efficient learning algorithm. In section 4 we describe how to set the model parameters from the data. In section 5, we study different learning schemes on synthetic data with simulated inter-individual variability. Finally, in section 6 we apply the method to learning a detailed population-level atlas of regions describing spontaneous activity as recorded in fMRI.

2 Multi-subject Decomposition Model for Brain Activity

Problem statement. We consider a dataset of brain signal time series of length n for S subjects, measured on p voxels: $\{\mathbf{Y}^s \in \mathbb{R}^{n \times p}, s = 1 \dots S\}$. We stipulate that the corresponding 3D images are the observation of k spatial latent factors $\mathbf{V}^s \in \mathbb{R}^{p \times k}$, that characterize functional processes or structured measurement artifacts, and associated time series $\mathbf{U}^s \in \mathbb{R}^{n \times k}$: $\mathbf{Y}^s \approx \mathbf{U}^s \mathbf{V}^{sT}$. We are interested in the study of resting state, or on-going activity, for which no experimental design can be used to model time-courses, thus we propose to learn \mathbf{U}^s and \mathbf{V}^s simultaneously, a problem known as *dictionary learning*, or linear signal decomposition [7,16].

Generative model. In the case of a multi-subject dataset, we give a hierarchical probabilistic model for dictionary learning. Following the standard dictionary learning model, the data observed for each subject is written as the linear combination of subject-specific dictionary elements, that are spatial maps \mathbf{V}_s . For resting-state brain activity, we do not model the loadings \mathbf{U}_s themselves, but their covariance.

$$\forall s \in \{1 \dots S\}, \mathbf{Y}^s = \mathbf{U}^s \mathbf{V}^{sT} + \mathbf{E}^s, \quad \mathbf{E}^s \sim \mathcal{N}(0, \sigma \mathbf{I}), \quad \mathbf{U}^s \sim \mathcal{N}(0, \Sigma_{\mathbf{U}}) \quad (1)$$

In addition, the subject-specific maps \mathbf{V}_s are generated from population-level latent factors, the spatial patterns written as brain maps \mathbf{V} :

$$\forall s \in \{1 \dots S\}, \mathbf{V}^s = \mathbf{V} + \mathbf{F}^s, \quad \mathbf{F}^s \sim \mathcal{N}(0, \zeta \mathbf{I}) \quad (2)$$

Finally, we specify the prior distribution on \mathbf{V} : $\mathcal{P}(\mathbf{V}) \propto \exp(-\xi \Omega(\mathbf{V}))$, where Ω is typically a norm or a quasi-norm.

Relation to existing models. With $\zeta = 0$, the model identifies \mathbf{V}^s with \mathbf{V} : all latent factors are the same across subjects. In this context, if the prior on \mathbf{V} is un-informative, the model boils down to a principal component analysis (PCA) on the concatenated \mathbf{Y}^s . For a Laplace prior, we recover probabilistic formulation of a standard sparse ℓ_1 -penalized PCA [21]. More generally, in this framework, sparsity-inducing priors give rise to a family of probabilistic projection models [1]. Our multi-subject model however differs from generalized canonical correlation analysis [15], and its sparse variants [1], as these approaches do not model subject-specific latent factors and thus do not allow for two levels of variance. Note that, for multi-subject studies, non-hierarchical models based on PCA and ICA impose orthogonality constraints on the loadings at the group level, and thus introduce a unnatural constraint on the \mathbf{U}^s across the different subjects.

ICA can be formulated in a maximum likelihood approach [4] and thus falls in the same general class of non-hierarchical dictionary learning models [16]. However, as ICA disregards explained variance, it leads to improper priors on \mathbf{V} and requires the use of a PCA pre-processing step to estimate the noise [3]. In neuroimaging, multi-subject dictionary learning using a fixed group model ($\zeta = 0$) in combination with ICA is popular, and called *concatenated ICA* [6]. In the experimental section of this paper, we will focus on the use of proper priors on \mathbf{V} based on sparsity-inducing norms Ω , such as the ℓ_1 norm. They are known to be efficient in terms of separating signal from noise, in the supervised settings [26], and lead to tractable optimizations that are convex, though non-smooth.

3 Optimization Strategy for Efficient Learning

We now present a new algorithm to efficiently estimate from the data at hand the model specified by Eq. (1) and (2). In the following, we call this problem Multi-Subject Dictionary Learning (MSDL). In the maximum a posteriori (MAP) estimation framework, we learn the parameters from the data by maximizing the sum of the log-likelihood of the data given the model, and penalization terms that express our hierarchical priors. In addition, as the variance of the group-level residuals in Eq. (2) could be arbitrarily shrunk by shrinking the norm of \mathbf{V} , we impose an upper bound on the norm of the columns of \mathbf{U}^s :

$$(\mathbf{U}^s, \mathbf{V}^s)_{s \in \{1 \dots S\}}, \mathbf{V} = \underset{\mathbf{U}^s, \mathbf{V}^s, \mathbf{V}}{\operatorname{argmin}} \mathcal{E}(\mathbf{U}^s, \mathbf{V}^s, \mathbf{V}), \quad \text{s.t. } \|\mathbf{u}_i^s\|_2^2 \leq 1 \quad (3)$$

$$\text{with } \mathcal{E}(\mathbf{U}^s, \mathbf{V}^s, \mathbf{V}) = \sum_{s=1}^S \frac{1}{2} \left(\|\mathbf{Y}^s - \mathbf{U}^s \mathbf{V}^{sT}\|_{\text{Fro}}^2 + \mu \|\mathbf{V}^s - \mathbf{V}\|_{\text{Fro}}^2 \right) + \lambda \Omega(\mathbf{V}),$$

where $\mu = \frac{\sigma}{\zeta}$ and $\lambda = \frac{\sigma}{\xi}$. The optimization problem given by Eq. (3) is not jointly convex in \mathbf{U}^s , \mathbf{V}^s , and \mathbf{V} , however it is separately convex in \mathbf{V}^s and $(\mathbf{U}^s, \mathbf{V})$. Our

¹ There exist noisy ICA approaches, but they all assume that the contribution of the noise to the observed data is small.

optimization strategy relies on alternating optimizations of \mathbf{V}^s , \mathbf{U}^s , \mathbf{V} , keeping other parameters constant. In the following we give the mathematical analysis of the optimization procedure; the exact operations are detailed in algorithm [1](#).

Following [17](#), we use a block coordinate descent, to minimize \mathcal{E} as a function of \mathbf{U}_s . Solving Eq. [\(3\)](#) as a function of \mathbf{V}^s corresponds to a ridge regression problem on the variable $(\mathbf{V}^s - \mathbf{V})^T$, the solution of which can be computed efficiently (line 9, algorithm [1](#)). Minimizing \mathcal{E} as a function of \mathbf{V} corresponds to minimizing $\sum_{s=1}^S \frac{1}{2} \|\mathbf{v}^s - \mathbf{v}\|_2^2 + \frac{\lambda}{\mu} \Omega(\mathbf{v})$ for all column vectors \mathbf{v} of \mathbf{V} . The solution is a proximal operator [8](#), as detailed in lemma [1](#).

Lemma 1. $\underset{\mathbf{v}}{\operatorname{argmin}} (\sum_{s=1}^S \frac{1}{2} \|\mathbf{v}^s - \mathbf{v}\|_2^2 + \gamma \Omega(\mathbf{v})) = \operatorname{prox}_{\gamma/S \Omega} \bar{\mathbf{v}}$, where $\bar{\mathbf{v}} = \frac{1}{S} \sum_{s=1}^S \mathbf{v}^s$.

The proof of lemma [1](#) follows from the fact that $\sum_{s=1}^S \|\mathbf{v}^s - \mathbf{v}\|_2^2 = S \sum_{s=1}^S \|\bar{\mathbf{v}} - \mathbf{v}\|_2^2 + \sum_{s=1}^S \|\bar{\mathbf{v}} - \mathbf{v}^s\|_2^2$, as the second term at the right hand side is independent from \mathbf{v} , the minimization problem simplifies to minimizing the first term, which corresponds to the problem solved by the proximal operator on $\bar{\mathbf{v}}$.

Algorithm 1. Solving optimization problem given in Eq. [\(3\)](#)

Input: $\{\mathbf{Y}^s \in \mathbb{R}^{n \times p}, s = 1, \dots, S\}$, the time series for each subject; k , the number of maps; an initial guess for \mathbf{V} .

Output: $\mathbf{V} \in \mathbb{R}^{p \times k}$ the group-level spatial maps, $\{\mathbf{V}^s \in \mathbb{R}^{p \times k}\}$ the subject-specific spatial maps, $\{\mathbf{U}^s \in \mathbb{R}^{n \times k}\}$ the associated time series.

```

1:  $E_0 \leftarrow \infty, E_1 \leftarrow \infty, i \leftarrow 1$  (initialize variables).
2:  $\mathbf{V}^s \leftarrow \mathbf{V}, \mathbf{U}_s \leftarrow \mathbf{Y}^s \mathbf{V} (\mathbf{V}^T \mathbf{V})^{-1}$ , for  $s = 1 \dots S$ 
3: while  $E_i - E_{i-1} > \varepsilon E_{i-1}$  do
4:   for  $s=1$  to  $S$  do
5:     for  $l=1$  to  $k$  do
6:       Update  $\mathbf{U}^s$ :  $\mathbf{u}_l^s \leftarrow \mathbf{u}_l^s + \|\mathbf{v}_l^s\|_2^{-2} (\mathbf{Y}^s (\mathbf{v}_l^s - \mathbf{U}^s \mathbf{V}^T \mathbf{v}_l^s))$ 
7:        $\mathbf{u}_l^s \leftarrow \mathbf{u}_l^s / \max(\|\mathbf{u}_l^s\|_2, 1)$ 
8:     end for
9:     Update  $\mathbf{V}^s$  (ridge regression):  $\mathbf{V}^s \leftarrow \mathbf{V} + (\mathbf{Y}^s - \mathbf{U}^s \mathbf{V}^T)^T \mathbf{U}^s (\mathbf{U}^s T \mathbf{U}^s + \mu \mathbf{I})^{-1}$ 
10:   end for
11:   Update  $\mathbf{V}$  using lemma 1:  $\mathbf{V} \leftarrow \operatorname{prox}_{\lambda/S\mu \Omega} (\frac{1}{S} \sum_{s=1}^S \mathbf{V}^s)$ .
12:   Compute value of energy:  $E_i \leftarrow \mathcal{E}(\mathbf{U}^s, \mathbf{V}^s, \mathbf{V})$ 
13:    $i \leftarrow i + 1$ 
14: end while

```

Choice of initialization. The optimization problem given by Eq. [\(3\)](#) is not convex, and thus the output of algorithm [1](#) depends on the initialization. As ICA applied to fMRI data extracts super Gaussian signals and thus can be used for sparsity recovery [25](#), we initialize \mathbf{V} with maps extracted with the fastICA algorithm [14](#), initialized with a random mixing matrix. However, as not all spatial maps estimated by ICA are super-Gaussian, we run ICA with an

increasing model order until it selects k maps with a positive kurtosis². Note that ICA also solves a non convex problem. We find empirically that this initialization procedure strongly reduces the number of iterations required for convergence compared to initialization with the results of a PCA or to random initialization.

Algorithmic complexity. The complexity of one iteration of algorithm 1 can be decomposed in the three contributions, corresponding to the update of each term. First, updating \mathbf{U}^s for each $s \in \{1 \dots S\}$ costs $\mathcal{O}((p+n)k^2 + knp)$. Second, updating \mathbf{V}^s requires $\mathcal{O}(k^3 + knp)$ operations. Last, the computational cost of updating \mathbf{V} is given by the cost of computing the proximal operator for $\Omega, T_\Omega(p)$, times the number of dictionary elements. Thus the total cost of one iteration is in $\mathcal{O}(S(knp + (p+n)k^2 + k^3) + kT_\Omega(p))$. This expression is linear in the total number of samples Sn . This is important in population studies as, unlike previous work, our approach uses the full multi-subject data set to perform simultaneously denoising and latent factor estimation. In addition, there exist closed-form expressions or efficient algorithms for computing the proximal operator for many interesting choices of Ω , in which case algorithm 1 also scales well with p and can be applied to high-resolution brain images. Note that, due to the size of the datasets, we use randomized projection algorithms to perform truncated SVDs required for initialization and model selection (detailed later).

Imposing smooth sparsity. Sparsity is typically induced with an ℓ_1 prior. However using the ℓ_1 norm disregards the fact that the penalized variables are actually spatial maps, hence have an inherent grid structure. In order to promote sparse and spatially coherent maps for \mathbf{V} , and consequently for the \mathbf{V}^s , we propose to use a Smooth-Lasso (SL) penalty [13]. The SL amounts to adding to the Lasso term an ℓ_2 penalty on the gradient, which straightforwardly yields:

$$\Omega_{\text{SL}}(\mathbf{v}) = \|\mathbf{v}\|_1 + \frac{1}{2}\mathbf{v}^T \mathbf{L} \mathbf{v} \quad (4)$$

where \mathbf{L} is the Laplacian operator defined on the 3D grid of voxels. The computation of the proximal operator associated to $\gamma\Omega_{\text{SL}}$ is detailed in Algorithm 2. It can be optimized with a fast first order method called FISTA [2] after noticing that the cost function to optimize can be divided in 2 terms. A convex smooth term with Lipschitz gradient formed by $\frac{1}{2}\|\mathbf{v} - \bar{\mathbf{v}}\|^2 + \frac{\gamma}{2}\mathbf{v}^T \mathbf{L} \mathbf{v}$ and a convex term $\gamma\|\mathbf{v}\|_1$. The Lipschitz constant of the smooth term is given by $1 + \gamma\|\mathbf{L}\|$, where $\|\mathbf{L}\|$ stands for the spectral norm of the Laplacian operator.

4 Model Selection and Choice of Parameters

In this section we detail how the different parameters of the model are set.

Setting the subject-level penalization. For a given choice of number of components k , we set the group-level penalization constant μ from the data,

² There is no guarantee that this procedure converges. We have observed empirically that, on fMRI data, high model-order ICA can extract at least 100 super-Gaussian maps. If the procedure does not converge, we suggest reducing k .

Algorithm 2. Proximal operator for the smooth-lasso with FISTA**Input:** Spatial map \mathbf{v} **Output:** Denoised map \mathbf{v}^*

```

1:  $\mathbf{z} = \mathbf{v}^* = \mathbf{v}, \tau = 1, 0 < \kappa < (1 + \gamma \|\mathbf{L}\|)^{-1}$ 
2: for  $l=1$  to  $k$  do
3:    $\mathbf{v}_o = \mathbf{v}^*$ 
4:    $\mathbf{v}^* = s_{\kappa\gamma}(\mathbf{z} - \kappa(\mathbf{z} - \mathbf{v} + \gamma\mathbf{L}\mathbf{z}))$ 
5:    $\tau_o = \tau$ 
6:    $\tau = \frac{1 + \sqrt{1 + 4\tau^2}}{2}$ 
7:    $\mathbf{z} = \mathbf{v}^* + \frac{\tau_o - 1}{\tau}(\mathbf{v}^* - \mathbf{v}_o)$ 
8: end for

```

s is the element-wise soft-thresholding operator: $s_{\kappa\gamma}(\cdot) = \text{sign}(\cdot) \max(|\cdot| - \kappa\gamma, 0)$ [\[11\]](#)

by computing estimates of the intra-subject variance σ and the inter-subject variance ζ . We first compute a lower bound e on $n\sigma$ using the variance of the residuals of a PCA of order k performed on each subject datasets \mathbf{Y}^s . Indeed a PCA gives the solution to Eq. [\(11\)](#) minimizing σ . Then, we note that, replacing Eq. [\(2\)](#) in Eq. [\(11\)](#), we have $\{\mathbf{Y}^s = \mathbf{U}^s\mathbf{V} + \mathbf{U}^s\mathbf{F}^s + \mathbf{E}^s, s = 1 \dots S\}$. Thus, we can have a lower bound f on the sum of square of $\{\mathbf{U}^s\mathbf{F}^s + \mathbf{E}^s, s = 1 \dots S\}$ by running a PCA on the concatenated $\{\mathbf{Y}^s, s = 1 \dots S\}$. If we consider that \mathbf{F}^s and \mathbf{E}^s are independent, we have $f \approx s(n\sigma + k\zeta)$, thus we set $\lambda = k/n(f/(se) - 1)^{-1}$.

Setting population-level penalization and model order. At the subject level, we can compute the likelihood of new data based on the probabilistic model given in Eq. [\(11\)](#). This model corresponds the probabilistic PCA model [\[23\]](#), and relies on the multivariate Gaussian likelihood:

$$\mathcal{L}(\mathbf{Y}^s | \mathbf{V}^s, \Sigma_{\mathbf{U}}, \sigma) = -\frac{1}{2} \log |\Sigma_{\text{model}}| - \frac{1}{2n} \text{tr}(\mathbf{Y}^s \Sigma_{\text{model}}^{-1} \mathbf{Y}^{sT}) + \text{cste}, \quad (5)$$

$$\text{with } \Sigma_{\text{model}} = \mathbf{V}^s \Sigma_{\mathbf{U}} \mathbf{V}^{sT} + \sigma \mathbf{I} \quad (6)$$

Note that while the matrix $\Sigma_{\mathbf{U}}$ can be very large ($p \times p$), it is of low rank (its rank is given by the rank of \mathbf{V}^s , k). Thus the above likelihood can be computed on large data at a low computational and memory cost using the Woodbury matrix identity and the matrix determinant lemma.

We set the amount of population-level penalization by choosing the parameter λ amongst a grid to maximize the likelihood of left-out data in a 3-fold cross-validation scheme. We apply stratified cross-validation: in each fold, we learn the model on two-thirds of each subject's dataset, and test on the left out third. We choose to split the data for each subject rather than splitting the population by leaving some subjects out, as we are interested in learning maps that give good models for each subject, rather than a model of variability across the population. We also apply cross-validation to select the model order k , i.e. the number of dictionary elements. However, as already noted in [\[20\]](#), setting a high model order may only lead to a saturation of the likelihood of left-out data, and not a decrease. In addition, at high model-order the model can learn patterns that account for subject-to-subject variability, as we will see on simulated data.

5 Simulation Study

Synthetic data generation. We generate latent spatial maps that match the spatial structure of functional networks or artifact present in the fMRI signal. From these maps, we generate observations by mixing them with random time series and adding random spatially-correlated Gaussian noise. To create the latent spatial maps, we use a blob model: each map is made of a few cones-shaped localized patterns, the position and size of which are chosen randomly (see Fig. 1). The number of blobs on each map is given by a binomial distribution corresponding to 3 trials with 50% success rate. Maps are generated one after the other with blob positions chosen randomly, but avoiding overlap between blobs across maps. In addition, we generate subject-specific latent spatial maps by adding an isotropic, Gaussian-distributed, jitter on the blobs position and width.

Empirical results on synthetic data. Synthetic datasets of 12 subjects, with 5 latent spatial maps, 150 time points, and 50×50 spatial grids were generated. In Fig. 1, we represent generated population maps for a smoothness of 2 pixels and a jitter of 3 pixels, as well as the corresponding estimates by different methods: thresholded ICA for sparse recovery [25], ℓ_1 -penalized sparse PCA (SPCA), and MSDL. On this dataset, the three methods find similar maps, but thresholded ICA displays more high-frequency noise. In Fig. 2, we display the subject-specific maps estimated by our method. We can see that the method makes a compromise between fitting the observed noisy subject-level data, and finding the same latent factors across subjects. As a result, the subject-specific maps capture the inter-subject variability, but also some observation noise. To quantify the quality of the recovery, we have generated 30 datasets, with varying amount of between-subject spatial jitter and spatial smoothness of the observation noise. For each of these datasets, we compute the best assignment matching the estimated maps with the ground truth using the Kuhn-Munkres algorithm [19] to maximize cross-correlation. In Fig. 3 we report the average cross-correlation across all maps and synthetic datasets for the population-level maps, but also for the subject-specific maps for which the ground truth is compared to the corresponding estimated subject-specific map in the case of the MSDL approach, and to the estimated population mean for other methods. We find that thresholded ICA is always the worst performer. If there is no spatial jitter across subjects, MSDL and SPCA perform similarly for the recovery of population-level maps, but SPCA outperforms MSDL for the recovery of subject-level maps. This can be explained by the fact that, in this case, the subject-level specificities learned are observation noise: our heuristic over-estimates μ . When there is indeed subject variability, MSDL outperforms only slightly SPCA for the estimation of population maps, but more clearly for the estimation of individual maps.

In Fig. 3, we report the likelihood of left-out data in a 3-fold cross-validation for varying model order, with a synthetic dataset comprising 5 spatial latent factors. We can see that if the subject-specific observation noise is not smooth, and there is no spatial jitter across subjects, the likelihood reaches a plateau

Fig. 1. Population-level maps: \mathbf{V} .

a. Ground truth

b. Maps estimated by thresholded ICA, the ground truth is outlined in black.

c. Maps estimated by sparse PCA

d. Maps estimated by our multi-subject dictionary learning model

e. Raw observations

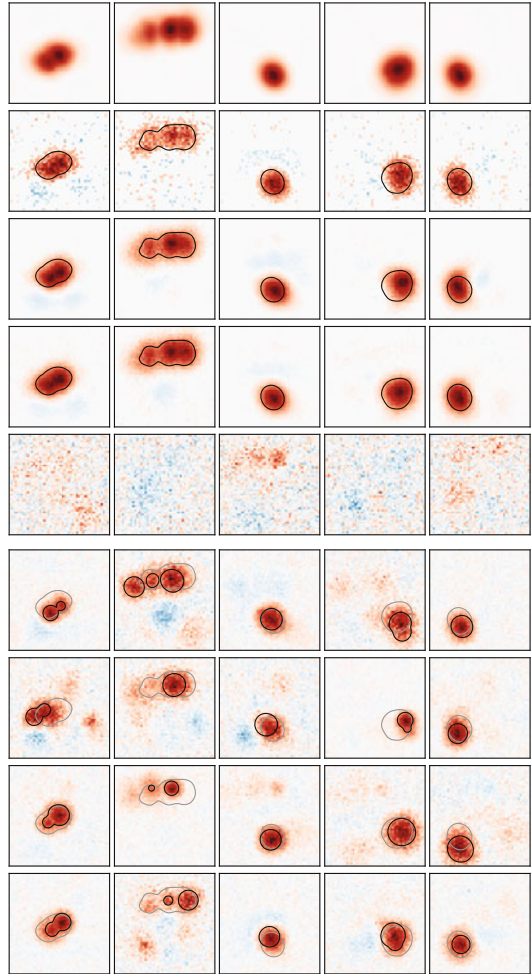


Fig. 2. Subject-specific maps estimated by our multi-subject dictionary learning model: each row corresponds to a different subject. The subject-specific ground truth is outlined in black, while the population-level ground truth is outlined in light gray.

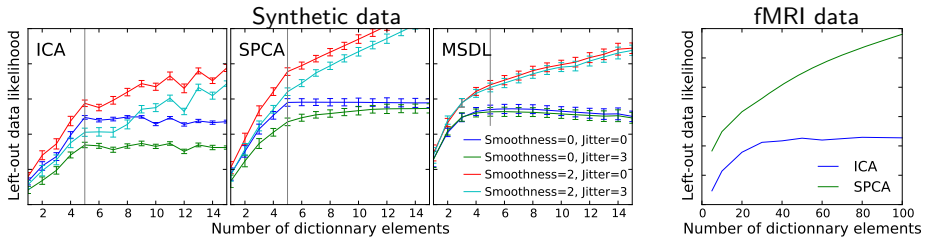
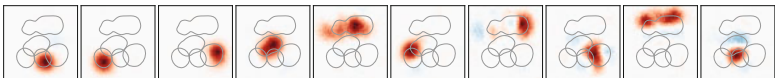


Fig. 3. Likelihood of left out data in a 3-fold stratified cross-validation with the different methods. Left: synthetic data for varying level of observation noise spatial smoothness and of subject-variability jitter; right: resting-state fMRI data.

Fig. 4. High model order.



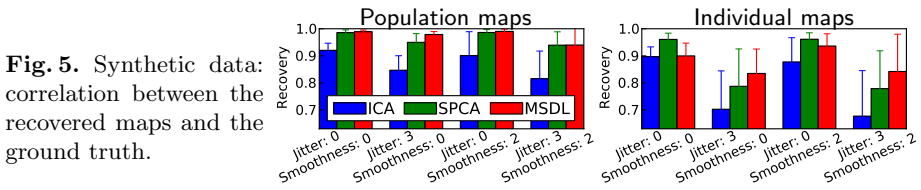


Fig. 5. Synthetic data: correlation between the recovered maps and the ground truth.

for a specified model order corresponding to the ground truth. However, if the observation noise is spatially smooth, and thus violates the i.i.d. hypothesis, the likelihood always increases with model order. In the presence of between-subject spatial jitter, the maps learned are structured and correspond to spatial gradients of the blobs (see Fig. 4), i.e. patterns that fit the between-subject variability.

6 An Atlas of Functional Regions for Spontaneous Activity

We apply our multi-subject dictionary learning method to an fMRI dataset of 20 healthy subjects scanned twice in a resting task, eyes closed. Each session is comprised of 244 brain volumes acquired with a repetition time of 2.4 s. After correction for slice timing differences, motion correction and inter-subject spatial normalization using SPM5, we extract the time series on a mask of the brain, resulting in roughly 500 samples per subject, and 25 000 features – a dataset of 2 Go. We center and variance normalize the time series before dictionary learning.

We have measured by 3-fold cross validation the likelihood of left-out data as a function of model order for ICA and SPCA, but because of lack of time and computing resource not for MSDL (see Fig. 3). We find that the likelihood saturates at a model order of 40 with ICA, thus in the following we choose $k = 40$. In Fig. 7 we display an outline at 33% of the maximum value for all the dictionary elements learned using MSDL. We find that these spatial maps segment the major sources of signal in the Echo-Planar Imaging volumes, that is the gray-matter functional regions that generate the Blood Oxygen-Level Dependent signal (Fig. 6 a and b), as well as sources of artifacts such as the ventricles, the circle of Willis (Fig. 6 c), or the white matter (Fig. 6 d). The estimated maps separate well these salient features from the background as they are mostly sparse. This is in contrast with ICA maps, such as those presented in 9 that are most-often thresholded to a high-value, hiding a noisy background as can be seen in unthresholded maps reported by 25.

For functional studies without known experimental design, such as resting-state, the interesting point is that the gray matter is divided in a set of localized spatial maps that can be seen as an atlas of functional regions present in brain spontaneous fully estimated from the fMRI data. In addition, this division of the brain is consistent with current anatomo-functional knowledge. For instance, in the primary areas, it follows the known topographic maps (see Fig. 7). It is also interesting to note that the parietal lobe, considered as a major functional hub, yields the finest parcellation. As a reference, we also report the corresponding

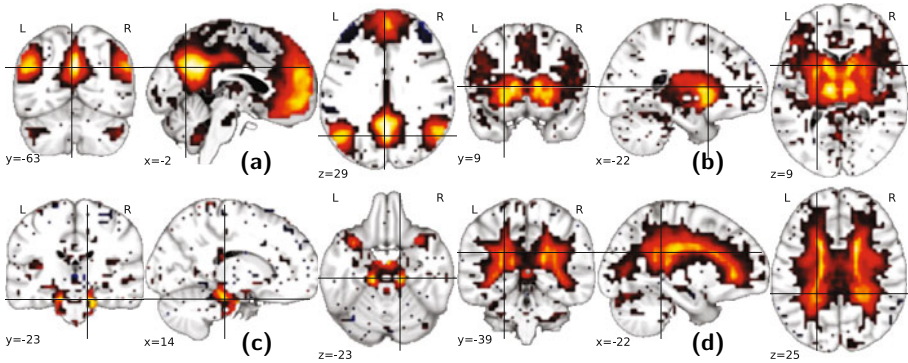


Fig. 6. Population-level maps estimated by MSDL: (a) default mode network, (b) thalamus, pallidum and caudate, (c) vascular system, including the circle of Willis, (d) white matter. The maps are not thresholded but reflect the sparsity pattern.

Fig. 7. Outlines at 33% of all dictionary elements estimated by MSDL for 2 different set of cutting planes. The motor system is divided in (1) dorsal, (2) lateral, and (3) ventral regions. Similarly, the visual system is divided in (4) a primary region centered on the Calcarine sulcus, overlapping with (5) a region centered on the striate cortex, and (6) extrastriate regions. (7), (8): fine details of the vascular system segmented in several maps.

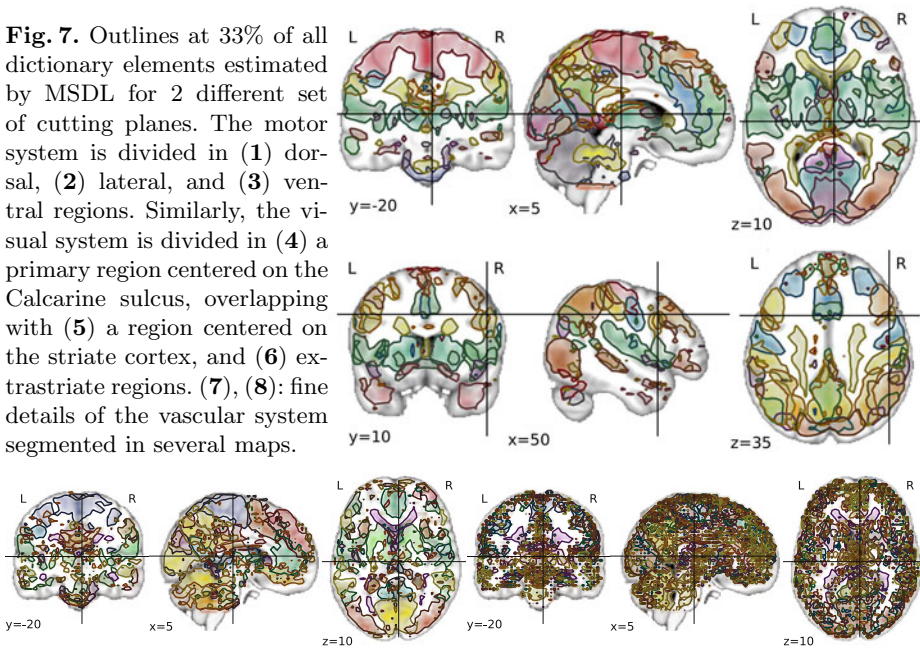
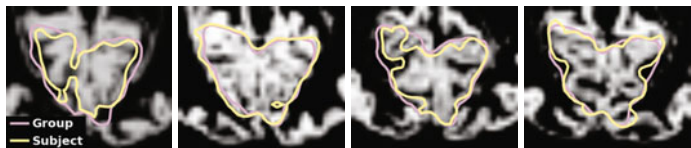


Fig. 8. Outlines at 33% of all ICA (left) and SPCA (right) dictionary elements.

Fig. 9. Dictionary element corresponding to the Calcarine sulcus overlaid on the gray matter map for different subjects.



spatial maps estimated with ICA and SPCA on Fig. 8. We find that the maps estimated by ICA outline similar regions but display more high spatial frequency noise with many different connected components. A few SPCA maps display a scattered, salt-and-pepper like sparsity pattern. This can be explained by the presence of highly spatially-correlated noise: the ℓ_1 penalization tends to choose only one voxel to represent a group of highly-correlated features.

In Fig. 9, we display the outline of the dictionary element corresponding to the Calcarine sulcus for 5 subjects, showing the population-level latent map in addition to the subject-specific map. We can see that the subject-specific map matches better the outline of a gray-matter segmentation –performed using SPM5 on anatomical T1 images– than the population-level map, although at no point anatomical scans were used in the dictionary-learning procedure.

7 Conclusion

The contributions of this work are two-fold. In a statistical learning context, we formulate a new problem, that of learning subject-specific latent factors that match an inter-subject model with explicit variability terms, and we give an efficient algorithm to solve this problem. Importantly, this algorithm separates the problem of learning loadings and subject-level latent factors, from denoising population-level latent factors using proximal operators. There is a rich literature concerned with proximal-based image denoising [8] and the problem we introduce should be further addressed with other proximal operators, such as total variation or structured sparsity [18] for which there exist efficient algorithms.

In a brain imaging context, we address the challenge of segmenting brain regions from spontaneous activity and pave the path for establishing a reference probabilistic functional atlas of this on-going brain activity. Unlike previous methods, our approach controls the amount of signal not fitted by the model at the subject level and at the population level. Using realistic simulations as well as resting-state fMRI data, we have shown that our procedure can *i) at the population level* extract contrasted spatial maps that are consistent with current anatomo-functional knowledge and *ii) at the individual level* adapt these maps to individual spatial configurations. Given the population-level atlas, individual maps for a new subject can be estimated at a small computational cost with the same subject-level model. This procedure thus gives a principled method for defining subject-specific regions to use in a functional-connectivity study, for instance to guide diagnosis or prognosis.

References

1. Archambeau, C., Bach, F.: Sparse probabilistic projections. In: Adv. NIPS (2008)
2. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2(1), 183 (2009)
3. Beckmann, C.F., Smith, S.M.: Probabilistic independent component analysis for functional magnetic resonance imaging. *Trans. Med. Im.* 23, 137 (2004)
4. Bell, A., Sejnowski, T.: An information-maximization approach to blind separation and blind deconvolution. *Neur. Comp.* 7, 1129 (1995)

5. Biswal, B., Mennes, M., Zuo, X., Gohel, S., Kelly, C., Smith, S., Beckmann, C., Adelstein, J., Buckner, R., Colcombe, S., et al.: Toward discovery science of human brain function. *Proc. Natl. Acad. Sci.* 107, 4734 (2010)
6. Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J.: A method for making group inferences from fMRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140 (2001)
7. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. *SIAM Review* 43, 129 (2001)
8. Combettes, P., Pesquet, J.: A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems* 24, 065014 (2008)
9. Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F.: Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci.* 103, 13848 (2006)
10. Daubechies, I., Roussos, E., Takerkart, S., Benharrosh, M., Golden, C., D'Ardenne, K., Richter, W., Cohen, J.D., Haxby, J.: Independent component analysis for brain fMRI does not select for independence. *Proc. Natl. Acad. Sci.* 106, 10415 (2009)
11. Donoho, D.: De-noising by soft-thresholding. *Trans. Inf. Theory* 41, 613 (1995), <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=382009>
12. Golland, P., Golland, Y., Malach, R.: Detection of spatial activation patterns as unsupervised segmentation of fMRI data. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 110–118. Springer, Heidelberg (2007)
13. Hebiri, M., Van De Geer, S.A.: The Smooth-Lasso and other $\ell_1 + \ell_2$ -penalized methods. ArXiv:1003.4885 (2010), <http://arxiv.org/abs/1003.4885v1>
14. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Networks* 13, 411 (2000)
15. Kettnering, J.R.: Canonical analysis of several sets of variables. *Biometrika* 58, 433 (1971)
16. Kreutz-Delgado, K., Murray, J., Rao, B., Engan, K., Lee, T., Sejnowski, T.: Dictionary learning algorithms for sparse representation. *Neur. Comp.* 15, 349 (2003)
17. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research* 11, 19 (2010)
18. Mairal, J., Jenatton, R., Obozinski, G., Bach, F.: Network flow algorithms for structured sparsity. In: *Adv. NIPS* (2010)
19. McHugh, J.: *Algorithmic Graph Theory*. Prentice-Hall, Englewood Cliffs (1990)
20. Minka, T.: Automatic choice of dimensionality for PCA. In: *Adv. NIPS*, p. 598 (2001)
21. Sigg, C., Buhmann, J.: Expectation-maximization for sparse and non-negative PCA. In: *Proc. ICML* (2008)
22. Smith, S., Fox, P., Miller, K., Glahn, D., Fox, P., Mackay, C., Filippini, N., Watkins, K., Toro, R., Laird, A., et al.: Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci.* 106, 13040 (2009)
23. Tipping, M., Bishop, C.: Probabilistic principal component analysis. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 611 (1999)
24. Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., Thirion, B.: Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010. LNCS*, vol. 6361, pp. 200–208. Springer, Heidelberg (2010)
25. Varoquaux, G., Keller, M., Poline, J., Ciuciu, P., Thirion, B.: ICA-based sparse features recovery from fMRI datasets. In: *ISBI*, p. 1177 (2010)
26. Wainwright, M.: Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming. *Trans. Inf. Theory* 55, 2183 (2009)

Activated Fibers: Fiber-Centered Activation Detection in Task-Based fMRI

Jinglei Lv¹, Lei Guo¹, Kaiming Li^{1,2}, Xintao Hu¹, Dajiang Zhu²,
Junwei Han¹, and Tianming Liu²

¹ School of Automation, Northwestern Polytechnical University, Xi'an, China
lvjinglei@gmail.com

² Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA, USA

Abstract. In task-based fMRI, the generalized linear model (GLM) is widely used to detect activated brain regions. A fundamental assumption in the GLM model for fMRI activation detection is that the brain's response, represented by the blood-oxygenation level dependent (BOLD) signals of volumetric voxels, follows the shape of stimulus paradigm. Based on this same assumption, we use the dynamic functional connectivity (DFC) curves between two ends of a white matter fiber, instead of the BOLD signal, to represent the brain's response, and apply the GLM to detect Activated Fibers (AFs). Our rationale is that brain regions connected by white matter fibers tend to be *more* synchronized during stimulus intervals than during baseline intervals. Therefore, the DFC curves for fibers connecting active brain regions should be positively correlated with the stimulus paradigm, which is verified by our extensive experiments using multimodal task-based fMRI and diffusion tensor imaging (DTI) data. Our results demonstrate that the detected AFs connect not only most of the activated brain regions detected via traditional voxel-based GLM method, but also many other brain regions, suggesting that the voxel-based GLM method may be too conservative in detecting activated brain regions.

Keywords: fMRI, DTI, activated fibers.

1 Introduction

Human brain is an inter-linked network that is wired by neuronal axons. In vivo DTI provides non-invasive mapping of the axonal fibers [1] that connect brain regions into structural networks. It is widely believed that human brain function is a network behavior, e.g., brain network oscillation in the resting state [2, 3] or synchronization of nodes within networks during task performance [4]. Therefore, it is natural and intuitive to incorporate structural network information into the study of human brain function. In our view, it is straightforward to consider a white matter fiber inferred from DTI tractography as the finest possible resolution of structural network that is composed of two gray matter voxels on its ends. Then, the level of functional synchronization of the fiber's two ends, measured by the temporal correlation of raw fMRI signals, can be used as the indicator of the fiber's functional state. For instance,

high functional synchronization means engagement in a certain task, while low functional synchronization indicates random behavior [4]. Hence, in this paper, we utilize the temporal functional synchronization of two end voxels of white matter fibers, measured by DFC, to represent the brain's responses to stimulus paradigm and apply the GLM to detect activated fibers.

The major innovations and contributions of this paper are as follows. 1) Instead of using raw fMRI BOLD signals to detect activated brain regions as used in many popular fMRI data analysis software packages such as FSL, SPM, and AFNI, we propose to use the DFC curves (represented by temporal correlation of fMRI BOLD signals at two ends of a fiber) to detect activated fibers and brain regions. In comparison, the raw fMRI signal measures BOLD activity at the voxels level, while DFC measures the *functional connectivity* of structurally connected brain regions. From a neuroscience perspective, the DFC reflects network-level response to paradigm stimuli, while fMRI raw signal reflects local voxel-level response to the stimuli. From a signal processing perspective, DFC is more robust to noises and has less non-stationarity, while raw fMRI BOLD signals has considerable noise and non-stationarity. 2) The framework is intuitive, effective and efficient. Our experimental results using multimodal working memory task-based fMRI and DTI data are promising. All of the computational components in this framework are based on well-established techniques such as functional connectivity measurement and the GLM method. Therefore, the framework can be easily deployed to many applications, once multimodal DTI and fMRI data is available.

2 Method

2.1 Overview of the Method

The proposed computational pipeline is summarized in Figure 1. Firstly, in step 1, we co-registered the task-based fMRI dataset into the DTI space using a linear registration method integrated in the FSL FLIRT (www.fmrib.ox.ac.uk/fsl/). For a given fiber tract, the gray matter (GM) voxels connected to its two ends were identified with the guidance of the brain tissue map derived from DTI dataset [7]. With the co-registered fMRI dataset, the fMRI BOLD time series of the identified GM voxels were attached to the fiber, as shown in Step 3. In general, the functional state of the fiber in response to input stimuli is measured by the functional connectivity between its two ends. Thus, in Step 4, a sliding window approach was used to capture the temporal dynamics of the functional connectivity, resulting in a functional connectivity time series and it is referred to as dynamic functional connectivity (DFC) hereafter. Similar to many existing approaches, the functional connectivity is defined as the temporal correlation between the fMRI BOLD signals. The widely used General Linear Model (GLM) [17] was then used to perform activation detection based on the DFC for all the fibers, as shown in Step 5. Finally, joint analysis (Step 6) was conducted between the activated fiber detection result and the conventional voxel-based activation detection result obtained in Step 2.

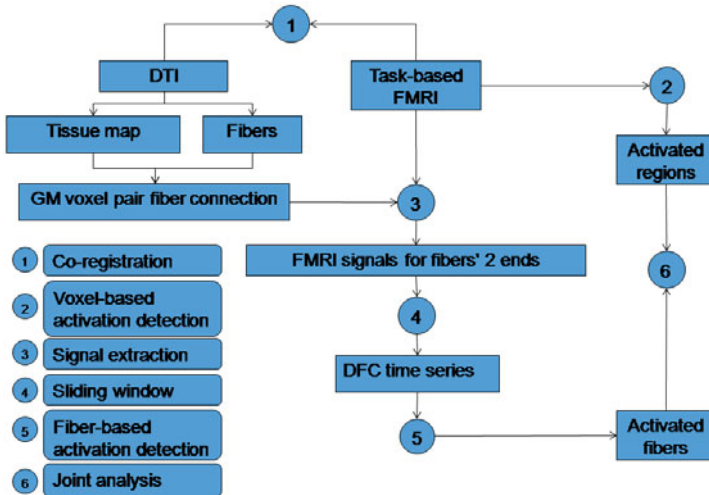


Fig. 1. The flowchart of the proposed computational pipeline for fiber-centered activation detection in task-based fMRI. The six steps are explained in the left bottom corner.

2.2 Data Acquisition and Preprocessing

In the OSPAN working memory task-based fMRI experiment [6], 23 subjects were scanned and fMRI images were acquired on a 3T GE Signa scanner. Briefly, acquisition parameters were as follows: 64×64 matrix, 4mm slice thickness, 220mm FOV, 30 slices, $\text{TR}=1.5\text{s}$, $\text{TE}=25\text{ms}$, $\text{ASSET}=2$. Each participant performed a modified version of the OSPAN task (3 block types: OSPAN, Arithmetic, and Baseline) while fMRI data was acquired. DTI data were acquired with dimensionality $128 \times 128 \times 60$, spatial resolution $2\text{mm} \times 2\text{mm} \times 2\text{mm}$; parameters were $\text{TR} 15.5\text{s}$ and $\text{TE} 89.5\text{ms}$, with 30 DWI gradient directions and 3 B0 volumes acquired.

The fMRI data was co-registered to the DTI space using a linear transformation via FSL FLIRT. For fMRI images, the preprocessing pipelines included motion correction, spatial smoothing, temporal prewhitening, slice time correction, global drift removal [5-8]. For DTI data, preprocessing included skull removal, motion correction and eddy current correction. Brain tissue segmentation was conducted on DTI data using a similar method in [9]. Since DTI and fMRI sequences are both echo planar imaging (EPI) sequences, the misalignment between DTI and fMRI images is much less than that between T1 and fMRI images [7]. The cortical surface was reconstructed from the brain tissue maps using in-house software [10]. The fiber tracking was performed using the MEDINRIA package [11].

2.3 Fiber Projection

It was reported in the literature that blood supply to white matter is significantly lower than that of the cortex (less than one fourth) [13]. Hence, the fMRI signals in white matter (WM) are less meaningful and our analysis will only focus on the investigation of gray matter (GM) connected by fiber tracts.

Unfortunately, the ends of fiber tracts are not necessarily located on the cortex. The reasons are at least two folds. First, the FA (fractional anisotropy) values around the boundaries of gray matter and white matter are relatively low and the tractography procedure might stop before reaching the cortex. As a result, the tracked fibers will not be within the cortical surface. Second, there is discrepancy in the brain tissue segmentation based on DTI data and the DTI tractography. In this case, the fiber could be either outside the cortex if the gray matter is over-segmented or inside the cortex if the gray matter is under-segmented, as reported in the literature [12].

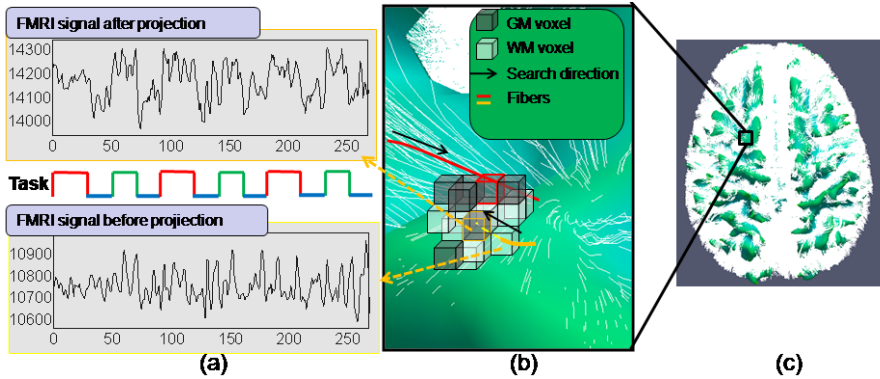


Fig. 2. Illustration of white matter fiber projection. The joint visualization of whiter matter fiber tracts (white curves) and the GM/WM cortical surface (green mesh) is shown on the right (c). The zoomed-in view of the black rectangle is shown in the middle (b). The fMRI BOLD signal of one end of a fiber before and after fiber projection is shown on the left (a). The stimulus paradigm curve is also shown in the middle of (a).

To project the fibers onto the GM cortex, we use the brain tissue map (WM, GM and CSF) derived from DTI dataset as guidance for accurate projection, as illustrated in Figure 2. In specific, denotes an end point of a fiber as $N_e(x_e, y_e, z_e)$ and the point adjacent to N_e as $N_{e-1}(x_{e-1}, y_{e-1}, z_{e-1})$. We define the forward direction as $\vec{d}_f = N_e - N_{e-1}$ and the backward direction \vec{d}_b as $-\vec{d}_f$. If N_e corresponds to a non-brain (outside the brain) or CSF voxel in the brain tissue map (e.g., the red fiber tract shown in Figure 2), we search along backward direction \vec{d}_b until reaching the gray matter. If N_e corresponds to a WM voxel in the brain tissue map (e.g., the orange fiber tract shown in Figure 2), we search along forward direction \vec{d}_f until reaching the GM. If N_e already corresponds to a GM voxel in the brain tissue map, no search is conducted. The search is conducted iteratively until at least one GM voxel can be found in the 1-ring neighborhood of the current seed point, or the number of iteration exceeds a given threshold. When multiple GM voxels exist, the closest one is used as the projected point. In very rare cases when the corresponding GM voxel cannot be found for the end point of a fiber, this fiber is considered as an outlier and discarded. After the fiber projection, two GM voxels were identified for a fiber at its two ends and the corresponding fMRI BOLD signals were attached to that fiber. A comparison for

extracted fMRI BOLD signals before and after projection is shown in Figure 2a. It is evident that the extracted fMRI signal from GM voxel after fiber projection has a similar shape as the input stimulus curve, while the fMRI signal extracted before fiber projection does not. This example demonstrates that projection of fibers to GM voxels is critical to extraction of meaningful fMRI BOLD signals.

2.4 Dynamic Functional Connectivity of Fiber-Connected Brain Regions

White matter axonal fibers, serving as structural substrates of functional connectivity among brain regions, have abundant functional features [3, 14]. In this paper, we define the dynamic functional connectivity (DFC) between a fiber’s two ends to monitor the state of the fiber under stimulus. The functional connectivity is quantified as the temporal Pearson correlation between the two fMRI BOLD signals attached to the fiber. A sliding window approach is adopted to capture the functional connectivity in the temporal domain, as illustrated in Figure 3.

Specifically, given a fMRI dataset with the total length of t (unit in volume) and the sliding window with the width of t_w , at each time point t_n , we extract a signal window with length t_w , centered at t_n to calculate the functional connectivity. Note that for the first and last few time points in the fMRI BOLD signal, the window length was shortened to calculate DFCs. After sliding over all the time points, the resulted DFCs have the same length as the original fMRI signals. Figure 3 provides illustrations of the DFC calculation procedure. The same DFC calculation procedure is performed for all of the fibers in the whole brain obtained from DTI tractography, and thus we convert voxel-based fMRI signals to fiber-centered DFC signals.

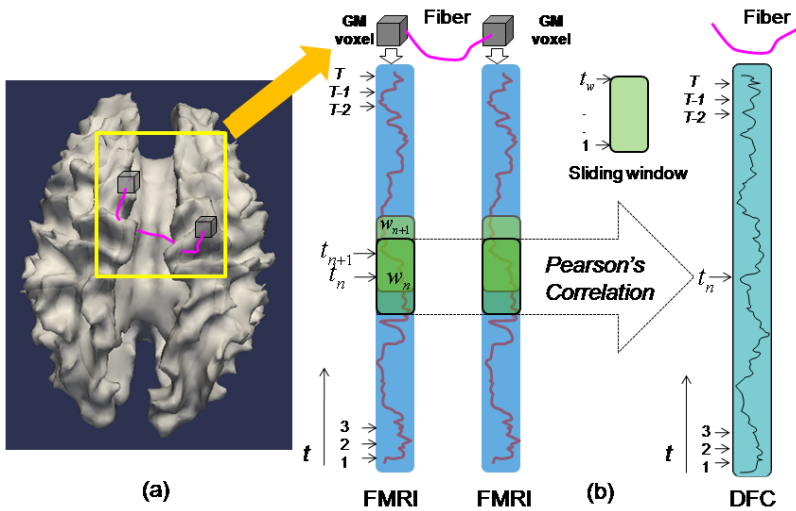


Fig. 3. Illustration of DFC calculation. (a). Illustration of extracting GM voxels (gray boxes) at the two ends of a fiber, represented by the purple curve. (b) A sliding window approach is applied on two raw fMRI signals and the DFC is calculated.

2.5 Activated Fibers Detection

The GLM model [17] has been widely used for task-based fMRI activation detection, which will be adopted here as well. In our fiber-based activation detection, we treat the DFC time series of fibers as inputs to the model for activation detection (Fig. 3), instead of the raw fMRI signals, and use FSL FEAT (<http://www.fmrib.ox.ac.uk/fsl/>) to generate the activation map of fibers for the task-based fMRI data. Fig. 4 provides an illustration of the computational pipeline for fiber-centered activation detection. In specific, we treated each fiber as a voxel and mapped all of the white matter fibers (around 30,000-50,000 in our dataset) to a virtual volume (Fig. 4). Then, the DFCs of all the fibers are stacked into a 4D volumetric images. By such conversion, FSL FEAT can be directly used for activation detection. After the activation detection, fibers whose corresponding virtual voxels were detected as activated are considered as activated (right bottom figure in Fig. 4).

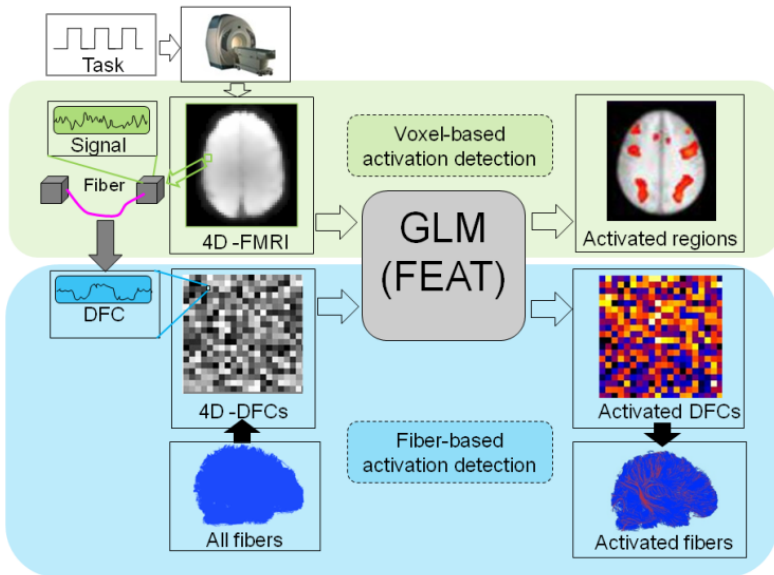


Fig. 4. Overview of the computational pipeline for fiber-centered activation detection. The voxel-based activation detection is shown in the top panel and the fiber-based activation detection is shown in the bottom panel.

In addition, the z-statistic value for each virtual voxel is mapped to the corresponding fiber as metric of its activation level. Different thresholds can then be used to achieve various activation maps. It should be noted that we performed GLM on the DFC of each fiber independently, as similar in the voxel-based GLM method in FSL FEAT. The statistical correlation procedures for false positive control such as the Gaussian random field theory [15] or false discovery rate method [16] have not been implemented in this paper yet, in that statistical modeling of correlation between fibers is still an open problem and will be left to our future work.

3 Experimental Results

In this section, the GLM was applied to detect activated voxels based on the task-based fMRI data and to detect activated fibers based on the framework described in sections 2.3-2.5. The width of the sliding window (t_w) used in DFC calculation is set as 15 times of TR in the following experiments.

3.1 Comparison of fMRI Signal and DFC

As a comparison study, we first calculated the Pearson correlation coefficients between the fMRI BOLD signals and the stimulus curve (convolved with hemodynamic response function), as well as the Pearson correlation coefficients between DFC curves and the stimulus. The distributions of the two sets of correlation coefficients over the entire brain are shown in Figure 5 as an example. It is noted that both of the statistics were conducted over the projected fMRI signals of all fiber ends. It is evident in Fig. 5 that the DFC curves have much higher correlations with the stimulus curve, as indicated by the higher accumulations in histogram bin with high correlations.

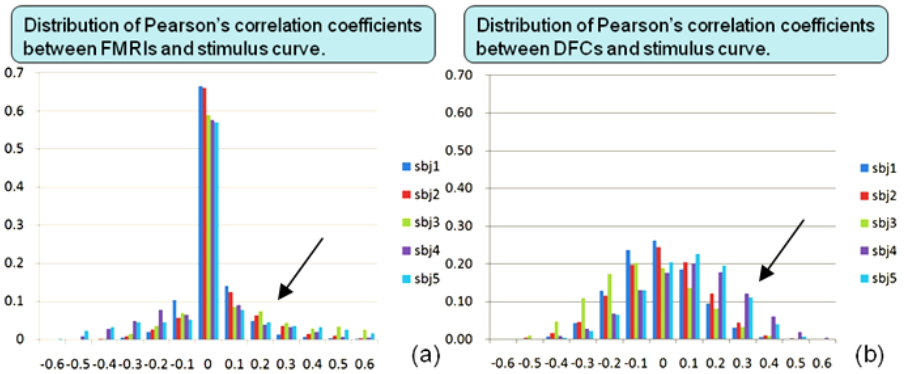


Fig. 5. Comparison of distributions of Pearson’s correlation coefficients. The horizontal axis represents the histogram bins and vertical axis is the Pearson correlation. (a) The distribution of the Pearson correlations between raw fMRI signals and the stimulus curve. (b) The distribution of the Pearson correlations between DFCs and the stimulus curve. Five subjects represented by different colors are shown in the figure.

Figure 6 visualizes one example, in which two GM voxels were activated in the voxel-based activation detection. The fMRI BOLD signals of the two GM voxels are shown in Figure 6a. The DFC of the activated fiber connecting the two GM voxels is shown in Figure 6b. In comparison, the fMRI signals show high level of noises, while the DFC shows much higher signal-to-noise ratio. In addition, the DFC curve shows a high level of stationarity during stimulus intervals (starting with the red and green circles in Figure 6b). Meanwhile, during resting time intervals (baseline), the DFC curve shows a significant valley (blue circles), which is much more stable and robust than the raw fMRI signals in Figure 6a. In short, the DFC curves are much more

synchronized with the stimulus curve than the raw fMRI signals. The results in Figure 5 and 6 support our previous premise that DFCs reflect brain's network-level responses to stimulus and provide significantly better contrasts between brain states during stimulus and baseline intervals in task-based fMRI.

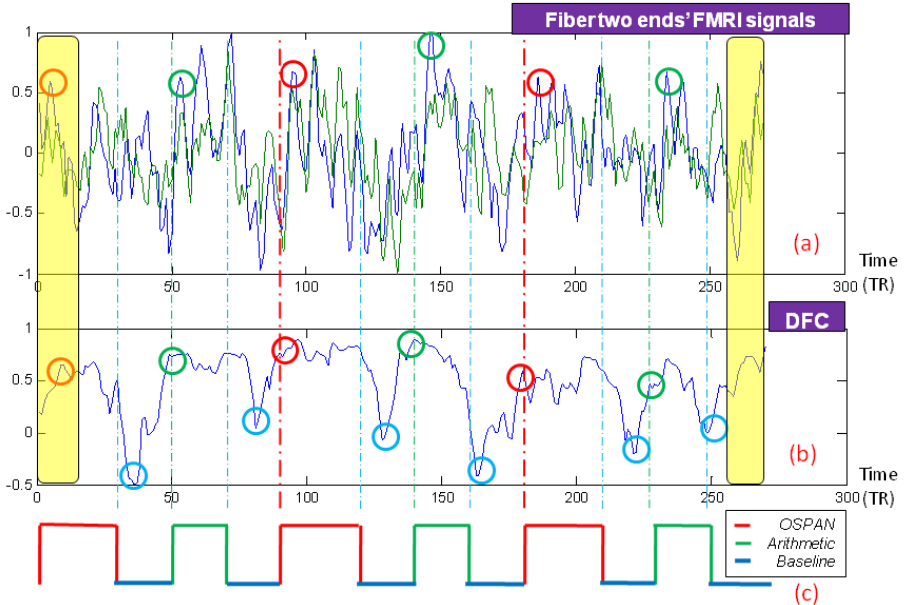


Fig. 6. Comparison of fMRI signals and DFC curve. (a) BOLD signals of two activated GM voxels that are connected by a fiber (one in blue, the other in green). (b) DFC curve of the fiber connecting the two GM voxels. (c) The time line of the input stimuli (OSPAN, arithmetic and resting). Circles in different colors correspond to different stimuli. Note that, as mentioned in section 2.3, the length of the sliding window for the start and end time points is relatively shorter, so the DFC curves in the yellow regions are not considered.

3.2 Global View of DFCs

In this subsection, we examine the global pattern of DFCs of all the fibers in the brain. Figure 7 presents a global view of DFCs for one subject (with around 38,000 fibers). The DFC curves are color-coded and stacked into a matrix, as shown in Figure 7a. We can see that significant portion of the fibers have similar temporal patterns in responses to the stimulus curve (Figure 7b). Meanwhile, the averaged DFC over all of the fibers in Figure 7b shows that the brain's overall response is in synchronization with the stimulus curves to some extent. Black pluses in Figure 7b indicate that when the stimulus changes, and the global state of the brain represented by the average DFCs also change sharply in response to the input stimulus curve. These results are reproducible in all of the cases we studied, and Figure 8 shows another example. The above results suggest that white matter fibers are much more synchronized during stimulus intervals than during baseline intervals, and DFCs are meaningful representations of the brain's responses to stimuli.

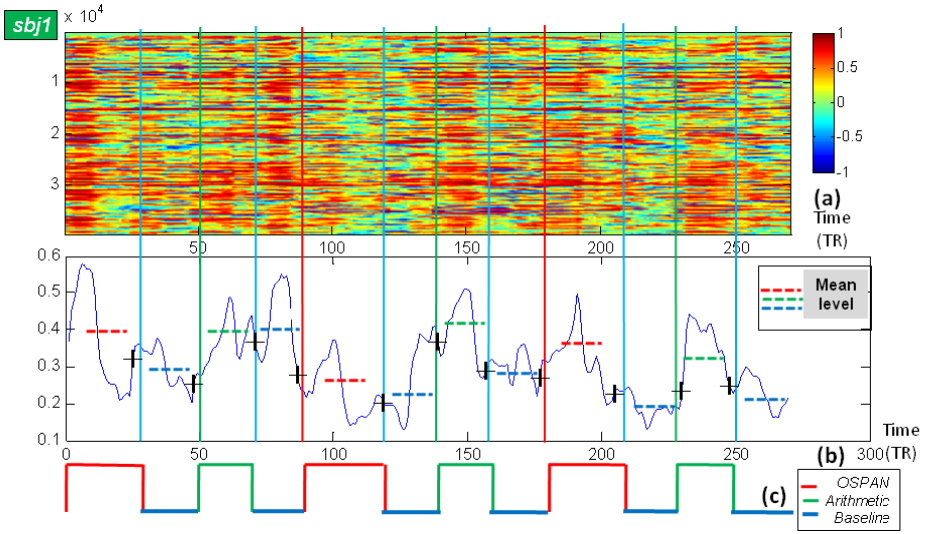


Fig. 7. Global view of DFCs of all the fiber tracts in the entire brain for one subject. (a) The color coded DFCs of all fibers. Color bar is on the right. Totally, about 38,000 fibers are stacked into each column. The vertical axis represents the index of fibers, and the horizontal axis is the timeline. (b) Average DFC over all fiber tracts. (c) Input stimulus. Three types of blocks are color-coded.

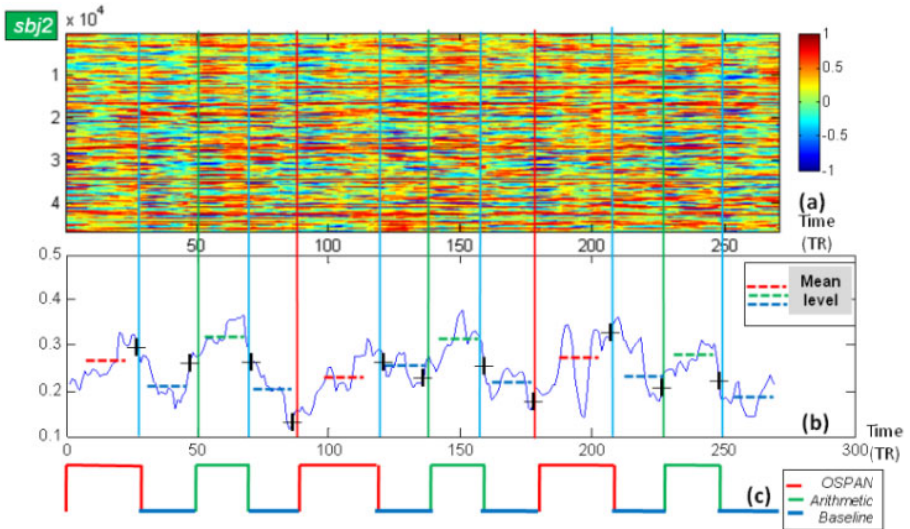


Fig. 8. Another example of global view of DFCs of all the fiber tracts (around 45,000 fibers) in the entire brain. The settings are similar as those in Fig. 7.

3.3 Statistics Analysis of Activated Fibers

In this subsection, we examine the spatial distribution of z-statistic values of the fibers in the whole brain. As an example, the histograms of the z-statistic values of all fibers in five brains for two types of stimulus (OSPAN and arithmetic contrasts in the working memory task [6]) are shown in Figure 9. It is evident that the global patterns of the histograms are quite similar across different types of stimuli, and the histogram

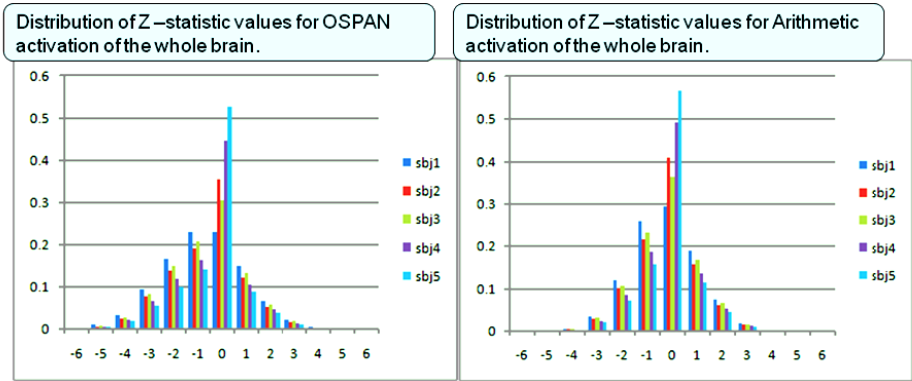


Fig. 9. Histograms of z-statistic values for two types of stimuli (OSPAN and Arithmetic respectively) for 5 subjects

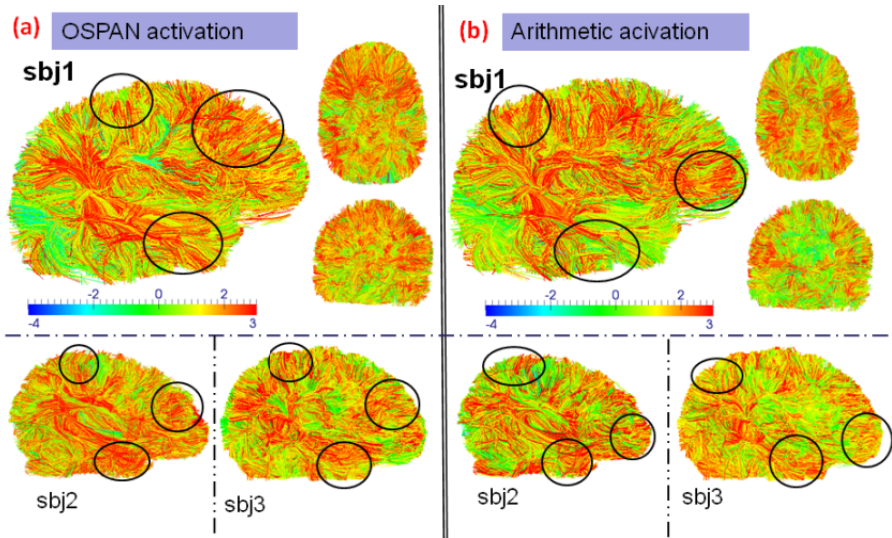


Fig. 10. Color-coded z-statistic values for fibers under two different stimuli. The color bars are in the middle. (a) Fibers under OSPAN stimulus. (b) Fibers under arithmetic stimulus. Both (a) and (b) show three subjects. Three views from different angles are shown for subject 1. Regions in black circles highlight example regions with consistent patterns across subjects under the same stimulus.

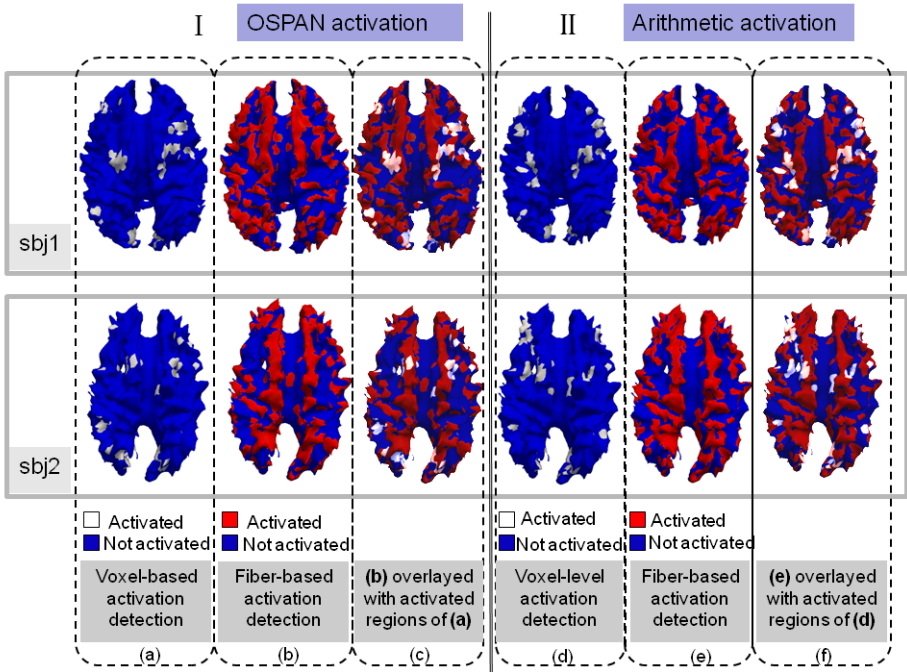


Fig. 11. Joint visualization of activated regions detected by voxel-based and fiber-centered GLM methods. Z-statistic value of 4 was used as threshold. (a)-(c): Activated regions for OSPAN stimulus resulted from the conventional voxel-based activation detection and our fiber-centered method. (d)-(f): Activated regions for arithmetic stimulus resulted from these two methods.

patterns for different brains are also quite similar. However, the spatial distributions of the z-statistic values for all the fibers in response to different stimuli are quite different, as shown in Figure 10. For better visualization, the fibers in the entire brain of three subjects are color-coded using the z-statistic values in Figure 10. The fibers in red have z-statistic values over 3. As shown in the exemplar regions highlighted by black circles, the z-statistic value maps corresponding to the two types of stimuli show different patterns (Figure 10a and 10b). This result is quite reasonable considering the brain regions that are activated in these two tasks, as shown in Figure 11 and in [6]. At the same time, activation patterns for the same type of stimuli are reasonably consistent across subjects, as demonstrated by the three subjects in Figure 10. The results in Figure 9 and 10 demonstrate the reliability of fiber-based activation detection in our methods.

3.4 Comparison of Fiber-Centered and Voxel-Based Activation Detections

Based on the activated fibers, we then detect activated brain regions via a straightforward mapping approach, i.e., any cortical or subcortical regions penetrated by the activated fibers are considered as activated brain regions. As an example, Figure 11

demonstrates that the activated fibers were mapped to cortical surface regions they connect to (red regions in Figure 11b and 11e). For comparison, the conventional voxel-based activations were also mapped onto the cortical surface (white regions in 11a and 11d). The voxel-based activated regions in Figure 11a and 11d were also overlaid on the activated regions in Figure 11b and 11e, respectively, as shown in Figure 11c and 11f. It is apparent that the activated regions detected in voxel-based method are a subset of the ones detected with our fiber-centered method, which partly demonstrates the validity of our methods. This result is reproducible in all of the cases we studied, suggesting that our fiber-centered method has better sensitivity in detecting activations than traditional voxel-based method.

Finally, we adopted different z-statistic value thresholds (from 3.5 to 5) to determine the activated fibers and thus activated brain regions in our method, and the results are shown in Figure 12. The activated regions by the voxel-based method are kept the same as those in Figure 11 and in [6]. It is apparent that increasing the z-statistic threshold will result in less activated fibers and brain regions, as shown in Figure 12a – 12d. It is striking that most of the regions obtained by the voxel-based method are consistently included in a subset of the detected brain regions by our method, suggesting the good sensitivity and robustness of our fiber-centered method. The results in Figure 12 further indicate that our fiber-centered method has better sensitivity than traditional voxel-based method. We believe the better sensitivity of our fiber-centered activation detection method is *not* due to false positives, because our extensive experimental results in Figure 6 – 10 consistently demonstrated that fibers are significantly more synchronized during stimulus intervals than during baseline intervals.

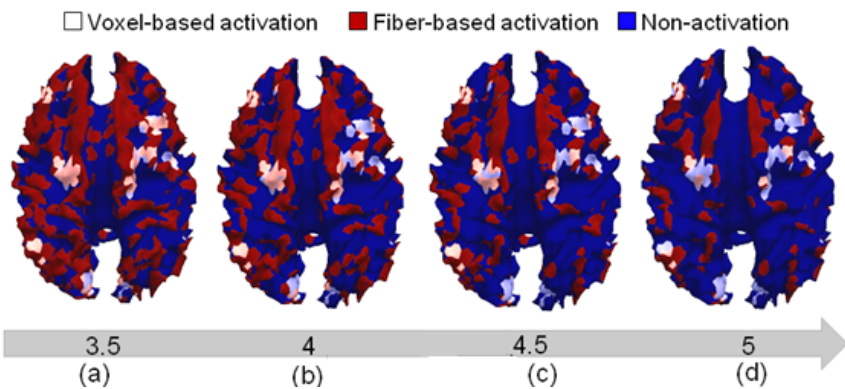


Fig. 12. Fiber-centered detection results using different z-statistic thresholds. (a)-(d): thresholds from 3.5 to 5, respectively. The activated regions by voxel-based method are overlaid on all of the results by fiber-centered methods.

4 Conclusion

We presented a novel framework for detection of activated fibers in task-based fMRI. In comparison with traditional voxel-based GLM activation detection, our framework

is able to detect more activated brain regions that are connected by white matter fibers with similar temporal patterns as the input stimulus paradigm. Our experimental results showed that many more fibers are significantly synchronized during task intervals, providing direction support to the concept of activated fibers. Then, the brain regions at the two ends of activated fibers are considered as participating in the task and thus are considered as activated regions. The experimental results in Figure 11 and 12 show that activated regions detected in traditional voxel-based GLM method are a subset of the activated regions detected by our method, partly supporting the validity of our framework. In the future, we will apply this framework to other multimodal task-based fMRI and DTI dataset, in order to further evaluate and validate this framework. Also, we will develop and evaluate statistical analysis methods to control the false positive detections in the activated fiber detection.

Acknowledgements

T Liu was supported by the NIH Career Award (NIH EB 006878) and the University of Georgia start-up research funding. Jinglei Lv was supported by the Doctorate Foundation of Northwestern Polytechnical University.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin-echo. *Journal of Magnetic Resonance Series B* 103(3), 247–254 (1994)
2. Biswal, B.B., Mennes, M., et al.: M Toward discovery science of human brain function. *PNAS* 107(10), 4734–4739 (2010)
3. Fox, M.D., Raichle, M.E.: Spontaneous fluctuations in brain activity observed with functional magnetic resonance imagin. *Nat. Rev. Neurosci.* 8(9), 700–711 (2007)
4. Friston, K.: Modalities, modes, and models in functional neuroimaging. *Science* 326(5951), 399–403 (2009)
5. Hu, X., Deng, F., Li, K., et al.: Bridging Low-level Features and High-level Semantics via fMRI Brain Imaging for Video Classification. *ACM Multimedia* (2010)
6. Faraco, C.C., Smith, D., Langley, J., et al.: Mapping the Working Memory Network using the OSPAN Task. *NeuroImage* 47(suppl. 1), S105 (2009)
7. Li, K., Guo, L., Li, G., et al.: Cortical surface based identification of brain networks using high spatial resolution resting state FMRI data. In: *IEEE International Conference on Biomedical Imaging: from Nano to Macro*, Rotterdam, Netherlands, pp. 656–659 (2010)
8. Lv, J., Guo, L., Hu, X., Zhang, T., Li, K., Zhang, D., Yang, J., Liu, T.: Fiber-centered analysis of brain connectivities using DTI and resting state FMRI data. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 143–150. Springer, Heidelberg (2010)
9. Liu, T., Li, H., Wong, K., et al.: Brain tissue segmentation based on DTI data. *NeuroImage* 38(1), 114–123 (2007)
10. Liu, T., et al.: Deformable Registration of Cortical Structures via Hybrid Volumetric and Surface Warping. *NeuroImage* 22(4), 1790–1801 (2004)
11. MedINRIA,
<http://www-sop.inria.fr/asclepios/software/MedINRIA/>

12. Zhang, D., et al.: Automatic cortical surface parcellation based on fiber density information. In: IEEE International Symposium on Biomedical Imaging (ISBI), Rotterdam, pp. 1133–1136 (2010)
13. Aviv, M., et al.: Cluster analysis of resting-state fMRI time series. *NeuroImage* 45, 1117–1125 (2009)
14. Honey, C.J., et al.: Predicting human resting-state functional connectivity from structural connectivity. *PNAS* 106(6), 2035–2040 (2009)
15. Worsley, K.J., et al.: A unified statistical approach for determining significant voxels in images of cerebral activation. *Human Brain Mapping* 4, 58–73 (1996)
16. Christopher, R., Genovese, N.A., Lazar, T.E.: Nichols: Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage* 15, 870–878 (2002)
17. Friston, K.J., et al.: Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Human Brain Mapping* 2, 189–210 (1995)

State–Space Models of Mental Processes from fMRI

Firdaus Janoos^{1,2}, Shantanu Singh¹, Raghu Machiraju¹,
William M. Wells², and Istvan Á. Mórocz²

¹ Dept. of Computer Science, The Ohio State University, USA

² Dept. of Radiology, Harvard Medical School, USA

Abstract. In addition to functional localization and integration, the problem of determining whether the data encode some information about the mental state of the subject, and if so, how this information is represented has become an important research agenda in functional neuroimaging. Multivariate classifiers, commonly used for brain state decoding, are restricted to simple experimental paradigms with a fixed number of alternatives and are limited in their representation of the temporal dimension of the task. Moreover, they learn a mapping from the data to experimental conditions and therefore do not explain the intrinsic patterns in the data. In this paper, we present a data–driven approach to building a spatio–temporal representation of mental processes using a state–space formalism, without reference to experimental conditions. Efficient Monte Carlo algorithms for estimating the parameters of the model along with a method for model–size selection are developed. The advantages of such a model in determining the mental–state of the subject over pattern classifiers are demonstrated using an fMRI study of mental arithmetic.

1 Introduction

The representation of information about the cognitive, perceptual or affective state of the subject from the metabolic traces of neural activity recorded in fMRI is an important research topic and is primarily investigated using *multivariate pattern recognition* (MVPR) [1]. The main advantage of the multivariate approach is that it does not make the naïve assumption that activity in voxels is independent, but rather searches for the information present in their interactions. Communication among neurons as well as larger functional units is the main basis of neural computation, and by not disregarding their interactions, multivariate methods may reveal more about the “neural code”.

However, these methods suffer from the following drawbacks: Firstly, as they learn a mapping from the distributed pattern of activity to labels describing stimuli or subject behavior, they are limited to explaining behavioral correlates and cannot discover patterns that may contain information about the internal mental state of the subject. Secondly, pattern classification approaches are restricted to experiments with a fixed number of alternatives (i.e. categorical valued stimuli) – not typical of natural perception. Generalization to complex paradigms with interval-valued parameters poses a significant methodological challenge [2]. Thirdly, they make the assumption that all fMRI scans with the same label have the same properties. Even spatio–temporal MVPR methods [3], which consider all the scans in a single block as the feature vector, are applicable only to block design experiments with fixed length blocks.

However, cognition recruits the entire brain and the underlying *mental processes* are fundamentally spatio-temporal in nature [4]. By neglecting the temporal evolution of brain function, such methods must necessarily compromise on representing all the information contained in the data. This is especially relevant for neurological and psychiatric disorders like dementia, schizophrenia, autism, multiple sclerosis, etc., or common learning disabilities like dyslexia or dyscalculia [9], where group–level analysis of spatial activity–maps has been inconclusive due to the high anatomical and functional variability observed even within the same group. It is widely acknowledged that the similarities and differences between the brain function of subjects may not reside in the spatial layout of the mental activity but more so in its temporal organization [4].

In an attempt to extract information about *mental processes* from fMRI, this paper presents a method to determine the *intrinsic spatio-temporal patterns* in the data during a complex cognitive experimental paradigm without reference to experimental variables. Here, we assume the presence of a set of abstract mental states, each corresponding to a characteristic configuration of neural / metabolic activity, that are revisited time and time again in a temporally ordered fashion during the performance of a mental task. Section 2 builds on the concept of the brain transitioning through a mental state–space using a *hidden Markov model* (HMM) [2] formalism. The model accounts for the hemodynamic response, which causes metabolic activity at one time–point to be observed in the next few fMRI scans, using a linear convolutional filter. The parameters of the model are estimated with *expectation maximization* (EM) and the correct model size is determined in an automated fashion by selecting the number of states that best describe the task being performed by the subject. This analysis is performed in a low–dimensional feature–space that concisely captures the salient spatial patterns in data.

In Section 3, the model is applied to an fMRI study of *developmental dyscalculia* (DDC) [9] to understand the effects of experimental parameters on the mental processes of DDC subjects vs. controls. Static activity–map based analysis for this paradigm failed to reveal systematic differences between the populations [9], primarily due to the high functional variability between DDC subjects as they solve an arithmetical task. By comparing the spatio–temporal patterns encoded in the SSM, we are able to observe group–level differences not identifiable through static spatial maps.

Related Work

Typically linear classifiers, such as correlation-based classifiers, single-layer perceptrons, linear discriminant analysis (LDA), linear support vector machines (SVMs) and Gaussian naïve Bayes (GNB) [11] have been used to learn the mapping from the distributed pattern of activity at one time–point to the prevailing experimental stimulus, due to simplicity of interpretation without significant loss of accuracy. The linear relationship between regressors, formed by convolving the stimuli with a *hemodynamic response function* (HRF), and the distributed patterns in the data have been studied using methods such as canonical correlation analysis (CCA) and multivariate Bayesian decoding (MVB) [3]. These methods assume a known mathematical relationship between stimulus and fMRI signal, which may be hard to define, especially in complex cognitive paradigms. Multivariate *unsupervised* approaches such as clustering [7] or matrix–based decompositions (e.g. ICA or PCA) are also popular [11], especially for

resting–state and non task–related data. The lack of a generative model of brain–function in these methods prevents neurophysiological interpretation of the results [11].

Markov model formalisms have been previously used in fMRI for determining the activation state of individual voxels in a univariate fashion [8]. Dynamic Bayesian networks and dynamic causal models [13] have been used to study the time-varying functional integration of a small number of pre–specified regions in the brain, from the interdependency structure of their average time-series.

In contrast, this paper present a novel data–driven multivariate method for the dynamical analysis of mental processes in a time-resolved fashion using a neurophysiologically inspired model for brain function that: (a) captures not only the spatial distribution but also the temporal ordering of activity; (b) can be applied to arbitrarily complex paradigms, not just experiments with fixed alternatives; (c) does not require specification of the mathematical relationship between experimental variables and fMRI signal; (d) permits discovery of internal mental states; (e) facilitates neurophysiological interpretation of the parameters; (f) allows comparison of the spatio–temporal patterns between subjects in their entirety. Comparing brain–function of subjects in abstract representation spaces rather than their spatial–maps directly has been shown to be a very powerful principle in psychology and neuroscience for revealing their commonalities and differences [5].

2 The State–Space Model (SSM)

A graphical description of the SSM with K hidden states is shown in Fig. 1(a). Here, $\mathbf{y}_t \in \mathbb{R}^D$ is the observed fMRI scan at time–point $t = 1 \dots T$ represented in the D –dimensional feature–space (FS), explained later. The corresponding experimental stimulus (not shown) is encoded by the vector \mathbf{s}_t . The underlying mental process is represented as a sequence of hidden states $x_t \in [1 \dots K]$, for $t = 1 \dots T$. The state marginal distribution is given by $\alpha = (\alpha_1 \dots \alpha_K)$, where $\Pr[x_t = k] = \alpha_k$. The transition probabilities $\Pr[x_{t+1} = k_2 | x_t = k_1] = \pi_{k_1, k_2}$ are given by the $K \times K$ stochastic matrix π . Each brain–state is associated with a characteristic distribution of neural (metabolic) activity represented by $\mathcal{N}(\mu_k, \Sigma_k)$ in FS coordinates.

This activity is observed in the fMRI signal \mathbf{y} via the effect of *hemodynamic response* (HR) of the brain, modeled by a linear *finite impulse response* (FIR) filter \mathbf{h} of length $L + 1$ TRs (typically set to a duration of 32s). It is obtained from the canonical *hemodynamic response function* (HRF) and has non-linear parameters γ controlling its delay, dispersion and ratio of onset-to-undershoot, with a prior density $p(\gamma) = \mathcal{N}(\mu_\gamma, \sigma_\gamma)$. The fMRI data, arising from the linear convolution of the activity patterns with the HRF, is modeled by $\mathbf{y}_t \sim \mathcal{N}\left(\sum_{l=0}^L \mathbf{h}_l \mu_{x_{t-l}}, \sum_{l=0}^L \mathbf{h}_l^2 \Sigma_{x_{t-l}} + \Sigma_\epsilon\right)$. Here, Σ_ϵ is the variance of a spatial noise term $\epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon)$. Note that the convolution is actually in voxel–space, but since the linear FS transform is commutative with convolution, it can be performed directly in FS.

Matrix transpose is denoted by the \top operator throughout the text. Introducing the following shorthand notation: $\vartheta_k \triangleq (\mu_k, \Sigma_k)$; $\vartheta \triangleq \{\vartheta_1 \dots \vartheta_K\}$; $\theta \triangleq \{\alpha, \pi, \vartheta, \Sigma_\epsilon\}$; $\mathbf{y}_{t_1 \dots t_2} \triangleq \{\mathbf{y}_{t_1} \dots \mathbf{y}_{t_2}\}$; $x_{t_1 \dots t_2} \triangleq \{x_{t_1} \dots x_{t_2}\}$; $\mu_{t-L \dots t} \triangleq \sum_{l=0}^L \mathbf{h}_l \mu_{x_{t-l}}$; $\Sigma_{t-L \dots t} \triangleq \sum_{l=0}^L \Sigma_{x_{t-l}} \mathbf{h}_l^2 + \Sigma_\epsilon$ and $p_\theta(\cdot) \triangleq p(\cdot | \theta, \mathbf{h}, K)$ the following probability model is obtained:

$$\begin{aligned}
 p(\mathbf{y}, x|\theta, \mathbf{h}, K) &= p(\mathbf{y}|x, \vartheta, \mathbf{h}, K)p(x|\alpha, \pi, K) \quad (1) \\
 p(x|\alpha, \pi, K) &= \prod_{t=1}^T p(x_t|x_{t-1}, \alpha, \pi) = \alpha_{x_1} \prod_{t=2}^T \pi_{x_t, x_{t-1}} \\
 p(\mathbf{y}|x, \vartheta, \mathbf{h}, K) &= \prod_{t=1}^T p(\mathbf{y}_t|x_{t-L} \dots x_t, \vartheta) = \prod_{t=1}^T \mathcal{N}(\mu_{t-L \dots t}, \Sigma_{t-L \dots t}).
 \end{aligned}$$

The convolution with the HRF introduces a dependency between states $x_{t-L} \dots x_t$, when conditioned on an observation \mathbf{y}_t . Although this violates the first-order Markov property required for the classical forward-backward recursions, its parameters can be efficiently estimated using the $L + 1$ -order Monte–Carlo algorithm described next.

The maximum likelihood (ML) estimate $\theta_{\text{ML}} = \arg \max_{\theta} \ln p(\mathbf{y}|\theta, K)$ is obtained using EM [2] which involves iterating the following two steps until convergence: **E-step**: $\mathcal{Q}(\theta, \theta^n) = \sum_x p(x|\mathbf{y}, \theta^n, \mathbf{h}, K) \ln p(\mathbf{y}, x, \theta|\mathbf{h}, K)$, and **M-step**: $\theta^{n+1} = \arg \max_{\theta} \mathcal{Q}(\theta, \theta^n)$.

E-Step

Using the relationship $p_{\theta^{(n)}}(x|\mathbf{y}) = p_{\theta^{(n)}}(\mathbf{y}, x)/p_{\theta^{(n)}}(\mathbf{y})$ and the fact that $p_{\theta^{(n)}}(\mathbf{y})$ is canceled out by the numerators and denominators in the M-Step updates (cf. eqn. 3), the conditional densities are replaced by their joint densities $p_{\theta^{(n)}}(\mathbf{y}, x_t)$, $p_{\theta^{(n)}}(\mathbf{y}, x_t, x_{t+1})$ and $p_{\theta^{(n)}}(\mathbf{y}, x_{t-L \dots t})$. These are calculated as:

$$\begin{aligned}
 p_{\theta}(\mathbf{y}, x_t) &= \sum_{x_{t+1-L \dots t-1}} a(x_{t+1-L \dots t})b(x_{t+1-L \dots t}) \\
 p_{\theta}(\mathbf{y}, x_t, x_{t+1}) &= \sum_{x_{t+1-L \dots t-1}} a(x_{t+1-L \dots t}) \cdot p_{\theta}(\mathbf{y}_{t+1}|x_{t+1-L \dots t+1})p_{\theta}(x_{t+1}|x_t) \cdot b(x_{t+2-L \dots t}) \\
 p_{\theta}(\mathbf{y}, x_t, x_{t+L}) &= a(x_{t, t+1-L}) \cdot p_{\theta}(\mathbf{y}_{t+L}, x_{t, t+L}) \cdot p_{\theta}(x_{t, t+L}) \cdot b(x_{t+1 \dots t+L}). \quad (2)
 \end{aligned}$$

where a and b are the *forward–backward recursion* terms:

$$\begin{aligned}
 a(x_{t+1-L \dots t}) &= p_{\theta}(\mathbf{y}_{1 \dots t}, x_{t+1-L \dots t}) = \sum_{x_{t-L}} p_{\theta^{(n)}}(\mathbf{y}_t|x_{t-L \dots t})p_{\theta^{(n)}}(x_t|x_{t-1}) \cdot a(x_{t-L \dots t-1}) \\
 b(x_{t+1-L \dots t}) &= p_{\theta}(\mathbf{y}_{t+1 \dots T}|x_{t+1-L \dots t}) = \sum_{x_{t+1}} p_{\theta^{(n)}}(\mathbf{y}_{t+1}|x_{t+1-L \dots t+1})b(x_{t+2-L \dots t+1}).
 \end{aligned}$$

The summations (i.e. expectations) over the densities of state–sequences $L + 1$ long of the form $\sum_{x_{t-L \dots t}} p_{\theta^{(n)}}(\mathbf{y}, x_{t-L \dots t})[\dots]$ in eqn. 3 and eqn. 2 are replaced with Monte–Carlo estimates, by *Gibbs sampling* from the distribution $p_{\theta^{(n)}}(\mathbf{y}, x_{t-L \dots t})$ with stochastic forward-backward recursions [12].

M-Step

Because of the inclusion of the FIR filter for the HRF, which violates the first-order Markov property of the state–sequence x when conditioned on an observation \mathbf{y}_t , the M-Step update equations take the following form:

$$\begin{aligned}
 \alpha_k^{n+1} &= \frac{p_{\theta^{(n)}}(x_1 = k|\mathbf{y})}{\sum_{k'=1}^K p_{\theta^{(n)}}(x_1 = k'|\mathbf{y})}, & \pi_{k_1, k_2}^{n+1} &= \frac{\sum_{t=2}^T p_{\theta^{(n)}}(x_t = k_1, x_{t+1} = k_2|\mathbf{y})}{\sum_{k'=1}^K \sum_{t=2}^T p_{\theta^{(n)}}(x_t = k_1, x_{t+1} = k'|\mathbf{y})}, \\
 \mu_k^{n+1} &= \sum_{k_0 \dots k_L} \mathbf{H}_{k, k_0 \dots k_L}^- \mu_{k_0 \dots k_L}^{n+1} & \text{and} & \Sigma_k^{n+1} = \sum_{k_0 \dots k_L} \mathbf{G}_{k, k_0 \dots k_L}^- \Sigma_{k_0 \dots k_L}^{n+1}. \quad (3)
 \end{aligned}$$

The updates to $\mu_{t-L\dots t}$ and $\Sigma_{t-L\dots t}$ (cf. eqn. [11](#)) for one particular assignment $\{k_0 \dots k_L\}$ of the sequence $L + 1$ states long are:

$$\begin{aligned}\mu_{k_0\dots k_L}^{n+1} &= \frac{\sum_{t=1}^T p_{\theta^{(n)}}(x_{t-L\dots t} = k_0 \dots k_L | \mathbf{y}) \mathbf{y}_t}{\sum_{t=1}^T p_{\theta^{(n)}}(x_{t-L\dots t} = k_0 \dots k_L | \mathbf{y})}, \\ \Sigma_{k_0\dots k_L}^{n+1} &= \frac{\sum_{t=1}^T p_{\theta^{(n)}}(x_{t-L\dots t} = k_0 \dots k_L | \mathbf{y}) \cdot (\mathbf{y}_t - \mu_{k_0\dots k_L}^{n+1})(\mathbf{y}_t - \mu_{k_0\dots k_L}^{n+1})^\top}{\sum_{t=1}^T p_{\theta^{(n)}}(x_{t-L\dots t} = k_0 \dots k_L | \mathbf{y})}.\end{aligned}$$

Here, \mathbf{H} and \mathbf{G} are the $K^{L+1} \times K$ convolution matrices that give the relationship between HRF coefficients \mathbf{h}_l , the activation pattern means μ_k for state k and the $\mu_{t-L\dots t}, \Sigma_{t-L\dots t}$ for any assignment of an $L + 1$ length state–sequence [12](#):

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_L + \dots \mathbf{h}_0 & 0 & \dots & 0 \\ \mathbf{h}_L + \dots \mathbf{h}_1 & \mathbf{h}_0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \mathbf{h}_{L-1} + \mathbf{h}_0 \\ 0 & 0 & \dots & \mathbf{h}_L + \mathbf{h}_0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{h}_L^2 + \dots \mathbf{h}_0^2 & 0 & \dots & 0 \\ \mathbf{h}_L^2 + \dots \mathbf{h}_1^2 & \mathbf{h}_0^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \mathbf{h}_{L-1}^2 + \mathbf{h}_0^2 \\ 0 & 0 & \dots & \mathbf{h}_L^2 + \mathbf{h}_0^2 \end{pmatrix}, \quad (4)$$

and $\mathbf{H}_{k,k_0\dots k_L}^-$ is the $(k, k_0 \dots k_L)$ -th element of the pseudo-inverse of \mathbf{H} , given by $\mathbf{H}^- = (\mathbf{H}^\top \mathbf{H})^- \mathbf{H}^\top$. Even though \mathbf{H} is an $K^{L+1} \times K$ matrix, it is extremely sparse with each column k of \mathbf{H} having only 2^{L+1} non-zero entries corresponding to those $\mu_{k_0\dots k_L}^{n+1}$ where $k \in \{k_0 \dots k_L\}$. Therefore, $\mathbf{H}^\top \mathbf{H}$ is computed in $\mathcal{O}(2^{L+1} K^2)$ time, and is inverted using the SVD pseudo-inverse. Similarly for \mathbf{G} .

The dependence of θ_{ML} on a specific HRF filter \mathbf{h} is removed by marginalizing out \mathbf{h} under a Laplace approximation to obtain a Bayesian estimate $\theta^* = \int_{\mathbf{h}} \theta_{\text{ML}}(\mathbf{h}) p(\mathbf{h}) d\mathbf{h}$, independent of \mathbf{h} . It is computed through Monte–Carlo integration by first sampling the parameter γ from $\mathcal{N}(\mu_\gamma, \sigma_\gamma)$, constructing $\mathbf{h}(\gamma)$, finding $\theta_{\text{ML}}(\mathbf{h})$ and then averaging over all samples.

Optimal State–Sequence Estimation

Given a set of parameters θ and observations \mathbf{y} , the most probable sequence of states $x^* = \arg \max_x \ln p_\theta(\mathbf{y}, x)$ is estimated by *backtracking* through the following recursive system: $\max_x \ln p_\theta(\mathbf{y}, x) = \max_{x_{t-L\dots T}} \eta_T$, where $\eta_t = \max_{x_{t-1}} [\ln p_\theta(\mathbf{y}_t, x_{t-L\dots t}) + \ln p_\theta(x_t | x_{t-1}) + \eta_{t-1}]$ and $\eta_1 = \ln p_\theta(\mathbf{y}_1 | x_1) + \ln p_\theta(x_1)$. The maximization over sequences of states $L + 1$ long $x_{t-L\dots t}$ is done using *iterated conditional modes* (ICM) [13](#), with random restarts.

Model–Size Selection

Model–size (i.e. K) selection can be done using Bayes factors, information theoretic criteria or reversible jump MCMC based methods [14](#), that strike a compromise between

¹ Consider one particular assignment of states $x_{t-L\dots t} = \{k_0 \dots k_L\}$ and let $\mu_{k_0\dots k_L} \triangleq \sum_{l=0}^L \mu_{k_l} \mathbf{h}_{L-l}$. In vector notation $\boldsymbol{\mu}_k^{(i)} \triangleq (\mu_1^{(i)} \dots \mu_K^{(i)})^\top$, the K -vector formed from the i -th element of $\mu_1 \dots \mu_K$ (each are of D -dimensions). Similarly, let $\boldsymbol{\mu}_{k_0\dots k_L}^{(i)}$ be the K^{L+1} -vector formed from the i -th element of $\mu_{1\dots 1} \dots \mu_{1\dots K}, \dots, \mu_{K\dots 1} \dots \mu_{K\dots K}$. Then in matrix notation, the convolution is $\boldsymbol{\mu}_{k_0\dots k_L}^{(i)} = \mathbf{H} \boldsymbol{\mu}_k^{(i)}$, and therefore $\boldsymbol{\mu}_k^{(i)} = \mathbf{H}^- \boldsymbol{\mu}_{k_0\dots k_L}^{(i)}$. Similarly for the i_1, i_2 -th element the $D \times D$ covariance matrix Σ_k .

model complexity $p(K)$ and *model evidence* $p(\mathbf{y}|K)$. In the absence of a domain–specific way to define model complexity, we adopt an alternative strategy where experimental conditions \mathbf{s}_t are used to select K that results in a *maximally predictive model*. The rationale behind this strategy is that fMRI data may contain multiple spatio–temporal patterns of both neurophysiological (such as default–network processes) and extraneous (such as respiratory, pulsatile, head–motion) origin, of which only task–related effects are of interest to the investigator. This criterion selects a model which has identified the patterns most relevant to the task. Although this step introduces a dependence between the experimental conditions and the model, the parameters themselves are estimated without reference to the task in an *unsupervised* fashion.

Let $x^{*,K}$ denote the optimal state–sequence estimate for an fMRI session \mathbf{y} produced by the model with K states and optimal parameters θ_K^* . And, let $\mathbf{s} = (\mathbf{s}_1 \dots \mathbf{s}_T)$ denote the corresponding experimental conditions recorded during the session. A *multinomial logistic regression* (MLR) classifier [2] with weights \mathbf{w} is trained to predict the state x_t at time t given stimulus \mathbf{s}_t according to $\Pr[x_t = k|\mathbf{s}_t] = \exp\{\mathbf{s}_t^\top \mathbf{w}_k\} / \sum_{k'=1}^K \exp\{\mathbf{s}_t^\top \mathbf{w}_{k'}\}$. The optimal K^* is then selected as that for which the error in predicting $x^{*,K}$ from experimental conditions $\text{ERR}_{\text{predict}}(x^{*,K}, K) \triangleq \mathbb{E} [1 - \Pr[x_t^{*,K}|\mathbf{s}_t]]$ is minimum. This is computed with M –fold cross–validation.

Spatial Maps

Spatial maps of the activation patterns for a specific value of the experimental stimuli \mathbf{s}_t is obtained by first computing the activation pattern mean $\mu_{\mathbf{s}_t} = \sum_{k=1}^K \Pr[X_t = k|\mathbf{s}_t] \mu_k$ and variance $\Sigma_{\mathbf{s}_t} = \sum_{k=1}^K \Pr[X_t = k|\mathbf{s}_t] \Sigma_k$, where $\Pr[X_t = k|\mathbf{s}_t]$ are the MLR probabilities for the condition \mathbf{s}_t . The z –score map is given by $\Sigma_{\mathbf{s}_t}^{-1/2} \mu_{\mathbf{s}_t}$ in feature–space which is then transformed back into a voxel–wise spatial map of activity.

The Feature–Space Φ

With the number of voxels $N \sim \mathcal{O}(10^5)$ orders of magnitude larger than the number of scans $T \sim \mathcal{O}(10^2)$, dimensionality reduction is necessary to prevent over–fitting. Typically, the fMRI data are represented by a subset of voxels or transformed into feature–spaces obtained using PCA, ICA or the original fMRI scans themselves (i.e. support vectors) [3]. Although dimensionality reduction can be performed in a supervised fashion by selecting features highly correlated with or most predictive of the experimental variables or through manual selection of regions-of-interest, such approaches are inherently biased towards the experimental variables against which the features are selected while ignoring intrinsic patterns in the data. When using unsupervised feature selection, it has been widely observed that the highest variance principle components (PCs) typically corresponded to motion and physiological noise, artifacts [11], while ICA does not offer a natural ordering of the components in terms of their saliency.

To represent the data in a low–dimensional space that retains intrinsic patterns, we use a feature–space obtained by the eigen–decomposition of the voxel–wise correlation matrix (as compared to the covariance matrix in PCA), which depends on the structure of the inter–relationships between the voxel time–series and not on their magnitudes. Dimensionality reduction is performed using a bootstrap analysis of stability [1]. The bootstrap generates a non–parametric estimate of the sampling distribution of a statistic (i.e. *bootstrap distribution*) from a single sample of the data by creating

multiple surrogate samples, of same size as the original sample, by *resampling with replacement* from the original sample. Bootstrap estimates of the voxel-wise correlation matrix are obtained by *block resampling* of the fMRI session to create a surrogate session. The stability of a basis vector $\phi^{(l)}$ is defined by its correlation coefficient $\rho^{(l)}(r_1, r_2) \triangleq |\langle \phi_{(r_1)}^{(l)}, \phi_{(r_2)}^{(l)} \rangle|$ across resamples r_1, r_2 of the session. Only those basis vectors $\phi^{(l)}$ have a certain percentage (typically 75%) of $\rho^{(l)}$ above a threshold τ_ρ are retained. The value of τ_ρ is set to that at which there is a sharp drop in dimensionality ($D \approx 100$ for our data-set).

3 Results

Methods and Materials

The method was applied on a study for *developmental dyscalculia* (DDC) [9] consisting of 20 control (10 female) and 13 DDC (6 female) subjects, who underwent fMRI while judging the incorrectness of multiplication results. In order to balance the group sizes, group-level analysis was done by selecting 10 subjects at random from each group and computing the statistics over multiple resamples.

In each trial of the *selfpaced, irregular paradigm*, two single-digit numbers (e.g. 4×5) were displayed visually for 2.5s. After an interval of 0.3s an incorrect solution (e.g. 27,23,12) was displayed for 0.8s. Subjects had *up to* 4s to decide, with a button press, if the answer was (a) *close* (within $\pm 25\%$ of the correct answer), (b) *too small* ($< 25\%$) or (c) *too big*. The next trial started after a rest of 1s, and each trial lasted 4–8.6s. For each $t = 1 \dots T$, the experimental conditions are described by the vector $s_t = (\text{Ph}, \text{LogPs}, \text{LogDiff})$, where Ph = 1, 2, 3 indicates if t is in the (1) *multiplication* or (2) *subtraction / judgement* or (3) *decision-making* phase of the experiment, $1 \leq \text{LogPs} \leq 10$ quantifies the product size of the multiplication task and $1 \leq \text{LogDiff} \leq 5$ quantifies the expected difficulty in judging the right answer. The data were acquired with a GE 3T MRI scanner with a quadrature head coil, using a BOLD sensitized 3D PRESTO pulse sequence with a volume scan time of 2.64s and resolution of $3.75 \times 3.75 \times 3.75 \text{mm}^3$. All subjects were scanned in two sessions for a total of ≈ 240 multiplication problems and 552 scans. The data were pre-processed to remove imaging artifacts, motion corrected, de-noised, and normalized to an atlas space. All further processing was done in the grey matter.

One SSM $\mathcal{M}_i = \{K, \theta^{*,K}, x^{*,K}\}$ was trained per subject and the following statistics calculated: **[ERR_{SSM}^{self}]** The “within-subject” prediction error-rate of the optimal state-sequence assigned to the fMRI run of one subject using the model for that subject (i.e. trained on the same data). **[ERR_{SSM}^{cross}]** The “between-subject” prediction error-rate of the optimal state-sequence assigned to the data of one subject using a model trained on the data of another subject. **[MI]** The *mutual information* between the state sequences generated for the same fMRI session y by the models \mathcal{M}_i and \mathcal{M}_j for two different subjects i and j . In general, the correspondence between the state labels of two different models is unknown. Comparing the state-sequences of the same data generated by the two models allows this correspondence to be determined. A higher MI indicates a higher level of agreement between the state-sequences of the same data when labeled by two different models, while an MI of zero indicates no correspondence. The MI between all $\binom{33}{2}$ pairs of subjects yields a pair-wise similarity matrix, which can then be

visualized in 2D through *multidimensional scaling* (MDS) as shown in Fig. 1(b) [5]. The dimensionality of the correlation derived feature–space (FS) Φ was set uniformly to $D = 100$ and we did not observe any significant differences in the statistics over the range of $D = 80$ to 150 across the subjects.

Although MVPR classifiers cannot be directly used with real–valued parameters, for comparative evaluation, we trained three linear SVM classifiers per subject: one to predict $\text{Ph} = 1, 2, 3$, one to predict LogPs quantized to two levels and one for LogDiff quantized to two levels. The within–subject $\text{ERR}_{\text{SVM}}^{\text{self}}$ and between–subject $\text{ERR}_{\text{SVM}}^{\text{cross}}$ error–rates for were measured with cross–validation and cumulated across the three SVMs. The pre–processed fMRI data were deconvolved with the canonical HRF when used with the SVMs. Among the other classifiers evaluated (viz. GNB, LDA and quadratic SVM), none significantly outperformed the linear SVM. The average values of these statistics are compiled in Table 1.

Table 1. Error Rates. The group–wise average and ± 1 standard error of mean (SEM) values for within–subject and between–subject prediction errors, the between–subject MI and the optimal number of states K^* respectively are tabulated. The last row shows the between–subject prediction errors and mutual information comparing control subjects and DDC subjects. The “chance–level” prediction error is ≈ 0.83 calculated by permuting the stimuli with respect to the scans. The “chance–level” MI is ≈ 0.8 bits.

Group	$\text{ERR}_{\text{SVM}}^{\text{self}}$	$\text{ERR}_{\text{SVM}}^{\text{cross}}$	$\text{ERR}_{\text{SSM}}^{\text{self}}$	$\text{ERR}_{\text{SSM}}^{\text{cross}}$	MI (bits)	K^*
Overall	0.42 ± 0.05	0.61 ± 0.09	0.36 ± 0.06	0.48 ± 0.07	3.15 ± 0.25	21.06 ± 4.53
Control	0.40 ± 0.04	0.56 ± 0.07	0.34 ± 0.05	0.41 ± 0.06	3.81 ± 0.12	19.67 ± 2.44
DDC	0.43 ± 0.06	0.62 ± 0.08	0.39 ± 0.07	0.52 ± 0.06	2.94 ± 0.15	23.18 ± 6.53
Ctrl. vs. DDC		0.64 ± 0.10		0.53 ± 0.08	2.72 ± 0.33	

Discussion

DDC is a specific learning disability affecting the acquisition of mathematical skills in children with otherwise normal general intelligence and has been attributed to either abnormalities in the numerical (non–verbal) quantity processing system localized in the left and right intraparietal regions, or to impairments of the verbally encoded mathematical–fact retrieval system localized in the frontal and temporal (language) regions. However, the neuronal substrates of this disorder are poorly understood, a fact complicated by difficulties in accurate diagnosis, heterogeneity of arithmetic deficits and the frequent association with dyslexia and attention disorders [9].

As seen from Table 1, a larger variation in model–sizes for the DDC group can be observed versus the controls. This points to a greater heterogeneity in the DDC data necessitating models with different sizes. Also, their consistently higher error–rate indicates the relative inaccuracy of their models versus the controls. Additionally, the MI between DDC subjects is on par with that between DDC and controls. This indicates that while a DDC model labels an fMRI run quite incongruently to a control model, it even labels the data differently as compared to another DDC model. This reaffirms the theory of high heterogeneity in the mental strategies adopted by DDC subjects.

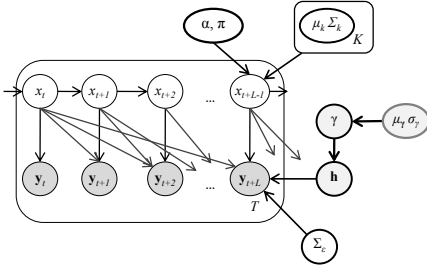
The group–level spatial maps estimated by the SSM with the FS Φ for the two groups are shown in Fig. 2. While the maps show overall recruitment patterns similar

to those observed with univariate general linear models, the time-resolved and multivariate nature of the SSM reveals foci that are stronger within and more distinct across the groups [9]. This result is expected according to the theory that DDC subjects differ from healthy adults not so much in the areas recruited during the task, but rather in their relative timings [9]. For example, during the first phase the DDC group shows lower activation in the right intra-parietal sulcus (rIPS), believed to contain mental number line, as compared to the controls. However, by the third phase the DDC group shows stronger recruitment as compared to the control group in the rIPS.

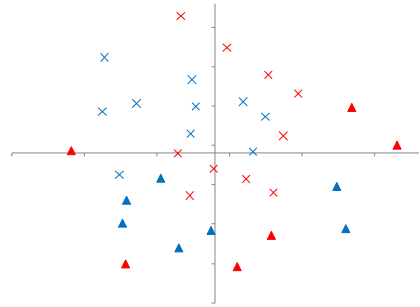
The specification of the SSM as a phenomenological model in terms of abstract mental-states allows comparing the spatio-temporal patterns between subjects in their entirety in this abstract representation [5] as shown in Fig. 1. This provides a visual representation of the *intrinsic* differences and commonalities between the mental process of the subjects. It should be noted that the labeling is applied *post hoc* after plotting all the subjects using MDS on the pairwise MI matrix and Here we see a clustering of the two populations, with a larger spread amongst the DDC subjects indicating higher heterogeneity versus the controls. The separation between the clusters for the two groups, was assessed as significant at a 0.80 confidence level using Cramer's test. Although not shown here, the effect of parametric variables on these clusterings can also be studied by computing the MI between subjects given a particular value of the stimulus s_t .

The SSM consistently outperforms the SVM in predicting the mental state of the subject. This is noteworthy given that the SSM learnt the parameters in an unsupervised fashion while the parameters of the SVM were specifically optimized for prediction error. By treating each fMRI scan independently, MVPR classifiers are unable to leverage the temporal structure in the data. Also, the SSM accounts for the hemodynamics in probabilistic fashion and marginalizes out their effects, while the SVM uses a point-estimate of the neural activation. Withal, the between-subject errors for the SVM were significantly (> 2 SEM) higher than the SSM. This points to the sub-optimality of comparing subjects based on static spatial maps alone without considering their temporal dynamics.

In conjunction with Φ , the SSMs and SVMs were also tested with the following PCA-based feature-spaces: **[FS:PCA-UNSUPER]** : The top 100 PCs accounting for $\approx 96\%$ of the variance; **[FS:PCA-SUPER-1]** : The 100 PCs maximally correlated with HRF-convolved regressors for Ph, LogPs and LogDiff identified through multiple regression; and **[FS:PCA-SUPER-2]** : The 100 PCs maximally correlated with only the HRF-convolved regressor for Ph. For the basis FS:PCA-UNSUPER with features selected in an unsupervised fashion, we observed drastic reduction in performance for both the SVM ($ERR_{SVM}^{self} = 0.53 \pm 0.08$ and $ERR_{SVM}^{cross} = 0.68 \pm 0.11$, overall) and SSM ($ERR_{SSM}^{self} = 0.51 \pm 0.09$ and $ERR_{SSM}^{cross} = 0.69 \pm 0.10$, overall). The poor accuracy of this basis is due to the lack of a specific relationship between the task selectivity of a component and its variance, and we observed that the highest variance PCs were usually of artifactual origin. Selecting PCs correlated with all the experimental variables in FS:PCA-SUPER-1 reduced the error rate of the SVM by $\approx 7.5\%$ as compared to that with Φ , while the error of the SSM increased by $\approx 6.9\%$. These results, with their < 1 SEM changes, indicate the optimality of Φ to capture task related patterns in the data.



(a) The State-Space Model (SSM)



(b) MDS Plot of Pair-wise MI

Fig. 1. The hidden brain-states are shown in Fig.(a) by x_t , the activation pattern is observed in the fMRI data $y_t \dots y_{t+L-1}$ after convolution with the hemodynamic response h . Fig.(b) shows the 33 subjects (20 control, marked with crosses and 13 DDC, marked with triangles) plotted in 2-D based on their pairwise MI. Red colored markers represent female subjects and blue represent males.

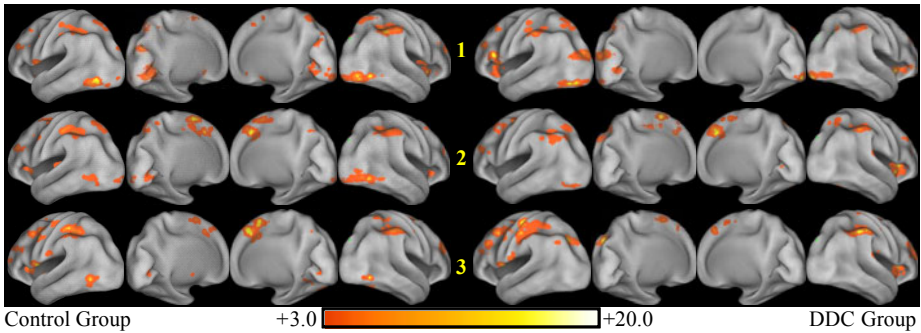


Fig. 2. Spatial Activity Maps. The average t -score maps of the two groups (voxel-wise group average divided by group std. dev.) are displayed on an inflated brain-surface (left lateral-posterior, left medial-posterior, right medial-posterior and right lateral-posterior). Each row shows the maps for one phase of the task and values $t < 3$ have been masked out for clarity.

The lack of significant improvement for the supervised basis FS:PCA-SUPER-1 may be because the exact coupling between experimental variables and fMRI signal (and therefore PC time-courses) is unknown and may be non-linear. The basis Φ is selected in an unsupervised fashion on a criteria of stability and is hence not affected by this ambiguity. For the supervised basis FS:PCA-SUPER-2, the SVM error-rate for Ph reduced by over 23.6% of that with Φ but the cumulative errors increased to $ERR_{SVM}^{self} = 0.64 \pm 0.07$ and $ERR_{SVM}^{cross} = 0.73 \pm 0.08$. Increased specificity for a particular variable came at the cost of reduced accuracy for the other variables FS:PCA-SUPER-2 had an even more deleterious effect on the SSM with accuracy at near chance levels. The SSM does not search for the effect of any particular experimental variable but rather for spatio-temporally coherent patterns, which may have been destroyed by this FS – indicating the limitations of a supervised FS in an unsupervised analysis.

4 Conclusion

In this paper, we introduced a state–space model (SSM) to represent the neural processes recorded in fMRI data. Its parameters, estimated in an unsupervised fashion, provide information about the *intrinsic mental states* of the subject and have a well defined interpretation in terms of the underlying phenomenological model. This is in contrast with the interpretation problems of MVPR methods which are based on criteria such as linear separability and maximum margin. The capability of the SSM to predict stimuli, which is *not of direct importance in experimental neuroscience* [3], is used instead as a strategy to select between models and to infer the presence of an effect of experimental variables on the data. Nonetheless, by learning the temporal organization of task–driven neural activity, this method exhibited prediction capability superior to MVPR methods. Furthermore, our approach allows comparison of the spatio–temporal patterns across subjects by comparing their models, and not just static activity–maps.

The model, however, in its current form does not provide a confidence level (p -value) for the effect of an experimental variable. Also, its application to default–state and non-task related fMRI studies would require an alternative model–size selection procedure that does use prediction error as a criterion. Therefore, given its advantages and disadvantages, we believe that it is a tool complementary to other methods that may provide new and different insights into mental processes at the individual and group level.

Acknowledgments

This work was partially funded by NAC grant NIH P41RR13218 and was supported in part by an allocation of computing time from the Ohio Supercomputer Center.

References

- [1] Bellec, P., Rosa-Neto, P., Lyttelton, O.C., Benali, H., Evans, A.C.: Multi-level bootstrap analysis of stable clusters in resting-state fMRI. *Neuroimage* 51(3), 1126–1139 (2010)
- [2] Bishop, C.M.: *Pattern Recognition and Machine Learning*, 1st edn. Springer, Heidelberg (2006); corr. 2nd printing edn. (October 2007)
- [3] Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, G., Penny, W., Ashburner, J.: Bayesian decoding of brain images. *Neuroimage* 39(1), 181–205 (2008)
- [4] Hari, R., Levänen, S., Raij, T.: Timing of human cortical functions during cognition: role of MEG. *Trends Cogn. Sci.* 4(12), 455–462 (2000)
- [5] Kriegeskorte, N., Mur, M., Bandettini, P.: Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst. Neurosci.* 2, 4 (2008)
- [6] Lanterman, A.D.: Schwarz, Wallace, and Rissanen: Intertwining themes in theories of model selection. *Int. Stat. Review* 69(2), 185–212 (2001)
- [7] Lashkari, D., Golland, P.: Exploratory fMRI analysis without spatial normalization. *Inf. Process Med. Imag.* 21, 398–410 (2009)
- [8] Makni, S., Beckmann, C., Smith, S., Woolrich, M.: Bayesian deconvolution of [corrected] fmri data using bilinear dynamical systems. *Neuroimage* 42(4), 1381–1396 (2008)

- [9] Morocz, I., Gross-Tsur, A., von Aster, M., Manor, O., Breznitz, Z., Karni, A., Shalev, R.: Functional magnetic resonance imaging in dyscalculia: preliminary observations. *Ann Neurology* 54(S7), S145 (2003)
- [10] Mourão-Miranda, J., Friston, K.J., Brammer, M.: Dynamic discrimination analysis: A spatial-temporal SVM. *Neuroimage* 36(1), 88–99 (2007)
- [11] O’Toole, A.J., Jiang, F., Abdi, H., Pénard, N., Dunlop, J.P., Parent, M.A.: Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. *J. Cog. Neurosci.* 19(11), 1735–1752 (2007)
- [12] Scott, S.L.: Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Am. Stat. Assoc.* 97(457), 337–351 (2002)
- [13] Stephan, K., Friston, K.: Analyzing effective connectivity with functional magnetic resonance imaging. *WIREs Cog. Sci.* 1, 446–459 (2010)

Functional Brain Imaging with M/EEG Using Structured Sparsity in Time-Frequency Dictionaries

Alexandre Gramfort^{1,2,3}, Daniel Strohmeier⁴, Jens Haueisen^{4,5,6},
Matti Hamalainen³, and Matthieu Kowalski⁷

¹ INRIA, Parietal team, Saclay, France

² LNAO/NeuroSpin, CEA Saclay, Bat. 145, 91191 Gif-sur-Yvette Cedex, France

³ Martinos Center, MGH Dept. of Radiology, Harvard Medical School, Boston, MA

⁴ Inst. of Biomedical Engineering and Informatics, Ilmenau University of Technology, Ilmenau, Germany

⁵ Biomagnetic Center, Dept. of Neurology, University Hospital Jena, Jena, Germany

⁶ Dept. of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia

⁷ Laboratoire des Signaux et Systèmes (L2S), SUPELEC (C-4-20), Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France

Abstract. Magnetoencephalography (MEG) and electroencephalography (EEG) allow functional brain imaging with high temporal resolution. While time-frequency analysis is often used in the field, it is not commonly employed in the context of the ill-posed inverse problem that maps the MEG and EEG measurements to the source space in the brain. In this work, we detail how convex structured sparsity can be exploited to achieve a principled and more accurate functional imaging approach. Importantly, time-frequency dictionaries can capture the non-stationary nature of brain signals and state-of-the-art convex optimization procedures based on proximal operators allow the derivation of a fast estimation algorithm. We compare the accuracy of our new method to recently proposed inverse solvers with help of simulations and analysis of real MEG data.

1 Introduction

Distributed source models in magnetoencephalography and electroencephalography (collectively M/EEG) use the individual anatomy derived from high-resolution anatomical Magnetic Resonance Images (MRI). They employ a dense grid of current dipoles on the automatically segmented cortical surface. Following Maxwell's equations, each dipole adds its contribution linearly to the measured signal leading to a linear solution to the forward problem.

However, the number of sources by far exceeds the number of M/EEG sensors, making the inverse problem ill-posed. Therefore, constraints using a priori knowledge based on the properties of real sources are necessary. Common priors are based on the Frobenius norm. More recently, sparsity-inducing priors such

as the ℓ_1 norm have been introduced to take into account the assumption that only a few brain regions are typically active during a cognitive task.

While wavelet decompositions and time-frequency (TF) analysis are commonly computed from M/EEG data to exhibit transient oscillatory signals, the characteristics of such decompositions are rarely employed as a prior to regularize the inverse problem.

In this contribution, we propose to use both of these *a priori* assumptions within the framework of the inverse problem, making the TF analysis on the sensors optional. To do so, we propose to use a structured prior based on the ℓ_{21} mixed-norm combined with a simple ℓ_1 norm. The prior is imposed on the coefficients of the TF decompositions using Gabor dictionaries.

Notation. We indicate vectors with bold letters, $\mathbf{a} \in \mathbb{R}^N$ (resp. \mathbb{C}^N) and matrices with capital bold letters, $\mathbf{A} \in \mathbb{R}^{N \times N}$ (resp. $\mathbb{C}^{N \times N}$). $\mathbf{a}[i]$ stands for the i^{th} entry in the vector. We denote $\|\mathbf{A}\|_{\text{Fro}}$ the Frobenius norm, $\|\mathbf{A}\|_{\text{Fro}}^2 = \sum_{i,j=1}^N |\mathbf{A}_{ij}|^2$, $\|\mathbf{A}\|_1 = \sum_{i,j=1}^N |\mathbf{A}_{ij}|$ the ℓ_1 norm, and $\|\mathbf{A}\|_{21} = \sum_{i=1}^N \sqrt{\sum_{j=1}^N |\mathbf{A}_{ij}|^2}$ the ℓ_{21} mixed norm. \mathbf{A}^T and $\mathbf{A}^{\mathcal{H}}$ denote a matrix transpose and a Hermitian transpose, respectively.

The inverse problem with time-frequency dictionaries. Given a linear forward operator $\mathbf{G} \in \mathbb{R}^{N \times P}$, also called lead field matrix or gain matrix, where N is the number of sensors and P the number of sources, the measurements $\mathbf{M} \in \mathbb{R}^{N \times T}$ (T number of time instants) are related to the source amplitudes $\mathbf{X} \in \mathbb{R}^{P \times T}$ by $\mathbf{M} = \mathbf{G}\mathbf{X}$.

Solving the forward problem consists of computing \mathbf{G} taking into account the electromagnetic properties of the head [11,10], whereas in the inverse problem one computes a best estimate of the neural currents \mathbf{X}^* based on the measurements \mathbf{M} . However, to accomplish this task, priors need to be imposed on \mathbf{X} . The most conventional prior assumes that its weighted ℓ_2 (Frobenius) norm is small. This corresponds to the family of Minimum-Norm (MN) inverse solvers [11,3]. Several alternative solvers based on ℓ_p norms with $p < 2$ have been also proposed. With $p \leq 1$, such priors promote sparse solutions [17,9]. Such priors however work on an instant by instant basis disregarding the oscillatory and non-stationary nature of electromagnetic brain signals. For this reason such solvers are usually employed following band-pass filtering of the data.

Beyond single instant solvers, various sparsity-promoting approaches have been proposed [20,7,26]. Although, they manage to capture the time courses of the activations, they implicitly assume that all active sources have non-zero activations throughout the analysis period. To go beyond this approach, we propose a solver where the sparsity of source configurations is promoted, but also where the time course of each active dipole is a linear combination of a few Gabor atoms. Our model can thus be expressed as:

$$\mathbf{M} = \mathbf{G}\mathbf{X} + \mathbf{E} = \mathbf{G}\mathbf{Z}\mathbf{\Phi}^{\mathcal{H}} + \mathbf{E} , \tag{1}$$

where $\mathbf{\Phi}^{\mathcal{H}} \in \mathbb{C}^{K \times T}$ is a dictionary of K Gabor atoms, $\mathbf{Z} \in \mathbb{C}^{P \times K}$ are the coefficients of the decomposition, and \mathbf{E} is additive white noise, $\mathbf{E} \sim \mathcal{N}(0, \lambda\mathbf{I})$. Note

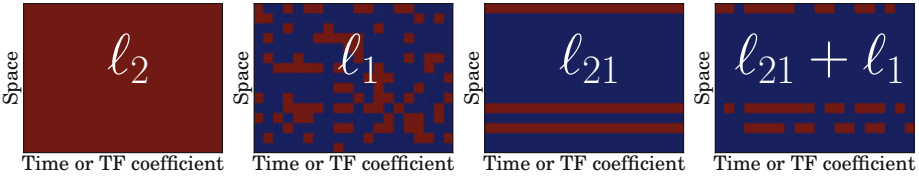


Fig. 1. Sparsity patterns promoted by the different priors (ℓ_2 no non-zero, ℓ_1 scattered and unstructured non-zero, ℓ_{21} block row structure, $\ell_{21} + \ell_1$ block row structure with intra-row sparsity). Red indicates non-zero coefficients.

that this last assumption can be justified for M/EEG data as it is possible to estimate the noise covariance matrix from pre-stimulation recordings and spatially whiten the data. Given a prior on \mathbf{Z} , $\mathcal{P}(\mathbf{Z}) \sim \exp(-\Omega(\mathbf{Z}))$, the maximum a posteriori estimate (MAP) is obtained by solving:

$$\mathbf{Z}^* = \arg \min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{M} - \mathbf{GZ}\Phi^H\|_{\text{Fro}}^2 + \lambda\Omega(\mathbf{Z}) \quad , \quad \lambda > 0 \quad . \quad (2)$$

If we consider $\Omega(\mathbf{Z}) = \|\mathbf{Z}\|_1$, (2) corresponds to a Lasso problem [23], also called Minimum Current Estimate (MCE) in the M/EEG literature [17], where features (or regressors) are spatio-temporal atoms. Similarly to the original formulation of MCE (*i.e.*, with no Φ), such a prior is likely to suffer from inconsistencies over time [20]. Indeed such a norm does not impose a structure for the non-zero coefficients, that are likely to be scattered all over \mathbf{Z}^* (see Fig. 1). Therefore, simple ℓ_1 priors do not guarantee that only a few sources are active during the time window of interest. To promote this, one needs to employ mixed-norms such as the ℓ_{21} norm [20]. By doing so, the estimates have a sparse row structure (see Fig. 1). However the ℓ_{21} prior on \mathbf{Z} does not produce denoised time series as it does not promote source estimates that are formed by a sum of a few Gabor atoms. In order to recover the sparse row structure, while simultaneously promoting sparsity of the decompositions, we propose to use a composite prior formed by the sum of ℓ_{21} and ℓ_1 norms. The prior then reads:

$$\Omega(\mathbf{Z}) = \rho\|\mathbf{Z}\|_1 + (1 - \rho)\|\mathbf{Z}\|_{21} \quad , \quad 0 < \rho < 1 \quad . \quad (3)$$

Gabor dictionaries. Here we briefly present some important properties of Gabor dictionaries (see [4] for more details). Given a signal observed over a time interval, its conventional Fourier transform estimates the frequency content but loses the time information. To analyze the evolution of the spectrum with time and hence the non-stationarity of the signal, Gabor introduced the windowed Fourier atoms which correspond to a short-time Fourier transform (STFT) with a gaussian window. In practice, for numerical computation, a challenge is to properly discretize the continuous STFT. The discrete version of the STFT is called the Gabor Transform. The setting we are considering is the finite dimensional one. Let $\mathbf{g} \in \mathbb{R}^T$ be a “mother” analysis window. Let $f_0 \in \mathbb{N}$ and $k_0 \in \mathbb{N}$ be the frequency

and the time sampling rate of the time-frequency plane generated by the STFT, respectively. The family of the translations and modulations of the mother window generates a family of Gabor atoms $(\phi_{mf})_{mf}$ forming the dictionary $\Phi \in \mathbb{C}^{T \times K}$. We denote the number of atoms by K . The atoms can be written

$$\phi_{mf}[n] = \mathbf{g}[n - mk_0]e^{\frac{i2\pi f_0fn}{T}}, \quad m \in \{0, \dots, \frac{T}{k_0} - 1\}, f \in \{0, \dots, \frac{T}{f_0} - 1\} .$$

If the product f_0k_0 is small enough, *i.e.*, the time-frequency plane is sufficiently sampled, the family $(\phi_{mf})_{mf}$ is a frame of \mathbb{R}^T , *i.e.*, one can recover any signal $\mathbf{x} \in \mathbb{R}^T$ from its Gabor coefficients $(\langle \mathbf{x}, \phi_{mf} \rangle) = \Phi^H \mathbf{x}$. For the rest of the paper we assume that this condition is satisfied.

More precisely, there exists two constants $A, B > 0$ such that

$$A\|\mathbf{x}\|_2^2 \leq \sum_{m,f} \langle \mathbf{x}, \phi_{mf} \rangle \leq B\|\mathbf{x}\|_2^2 . \tag{4}$$

When $A = B$, the frame is *tight*, and if $A = B = 1$ then the frame is an orthogonal basis. The Balian-Low theorem says that it is impossible to construct a Gabor frame which is a basis. Consequently, a Gabor transform is redundant or overcomplete and there exists an infinitely number of ways to reconstruct \mathbf{x} from a given family of Gabor atoms. In the following, Φ is a frame.

The canonical reconstruction of \mathbf{x} from its Gabor coefficients requires a canonical dual window, denoted by $\tilde{\mathbf{g}}$. Following (4) to define $(\tilde{\phi}_{mf})_{mf}$ we have:

$$\mathbf{x} = \sum_{m,f} \langle \mathbf{x}, \phi_{mf} \rangle \tilde{\phi}_{mf} = \sum_{m,f} \langle \mathbf{x}, \tilde{\phi}_{mf} \rangle \phi_{mf} = \Phi^H \mathbf{x} \tilde{\Phi} = \tilde{\Phi}^H \mathbf{x} \Phi ,$$

where $\tilde{\Phi}$ is the Gabor dictionary formed with the dual windows. When the frame is tight, then we have $\tilde{\mathbf{g}} = \mathbf{g}$, and more particularly we have $\Phi \Phi^H = \|\Phi\| \mathbf{Id}$. The representation being redundant, for any $\mathbf{x} \in \mathbb{R}^T$ one can find a set of coefficients z_{mf} such that $\mathbf{x} = \sum_{m,f} z_{mf} \phi_{mf}$, while the z_{mf} verify some suitable properties dictated by the application. For example, it is particularly interesting for M/EEG to find a sparse representation of the signal.

In practice, the Gabor coefficients are computed using the Fast Fourier Transform (FFT). The synthesis operation (and then, the inverse transform with the appropriate window) is accomplished with the inverse FFT and overlap-add techniques. Such analysis and synthesis operations are efficiently implemented in the LTFAT Matlab toolbox² [22].

Related work. Time-frequency analysis is commonly used in the context of M/EEG both in the sensor and source space, but rarely integrated with the solution of the inverse problem. Some earlier contributions, such as [5,8,16], apply a 2-step approach. First TF atoms are estimated from sensor data, typically

¹ We can however say nothing about $\Phi^H \Phi$ in general.

² <http://ltfat.sourceforge.net/>

with greedy approaches like Matching Pursuit. Subsequently, the inverse problem is solved on the selected components using parametric [8], scanning [16] or distributed methods [5]. Such methods suffer from several limitations. They implicitly assume that the source waveforms correspond to single TF atoms, while real brain signals are better modeled by a combination of atoms. In addition, estimation errors made in the first step have a direct impact on the accuracy of the source estimates. This is a particularly critical issue since the first step does not take into account the biophysics of the problem, *i.e.*, the solution of the forward problem.

Spatial sparsity of source configurations has also been a recurrent assumption to improve the resolution of the M/EEG inverse problem. Recently, priors based on the ℓ_{21} mixed-norm have been proposed to achieve rotation invariance [12] and to recover spatially sparse while temporally smooth solutions [20]. However, in [20], a part of the temporal smoothness is obtained by filtering the data and by using temporal basis functions obtained with an SVD. Alternatively, a sparsity-inducing Bayesian formulation of the inverse problem has been proposed [7,26]. However, these approaches make the strong assumption that the source time courses are stationary. For example, the estimation crucially depends on the time interval considered. Also the solutions obtained by these solvers are invariant with respect to the permutation of the columns of \mathbf{M} , *i.e.*, the temporal sequence of the data is immaterial.

In [24], an inverse solver that models the transient and non-stationary responses in M/EEG is proposed. A probabilistic model with wavelet shrinkage is employed to promote spatially smooth time courses. The estimation however relies on model approximations with no guarantee on the solution obtained. The most related work to ours, beyond the field of M/EEG, is probably [18] where sparsity is also promoted on the TF decompositions. The related optimization problem, is however solved with a truncated Newton method which only applies to differentiable problems. The non-differentiability of the cost function is tackled by using smooth approximation in the minimization. Moreover, Newton methods are known to be fast in the neighborhood of the solution, but little is known about the global convergence rate. In [19], it is proved that a suitable Newton technique has the same rate of convergence as the accelerated first order schemes like one we are employing below.

In this contribution, we do not address the problem of learning spatial basis functions such as in [25,2] as doing so makes the cost function non-convex which affects the speed of convergence of the algorithm and also makes the solvers dependent on the initialization.

2 Optimization Strategy

The procedure we propose is based on first-order schemes that handle the optimization of any cost function \mathcal{F} if it can be written as a sum of two terms: 1 smooth convex term f_1 with Lipschitz gradient and 1 convex term f_2 , potentially non-differentiable [1]: $\mathcal{F}(\mathbf{Z}) = f_1(\mathbf{Z}) + f_2(\mathbf{Z})$. The cost function in (2) belongs

to this category. However, we need to be able to compute the proximal operator associated to f_2 .

Definition 1 (Proximity operator). *Let $\varphi : \mathbb{R}^M \rightarrow \mathbb{R}$ be a proper convex function. The proximity operator associated to φ , denoted by $\text{prox}_\varphi : \mathbb{R}^M \rightarrow \mathbb{R}^M$ reads:*

$$\text{prox}_\varphi(\mathbf{Z}) = \arg \min_{\mathbf{V} \in \mathbb{R}^M} \frac{1}{2} \|\mathbf{Z} - \mathbf{V}\|_2^2 + \varphi(\mathbf{V}) .$$

In the case of the composite prior in (3), the proximity operator is given by the following lemma.

Lemma 1 (Proximity operator for $\ell_{21} + \ell_1$). *Let $\mathbf{Y} \in \mathbb{C}^{P \times K}$ be indexed by a double index (p, k) . $\mathbf{Z} = \text{prox}_{\lambda(\rho)\|\cdot\|_1 + (1-\rho)\|\cdot\|_{21}}(\mathbf{Y}) \in \mathbb{C}^{P \times K}$ is given for each coordinates (p, k) by*

$$Z_{p,k} = \frac{Y_{p,k}}{|Y_{p,k}|} (|Y_{p,k}| - \lambda\rho)^+ \left(1 - \frac{\lambda(1-\rho)}{\sqrt{\sum_k (|Y_{p,k}| - \lambda\rho)^2}} \right)^+ .$$

where for $x \in \mathbb{R}$, $(x)^+ = \max(x, 0)$, and by convention $\frac{0}{0} = 0$.

This result is a corollary of the proximity operator derived for hierarchical group penalties recently proposed in [15]. The penalty described here can indeed be seen as a 2-level hierarchical structure, and the resulting proximity operator reduces to successively applying the ℓ_{21} proximity operator then the ℓ_1 one.

The pseudo code is provided in Algorithm 1. The Lipschitz constant \mathcal{L} of the gradient of the smooth term in (2) is given by the square of the spectral norm of the linear operator $\mathbf{Z} \rightarrow \mathbf{G}\mathbf{Z}\Phi^{\mathcal{H}}$. We estimate it with the power iteration method.

Algorithm 1. FISTA with TF Dictionaries

Input: Measurements \mathbf{M} , lead field matrix \mathbf{G} , regularization parameter $\lambda > 0$ and I the number of iterations.

Output: \mathbf{Z}^*

- 1: Auxiliary variables : \mathbf{Y} and $\mathbf{Z}_o \in \mathbb{R}^{P \times K}$, and τ and $\tau_o \in \mathbb{R}$.
 - 2: Estimate the Lipschitz constant \mathcal{L} with the power iteration method.
 - 3: $\mathbf{Y} = \mathbf{Z}^* = \mathbf{Z}$, $\tau = 1$, $0 < \mu < \mathcal{L}^{-1}$
 - 4: **for** $i = 1$ to I **do**
 - 5: $\mathbf{Z}_o = \mathbf{Z}^*$
 - 6: $\mathbf{Z}^* = \text{prox}_{\mu\lambda\Omega} (Y + \mu\mathbf{G}^T(\mathbf{M} - \mathbf{G}\mathbf{Y}\Phi^{\mathcal{H}})\Phi)$
 - 7: $\tau_o = \tau$
 - 8: $\tau = \frac{1 + \sqrt{1 + 4\tau^2}}{2}$
 - 9: $\mathbf{Y} = \mathbf{Z}^* + \frac{\tau_o - 1}{\tau}(\mathbf{Z}^* - \mathbf{Z}_o)$
 - 10: **end for**
-

Source models with unconstrained orientations. When the source orientations given by the normals of the cortical mesh cannot be trusted, it can be interesting to relax this constraint by placing three orthogonal sources at each spatial location. However, the TF composite prior also needs to be adapted. Assuming that each source is indexed by a spatial location i and an orientation $o \in \{1, 2, 3\}$, the ℓ_1 and ℓ_{21} norms read:

$$\|\mathbf{Z}\|_1 = \sum_{ik} \sqrt{\sum_{o=1}^3 |\mathbf{Z}[(i, o), k]|^2} \quad \text{and} \quad \|\mathbf{Z}\|_{21} = \sum_i \sqrt{\sum_{ko} |\mathbf{Z}[(i, o), k]|^2},$$

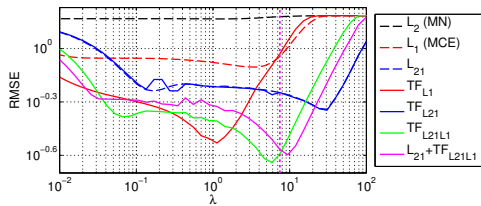
where k indexes the TF coefficients. It amounts to grouping the orientations in a common ℓ_2 norm such as in [20,12].

Implementation. Algorithm 1 requires to compute Gabor transforms at each iteration which can be computationally demanding. However, due to the ℓ_{21} sparsity inducing prior, only a few rows of \mathbf{Z} have non-zeros coefficients. The Gabor transform is therefore computed for only a limited number of rows, equivalently a small number of active sources. This makes the computation of $\mathbf{Y}\Phi^t$ (cf. Algorithm 1 line 6) much faster.

In order to significantly reduce the computation time of sparse regression problems, as the one presented here, a common strategy in machine learning is to use an *active-set* approach [21]. Intuitively, if one can verify the optimality of a solution (typically with Karush-Khun-Tucker (KKT) conditions), one can start by solving a small subproblem and then check if the solution obtained is optimal for the full problem. In the context of M/EEG, it consists in solving the inverse problem with a small set of sources, assuming the others have zero activation. This is particularly interesting when processing real M/EEG data for which P can be up to 30000, whereas only at most a few hundred sources are likely to be active. Whereas KKT optimality conditions can be efficiently verified with a ℓ_{21} penalty, it is not the case anymore with (3). To limit the computation time, we propose to address the problem in two steps. In a first step, the inverse problem is solved with a ℓ_{21} prior, using a small value for λ and an active set strategy. Using a small λ makes the active set larger than necessary, so that the active sources form a subset of it. Then (2) is solved with the composite prior on the restricted source space. By doing so, the computation on real data, such as those presented in Section 3.2, takes a few minutes on a standard laptop computer for given values of the regularization parameters.

Model selection. Model selection amounts to setting the parameters λ and ρ . As a principled way to do this, we use a k -fold cross-validation (CV) procedure in which the signal of sensors left out is predicted from the solution estimated using the remaining sensors. The best parameters are the ones that lead to the smallest average root mean square error (RMSE) between measurements and predictions across the folds. In practice, we use a 4-fold CV with a logarithmic grid of 40 values for λ . To limit computation, we employed a fixed $\rho = 0.1$, since our experiments proved it to be a good default choice.

Fig. 2. Comparison of RMSE in the source space as a function of λ (SNR=6dB). Dashed lines correspond to results with no TF. TF priors improve the reconstruction and the best accuracy is obtained with the TF $\ell_{21} + \ell_1$ prior. The 2-steps approach gives an almost optimal accuracy. The vertical dashed line gives the λ estimated by CV.



3 Results

In the following, we first evaluate the accuracy of our solver on toy data and a simulated EEG dataset. We then apply our solver to experimental MEG data.

3.1 Simulation Study

In order to have a reproducible and reasonably fast comparison of the different priors, we generated a small artificial dataset with 20 electrodes and 200 sources. 4 of these sources were randomly selected to be active. The ECD waveforms (cf. Fig. 3(a)) represent 1 high and 3 low frequency components. The time course of the oscillatory high frequency component is modeled by a Gabor atom, whereas the time courses of the low frequency components were obtained from a somatosensory evoked potential study [14] by fitting manually ECDs to the P15, N20 and P23 components. To make the comparison of the priors independent of the forward model and the sources spatial configuration, the linear forward operator was a random matrix, whose columns are normalized to 1. White gaussian noise was added to the signals to achieve a desired signal-to-noise ratio (SNR). Following the notation of (11), we define SNR as $20 \log_{10}(\|M\|_{\text{Fro}}/\|E\|_{\text{Fro}})$.

Figure 2 presents the RMSE on the estimation for different solvers as a function of λ ($\text{RMSE} = \|\mathbf{X}_{\text{sim}} - \mathbf{X}_{\Omega}^*\|_{\text{Fro}}^2$). λ was chosen on a logarithmic grid from 10^{-2} to 10^2 and ρ was fixed to 0.1. The Gabor dictionary is tight, constructed with a 128 samples long window \mathbf{g} with $k_0 = 4$ samples time shift and $f_0 = 1$ sample frequency shift. Results show that the composite TF prior outperforms the other priors, while the 2-steps approach gives an almost optimal accuracy with a λ estimated by CV. Figure 3 shows the reconstructions for the best λ according to Fig. 2 for the ℓ_1 , ℓ_{21} and the TF composite priors. It can be observed, that the inverse method with the composite TF prior is able to reconstruct the smooth time course of the simulated sources contrary to ℓ_1 and ℓ_{21} priors.

The TF composite prior was then challenged on a realistic EEG configuration with a 4-shell spherical head model (radii 0.90, 0.92, 0.96 and 1) and 60 electrodes placed according to the international 10-5 electrode system. The source waveforms were the same as before. The source space in Fig. 4 consisted of 152 sources in a regular grid (0.2 spacing) inside the innermost sphere. Source orientations were randomly selected. For depth compensation, a source covariance

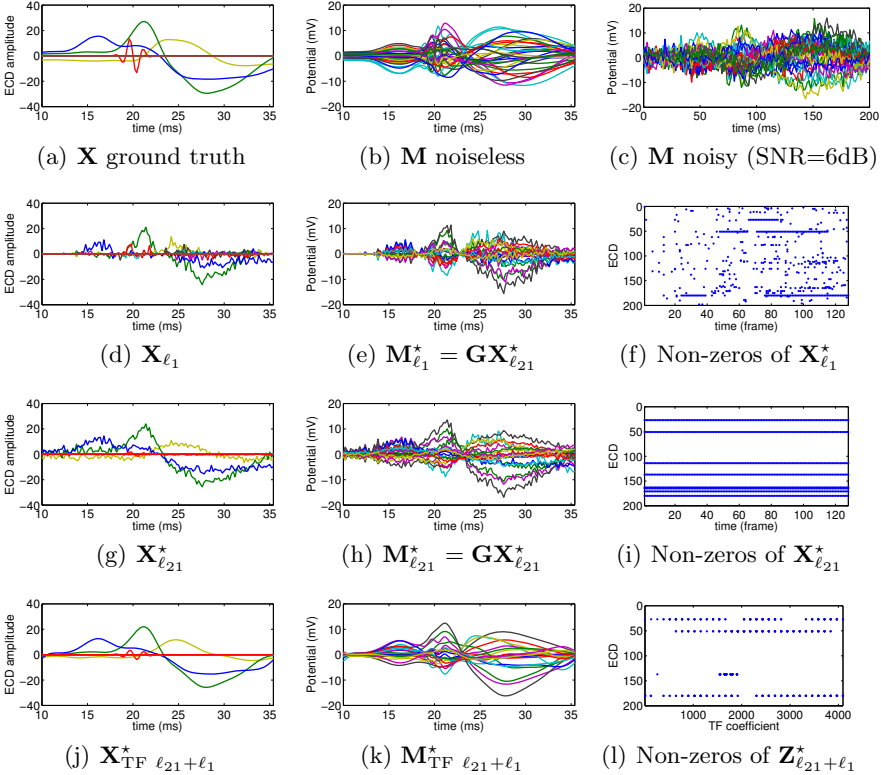


Fig. 3. Simulations results with $SNR = 6\text{ dB}$. (a) simulated source activations. (b) Noiseless simulated measurements. (c) Simulated measurements corrupted by noise. (d-e-f) Estimation with ℓ_1 prior. (g-h-i) Estimation with ℓ_{21} prior [20]. (j-k-l) Estimation with composite TF prior. (f-i-l) show the sparsity patterns obtained by the 3 different priors as explained in Fig. 1. Result (j) shows how the composite TF prior improves over (d) and (g). (l) presents also a higher level of sparsity compared to (f) and (i).

based weighting method described in [13] was applied. Fig. 4 shows the head model and reconstructions obtained with ℓ_{21} and the TF $\ell_{21} + \ell_1$ priors. Even if the performance drops down due to the limited spatial resolution of EEG, the TF composite prior gives the best RMSE and is able to reconstruct and separate the high frequency component.

3.2 Experimental Results with MEG Data

We also applied our method to somatosensory MEG data. In this experiment, the right median-nerve was stimulated at the wrist with 0.2 ms constant current pulses above the motor threshold. The inter-stimulus interval was random between 3 - 12 s in an event-related design. MEG data were acquired using a 306-channel Neuromag Vectorview system. The signals were recorded with a bandpass of

Fig. 4. Results with real EEG lead field (SNR=3dB). The 4 dipoles are color coded. Magenta dots show the 3D grid of sources. Dark dots show the EEG sensors locations. Contrary to TF $\ell_{21} + \ell_1$, ℓ_{21} fails to recover the deep green dipole time course.

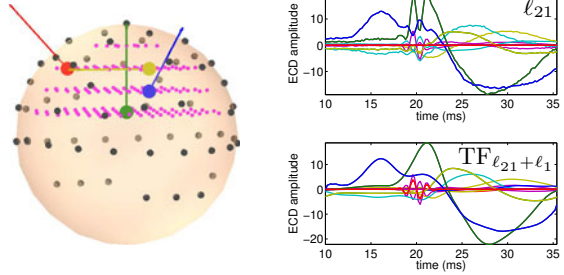
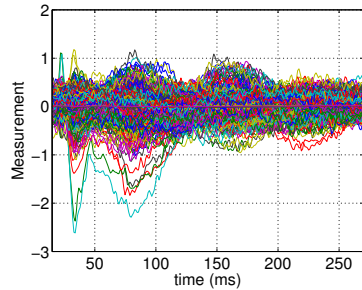
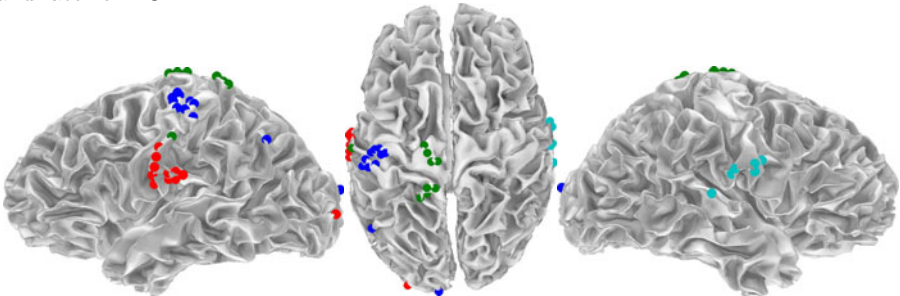


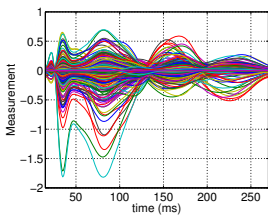
Fig. 5. Results obtained with the TF composite prior $\ell_{21} + \ell_1$ applied to somatosensory MEG data. Estimation was performed with unconstrained orientations on a set of 8195 cortical locations ($\mathbf{G} \in \mathbb{R}^{306 \times 24585}$). Estimation leads to 37 active brain locations that have been clustered into 4 groups matching known functional regions. Clustering was done using k-means based on the source activation time courses. (e) illustrates the cascade of activation starting from cS1, to both S2 cortices and later cPPC.



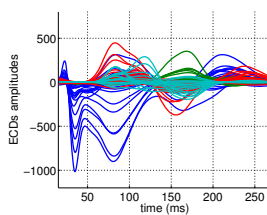
(a) MEG data: $\mathbf{M} \in \mathbb{R}^{306 \times 256}$



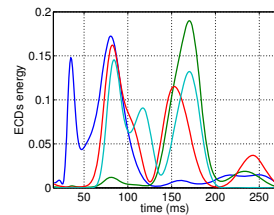
(b) Locations of the estimated active ECDs. ECDs clusters are color coded: Dark blue matches cS1, red cS2, light blue iS2, green cPPC.



(c) $\mathbf{M}^* = \mathbf{G}\mathbf{X}^*$ (denoised sensors data)



(d) $\mathbf{X}^* \in \mathbb{R}^{24585 \times 256}$



(e) Average energy density over time in each cluster.

0.01 - 250 Hz, digitized at 1004 samples/s and averaged offline triggered by the stimulus onset. All epochs containing EOG signals higher than 150 μV peak-to-peak amplitude were discarded from the averages, resulting in 68 averaged epochs. For source estimation, the noise-covariance matrix was estimated from the baseline period of 200 ms before stimulus onset in the raw data. The sources were estimated assuming unconstrained orientations. The Gabor dictionary is tight, constructed with a 256 samples ($\simeq 256$ ms) long window \mathbf{g} with $k_0 = 16$ samples time shift and $f_0 = 1$ sample frequency shift. Results are presented in Fig. 5.

The first activation of the contralateral primary somatosensory cortex (cS1) peaks around 20 ms and lasts up to 100 ms; then the secondary somatosensory cortices (contralateral cS2, ipsilateral iS2) activate around 70 ms and lasts up to 200 ms. The posterior parietal cortex (PPC) starts to activate at 70 ms with a more significant activation between 140 and 200 ms. This is consistent with the understanding of PPC, also known as the parietal association area, which is known to be higher in the hierarchy of cognitive processing [6].

4 Conclusions

In this work, we showed how physiologically motivated priors for brain activations can be accounted for in a mathematically principled framework in M/EEG source analysis. Using a composite prior, the sparsity of spatial patterns, the temporal smoothness, and the non-stationarity of the source signals were well recovered. Thanks to the structure of the cost function considered, mainly its convexity, an efficient optimization strategy was proposed. The problem being convex, the solver is not affected by improper initialization and cannot be trapped in local minima. Simulations indicated benefits of the approach over alternative solvers, while results with well understood MEG data confirm the accuracy of the reconstruction with real signals. Both results show that our solver is a promising new approach for mining M/EEG data.

Further work will investigate the impact of the choice of the time-frequency dictionary and the compromise between time and frequency resolution of the Gabor atoms, leading eventually to the use of a union of dictionaries.

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* 2(1), 183–202 (2009)
2. Bolstad, A., Veen, B.V., Nowak, R.: Space-time event sparse penalization for magneto-/electroencephalography. *NeuroImage* 46(4), 1066–1081 (2009)
3. Dale, A., Liu, A., Fischl, B., Buckner, R.: Dynamic statistical parametric neurotechnique mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67 (2000)
4. Daubechies, I.: Ten lectures on Wavelets. *SIAM-CBMS Conferences Series* (1992)
5. Durka, P.J., Matysiak, A., Montes, E.M., Valdés-Sosa, P., Blinowska, K.J.: Multichannel matching pursuit and EEG inverse solutions. *Journal of Neuroscience Methods* 148(1), 49–59 (2005)

6. Forss, N., Hari, R., Salmelin, R., Ahonen, A., Hamalainen, M., Kajola, M., Knuutila, J., Simola, J.: Activation of the human posterior parietal cortex by median nerve stimulation. *Exp. Brain Res.* (99), 309–315 (1994)
7. Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J.: Multiple sparse priors for the M/EEG inverse problem. *Neuroimage* 39(3), 1104–1120 (2008)
8. Geva, A.B.: Spatio-temporal matching pursuit (SToMP) for multiple source estimation of evoked potentials. In: *Electrical and Electronics Eng.* pp. 113–116 (1996)
9. Gorodnitsky, I., George, J., Rao, B.: Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and Clinical Neurophysiology* (January 1995)
10. Gramfort, A., Papadopoulos, T., Olivi, E., Clerc, M.: OpenMEEG: opensource software for quasistatic bioelectromagnetics. *BioMed Eng. OnLine* 9(1), 45 (2010)
11. Hämäläinen, M., Ilmoniemi, R.: Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32(1), 35–42 (1994)
12. Haufe, S., Nikulin, V.V., Ziehe, A., Müller, K.R., Nolte, G.: Combining sparsity and rotational invariance in EEG/MEG source reconstruction. *NeuroImage* 42(2), 726–738 (2008)
13. Haufe, S., Tomioka, R., Dickhaus, T., Sannelli, C., Blankertz, B., Nolte, G., Müller, K.R.: Large-scale EEG/MEG source localization with spatial flexibility. *NeuroImage* 54(2), 851–859 (2011)
14. Jaros, U., Hilgenfeld, B., Lau, S., Curio, G., Hauelsen, J.: Nonlinear interactions of high-frequency oscillations in the human somatosensory system. *Clin. Neurophysiol.* 119(11), 2647–2657 (2008)
15. Jenatton, R., Mairal, J., Obozinski, G., Bach, F.: Proximal methods for hierarchical sparse coding. In: *ICML* (2010)
16. Lelic, D., Gratkowski, M., Valeriani, M., Arendt-Nielsen, L., Drewes, A.M.: Inverse modeling on decomposed electroencephalographic data: A way forward? *Journal of Clinical Neurophysiology* 26(4), 227–235 (2009)
17. Matsuura, K., Okabe, Y.: Selective minimum-norm solution of the biomagnetic inverse problem. *IEEE Trans. Biomed. Eng.* 42(6), 608–615 (1995)
18. Model, D., Zibulevsky, M.: Signal reconstruction in sensor arrays using sparse representations. *Signal Processing* 86(3), 624–638 (2006)
19. Nesterov, Y., Polyak, B.: Cubic regularization of newton’s method and its global performance. *Mathematical Programming* 108(1), 177–205 (2006)
20. Ou, W., Hämäläinen, M., Golland, P.: A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* 44(3), 932–946 (2009)
21. Roth, V., Fischer, B.: The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In: *ICML*, pp. 848–855 (2008)
22. Soendergard, P., Torrésani, B., Balazs, P.: The linear time frequency toolbox. Tech. rep., Technical University of Denmark (2009)
23. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Serie B* 58(1), 267–288 (1996)
24. Trujillo-Barreto, N.J., Aubert-Vázquez, E., Penny, W.D.: Bayesian M/EEG source reconstruction with spatio-temporal priors. *Neuroimage* 39(1), 318–335 (2008)
25. Valdés-Sosa, P.A., Vega-Hernández, M., Sánchez-Bornot, J.M., Martínez-Montes, E., Bobes, M.A.: EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. *HBM* 30(6), 1898–1910 (2009)
26. Wipf, D., Nagarajan, S.: A unified Bayesian framework for MEG/EEG source imaging. *Neuroimage* 44(3), 947–966 (2009)

Generalized Sparse Regularization with Application to fMRI Brain Decoding

Bernard Ng and Rafeef Abugharbieh

Biomedical Signal and Image Computing Lab, UBC, Canada
bernardyng@gmail.com

Abstract. Many current medical image analysis problems involve learning thousands or even millions of model parameters from extremely few samples. Employing sparse models provides an effective means for handling the curse of dimensionality, but other propitious properties beyond sparsity are typically not modeled. In this paper, we propose a simple approach, generalized sparse regularization (GSR), for incorporating domain-specific knowledge into a wide range of sparse linear models, such as the LASSO and group LASSO regression models. We demonstrate the power of GSR by building anatomically-informed sparse classifiers that additionally model the intrinsic spatiotemporal characteristics of brain activity for fMRI classification. We validate on real data and show how prior-informed sparse classifiers outperform standard classifiers, such as SVM and a number of sparse linear classifiers, both in terms of prediction accuracy and result interpretability. Our results illustrate the added-value in facilitating flexible integration of prior knowledge beyond sparsity in large-scale model learning problems.

Keywords: brain decoding, fMRI classification, prior-informed learning, sparse optimization, spatiotemporal regularization.

1 Introduction

Recent years witnessed a surging interest in exploiting sparsity [1-9] to handle the ever-increasing scale and complexity of current medical image analysis problems [10,11]. Oftentimes, one is faced with exceedingly more predictors than samples. Under such ill-conditioned settings, incorporating sparsity into model learning proved to be of enormous benefits. In particular, enforcing sparsity enables model parameters associated with irrelevant predictors to be implicitly removed, i.e. shrunk to exactly zero [1], which reduces overfitting thus enhances model generalizability. Learning parsimonious models by imposing sparsity also simplifies result interpretation [1], which is of utmost importance in most medical studies.

Since the advent of the least absolute shrinkage and selection operator (LASSO) regression model [1], where Tibshirani showed that penalizing the l_1 norm induces sparsity in the regression coefficients, numerous powerful variants were subsequently proposed [2-9]. Zou and Hastie, for instance, proposed the elastic net penalty [2], which retains the sparse property of LASSO but additionally encourages correlated predictors to be jointly selected in a data-driven manner. For applications where a

natural grouping of the predictors exists, Yuan and Lin proposed the group LASSO penalty [3], which sparsely selects subsets of predefined groups with non-zero weights assigned to all predictors within the selected groups. To re-enable predictor-level sparsity back into group LASSO, Sprechmann et al. proposed combining the LASSO and group LASSO penalties under the name, hierarchical LASSO [4], also known as sparse group LASSO [5]. Other extensions include collaborative hierarchical LASSO [4] and overlapped group LASSO [6] among many others. All these models provide effective means for imposing structural constraints on the model parameters [4], but lack the flexibility for incorporating potentially advantageous problem-specific properties beyond sparsity [7]. For example, adjacent pixels in an image are typically correlated. Model parameters associated with these pixels should thus be assigned similar magnitudes to reflect the underlying correlations. However, merely enforcing sparsity does not model such associations between predictors.

To encourage smoothness in model parameters, in addition to sparsity, Tibshirani et al. proposed the fused LASSO model [8], which combines the LASSO penalty with a term that penalizes the l_1 norm of the differences between model parameters of adjacent predictors. To extend beyond smoothness, Tibshirani and Taylor recently proposed the Generalized LASSO model [7], which penalizes the l_1 norm of a weighted combination of the model parameters. By varying the choice of weights, Generalized LASSO facilitates numerous applications, such as trend filtering, wavelet smoothing, and outlier detection [7]. In a previous work [9], we proposed an extension of LASSO that also enables such modeling flexibility but is much simpler to optimize. Specifically, we proposed adding a generalized ridge penalty (l_2 norm of a weighted combination of the model parameters) to the LASSO regression model and showed that the resulting optimization problem can be efficiently minimized with existing LASSO solvers [12-15].

In this paper, we propose a simple yet effective approach for incorporating prior knowledge into a wide collection of sparse linear models, as motivated by our previous work [9]. We refer to our approach as generalized sparse regularization (GSR). In contrast to [7] and [9], we show that GSR is applicable to a much broader set of sparse linear models than just the LASSO regression model. The adaptability of GSR to such a wide range of models stems from how any l_2 norm penalty can be merged into an l_2 data fitting loss through a simple augmentation trick. Thus, adding a generalized ridge penalty to any sparse linear models with an l_2 data fitting loss, as commonly used in regression models [1-9], enables problem-specific properties to be easily integrated while preserving the functional form of the original optimization problem that these models entail. This desirable property of GSR facilitates the direct deployment of a wealth of existing sparse optimizers [12-15].

To demonstrate the power of GSR, we apply GSR to a large-scale functional magnetic resonance imaging (fMRI) classification problem, where only tens of samples are available for training a classifier with several tens of thousands of coefficients. Recent research in this area of fMRI analysis has mainly focused on exploitation of sparsity-enforcing techniques, such as sparse logistic regression [16,17], elastic net [18], and group LASSO [19] among others [20], to mitigate the curse of dimensionality. However, merely enforcing sparsity does not promote spatial smoothness in classifier weight patterns [9], which deviates from how spatially proximal voxels tend to display similar level of brain activity [21]. We previously proposed remedying this

limitation by modeling spatial correlations using a generalized ridge penalty [9]. Recently, van Gerven et al. proposed a Bayesian formulation for incorporating a spatiotemporal prior, where the authors opted to model uncertainty by estimating the posterior probabilities of the classifier weights as opposed to obtaining sparse weightings through finding the maximum a posterior solution. [22]. In this work, we model other characteristics of brain activity, in addition to spatial correlations, by exploiting the flexibility of GSR. In particular, we apply GSR to build anatomically-informed sparse classifiers that simultaneously model the intrinsic spatiotemporal structure in fMRI data, and explore whether incorporating such additional prior knowledge can further enhance prediction accuracy and result interpretation.

2 Methods

2.1 Overview of Sparse Linear Models

In this section, we focus on the problem of regression, since most sparse linear models [1-9] are inherently designed for such application, and defer discussion of how the sparse linear models described in this section can be employed for classifier learning in Section 2.3. Consider the standard least square regression problem:

$$\hat{a} = \min_a \|y - Xa\|_2^2, \quad (1)$$

where y is an $N \times 1$ response vector, X is an $N \times M$ predictor matrix, a is an $M \times 1$ coefficient vector, N is the number of samples, and M is the number of predictors. The closed-form solution of (1) is given by:

$$\hat{a} = (X^T X)^{-1} X^T y. \quad (2)$$

When $N \ll M$, which is typical in many medical imaging problems [10,11], $(X^T X)^{-1}$ is ill-conditioned. Thus, direct estimation of \hat{a} using (2) usually results in overfitting. To obtain a more generalizable estimate of \hat{a} , a common strategy is to employ regularization. In particular, Tibshirani proposed enforcing sparsity on a to achieve the dual objective of reducing overfitting and enhancing interpretability [1]:

$$\hat{a} = \min_a \|y - Xa\|_2^2 + \alpha \|a\|_1, \quad (3)$$

where $\alpha \geq 0$. The model (3) is commonly referred to as the LASSO regression model, where Tibshirani showed that penalizing the l_1 norm induces sparsity on a [1].

The success of the LASSO regression model in numerous applications resulted in an explosion of research on sparse linear models [14]. Although many LASSO variants involve only a simple addition of other penalty terms to (3), the modeling power that these extensions facilitate proved substantial. For example, Zou and Hastie proposed adding a ridge penalty to (3), which is known as the elastic net model [2]:

$$\hat{a} = \min_a \|y - Xa\|_2^2 + \alpha \|a\|_1 + \beta \|a\|_2^2, \quad (4)$$

where $\beta \geq 0$. This model has two key advantages over LASSO. First, in contrast to LASSO, the number of non-zero elements in a is no longer limited by the number of samples [2]. This property is especially important in medical imaging applications, since the number of samples is often much smaller than the number of predictors. Second, in cases where the predictors are correlated, elastic net tends to jointly select the correlated predictors, whereas LASSO would arbitrarily select only one predictor among each correlated set [2].

Another widely-used extension of LASSO is the group LASSO, proposed by Yuan and Lin for applications where a natural grouping of the predictors is present [3]:

$$\hat{a} = \min_a \|y - Xa\|_2^2 + \alpha \|a_g\|_{2,1}, \quad (5)$$

$$\|a_g\|_{2,1} = \sum_{h=1}^H \|a_{g_h}\|_2, \quad (6)$$

where a_{g_h} are the coefficients associated with predictors belonging to the predefined group g_h , and $h \in \{1, \dots, H\}$ where H is the number of predefined groups [3]. As evident from (6), $\|a_g\|_{2,1}$ is exactly the l_1 norm of $\|a_{g_h}\|_2$. Thus, minimizing $\|a_g\|_{2,1}$ encourages sparse subsets of groups to be selected with non-zero coefficients assigned to all predictors within each selected group. However, since not all predictors within a group are necessarily relevant, Sprechmann et al. [4] proposed the hierarchical LASSO model, also known as sparse group LASSO [5]:

$$\hat{a} = \min_a \|y - Xa\|_2^2 + \alpha \|a_g\|_{2,1} + \beta \|a\|_1, \quad (7)$$

which is essentially group LASSO combined with a LASSO penalty. Reintroducing the LASSO penalty encourages internal sparsity such that only a sparse subset of predictors within each selected group is chosen [4]. This property of (7) is particularly useful in cases where one is uncertain about the exact group assignment.

To extend (7) to scenarios where multiple sets of samples, y^s , are generated from different processes, e.g. when observations are collected from multiple subjects [11], but associated with the same set of predictors, e.g. using the same set of semantic features to predict brain activation patterns of different subjects [11], Sprechmann et al. [4] proposed the collaborative hierarchical LASSO model:

$$\hat{a} = \min_a \|Y - XA\|_F^2 + \alpha \sum_{k=1}^K \|A_{g_h}\|_F + \beta \sum_{s=1}^S \|a^s\|_1, \quad (8)$$

where S is the number of processes, $Y = (y^1, \dots, y^S)$, $A = (a^1, \dots, a^S)$, A_{g_h} are the rows of A belonging to group g_h , and a^s is the coefficient vector of process s . $\|\cdot\|_F$ denotes the Frobenius norm. Minimizing (8) promotes the same groups to be selected across processes with the chosen sparse subset of predictors within each selected group being potentially different [4]. This model thus enables information across processes to be shared while providing flexibility in handling inter-process variability by allowing sparsity patterns within the selected groups to vary across processes.

All of the LASSO extensions above, as well as others not discussed in this paper due to space limitations, provide effective means for modeling data structure through

promoting groups of predictors to be jointly selected. However, the associations between predictors are ignored in these models, i.e. associated predictors may be jointly selected but the coefficients assigned to these predictors can greatly vary. Since the associations between predictors can be an important source of information for further constraining the ill-conditioned problem (1), we propose a simple approach to incorporate predictor associations into all of the models above, as discussed next.

2.2 Generalized Sparse Regularization

To integrate properties beyond sparsity into sparse linear models as those described in Section 2.1, we propose complementing these models with the following penalty:

$$J_{GSR}(a) = \|\Gamma a\|_2^2, \tag{9}$$

where $\Gamma = (\gamma_1, \dots, \gamma_R)^T$ is an $R \times M$ penalty matrix for modeling the associations between predictors, γ_r is an $M \times 1$ vector encoding the penalty for each association r , and R is the number of associations being modeled. For instance, if predictors x_p and x_q are associated with each other, say by virtue of being spatially adjacent voxels, then a_p and a_q should presumably be assigned similar values to reflect this intrinsic association. This can be accomplished by, e.g. setting γ_{pq} to 1 and γ_{qp} to -1 such that differences in a_p and a_q are penalized. Many other context-specific properties can be analogously modeled by varying Γ as discussed in Section 2.4 and [7].

One may envisage other penalties to facilitate similar modeling flexibility. However, not all penalties added to the models in Section 2.1 will result in a practical optimization problem. The critical question is thus whether the optimization problem resulting from integrating (9) into models in Section 2.1 can be efficiently minimized. To address this question, we draw upon a basic property of the Euclidean norm:

$$\sum_{\omega=1}^{\Omega} \|z_{\omega}\|_2^2 = \|z\|_2^2, \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_{\Omega} \end{pmatrix}, \tag{10}$$

where z_{ω} is a vector of arbitrary length. Since models (3), (4), (5) and (7) in combination with (10), can all be written in the form:

$$\hat{a} = \min_a \|y - Xa\|_2^2 + \lambda \|\Gamma a\|_2^2 + J(a), \tag{11}$$

where $\lambda \geq 0$ and $J(a)$ is the sparse penalty term of the respective models, invoking (10) on (11) results in the following optimization problem:

$$\hat{a} = \sqrt{1 + \lambda} \min_{\tilde{a}} \|\tilde{y} - \tilde{X}\tilde{a}\|_2^2 + J(\tilde{a}), \tag{12}$$

$$\tilde{y} = \begin{pmatrix} y \\ 0 \end{pmatrix}, \quad \tilde{X} = (1 + \lambda)^{-\frac{1}{2}} \begin{pmatrix} X \\ \sqrt{\lambda} \Gamma \end{pmatrix}, \tag{13}$$

which has the exact same functional form as the optimization problems of the original sparse linear models. Thus, existing solvers of the respective sparse linear models [12-15] can be directly employed to efficiently minimize (11).

For the multi-process model (8), if the associations between predictors are similar across processes, then the same matrix augmentation trick can be applied with the zero vector in (13) replaced by an $R \times S$ zero matrix. If associations between predictors vary across processes, a minor modification is needed:

$$\tilde{y} = \begin{pmatrix} \text{vector}(Y) \\ 0 \end{pmatrix}, \quad \tilde{X} = (1 + \lambda)^{-\frac{1}{2}} \begin{pmatrix} X & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & X \\ & & & \sqrt{\lambda}\Gamma \end{pmatrix}, \quad (14)$$

where $\text{vector}(\cdot)$ is an operator that stacks the columns of its argument into a vector, and a in the first term of (11) is now $\text{vector}(A)$. We highlight that, in addition to enabling different predictor associations to be modeled for each process, (14) permits modeling of predictor associations *across* processes. This modification of the matrix augmentation trick thus builds even greater flexibility into (8).

2.3 GSR for Classification

The models described in Section 2.1 are inherently designed for regression problems with y being a continuous variable. To extend the GSR-augmented sparse regression models to the classification setting without altering the functional form of the associated optimization problems, we employ the spectral regression technique [23]:

Step 1. Learn the constraint-free optimal projection of the training data X , e.g. using graph embedding (GE) [24]:

$$Wy = \lambda Dy, \quad (15)$$

where y is the projection of X on the subspace defined by W [24]. W_{ij} models the intrinsic relationships between samples i and j of X , and D is a diagonal matrix with $D_{ii} = \sum_j W_{ij}$. We note that the key advantage of GE is that it enables various subspace learning algorithms to be used as classifiers by simply varying W [24].

Step 2. Determine the classifier weights, a , such that y and Xa are as similar as possible under the desired constraints, $Q(a)$:

$$\hat{a} = \min_a \|y - Xa\|_2^2 + Q(a). \quad (16)$$

Clearly, setting $Q(a)$ to $\lambda \|\Gamma a\|_2^2 + J(a)$ results in our proposed GSR model (11). Hence, exploiting spectral regression in combination with GSR facilitates sparsity and predictor associations to be jointly integrated into classifier learning.

2.4 Application to fMRI Spatiotemporal Classification

Given $N \times M \times 1$ feature vectors, x_i , forming the rows of an $N \times M$ predictor matrix, X , the general classification problem involves finding the corresponding $N \times 1$ label vector, l , containing the class label l_i of x_i . We assume here that each feature x_{ip} can be naturally assigned to a group $g_h \in G = \{g_1, \dots, g_H\}$. In the context of spatiotemporal fMRI classification, we treat the signal intensity of each brain voxel p at time t_k within a trial as a feature, and all brain volumes within a trial of the same experimental condition as a sample x_i , as illustrated in Fig. 1.

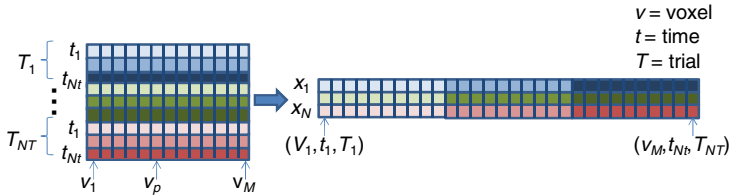


Fig. 1. Predictor matrix. Brain volumes within the same trial are concatenated and treated as a single sample x_i . N_T is the number of volumes within a trial and N_T is the number of trials for a particular experimental condition.

Our goal is thus to determine to which condition, l_i , does each concatenated set of brain volumes x_i belongs. Since the brain is known to be functionally organized into specialized neural regions [25], this provides a natural grouping of the voxels.

To model this modular property of the brain and the spatiotemporal correlations in brain activity, we build an anatomically-informed spatiotemporally smooth sparse linear discriminant analysis (ASTSLDA) classifier by first solving for y in (15) with:

$$W_{ij} = \begin{cases} 1/m_c, & l_i = l_j = c \\ 0, & \text{otherwise} \end{cases}, \tag{17}$$

where m_c is the number of samples in class c [23]. We then optimize (11) with $J(a)$ set to the group LASSO penalty (6), and the signal intensity of the voxels within each brain region of interest (ROI) at each time point t_k treated as a group (Section 3). Γ is set as the spatiotemporal Laplacian operator:

$$\Gamma_{p_k q_s} = \begin{cases} -1, & q_s \in N_{p_k} \\ 0, & \text{otherwise} \end{cases}, \quad \Gamma_{p_k p_k} = - \sum_{q_s \neq p_k} \Gamma_{p_k q_s}, \tag{18}$$

where N_{p_k} is the spatiotemporal neighborhood of voxel p at time t_k . Specifically, 6-connected spatial neighbors and signal intensity of voxel p itself at adjacent time points are defined as the spatiotemporal neighbors.

We can easily build other sparse LDA variants in an analogous manner. To build an anatomically-informed spatially smooth sparse LDA (ASSLDA) classifier, we employ (18) but with N_{p_k} being the 6-connected spatial neighborhood of voxel p . To build a sparse LDA classifier that is only anatomically-informed (ASLDA), we set λ in (11) to 0 with $J(a)$ being the group LASSO penalty (6). Voxel-level sparse LDA

classifiers that incorporates a spatiotemporal prior (STSLDA), a spatial prior (SSLDA), and no prior (SLDA) can be built in a similar manner as their anatomically-informed counterparts, but with $J(a)$ in (11) being the LASSO penalty (3). LDA classifier with elastic net penalty (ENLDA) can also be built from (11) by setting Γ to identity and $J(a)$ as the LASSO penalty (3). Our proposed model (11) thus encompasses many of the widely-used sparse linear models.

3 Materials

The publicly available StarPlus database [26] was used for validation. We provide here a brief description of the data. Details can be found in [26,27]. In the StarPlus experiment, all subjects performed 40 trials of a sentence/picture matching task. In each trial, subjects were required to look at a picture (or sentence) followed by a sentence (or picture), and then decide whether the sentence (picture) correctly described the picture (sentence). The first stimulus was presented for 4 s followed by a blank screen for 4 s. The second stimulus was then presented for up to 4 s followed by a 15 s rest period. In half of the trials, the picture preceded the sentence, and vice versa. fMRI brain volumes were acquired from 13 normal subjects at a TR of 500 ms, but only 6 of the subjects' data are available online [26]. Each subject's dataset comprised voxel time courses within 25 ROIs that were chosen by neuroscience experts, resulting in approximately 5000 voxels per subject. Inter-subject differences in the number of voxels were due to anatomical variability. Motion-correction and temporal detrending were applied on the voxel time courses to account for head motions and low frequency signal drifts. No spatial normalization was performed.

4 Results and Discussion

In this work, we explored the implications of incorporating prior knowledge into model learning using GSR on a large-scale fMRI classification problem. We treated the signal intensity of each voxel at each time point within a trial as a feature, and all brain volumes within the same trial of each experimental condition as a sample. The classification task was thus to discriminate sets of brain volumes associated with a picture from those associated with a sentence. To account for the delay in the hemodynamic response (HDR) [28], we only used the 8 brain volumes collected 4 s after stimulus onset within each trial. Each sample thus consisted of approximately 40,000 features (i.e. roughly 5000 voxels \times 8 volumes, see Section 3) with 40 samples per class (i.e. 40 trials per class) for each subject. For comparison, we contrasted ASTSLDA against LDA, linear SVM, SLDA, ENLDA, SSLDA [9], ASLDA, and ASSLDA. We employed the spectral projected gradient technique (SPG) [13] to minimize the optimization problem of the respective classifier models. SPG was chosen for its ability to efficiently handle large-scale problems (classifiers with \sim 40,000 coefficients in this study). All contrasted classifiers could be learned within a few seconds using SPG for a fixed set of parameter values λ and α (β was not involved in the classifier models contrasted in this work). Ten-fold nested cross validation was used [27] to select λ and α , and estimate the prediction accuracy. The

average prediction accuracies over subjects are shown in Fig. 2. An axial slice of the classifier weight patterns corresponding to 5.5 s and 6 s after stimulus onset (i.e. when HDR typically peaks) was also plotted (Fig. 3) for result interpretations. Only 3 exemplar subjects were included due to space limitations.

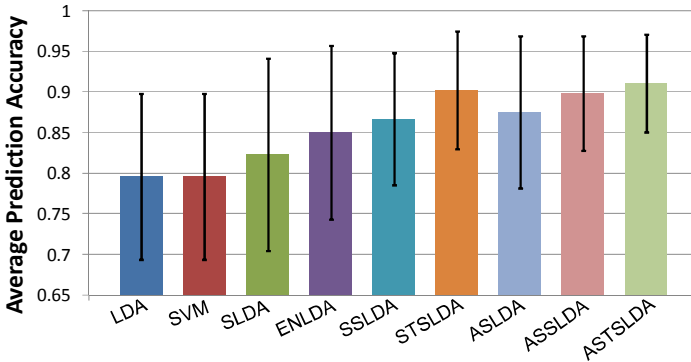


Fig. 2. Prediction accuracy comparisons. ASTSLDA outperformed all other contrasted methods with an average accuracy of 91%.

LDA resulted in the worst average prediction accuracy, which was likely due to overfitting. The classifier weight patterns also appeared spatially-random. Using SVM led to similar predictive performance and randomly-distributed weight patterns. Reducing overfitting using SLDA increased accuracy, but the weight patterns seemed overly-sparse. The reason was likely due to LASSO's constraint on the number of non-zero elements in the classifier weight vector a , i.e. restricted by the sample size [2]. In fact, since we treated the signal intensity of the brain volumes within a trial as a single feature vector, weights would be spread across time, resulting in even sparser spatial patterns than what would be obtained in the conventional case where each brain volume is taken as a sample. Also, we observed little consistency in the SLDA weight patterns across subjects and time.

Alleviating LASSO's constraint on the number of non-zero elements in a using ENLDA [2] improved prediction accuracy over SLDA. However, the weight patterns remained overly-sparse, which demonstrate the highly-overlooked fact that higher predictive accuracy does not necessarily translate to more neurologically-sensible weight patterns. In contrast, modeling spatial correlations using SSLDA resulted in higher prediction accuracy and spatially smoother weight patterns that better conform to how spatially proximal voxels tend to be correlated [21]. Our results thus illustrate the added-value of incorporating prior knowledge, both in terms of predictive performance and result interpretability. These benefits of exploiting prior knowledge were further exemplified using STSLDA, which obtained similar spatially smooth patterns but a further increase in prediction accuracy. This increase likely arose from how STSLDA pooled information across brain volumes through penalizing discrepancies in classifier weights between spatiotemporal neighbors.

Accounting for the modular structure of the human brain using ASLDA improved prediction accuracy over the non-anatomically-informed classifiers except STSLDA,

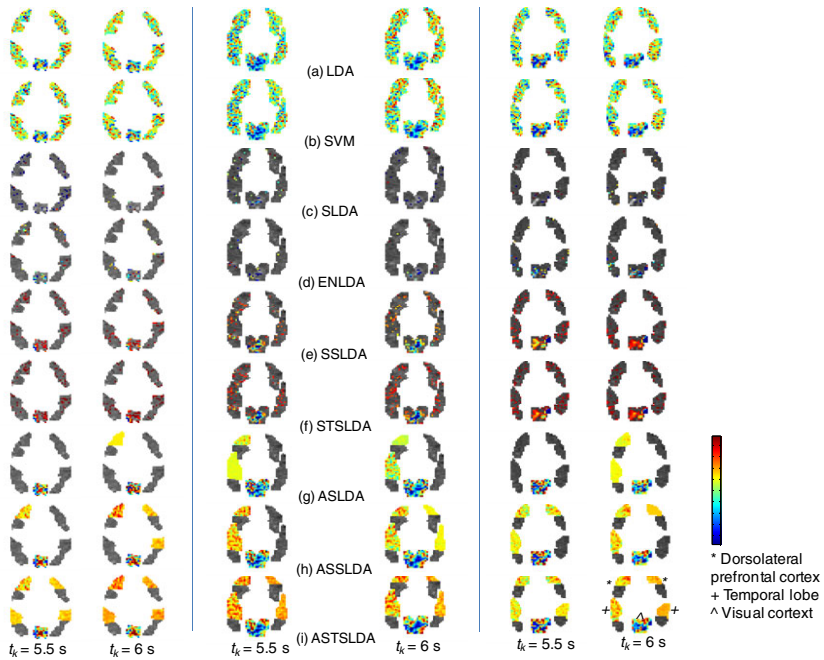


Fig. 3. Classifier weight patterns. Weight patterns 5.5 s and 6 s after stimulus onset for 3 exemplar subjects. (a) LDA and (b) SVM resulted in randomly-distributed weight patterns. (c) SLDA generated overly-sparse patterns, which was mildly improved with (d) EN-LDA. (e) SSLDA and (f) STSLDA produced spatially smoother patterns. (g) ASLDA and (h) ASSLDA weight patterns displayed little consistency across time and subjects. (i) ASTSLDA patterns were more consistent across subjects and time compared to the contrasted classifiers.

which again demonstrate the importance of constraining the model learning problem with prior knowledge (i.e. through grouping of associated features) to handle the curse of dimensionality. However, the weight patterns displayed little consistency between subjects and across time points. Modeling spatial correlations using ASSLDA resulted in higher accuracy than ASLDA, but it too obtained (slightly) lower accuracy than STSLDA. These results further support that exploiting the commonality between adjacent brain volumes, in addition to modeling spatial correlations within each brain volume, can greatly improve predictive performance.

Modeling both the modular property of the brain and the spatiotemporal characteristics of brain activity using ASTSLDA resulted in the best overall predictive performance with an average accuracy of 91%, which is an 11% increase over LDA and SVM. ASTSLDA also achieved the lowest variability in prediction accuracy, thus demonstrating stable performance despite the considerable inter-subject variability often seen in fMRI studies [21]. Moreover, higher consistency in weight patterns was observed across both subjects and time, with weights correctly assigned to the dorsolateral prefrontal cortex, which is responsible for verification task (e.g. decide if a sentence matches a picture) [29], as well as the visual cortex along the calcarine fissure and the temporal lobe, which pertain to picture/sentence discrimination [30].

5 Conclusions

We proposed GSR, a general approach for enabling properties beyond sparsity to be incorporated as an integral part of sparse model learning. By exploiting a basic property of the Euclidean norm, we showed that GSR can be directly applied to many widely-used sparse linear models without altering the functional form of their respective optimization problems. GSR hence facilitates greater modeling flexibility without the need for devising new complex optimization routines. We validated GSR on a large-scale classification problem and demonstrated how jointly modeling the modular nature of the human brain and the intrinsic spatiotemporal structure of brain activity can substantially improve prediction accuracy over standard techniques, such as LDA and SVM. An important extension of our current work will be to employ hierarchical LASSO in combination with GSR to model how some voxels within an ROI may be irrelevant for task discrimination. Also, exploiting the commonality between subjects within a group by integrating GSR into the collaborative hierarchical model may further improve prediction, which we intend to explore.

References

1. Tibshirani, R.: Regression Shrinkage and Selection via the LASSO. *J. Royal Stat. Soc. Series B* 58, 267–288 (1996)
2. Zou, H., Hastie, T.: Regularization and Variable Selection via the Elastic Net. *J. Royal Stat. Soc. Series B* 67, 301–320 (2005)
3. Yuan, M., Lin, Y.: Model Selection and Estimation in Regression with Grouped Variables. *J. Royal Stat. Soc. Series B* 68, 49–67 (2006)
4. Sprechmann, P., Ramirez, I., Sapiro, G.: Collaborative Hierarchical Sparse Modeling. Technical report, arXiv:1003.0400v1 (2010)
5. Friedman, J., Hastie, T., Tibshirani, R.: A Note on the Group LASSO and a Sparse Group LASSO. Technical report, arXiv:1001.0736v1 (2010)
6. Jacob, L., Obozinski, G., Vert, J.P.: Group Lasso with overlaps and graph Lasso. In: *Proc. Int. Conf. Mach. Learn.*, pp. 433–440 (2009)
7. Tibshirani, R., Taylor, J.: The Solution Path of the Generalized Lasso. *Ann. Stat.* (in press)
8. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and Smoothness via the Fused Lasso. *J. Royal Stat. Soc. Series B* 67, 91–108 (2005)
9. Ng, B., Vahdat, A., Hamarneh, G., Abugharbieh, R.: Generalized sparse classifiers for decoding cognitive states in fMRI. In: Wang, F., Yan, P., Suzuki, K., Shen, D. (eds.) *MLMI 2010. LNCS*, vol. 6357, pp. 108–115. Springer, Heidelberg (2010)
10. Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri, M.: Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* 286, 531–536 (1999)
11. Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., Just, M.A.: Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science* 320, 1191–1195 (2008)
12. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least Angle Regression. *Ann. Stat.* 32, 407–499 (2004)
13. van den Berg, E., Friedlander, M.P.: Probing the Pareto Frontier for Basis Pursuit Solutions. *SIAM J. Sci. Comput.* 31, 890–912 (2008)

14. Friedman, J., Hastie, T., Tibshirani, R.: Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Software* 33, 1–22 (2010)
15. Schmidt, M., Fung, G., Rosales, R.: Optimization Methods for L1-Regularization. Technical report, the University of British Columbia (2009)
16. Yamashita, O., Sato, M., Yoshioka, T., Tong, F., Kamitani, Y.: Sparse Estimation Automatically Selects Voxels Relevant for the Decoding of fMRI Activity Patterns. *NeuroImage* 42, 1414–1429 (2008)
17. Ryali, S., Supekar, K., Abrams, D.A., Menon, V.: Sparse Logistic Regression for Whole-brain Classification of fMRI Data. *NeuroImage* 51, 752–764 (2010)
18. Carroll, M.K., Cecchi, G.A., Rish, I., Garg, R., Rao, A.R.: Prediction and Interpretation of Distributed Neural Activity with Sparse Models. *NeuroImage* 44, 112–122 (2009)
19. van Gerven, M., Takashima, A., Heskes, T.: Selecting and Identifying Regions of Interest Using Groupwise Regularization. In: *NIPS Workshop on New Directions in Statistical Learning for Meaningful and Reproducible fMRI Analysis* (2008)
20. Michel, V., Eger, E., Keribin, C., Thirion, B.: Multi-class sparse bayesian regression for neuroimaging data analysis. In: Wang, F., Yan, P., Suzuki, K., Shen, D. (eds.) *MLMI 2010. LNCS*, vol. 6357, pp. 50–57. Springer, Heidelberg (2010)
21. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.B.: Dealing with the Shortcomings of Spatial Normalization: Multi-subject Parcellation of fMRI Datasets. *Hum. Brain Mapp.* 27, 678–693 (2006)
22. van Gerven, M., Cseke, B., de Lange, F.P., Heskes, T.: Efficient Bayesian Multivariate fMRI Analysis Using a Sparsifying Spatio-temporal Prior. *NeuroImage* 50, 150–161 (2010)
23. Cai, D., He, X., Han, J.: Spectral Regression: A Unified Approach for Sparse Subspace Learning. In: *Proc. IEEE Int. Conf. Data Mining*, pp. 73–82 (2007)
24. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph Embedding and Extension: A General Framework for Dimensionality Reduction. *IEEE Trans. Pat. Ana. Machine Intell.* 29, 40–50 (2007)
25. Fodor, J.A.: *The Modularity of the Mind*, pp. 2–47. MIT, Cambridge (1983)
26. <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>
27. Mitchell, T., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., Just, M., Newman, S.: Learning to Decode Cognitive States from Brain Images. *Mach. Learn.* 57, 145–175 (2004)
28. Liao, C.H., Worsley, K.J., Poline, J.B., Aston, A.D., Duncan, G.H., Evans, A.C.: Estimating the Delay of the fMRI Response. *NeuroImage* 16, 593–606 (2002)
29. Manentiab, R., Cappab, S.F., Rossiniac, P.M., Miniussiad, C.: The Role of the Prefrontal Cortex in Sentence Comprehension: An rTMS Study. *Cortex* 44, 337–344 (2008)
30. Vandenberghe, R., Price, C., Wise, R., Josephs, O., Frackowiak, R.S.J.: Functional Anatomy of a Common Semantic System for Words and Pictures. *Nature* 383, 254–256 (1996)

A Multi-scale Kernel Bundle for LDDMM: Towards Sparse Deformation Description across Space and Scales

Stefan Sommer¹, Mads Nielsen^{1,2}, François Lauze^{1,2}, and Xavier Pennec³

¹ Dept. of Computer Science, Univ. of Copenhagen, Denmark
sommer@diku.dk

² Synarc Imaging Technologies, Rødovre, Denmark

³ Asclepios Project-Team, INRIA Sophia-Antipolis, France

Abstract. The Large Deformation Diffeomorphic Metric Mapping framework constitutes a widely used and mathematically well-founded setup for registration in medical imaging. At its heart lies the notion of the regularization kernel, and the choice of kernel greatly affects the results of registrations. This paper presents an extension of the LDDMM framework allowing multiple kernels at multiple scales to be incorporated in each registration while preserving many of the mathematical properties of standard LDDMM. On a dataset of landmarks from lung CT images, we show by example the influence of the kernel size in standard LDDMM, and we demonstrate how our framework, LDDKBM, automatically incorporates the advantages of each scale to reach the same accuracy as the standard method optimally tuned with respect to scale. The framework, which is not limited to landmark data, thus removes the need for classical scale selection. Moreover, by decoupling the momentum across scales, it promises to provide better interpolation properties, to allow sparse descriptions of the total deformation, to remove the trade-off between match quality and regularity, and to allow for momentum based statistics using scale information.

Keywords: diffeomorphic registration, computational anatomy, LDD-KBM, LDDMM, scale, sparsity, kernels, momentum, landmarks.

1 Introduction

Among the many methods for non-rigid registration in medical imaging, the Large Deformation Diffeomorphic Metric Mapping framework (LDDMM) has the benefit of both providing good registrations and drawing strong theoretical links with Lie group theory and evolution equations in physical modeling [5,17]. The mathematical foundation ensures that optimization procedures find diffeomorphic optima and allows statistics to be performed on the results of registrations [9,15].

Diffeomorphisms in the LDDMM framework are regularized by a norm, which not only greatly influences the computed registration but also affects subsequent

statistics. The norm is often connected to a kernel, and since deformation frequently occurs at different scales, the choice of appropriate kernel scale will in such cases involve compromises between regularity and registration quality. In addition, in order to reduce computational complexity and provide meaningful statistics, it would be desirable to introduce sparsity in the framework which further complicates the choice of kernels.

In this paper, we propose a generalization of LDDMM allowing multiple kernels at multiple scales to be incorporated into the registration. We extend the theory by introducing a multi-scale kernel bundle, and the resulting Large Deformation Kernel Bundle Mapping (LDDKBM) framework keeps the strong mathematical foundation while giving significant benefits. Unlike previous methods, our construction explicitly decouples the momentum across scales allowing the algorithm to adapt to the different scales in the input data. This makes classical scale selection unnecessary and removes the common trade-off between match quality and regularity of the diffeomorphism. The ability to describe the deformation at precisely the right scales promises to allow for sparse representations and statistics incorporating scale information.

1.1 Deformation at Multiple Scales; An Example

Figure 1 shows a simple example of landmark matching. In order to register the points, movement is needed at both large and small scales, and the standard LDDMM algorithm faces a compromise between regularity of the diffeomorphism and match quality. The registration in the fourth image is performed by the method developed in this paper incorporating kernels at different scales, and the result shows how the trade-off between regularity and quality has been removed. Furthermore, no cross validation for choosing the kernel size is needed and contrary to previous methods, the computed momentum is represented only at the appropriate scales.

1.2 Related Work

Besides LDDMM, many methods for non-rigid registration are in use today including elastic methods [11], parametrizations using static velocity fields [18] and the demons algorithm [13,16]. The deformable template model pioneered by Grenander in [7] and the flow approach by Christensen et al. [4] was paramount in the development of LDDMM together with the theoretical contributions of Dupuis et al. and Trouvé [6,14]. Algorithms for computing optimal diffeomorphisms have been developed in [2], and [15] uses the momentum representation for statistics and develops a momentum based algorithm for the landmark matching problem. The review paper [18] and the book [17] provide excellent overviews of the theory and applications of LDDMM in medical imaging.

The role of the kernel and deformation at different scales have been addressed by Risser et al. in [12], where the authors propose a multi-kernel LDDMM approach which constructs new kernel shapes by adding Gaussian kernels. Our method differs from this approach by extending LDDMM to allow decoupling of the energy and momentum at each scale, and it therefore enables the algorithm

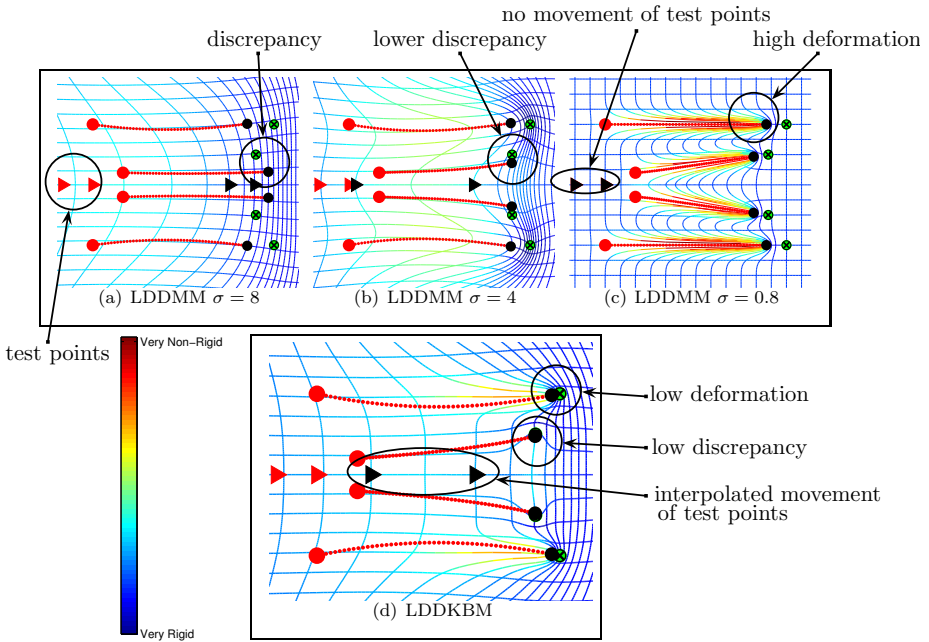


Fig. 1. Matching four landmarks (red) to four landmarks (green, crossed) with results (black) for Gaussian kernels of difference sizes (scale σ in grid units). Top row: standard LDDMM, bottom row: the proposed LDDKBM method. An initially square grid is shown deformed by each diffeomorphism along with two test points (red triangles) transported by the diffeomorphism to the black triangles. The grids are colored with the trace of Cauchy-Green strain tensor (log-scale). The compromise between regularity and match quality is visible for the standard LDDMM method. The poor match for the smallest scale is due to the lack of “carpooling” effect [10]. Notice that for standard LDDMM with the smallest scales, the test points are not moved at all or split up. The proposed method obtain both a regular diffeomorphism and a good match, and the test points are transported in an arguably natural way.

to select the appropriate deformation at each scale individually. The differences between the methods are detailed later in the paper.

1.3 Content and Outline

In the next section, we give an overview of the LDDMM framework and its use in non-rigid registration in medical imaging. In particular, the focus will be on the role of the kernel. We then progress to developing our key contribution, the multi-scale kernel bundle framework LDDKBM, and highlight the important properties of the construction. We will show how the decoupling of scales differentiates the construction from previous methods and promises sparse representations. The evaluation of the method on a dataset of landmarks obtained from CT

lung images will show the automatic incorporation of the appropriate scale and indicate why the method allows sparsity. The paper thus contributes by

- (1) highlighting the role of the kernel in LDDMM and showing how the registration results depend on the choice of kernel,
- (2) developing the multi-scale kernel bundle generalization LDDKBM to automatically incorporate multiple scales in the registration,
- (3) and showing how the developed method removes the need for the classical scale selection while promising improved momentum based statistics and sparse representations.

We note that while the evaluation is performed on landmark data, the LDDKBM framework is very general, and it will apply to different types of data as well. We are currently working on an implementation for image registration, and we consider this paper a first step which eventually will lead to the method being useful for clinical applications.

2 LDDMM and Kernels

The Large Deformation Diffeomorphic Metric Mapping framework provides a well-defined metric structure on spaces of diffeomorphisms and gives convenient ways to parametrize such spaces. We here give a brief overview of the LDDMM construction before going into details about the use of kernels in the framework. Landmarks will be used for the examples, though both the LDDMM framework and the kernel bundle we will develop also apply to images, curves, surfaces, and tensors.

Registration of geometric objects is often performed by defining an action of diffeomorphisms on the objects before searching for diffeomorphisms matching the objects through the action. For example, in order to register landmarks x_1, \dots, x_N and y_1, \dots, y_N in \mathbb{R}^d , $d = 2, 3$, we search for a diffeomorphism $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\varphi(x_i) = y_i$. Frequently, a perfect match is not possible or even not desirable because noisy data may force the diffeomorphism to be highly irregular. Instead, we search for φ minimizing

$$E(\varphi) = E_1(\varphi) + \lambda U(\varphi) \quad (1)$$

where $E_1(\varphi)$ is a regularization measure, $U(\varphi)$ a measure of the quality of the match, and $\lambda > 0$ a weight. A simple and often used choice for U are the L^2 error which takes the form $U(\varphi) = \sum_{i=1}^N \|\varphi(x_i) - y_i\|^2$ for landmarks. In the LDDMM framework, the regularization measure $E_1(\varphi)$ is defined as the minimum energy of paths of diffeomorphisms transporting the identity Id_Ω to φ , i.e.

$$E_1(\varphi) = \min_{v_t \in V, \varphi_{0t}^v = \varphi} \int_0^1 \|v_s\|_V^2 ds \quad (2)$$

with $\|\cdot\|_V$ being a right invariant Riemannian metric on the tangent space V of derivatives of such paths, and φ_{0t}^v denoting the path starting at Id_Ω with

derivative $\partial_t \varphi_{0t}^v = v_t \circ \varphi_{0t}^v$. This norm is chosen to penalize highly varying paths and, therefore, a low value of $E_1(\varphi)$ implies that the path to reach φ , and hence φ itself, is regular. In addition, the norm and tangent space V is chosen to ensure several important properties, which include ensuring a minimizer for E exists and that integration in V is well-defined. The latter property enables us to represent many diffeomorphisms as endpoints of paths φ_{0t}^v .

2.1 The Tangent Space

The norm on the tangent space V is most often chosen to ensure V is a reproducing kernel Hilbert space [17]. If the domain Ω is a subset of \mathbb{R}^d so that V is a subset of the maps $\Omega \rightarrow \mathbb{R}^d$, one can with an appropriate norm $\|\cdot\|_V$ on V show the existence of a kernel $K : \Omega \times \Omega \rightarrow \mathbb{R}^{d \times d}$ so that, for any constant vector $a \in \mathbb{R}^d$, the vector field $K(\cdot, x)a \in V$, and $\langle K(\cdot, x)a, K(\cdot, y)b \rangle_V = a^T K(x, y)b$ for all points $x, y \in \Omega$ and all vectors $a, b \in \mathbb{R}^d$. This latter property is denoted the reproducing property and it provides a way to evaluate the inner product on the span of the kernel elements. Tightly connected to the norm and kernels is the momentum, which is named so because of its connection to momentum in classical mechanics. The momentum operator L on V is defined by

$$\langle Lv, w \rangle_{L^2(\Omega)} = \int_{\Omega} (Lv(x))^T w(x) dx = \langle v, w \rangle_V$$

and hence it connects the inner product on V with the inner product in $L^2(\Omega)$. As we will see in the landmark case, the value Lv might be singular and in fact $L(K(\cdot, y)a)(x)$ is the Dirac measure $\delta_y(x)a$. These relations lead to convenient equations for minimizers of the energy (II). In particular, the EPDiff equations for the evolution of the momentum a_t for optimal paths assert that if φ_t is a path minimizing $E_1(\varphi)$ with $\varphi_1 = \varphi$ minimizing $E(\varphi)$ and v_t is the derivative of φ_t then v_t satisfies the system

$$v_t = \int_{\Omega} K(\cdot, x)a_t(x) dx ,$$

$$\frac{d}{dt}a_t = -Da_t v_t - a_t \nabla \cdot v_t - (Dv_t)^T a_t .$$

The first equation connects the momentum a_t with the velocity v_t . The EPDiff equations reduces to particular forms for several objects. For landmarks x_1, \dots, x_N , the momentum will be concentrated at $\varphi_t(x_i)$ as Dirac measures $a_{i,t} \delta_{\varphi_t(x_i)}$ leading to the finite dimensional system of ODE's

$$v_t = \sum_{l=1}^N K(\cdot, \varphi_t(x_l))a_{l,t} , \quad \frac{d}{dt}\varphi_t(x_i) = v_t(\varphi_t(x_i)) ,$$

$$\frac{d}{dt}a_{i,t} = - \sum_{l=1}^N D_1 K(\varphi_t(x_i), \varphi_t(x_l))a_{i,t}^T a_{l,t} .$$

(3)

2.2 Kernels

There is some freedom in the choice of kernel or, equivalently, the operator L but, in lack of general models for e.g. deformations of organs to be registered [11], it is hard to give satisfactory modelling arguments specifying the kernel shapes. Rotational and translational invariance is commonly assumed [17] and the Gaussian kernel $K(x, y) = \exp(-\frac{|x-y|^2}{\sigma^2})\text{Id}_d$ is a convenient and often used choice. The scaling factor σ is not limited to Gaussian kernels and a scaling factor needs to be determined for both Gaussian kernels and many other proposed kernel shapes.

Larger scales lead in general to higher regularization and smoother diffeomorphisms, whereas smaller kernels penalize higher frequencies less and often gives better matches. This phenomenon is in particular apparent for objects with sparse information, such as the point matching problem (3) illustrated in Figure 1 and images with e.g. areas of constant intensity.

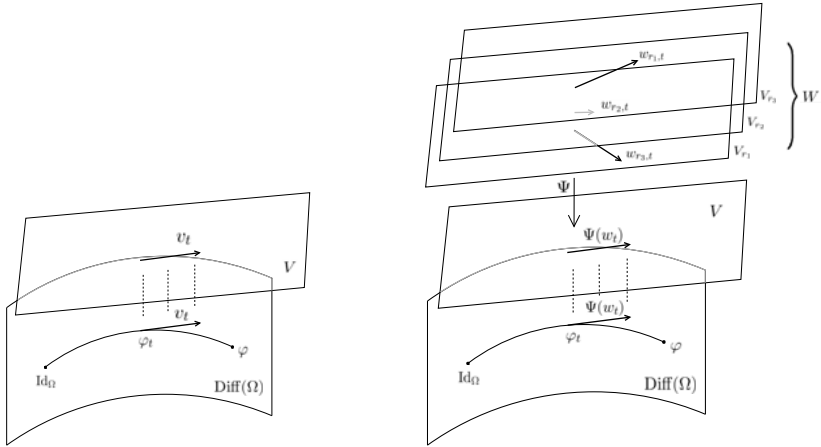
3 LDDKBM: A Multi-scale Kernel Bundle

To incorporate multiple kernels in the framework, we extend LDDMM by a multi-scale kernel bundle. The construction, which we denote LDDKBM (Large Deformation Diffeomorphism Kernel Bundle Mapping) is very general and allows both multiple kernel shapes and scales though varying the scale of the kernel constitutes the most obvious use. We start by outlaying the framework and later discuss how it deviates from other proposals incorporating scale space ideas.

As described in the previous sections, a time varying vector field v_t in the tangent space V generates a path φ_{0t}^v of diffeomorphisms. This provides a view of the space of diffeomorphisms G_V obtainable in this way as a manifold having V as its tangent space at all points, confer Figure 2(a). The norm $\|\cdot\|_V$ works as a metric making G_V a Riemannian manifold. In this view, the regularization factor E_1 is the energy of the path φ_t , and this energy is tightly connected to the kernel arising from the norm on V .

In order to use more kernels, we extend the tangent space V to the family $W = \{V_r\}_{r \in I_W}$ of Hilbert spaces V_r parametrized by scalars r in a set I_W so that, for each r , V_r is a subspace of the tangent space V but equipped with a different metric. The key point here is to allow the norms $\|\cdot\|_{V_r}$ on V_r and connected kernels K_r to vary with r . In this way, the tangent bundle $G_V \times V$ is extended to a vector bundle $G_V \times W$ which we denote the multi-scale kernel bundle. The space W inherits the vector space structure from each V_r , and it is a pre-Hilbert space with norm $\|w\|_W^2 = \int_{I_W} \|w_r\|_{V_r}^2 dr$. Using the norm, we define the energy E_1^W of paths in W by $E_1^W(w_t) = \int_0^1 \|w_s\|_W^2 ds$.

With appropriate conditions on V_r and the norms $\|\cdot\|_{V_r}$, W is a Hilbert space and, by possibly restricting to a subspace, we can pass from W to V using the map $\Psi : W \rightarrow V$ defined by integration $\Psi(w) = \int_{I_W} w_r dr$ for elements $w = \{w_r\}_r \in W$. Since we as usual have a connection between paths in V and



(a) In standard LDDMM, a path on the manifold $\text{Diff}(\Omega)$ is generated by integrating the time varying vector field v_t in the tangent space V . (b) A path w_t in the vector space W , here constructed from V_{r_1} , V_{r_2} , and V_{r_3} , sum through the map Ψ to a vector field $\Psi(w_t)$, which in turn generates a path on $\text{Diff}(\Omega)$.

Fig. 2. The manifold view of standard LDDMM and LDDKBM

paths on the manifold G_V , we get using Ψ a similar relation between paths $w_t = \{w_{r,t}\}_r$ in W and paths in G_V by

$$w_t \mapsto \varphi_{0t}^{\Psi(w)}, \tag{4}$$

i.e. $\varphi_{0t}^{\Psi(w)}$ is the path starting at Id_Ω with derivative $\partial_t \varphi_{0t}^{\Psi(w)} = \Psi(w_t) \circ \varphi_{0t}^{\Psi(w)}$. The path energy E_1^W gives a regularization measure on diffeomorphisms

$$E_1^W(\varphi) = \min_{w_t \in W, \varphi_{01}^{\Psi(w)} = \varphi} \int_0^1 \|w_s\|_W^2 ds \tag{5}$$

which, together with a quality of match measure $U(\varphi)$, allows a reformulation of the registration problem as the search for a diffeomorphism minimizing

$$E^W(\varphi) = E_1^W(\varphi) + \lambda U(\varphi) . \tag{6}$$

The above formulation should be compared with the standard LDDMM formulation (1) using the regularization (2). It is immediately clear that the proposed method is an extension of standard LDDMM, since the original regularization is the special case with only one scale and hence $W = V$.

We note that the parameter space I_W can be a compact interval or finite set of scalars in which case the integral reduces to just a sum. Often, it will be an interval specifying a scale range, and a practical implementation will discretize the interval into a finite set of scalars.

3.1 Evolution Equations

Many of the properties of standard LDDMM are retained in the LDDKBM construction. Importantly, the evolution equations for solutions of the registration problem take on a form very similar to the EPDiff equations; if $\Psi(w_t)$ is the derivative of the path of diffeomorphisms φ_t minimizing (5) with $\varphi = \varphi_1$ minimizing (6) then

$$\begin{aligned}
 w_{r,t} &= \int_{\Omega} K_r(\cdot, x) a_{r,t}(x) dx , \\
 \frac{d}{dt} a_{r,t} &= \int_{I_W} -D a_{r,t} w_{s,t} - a_{r,t} \nabla \cdot w_{s,t} - (D w_{s,t})^T a_{r,t} ds .
 \end{aligned}
 \tag{7}$$

with $a_{r,t}$ being the momentum for the part w_r of w . In essence, we just integrate the standard EPDiff equations over the parameter space I_W to obtain the evolution of the momentum at each scale. The system implies that the momentum conservation property holds with the kernel bundle construction.

An important difference from the standard framework relates to the energy along optimal paths. The relation to geodesics in standard LDDMM suggests that the norm $\|v_t\|_V$ is constant in t when v_t is optimal for $E_1(\varphi)$. This is in fact the case for standard LDDMM. For LDDKBM, momentum is conserved along optimal paths of $E_1^W(\varphi)$ though $\|w_t\|_W$ is not constant. This occurs because the new energy is not directly related to a metric in the Riemannian sense.

3.2 Scale Decoupling and Relation to Other Approaches

The energy E_1^W measures the contribution of each scale independently, and the total energy is the sum of the energy at each scale. Because of this, the momentum components can vary over scale, and the method may use any combination of small and large scale features at each spatial location to obtain the best match with least energy at each scale.

In contrast to this, the simultaneous coarse and fine method developed by Risser et al. in [12] builds a kernel by summing Gaussians of different scale. This effectively changes only the shape of the kernel and does not allow different momentum at different scales. Therefore, the method works simultaneously in the sense that no decoupling between scales is present, whereas the method proposed here exactly aims at decoupling the scales. The decoupling will be clear in the experiments and we believe it is desirable for several reasons. We expect better matches because movement at one scale does not imply undesired movement at all other scales. This will also allow sparsity to enter the representation: since movement is represented only at the right scale, the momentum components at all unrelated scales can be zero. Though we currently do not directly enforce sparsity, we will see this effect visually in the experiments. Last, doing statistics on the momentum as proposed in [15] will result in information at each scale because of the decoupling. With the simultaneous coarse and fine method, no scale-specific momentum is available.

3.3 Implementation

As for standard LDDMM, the choice of algorithm for optimizing (6) depends very much on the objects to be matched. For the experiments in this paper, we will match landmarks to which the shooting method of [15] applies. The algorithm does a second order optimization while integrating along critical paths specified by the system (3). We implement a much similar algorithm for the finite dimensional landmark version of the system (7). Because the energy along optimal paths is not constant, the energy must be integrated over the point and velocity trajectories and this rules out the most simple 2nd-order schemes. Therefore, at this point, we use a gradient descent algorithm but we are in the process of developing both 2nd-order and fast 1st-order algorithms.

4 Experiments

We perform experiments on the publicly available DIR [3] dataset of lung CT images and manually annotated landmarks. The dataset consists of five cases of CT images for different stages of the inhale and exhale phases and annotated landmarks for the maximum inhale and exhale phases, respectively. The images and landmarks are available on grids with voxel size varying slightly between the cases but close to $1 \times 1 \times 2.5$ mm. Further details can be found in the reference.

For each case, the 300 publicly available landmarks, x_1^I, \dots, x_{300}^I for the maximum inhale phases and x_1^E, \dots, x_{300}^E for the maximum exhale phase, correspond pairwise. We will drive the registration using these landmarks and not incorporate image information at this point. We measure the fiducial registration error (FRE) of each computed registration. Since a small FRE indicates a good match on the registered landmarks only, we need an additional way of comparing the quality of the diffeomorphism. We do this by computing the match on random subsets of the landmarks and compute the target registration error (TRE) on the landmarks not included in the match. Choosing kernel scale based on this evaluation is akin to performing cross-validation tuning in machine learning.

4.1 Setup and Results

We choose random subsets of 75 landmarks to drive the registration. For each such choice of subset S and each of the five patient cases, we compute the FRE $(\sum_{j \in S} \|\varphi(x_j^I) - x_j^E\|^2)^{1/2}$ and the TRE $(\sum_{j \notin S} \|\varphi(x_j^I) - x_j^E\|^2)^{1/2}$. We find the relative size of these measures against the respective values before the match, and average over the patients and different choices of subsets. This setup is performed for standard LDDMM with Gaussian kernels with scale ranging between 10 and 170 voxels and with LDDKBM with five scales in the same range. For both methods, we let $\lambda = 8$ for the energy weight in (1) and (6), and, for the kernel bundle, we let each scale contribute with equal weight. The theoretical influence of these weights remains to be explored, though, from a practical point of view, we have tried different values without significant effects on the results.

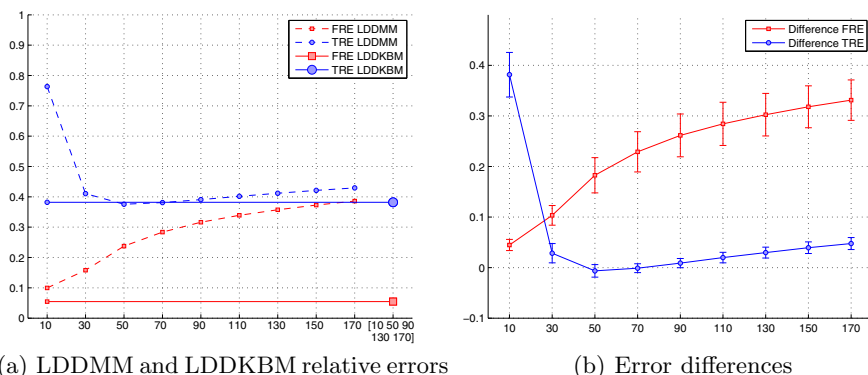


Fig. 3. (a) Average relative FRE and TRE for different kernel scales (standard LDDMM) and LDDKBM (horizontal line and rightmost). Values of 0 indicate perfect matches and 1 indicates no reduction in error compared to before the match. Labels on the horizontal axis are kernel scale in voxels (9 different scales for standard LDDMM and the interval [10 170] discretized in five scales for LDDKBM). (b) FRE and TRE for standard LDDMM (scale in voxels on horizontal axis) subtracted the respective errors for LDDKBM with interval the [10 170] discretized in five scales. Positive values indicate superior performance of the proposed method. Error bars show standard deviation of the results.

As the results in Figure 3(a) shows, the FRE increases with scale for standard LDDMM. The LDDKBM method produces an FRE of around half the FRE for the smallest scale for standard LDDMM. Similarly, we see that the TRE decreases with increasing scale up to a scale of 50 units after which it starts increasing. This indicates that a kernel scale of 50 units will be appropriate for standard LDDMM. As displayed in Figure 3(b), the LDDKBM method attains an error well within one standard deviation of the best standard LDDMM result. Performing t-tests on the 5% significance level of the hypotheses that the methods have equal mean results in a failure to reject the hypotheses for the best standard LDDMM results, and thus we can only conclude a difference between the methods for the scales where LDDKBM clearly outperform standard LDDMM. The LDDKBM method therefore *automatically* adapts to the right scales. Furthermore, the results indicate that the same quality of match can be reached with less data since we potentially could use the entire dataset to drive the registration. Manual scale selection will allow only a part of the data as input for the registration as the rest is needed to select the kernel scale.

To show the decoupling between scales, Figure 4 displays the L_2 -norm of w_t at $t = 0$ for three scales of LDDKBM and v_t at $t = 0$ with the best scale for standard LDDMM. The decoupling is apparent, and the expected sparsity is visible in particular for the smallest scales where movement is very localised.

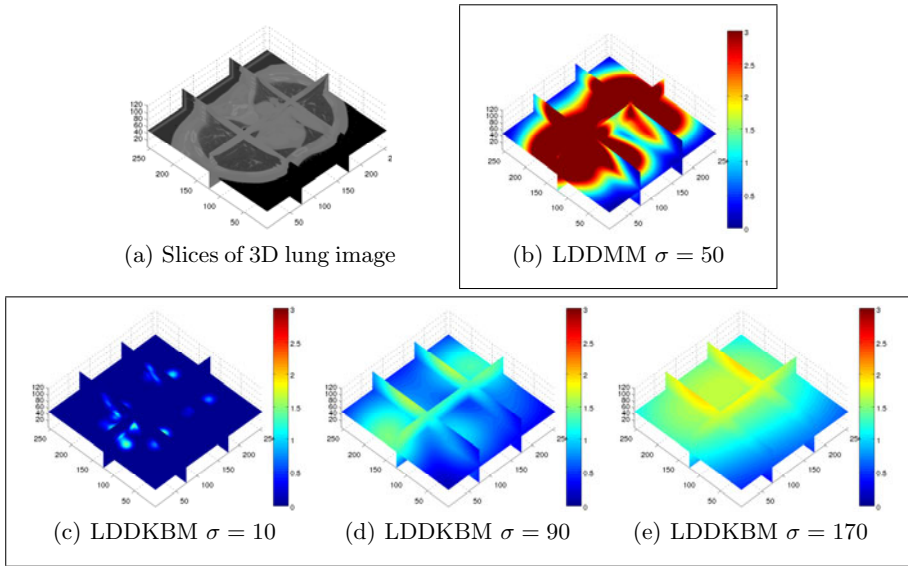


Fig. 4. Slices of 3D lung image and initial vector fields generating registration. Upper row, left to right: 3D image of lung to be registered, L^2 -norm of vector field v_t generating φ at $t = 0$ for scale 50 (LDDMM). Lower row, left to right: L^2 -norm of w_t generating φ at $t = 0$ for scales 10, 90, 170 (LDDKBM). Even for the best scale, standard LDDMM is not localized with high energy to account for the warp across scales, while LDDKBM use the appropriate scales localized and sparsely for the smallest scales.

5 Conclusion and Outlook

The multi-scale kernel bundle framework LDDKBM extends the LDDMM framework by incorporating multiple kernels at multiple scales in the registration. The method decouples scales and thereby allows the registration to act at the right scales only. As shown in the experiments, the method removes the need for classical scale selection by automatically, without user intervention or cross-validation, adapting to the right scales, and it thereby provides as good results as standard LDDMM optimally tuned with respect to scale. The decoupling of scales and the ability of the method to adapt at different scales are visualized and LDDKBM promises to allow for sparse description of deformation and momentum based statistics incorporating scale information.

The method opens up many directions for future work. Extending existing LDDMM algorithms to LDDKBM with non-landmark data is clearly important, and using the smoothness properties of different scales to speed up such algorithms will be beneficial. We expect to make use of the improved representation to compute sparse systems which will further improve computation efficiency and statistical results. Enforcing sparsity through sparse priors will be an important step for this.

References

1. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006)
2. Beg, M.F., Miller, M.I., Trounev, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV* 61(2), 139–157 (2005)
3. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology* 54(7), 1849–1870 (2009)
4. Christensen, G., Rabbitt, R., Miller, M.: Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing* 5(10) (2002)
5. Cotter, C.J., Holm, D.D.: Singular solutions, momentum maps and computational anatomy. nlin/0605020 (May 2006)
6. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching (1998)
7. Grenander, U.: *General Pattern Theory: A Mathematical Study of Regular Structures*. Oxford University Press, USA (1994)
8. Hernandez, M., Bossa, M., Olmos, S.: Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision* 85(3), 291–306 (2009)
9. Joshi, S., Davis, B., Jomier, B.M., Guido Gerig, B.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
10. Micheli, M.: *The differential geometry of landmark shape manifolds: metrics, geodesics, and curvature*. Ph.D. thesis, Brown University, Providence, USA (2008)
11. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian elasticity: A statistical regularization framework for non-linear registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 943–950. Springer, Heidelberg (2005)
12. Risser, L., Vialard, F.-X., Wolz, R., Holm, D.D., Rueckert, D.: Simultaneous fine and coarse diffeomorphic registration: Application to atrophy measurement in alzheimer’s disease. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 610–617. Springer, Heidelberg (2010), PMID: 20879366
13. Thirion, J.: Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis* 2(3), 243–260 (1998)
14. Trounev, A.: *An infinite dimensional group approach for physics based models in patterns recognition* (1995)
15. Vaillant, M., Miller, M., Younes, L., Trounev, A.: Statistics on diffeomorphisms via tangent space representations. *NeuroImage* 23(suppl. 1), S161–S169 (2004)
16. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage* 45(suppl. 1), 61–72 (2009)
17. Younes, L.: *Shapes and Diffeomorphisms*. Springer, Heidelberg (2010)
18. Younes, L., Arrate, F., Miller, M.I.: Evolutions equations in computational anatomy. *NeuroImage* 45(1, suppl. 1), S40–S50 (2009)

Automatic Part Selection for Groupwise Registration

Pei Zhang and Timothy F. Cootes

Imaging Sciences, School of Cancer and Enabling Sciences, The University of Manchester, UK
Pei.Zhang-2@postgrad.manchester.ac.uk,
tim.cootes@manchester.ac.uk

Abstract. Groupwise non-rigid image registration plays an important role in medical image analysis. As local optimisation is largely used in such techniques, a good initialisation is required to avoid local minima. Although the traditional approach to initialisation—affine transformation—generally works well, recent studies have shown that it is inadequate when registering images of complex structures. In this paper we present a more sophisticated method that uses the sparse matches of a parts+geometry model as the initialisation. The choice of parts is made by a voting scheme. We generate a large number of candidate parts, randomly construct many different parts+geometry models and then use the models to select the parts with good localisability. We show that the algorithm can achieve better results than the state of the art on three different datasets of increasing difficulty. We also show that dense mesh models constructed during the groupwise registration process can be used to accurately annotate new images.

1 Introduction

Groupwise non-rigid registration has become an important tool in medical image interpretation [12,3,15,2,5]. It is widely used for such tasks as building statistical shape models, anatomical atlas construction and data fusion. Most approaches to groupwise non-rigid registration involve local optimisation, and are thus sensitive to initialisation. Poor starting points can easily lead the algorithm to get stuck at local minima. It is common to start from the affine transformation which best aligns the data [12,3,15,2,5]. However, for complex structures, such as radiographs of the human hand, it has been shown that this does not work well (e.g. fingers are highly misaligned, leading to a very blurred mean.) [11,6]. Hence, a more robust and sophisticated initialisation is desirable.

One possible solution is to locate a sparse set of points on each image and use them to do the initialisation [13,9,11,6]. The approach is to construct a model with a set of distinctive parts¹ and use the model to search the images, finding the best matches on each. The set of the best matches is then used to start a groupwise non-rigid registration algorithm. The distinctive parts can be obtained either manually [9,11] or automatically [13,16]. Each part has a set of match candidates on each image and each candidate represents a possible location of the part. The model, which is a sparse representation of the object, can be a Point Distribution Model (PDM) [13] or a parts+geometry model [9,11,6]. In earlier work [16] it is shown that the latter can lead to a more robust set of starting points than the former.

¹ Typically a fixed shaped region (e.g. a rectangle or an ellipse) placed around a feature point.

In [16] we automatically generate a number of candidate parts and used a Genetic Algorithm (GA) to select the parts+geometry model which best localised the starting points. Due to the complexity of the algorithm, we only considered a subset of candidate parts—those that had good localisability, which was selected based on the response surface of the parts. This pre-selection may fail to include useful parts, and does not guarantee that the chosen parts can give consistent and unambiguous matches. Another drawback of this method is that one has to specify the number of parts to use. This is a common problem of the methods of automatic construction of parts+geometry models [11, 6]. As shown in [16], this parameter can significantly affect the accuracy of the following groupwise non-rigid registration. Automatic determining the best number of parts is thus desirable.

In this paper we use a voting strategy to explicitly explore the localisability of the whole set of candidate parts. Compared with [16], our new method is simple, fast and accurate. The key idea is to construct a large number of parts+geometry models, each describing a different, but overlapping, subset of parts. Each such model votes for candidates for each of its parts on every image. Thus for every potential part (and there may be thousands), we record potential matches on each image, and store the votes for each match. By analysing the distributions of such votes we can select the most reliable parts—those whose votes are most consistent. We then use a minimum description length (MDL) criterion to select a subset of such parts, which is best able to define the starting points across the region of interest. Another advantage of our method is that it can automatically determine the number of parts to use.

The voting approach introduces robustness to the system. By constructing random combinations and voting we reduce the danger that parts with poor localisability will degrade the overall performance of the parts+geometry model, as their influence will be diluted. The final MDL based selection process removes parts which are consistent but incorrect.

In the following we elaborate the method and demonstrate its application to three different datasets of increasing difficulty. The output of the system is a parts+geometry model which is able to accurately locate sparse correspondences. When applied to the training set itself it finds points which are shown to be an excellent initialisation for a groupwise non-rigid registration algorithm [5]. Experimental results show that our algorithm can give significantly more accurate final dense correspondences than either an affine initialisation or that using the GA based method [16], even for an apparently simple dataset where affine initialisation might be considered sufficient.

We also show that the parts+geometry model can be used to initialise an elastic mesh model when applied to new images, which can achieve accurate annotation—a further confirmation of the efficacy of the approach.

2 Methodology

Given a set of training images, we seek to select a small number of parts, together with geometric relationships, which can be used to efficiently locate a set of starting points of good correspondences across the images. The approach is to generate a large number of candidate parts and find their associated matches on every image. We then use a voting

strategy to evaluate the localisability of each part, and thus to rank them. We then select a non-overlapping combination of the best parts to form candidate models, and use an MDL criterion to select the best overall combination of parts and geometry. Below we describe each step. Note that the descriptions given in Sect. 2.1 and Sect. 2.2 are rather brief. For more details, the reader is referred to [1] and [16].

2.1 Candidate Parts

Suppose we have a set of images $\{I_i | i = 1, \dots, N_I\}$ and a set of candidate parts $\{P_j | j = 1, \dots, N_P\}$. The candidate parts can be generated by running an interest point operator over one or more images [13], or simply taking a large number of overlapping samples from a representative image at a range of scales [16]. We then search every image, I_i , with every part, P_j , to get a set of n_{ij} candidate matches, $\{\mathbf{p}_{ijk} | k = 1, \dots, n_{ij}\}$, where $\mathbf{p} = (\mathbf{x}, s, \theta)$ defines the position, scale and orientation of the candidate match.

The goal is to select a small set of parts which can define an optimal set of starting points. By optimal we mean that those points have the best correspondence. Clearly some parts will be useful for establishing accurate correspondence across the set, some will have large numbers of spurious matches and thus distract the algorithm, and some may not be present at all on many of the images. We thus need a method of selecting those parts which can find correspondence the most effectively.

One approach, widely used in SIFT matching [14], is to select those parts which only have one strong match per image. However in any object composed of repeating structures (such as the hands or spines), useful features are likely to have multiple good matches. In the following we describe a method of finding the best match for each part on each image by voting with a large number of parts+geometry models. Reliable matches tend to be those most voted for, and good parts tend to be those for which the votes are most consistent.

2.2 Parts+geometry Models

The parts+geometry model represents the object of interest using the parts together with a collection of geometric relationships between pairs of parts, defined by a set of arcs ω . Let $\mathcal{P}_S = \{P_j | j \in S \subset \{1, \dots, N_P\}\}$ be a set of $m = |S|$ parts drawn from the full set (where $|\cdot|$ denotes cardinality). Following the technique outlined in [1], we can automatically build a parts+geometry model \mathcal{G}_S for \mathcal{P}_S . The idea is to use minimum spanning tree to initialise the model, search with this model on the remaining images and then rebuild the model from the best matches.

Let $f_j(\mathbf{p}_{jk})$ be the cost associated with a part P_j having pose \mathbf{p}_{jk} (we drop the i subscript for clarity). The relationship between the parts P_j and $P_{j'}$ can be represented in the cost function, $f_{jj'}(\mathbf{p}_{jk}, \mathbf{p}_{j'k'}) = -\log \Pr(\mathbf{x}_{jk} - \mathbf{x}_{j'k'})$. To find the best match of \mathcal{G}_S on an image, we find the match candidate k for each P_j which minimises

$$\sum_{j \in S} f_j(\mathbf{p}_{jk}) + \alpha \sum_{(j, j') \in \omega} f_{jj'}(\mathbf{p}_{jk}, \mathbf{p}_{j'k'}), \quad (1)$$

where $\omega \subseteq S \times S$ and α affects the relative importance of parts and geometry matches.

If the graph is densely connected, such as that used by Donner et al. [8], then complex solvers are required. However, for simpler topology there are fast, guaranteed optimal solutions. We use a method which is similar to that used in [10], in which a network is created where each node can be thought of as having at most two parents. The optimal solution for this can be obtained in $O(mK^3)$ time, where K is the average number of candidates per part.

2.3 Voting for Sparse Correspondences

Now we describe a voting strategy to select the parts which contribute to effective models. The idea is that if a match candidate for a part is selected by many different parts+geometry models, it is likely to be the best match of the part (Fig. 1). Furthermore, if there are significantly more votes for the best match compared to alternatives, then it suggests that the part is a good one. On the other hand, if there are several candidate positions with similar numbers of votes, it suggests that the part is ambiguous. A part which consistently selects the “good” matches on many images is likely to have good localisability. We formalise this as follows.

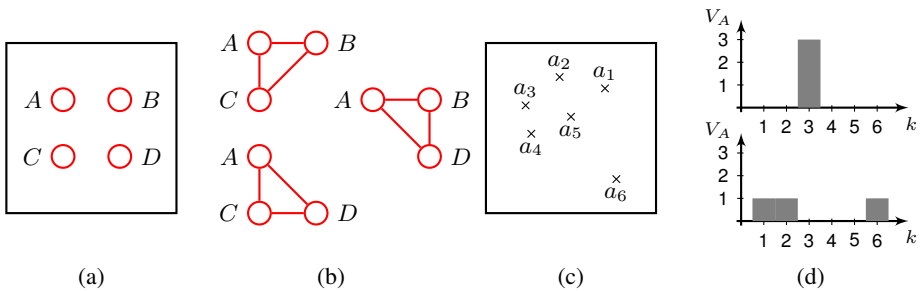


Fig. 1. (a) Consider four different parts A, B, C and D in an image. (b) Given $m = 3$, we can construct three different parts+geometry models with part A . (c) We use these models to search another image and each of them will localise a match candidate a_k for part A . (d) If all three models choose the same match candidate, say a_3 , then it is likely to be the best match of A and there will be a peak of votes at a_3 (top). If such kind of peak is observed on most of images, it suggests that part A is good at localisation. If the models do not agree with each other, the votes may be approximately uniformly distributed amongst the match candidates (bottom). It is thus hard to tell which one is the best match. And if this occurs quite often, we may conclude that part A cannot be reliably located.

Given a set of parts, we construct at least r different parts+geometry models for each part. Each such model contains a random number, m , of parts. We use each model to identify the best match on each image for each part, and record votes for each candidate match. Let V_{ijk} be the number of votes for candidate k of P_j on I_i —the number of times that the candidate k is selected as the best match for the part P_j . Let $T_{ij} = \sum_k V_{ijk}$ be the total votes cast by all parts+geometry models for the position of P_j on I_i . Then $L_{ijk} = V_{ijk}/T_{ij}$ is an estimate of the probability that candidate k is the correct match. The candidate $\mathbf{p}_{ij\hat{k}}$, where $\hat{k} = \arg \max_k L_{ijk}$, is then the best match

for P_j on I_i . If $L_{ij\hat{k}}$ is near 1, then it suggests that many different models voted for this location, and thus P_j has good localisability on I_i . If it is near zero, it suggests that the models voted for a range of different candidates, with no one standing out— P_j cannot be reliably located on I_i . The localisability of a given part can therefore be measured using $Q_j = \frac{1}{N_I} \sum_i L_{ij\hat{k}}$, the average probability of the best match over the set of images. We can then rank the parts by Q_j —those with the highest values are likely to be the most reliable.

2.4 Selecting a Subset of Parts

Assume now that we have ranked all parts by Q_j and renumbered so that $Q_j \geq Q_{j+1}$. We only consider the top M parts and discard the rest. We use the following greedy algorithm to select an initial subset \mathcal{P}_S .

1. $S \leftarrow \{1\}, j \leftarrow 2$
2. while ($Q_j > t_Q$ and $j \leq M$)
 - (a) if P_j does not overlap with any parts in \mathcal{P}_S , then $S \leftarrow j$
 - (b) $j \leftarrow j + 1$

We define two parts to be overlapping if their mean separation is less than the average radii of the two parts. This constraint ensures that the selected parts are well separated.

2.5 Evaluating a Set of Parts

Each part P_j defines a set of corresponding points, the one on each image that maximises $L_{ij\hat{k}}$. Thus \mathcal{P}_S defines a set of such points on each image. We measure the quality of the resulting correspondences by evaluating how effectively a model built from the images with those points can explain the original image set—essentially a description length. Such measures are widely used in groupwise non-rigid registration techniques [7][15]. We use the method presented in [16] to compute the description length, $\mathcal{L}(\mathcal{P}_S)$, of correspondences resulting from \mathcal{P}_S .

2.6 Extending the Model

To ensure only reliable parts are included in the initial \mathcal{P}_S , the threshold t_Q is set quite high, and thus some relevant areas of the image may be missed. We thus use the following steps to extend \mathcal{P}_S with extra parts if they can further improve the description length:

```
for  $j \leftarrow 1$  to  $M$ 
  if  $P_j$  does not overlap with  $\forall P_{j'} \in \mathcal{P}_S$  and  $\mathcal{L}(\mathcal{P}_S \cup \{P_j\}) < \mathcal{L}(\mathcal{P}_S)$ 
    then  $S \leftarrow j$ 
```

The result of this process is a set of parts, the number of which is automatically chosen by the algorithm. We use the set of parts to build a parts+geometry model. The matches of the model on each image define a sparse set of points, which can be used to initialise a dense groupwise correspondence algorithm (see [16] for details). The model can also be used to search new images to locate equivalent structures.

3 Annotating New Images

After groupwise non-rigid registration, each training image is annotated with a dense set of mesh control points, which define the correspondence between a model reference frame and the images. If the mesh points are well corresponded, we should be able to build models from them which can accurately label new images.

This can be achieved by virtue of the parts+geometry model learned in the training phase. We can use it to perform a global search to locate a sparse set of points on the new image. These points can be used to initialise denser models. Although we have achieved encouraging results using an Active Appearance Model (AAM) [4] trained on the data, we find that the global shape constraints in such models limit the final search accuracy in this case, and that further local refinement is necessary. However, we also find that the sparse points located by the parts+geometry model are sufficiently accurate to initialise a simple elastic mesh model similar to that used during the registration stage (see below), which can achieve annotation accuracy comparable to that of registration and still run quickly.

3.1 Elastic Mesh Model

We use an elastic mesh model to estimate the dense correspondence field. A triangular mesh is defined in the mean reference frame produced by the groupwise non-rigid registration. We use a Thin-Plate Spline to project the model mesh points from the reference frame to the target. We then use a variant of non-rigid registration to optimise the positions of the points of the mesh.

Let $\mathbf{z} = (x_1, y_1, x_2, y_2, \dots, x_N, y_N)^T$ be the positions of the mesh points. Let $\bar{\mathbf{z}}$ represent the positions of the control points in the reference image frame. The piecewise linear warp from the reference image to the target image is given by $\mathbf{y}' = W(\mathbf{y} : \mathbf{z})$. Let R be the region defined by the mesh in the reference image.

We define a quality of fit measure as

$$F(\mathbf{z}) = \sum_{\mathbf{y} \in R} |I(W(\mathbf{y} : \mathbf{z})) - I_R(\mathbf{y})| / \sigma_r + \beta F_s(\mathbf{z}) \quad (2)$$

where I_R is the reference image (constructed during registration of the training set), σ_r is an estimate of the noise of the residuals, $F_s(\mathbf{z})$ is a shape regularisation term and β is a weighting constant (set to 0.1 in all experiments). In the following we use a locally elastic form $F_s(\mathbf{z}) = \sum_{i=1}^N |\mathbf{z}_i - \hat{\mathbf{c}}_i|^2 / \sigma_z^2$, where σ_z^2 is the variance of $|\mathbf{z}_i - \hat{\mathbf{c}}_i|$, and $\hat{\mathbf{c}}_i$ is the position predicted for \mathbf{z}_i given the position of its neighbours. For an internal point in a regular mesh, this is the centre of gravity of its neighbours, elsewhere it is a suitably weighted sum of its neighbours.

A gradient descent approach is used to optimise the control point positions. The derivatives of $F(\mathbf{z})$ can be calculated efficiently by displacing each point in turn, as moving one point only affects the nearby triangles. During the training phase we generated a sequence of reference images of increasing resolution and accuracy as the registration progressed. Here we again use a multi-resolution approach, optimising the mesh so as to match to each reference image (increasing in resolution) in turn.

4 Experimental Results

We demonstrate the approach on three different datasets² of increasing difficulty:

- (1) 200 digital micrographs of female fly wings. Each image has a size of 1280×1022 pixels and is marked with 15 points by human expert (Fig. 2a);
- (2) 200 radiographs of the hands of children (aged between 10-13), taken as part of study into bone ageing. The image size varies across the set, with a height ranging from 1000 to 1700 pixels. Each image has 37 manually labelled landmarks. The resolution of this set of images is 0.2 mm per pixel (Fig. 2b);
- (3) 200 radiographs of the lumbar spine. The image height varies from 1500 to 1700 pixels. Each image has 337 manual landmarks placed around the outline of the vertebrae. This set of images has a resolution of 0.1694 mm per pixel (Fig. 2c).

Examples of these three datasets and their manual landmarks are given in Fig. 2. In all experiments the manual annotations are only used to evaluate the performance of the automatic system.

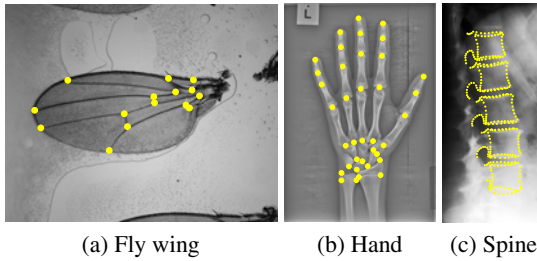


Fig. 2. Examples of the datasets and associated landmarks used in the experiment

4.1 Performance of the Voting Based Method

We compare our voting based method with the GA based method [16]. We randomly selected 100 images from the three sets respectively as the training sets. We used the method presented in [16] to generate the candidate parts and search for their match candidates. We automatically constructed over 1000, 1900 and 900 part models for the fly wings, hands and spines, and retained the best 250, 500 and 100 respectively. We fed those three subsets to the GA based method to select the best parts+geometry models, for a range of different numbers of parts. The resulting sparse points were used to initialise a groupwise non-rigid registration algorithm [5].

To evaluate the accuracy of the result we compare with a manual annotation. We used the resulting dense correspondences to warp each set of manual landmarks into a reference frame, computing their mean. We then projected this mean back to each individual

² Dr. Chris Klingenberg provided the *Drosophila* wing images and markup. The hand and spine radiographs were provided by Prof. Judith E. Adams, Dr. Elisa M. Pacheco and Dr. Teik Oh, with annotations from Dr. Steve A. Adeshina, Dr. Martin G. Roberts.

image and calculated the mean absolute difference between each mean point and the original manual annotation. For the spines, we computed the mean distance between each mean point and the nearest curve on the manual annotation—due to “sliding” along curves in the manual annotation. This is found to be a more useful measure of the performance than direct point-to-point distance.

For the voting based method, we consider the whole sets of candidate parts. We generated a number of parts+geometry models with random m , repeating until each part has been considered in at least $r = 100$ different models. We computed the average probability measure, Q , for each part, and retained those with $Q > t_R$ (Fig. 3a). We chose an initial subset of non-overlapping parts with $Q > t_Q$ (Fig. 3b), and extended the set with further parts which may better encode the entire training set in terms of an MDL principle (Fig. 3c). In the experiment we set $t_R = 0.9$ and $t_Q = 0.95$ for the hands and fly wings, $t_R = 0.5$ and $t_Q = 0.55$ for the spines (lower thresholds are required for this challenging dataset as few parts give consistently high values of Q). To evaluate the merit of the added parts, we constructed two parts+geometry models: one is built from the initial set (Fig. 3b) and the other is from the extended set (Fig. 3c-d).

Table 1 shows the accuracy of the correspondence on the three datasets after the dense registration, initialised using different methods. The voting based method outperforms both the simple affine initialisation and the GA based method in each case. An interesting observation is that for the fly wings, affine can give quite good correspondences. However, a significant improvement can still be obtained even when using a 5-part parts+geometry model. This demonstrates that initialisation is crucial to the

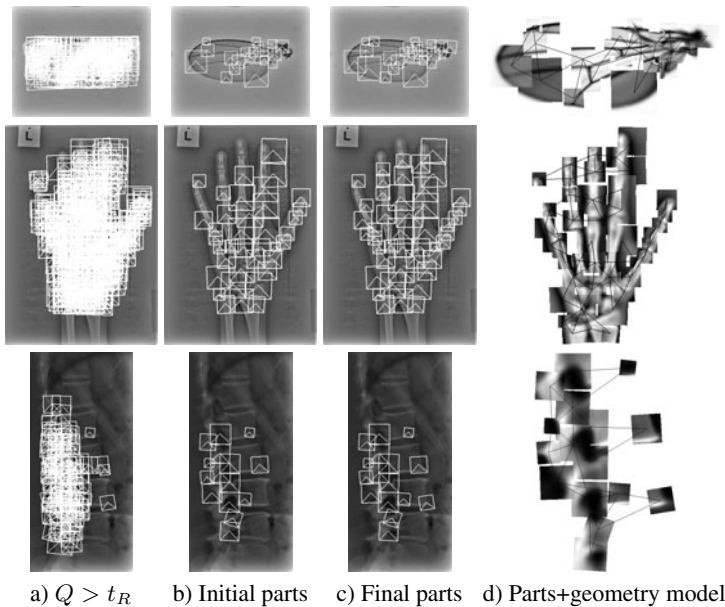


Fig. 3. Results of the voting based method on the three datasets. From top to bottom, fly wing, hand and spine.

Table 1. Point location errors of the dense correspondence

a) Fly wings (pixels)

b) Hands (mm)

Method	m	Mean \pm s.e.	Med.	90%	Method	m	Mean \pm s.e.	Med.	90%
Affine	-	3.5 \pm 1.7	1.8	2.5	Affine	-	18.5 \pm 2.0	9.9	41.8
GA	5	2.2 \pm 0.09	1.8	3.4	GA	10	1.4 \pm 0.08	1.2	2.6
	10	1.8 \pm 0.09	1.5	2.7		20	1.2 \pm 0.08	0.9	2.0
	15	1.6 \pm 0.1	1.4	2.2		30	1.2 \pm 0.09	0.9	1.9
	20	1.8 \pm 0.09	1.5	2.8		40	1.0 \pm 0.05	0.8	1.5
	25	1.9 \pm 0.08	1.7	3.1		50	1.0 \pm 0.06	0.8	1.6
Voting	Initial (19)	1.2 \pm 0.03	1.1	1.6	Voting	Initial (37)	1.1 \pm 0.06	0.9	1.5
	Extended (25)	1.2\pm0.03	1.2	1.5		Extended (44)	0.9\pm0.03	0.8	1.2

c) Spines (mm)

Method	m	Mean \pm s.e.	Med.	90%
Affine	-	37.6 \pm 2.9	28.8	74.8
GA	5	6.7 \pm 0.9	2.4	25.0
	10	3.2 \pm 0.5	1.9	4.1
	15	4.1 \pm 0.5	2.6	5.1
	20	4.1 \pm 0.5	2.6	5.2
Voting	Initial (15)	3.4 \pm 0.4	2.2	4.6
	Extended (16)	3.0\pm0.4	2.1	3.8

performance of groupwise non-rigid registration, and that a sophisticated initialisation is desirable even for those images where a simple one may appear to work.

For the fly wings we find that the parts+geometry model built from the initial set works as well as the extended model. However the extended model works better on the hands and spines. This indicates that for the objects with limited shape variation, such as the the fly wings, a reasonable choice of t_Q is good enough to select useful parts. The resulting parts+geometry model can well capture the morphological features of the object. Adding more parts using the MDL criterion does not impair the localisability of the model, but introduces redundant information and increases its complexity. When the object consists of considerable morphological variation, the MDL principle, however, does help improve the representative ability of the geometric model and thus localisability. An example can be seen in the spine experiment. Compared with the initial set of parts, only one more part is added to the final set, leading to a more detailed representation of local shape and appearance changes and a more accurate result.

We give examples of the projection of the mean annotation onto individual images in Fig. 4. This shows typical results of the performance of the models computed with the voting based algorithm. The method can deal well with the challenging hand set, though mistakes may be occasionally made on some images. For the spine set, we get a good overall fit in many images, though it is much more challenging as the shape of vertebrae can change dramatically and large rotation exists. The model works well on every image in the fly wing set.

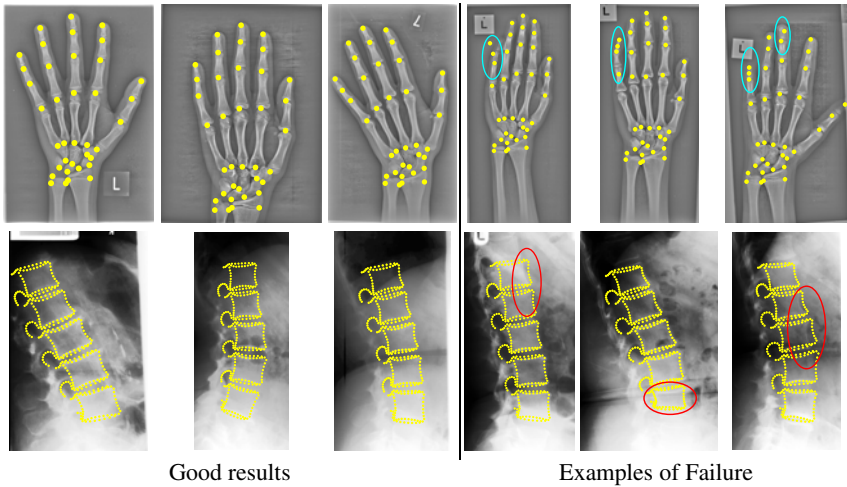


Fig. 4. Projection of average points onto different images in the hand and spine sets (using the parts+geometry models built from the extended sets). Failures for the hand set are indicated by cyan ellipses while those for the spines are indicated by red ellipses.

4.2 Influence of Important Parameters

To investigate the effect of the two parameters in the voting based method—the lower bound of the number of models per part, r , and the number of images used to construct the models N_I —we repeated the above experiment with the same set of images on the hand database. We first used all images as the training set and ran the algorithm again for a range of different r . We then fixed $r = 100$ and ran the algorithm on different numbers of images. The resulting parts+geometry models were used to find correspondence on the rest of the images.

Figure 5 shows the impact of the two parameters on the accuracy of the resulting dense correspondence. Even with r as low as 10 the voting based method can still give good results. The running time of the voting stage is linear in r . Compared with the best result obtained using the GA based algorithm (40 parts), our method only needs 2.5 hours to run with $r = 10$ on an 8-node IBM Bladecentre cluster (Xeon 2.0GHz) while the GA one needs around 9 hours. As there is a linear relationship between the algorithm and the number of the training images, using fewer images can also speed up the algorithm. A similar pattern is also found on the fly wing dataset.

4.3 Annotating Unseen Images

We used the same training sets used in Sect. 4.1 to build the statistical models described in Sect. 3 and used them to annotate separate test sets, each containing 100 images. To evaluate the performance of the elastic mesh model, we warped the manual annotation of each training image to a reference frame, computing the mean. We then projected the mean onto each test image using the deformation field defined between the reference frame and the found mesh points. As before, we calculated the point-to-point location

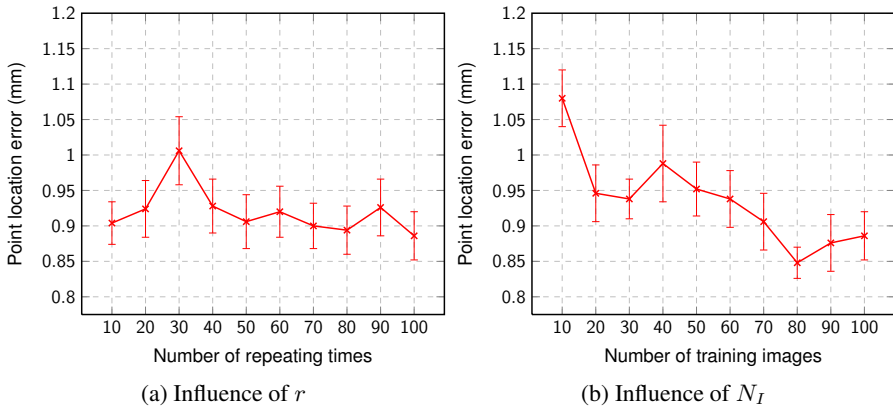


Fig. 5. The influence of the two parameters on the accuracy of groupwise non-rigid registration initialised by the voting based algorithm.

Table 2. Annotation errors for the unseen images

Dataset	Mean \pm s.e.	Med.	90%
Fly wings (pixels)	1.5 \pm 0.2	1.2	2.7
Hands (mm)	1.4 \pm 0.1	1.4	2.3
Spines (mm)	3.9 \pm 0.6	2.2	6.0

error for the fly wings and hands, and point-to-curve error for the spines. The results are shown in Table 2. The accuracy of annotation is comparable to that of registration, demonstrating that the dense correspondences are well established.

5 Conclusions and Future Work

We have described a method that can effectively initialise groupwise non-rigid registration on datasets of simple or complex structures. This is achieved by first selecting a subset of parts with good localisability from a large number of candidates parts using a voting strategy. A parts+geometry model is then built with the chosen parts and is used to find sparse correspondences across the images. The basic algorithm is quite general, and could work with any form of part model.

Extensive experiments show that our algorithm is able to achieve better registration than the state of the art on three different datasets, two of them being particularly challenging. It can also run much faster without loss of accuracy.

To further demonstrate the performance of the approach we use the resulting dense correspondences to build statistical models to label new images. We show that robust, accurate annotations can be obtained by using a simple elastic mesh model initialised by the parts+geometry models constructed by the method. The annotation accuracy is similar to that obtained on the training set by dense registration.

Although simply setting threshold is working well for all datasets reported, in future we will investigate more sophisticated method to select the best subset of parts. More highly connected graphs will also be studied as currently we use a relatively simple topology for the graphical model in order to allow efficient global optimisation.

References

1. Adeshina, S.A., Cootes, T.F.: Constructing part-based models for groupwise registration. In: Proceedings of ISBI, pp. 1073–1076 (2010)
2. Balci, S.K., Golland, P., Shenton, M., Wells, W.M.: Free-form B-spline deformation model for groupwise registration. In: Proceedings of MICCAI Statistical Registration Workshop, pp. 23–30 (2007)
3. Bhatia, K.K., Hajnal, J.V., Puri, B.K., Edwards, A.D., Rueckert, D.: Consistent groupwise non-rigid registration for atlas construction. In: Proceedings of ISBI, pp. 908–911 (2004)
4. 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)
5. Cootes, T.F., Twining, C.J., Petrović, V., Babalola, K.O., Taylor, C.J.: Computing accurate correspondences across groups of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 1994–2005 (2010)
6. Crandall, D.J., Huttenlocher, D.P.: Weakly supervised learning of part-based spatial models for visual object recognition. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 16–29. Springer, Heidelberg (2006)
7. Davies, R.H., Twining, C.J., Cootes, T.F., Taylor, C.J.: A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging* 21, 525–537 (2002)
8. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: *BMVC*, vol. 2, pp. 1080–1089 (2007)
9. Donner, R., Wildenauer, H., Bischof, H., Langs, G.: Weakly supervised group-wise model learning based on discrete optimization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 860–868. Springer, Heidelberg (2009)
10. Felzenszwalb, P.F., Huttenlocher, D.P.: Representation and detection of deformable shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(2), 208–220 (2005)
11. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: Proceedings of CVPR (2003)
12. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
13. Langs, G., Donner, R., Peloschek, P., Bischof, H.: Robust autonomous model learning from 2D and 3D data sets. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 968–976. Springer, Heidelberg (2007)
14. Lowe, D.: Object recognition from scale invariant features. In: Proceedings of ICCV, vol. 2, pp. 1150–1157 (1999)
15. Twining, C.J., Cootes, T.F., Marsland, S., Petrovic, V., Schestowitz, R., Taylor, C.J.: A unified information-theoretic approach to groupwise non-rigid registration and model building. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 1–14. Springer, Heidelberg (2005)
16. Zhang, P., Adeshina, S.A., Cootes, T.F.: Automatic learning sparse correspondences for initialising groupwise registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 635–642. Springer, Heidelberg (2010)

Temporal Groupwise Registration for Motion Modeling

Mehmet Yigitsoy*, Christian Wachinger*, and Nassir Navab

Computer Aided Medical Procedures (CAMP), Technische Universität München
{yigitsoy, wachinger, navab}@in.tum.de

Abstract. We propose a novel method for the registration of time-resolved image sequences, called Spatio-Temporal grOupwise non-rigid Registration using free-form deforMations (STORM). It is a *groupwise* registration method, with a group of images being considered simultaneously, in order to prevent bias introduction. This is different from pairwise registration methods where only two images are registered to each other. Furthermore, STORM is a *spatio-temporal* registration method, where both, the spatial and the temporal information are utilized during the registration. This ensures the smoothness and consistency of the resulting deformation fields, which is especially important for motion modeling on medical data. Moreover, popular free-form deformations are applied to model the non-rigid motion. Experiments are conducted on both synthetic and medical images. Results show the good performance and the robustness of the proposed approach with respect to outliers and imaging artifacts, and moreover, its ability to correct for larger deformation in comparison to standard pairwise techniques.

1 Introduction

In radiation therapy, target localization is one of the biggest challenges, since the exposition of healthy tissue to ionizing radiation should be kept as low as possible [5]. Organ motion due to respiration, however, can lead to inaccuracies during target localization [8]. Modeling and analyzing organ motion is therefore important for such applications. Thanks to dynamic imaging, which provides time-resolved images of the anatomy, the modeling of organ motion becomes feasible. Time-resolved images are extensively used to study cardiac [10, 11, 16], lung [3, 4, 8] and liver [17, 26] motion.

The creation of motion models necessitates the non-rigid alignment of time-resolved volumes. The standard approach is, first, to perform pairwise registrations, and second, to combine the deformation fields to create the motion model. This can result in inconsistent and non-smooth motion fields along the temporal direction [22]. In applications, such as radiotherapy, it is, however, crucial to ensure smoothness and consistency over time to estimate the delivered dose accurately at any time point in the image sequence [14]. Further, the pairwise registration approach has the disadvantage that either, all registrations are performed towards one target leading to large displacements fields, or, registrations between adjacent volumes are calculated causing the accumulation of errors in the image sequence [4].

Methods that address the issue of incorporating temporal information [1, 13] still have the problem of fixed reference image selection. Although there are spatio-temporal

* Joint first authors.

methods that avoid the selection of a reference image in the sequence, either their main focus is to register two different image sequences [15, 16], or to register two image sequences with the second being the replication of the reference image [22]. On the other hand, groupwise registration techniques are proposed for atlas construction that address the choice of the reference image [9, 2, 30]. Although, their application areas are quite different from motion modeling, the main goal of these approaches is to reduce the bias introduced by a fixed reference image [9, 6].

In this paper, we propose a new registration method for the deformable alignment of time-resolved images. Instead of applying pairwise registrations, we perform a fully simultaneous, groupwise registration of all images in the sequence. This eliminates the bias introduced by reference image selection and the error accumulation during the sequential registration. The second novelty of our approach addresses the need for a smooth deformation field. Instead of having a separate deformation field for each image, we create one deformation field that is one dimension higher than the image dimension. Since we work with free-form-deformations (FFD) based on B-splines, this intrinsically ensures that we obtain a smooth deformation field at each update step, also along the temporal direction. This leads to further advantages for the registration method because it is more robust to outliers, can handle large deformations, and allows for temporal interpolation, which differentiates it from existing approaches.

2 Related Work

Approaches aiming at the registration of time-resolved image sequences mainly focus on 4D modeling of organ motion of heart, liver or lung. Klein et al. [11] use 4D deformable image registration to compensate for the cardiac motion in positron emission tomography (PET) images, following a pairwise approach. Rohlfing et al. [17] apply intensity-based deformable registration on gated magnetic resonance (MR) images to model liver motion during a respiratory cycle. They register each volumetric phase image to a reference image by using the displacement field computed for the previous volumetric image as the initial deformation. Sarrut et al. [20] simulate a 4D computed tomography (CT) image of the thorax by first registering end-diastole and end-systole images and then by linearly interpolating images in between. These methods either do not use temporal information or fix a reference frame during the registration. Sundar et al. [22] addresses the temporal smoothness problem by considering motion estimation from cine MRI cardiac images as the registration of two 3D image sequences. The second sequence was, however, created by replicating the end-diastolic image. This means that although this method addresses the problem of temporal smoothness, the existence of the reference image still remains an issue. Wu et al. [29] propose a groupwise registration method for longitudinal image sequences where a smoothing kernel is used to assure temporal consistency.

Simultaneous registration approaches align groups of images without the necessity of reference selection. Learned-Miller [12] proposes the congealing framework for image registration, which sums up the entropy of pixel stacks over the image. Zöllei et al. [30] apply this method to the affine registration of population data. Balci et al. [2] extend this method by including free-form deformations to the deformable registration

of a group of brain images. Wachinger et al. [27] perform simultaneous registration by accumulating pairwise estimates for the similarity estimation. All these methods for simultaneous registration are not designed for motion estimation, and consequently, do not take temporal information into account. Bystrov et al. [3] try to address both aspects, temporal smoothness and reference image selection, but the method is limited to local motion trajectories.

3 Method

Assuming n images I_1, \dots, I_n , with $I_i: \Omega \rightarrow \mathbb{R}, \Omega \subset \mathbb{R}^N$, we stack the images and create an image I of dimension $N + 1$, $I = [I_1, I_2, \dots, I_n]$, with $I: \Omega \times \{1, \dots, n\} \rightarrow \mathbb{R}$. The transformation $\mathbf{T}: \mathbb{R}^{N+1} \rightarrow \mathbb{R}^{N+1}$ is defined as

$$\mathbf{T}_{\Phi}(\mathbf{x}) = \mathbf{x} + \mathbf{D}_{\Phi}(\mathbf{x}) \quad (1)$$

with the spatial coordinate $\mathbf{x} = [x_1, x_2, \dots, x_{N+1}] \in I$ and \mathbf{D}_{Φ} the deformation field with deformation parameters Φ . We denote the warped images with $I^{\downarrow}(\mathbf{x}) = I(\mathbf{T}(\mathbf{x}))$. In the following sections, we describe the main components of our registration method: transformation model, similarity metric, and optimization method.

3.1 Transformation Model

We use FFD B-Splines to model non-rigid motion [19], with an additional dimension along the temporal direction. For the temporal direction, we can set B-Spline functions of a different order or we can adapt the spacing between the control points to model the desired smoothness along the temporal direction. Φ is defined as a grid of $n_1 \times n_2 \times \dots \times n_{N+1}$ control points on image I with spacings $\delta_1, \delta_2, \dots, \delta_{N+1}$. We access a control point with $\mathbf{i} = [i_1, i_2, \dots, i_{N+1}]$. Further, using cubic B-splines we obtain

$$D_{\Phi}(\mathbf{x}) = \sum_{a_1=0}^3 \sum_{a_2=0}^3 \dots \sum_{a_{N+1}=0}^3 B_{a_1}(r_1) B_{a_2}(r_2) \dots B_{a_{N+1}}(r_{N+1}) \cdot \Phi(\mathbf{i} + \mathbf{a}) \quad (2)$$

where

$$i_1 = \left\lfloor \frac{x_1}{\delta_1} \right\rfloor - 1, i_2 = \left\lfloor \frac{x_2}{\delta_2} \right\rfloor - 1, \dots, i_{N+1} = \left\lfloor \frac{x_{N+1}}{\delta_{N+1}} \right\rfloor - 1, \quad (3)$$

$$r_1 = \frac{x_1}{\delta_1} - \left\lfloor \frac{x_1}{\delta_1} \right\rfloor, r_2 = \frac{x_2}{\delta_2} - \left\lfloor \frac{x_2}{\delta_2} \right\rfloor, \dots, r_{N+1} = \frac{x_{N+1}}{\delta_{N+1}} - \left\lfloor \frac{x_{N+1}}{\delta_{N+1}} \right\rfloor, \quad (4)$$

B_0 to B_3 are cubic B-spline basis functions, and $\mathbf{a} = [a_1, a_2, \dots, a_{N+1}]$. For the interpolation of $N + 1$ dimensional data, we have $\Phi(\mathbf{i}) = (\phi_{x_1}^{\mathbf{i}}, \phi_{x_2}^{\mathbf{i}}, \dots, \phi_{x_{N+1}}^{\mathbf{i}})^{\top}$.

3.2 Accumulated Pairwise Estimates

We apply accumulated pairwise estimates (APE) [27] as similarity measure. It was introduced as a framework for groupwise similarity metrics. In order to describe it in the scope of this temporally extended framework, we use the following notation. As before,

we consider \mathbf{x} to be a point in $\Omega \times \{1, \dots, n\}$. Further, be $\mathbf{v} \in \Omega$ and $g, h \in \{1, \dots, n\}$. The point \mathbf{x} that is related to the spatial location \mathbf{v} and the temporal location g is denoted by $\mathbf{x}_v^g = [\mathbf{v}, g]$. Therefore, we can access the same pixel by $I_g(\mathbf{v}) = I(\mathbf{x}_v^g)$, leading to the cost function

$$\mathcal{C}(\Phi) = \sum_{g=1}^n \sum_{\substack{h=1 \\ h \neq g}}^n \|I_g^\downarrow - I_h^\downarrow\|^2 \quad (5)$$

$$= \sum_{g=1}^n \sum_{\substack{h=1 \\ h \neq g}}^n \sum_{\mathbf{v} \in \Omega} (I(\mathbf{T}\Phi(\mathbf{x}_v^g)) - I(\mathbf{T}\Phi(\mathbf{x}_v^h)))^2. \quad (6)$$

3.3 Optimization

Simultaneous registration leads to a computationally challenging optimization problem because of the increase of the parameter space and the more complex evaluation of the similarity metric. For this reason, we use a gradient-based optimization procedure in combination with a stochastic sampling in the spatial domain [25, 30, 2]. The stochastic sampling significantly reduces the computational cost, where we randomly select 10% of the total pixels in Ω in each iteration on which the similarity is calculated. It is further noted in [25] that the noisy estimate of the gradient can reduce the problem of local minima. The update equation is

$$\Phi \leftarrow \Phi + \tau \cdot \nabla \mathcal{C} \quad (7)$$

with the step length τ , which is determined by a line search following the Wolfe conditions. The partial derivative with respect to control point ϕ^i along direction x_1 , with other directions analogously, is

$$\frac{\partial \mathcal{C}(\Phi)}{\partial \phi_{x_1}^i} = \frac{\partial}{\partial \phi_{x_1}^i} \sum_{g=1}^n \sum_{\substack{h=1 \\ h \neq g}}^n \sum_{\mathbf{v} \in \Omega} (I^\downarrow(\mathbf{x}_v^g) - I^\downarrow(\mathbf{x}_v^h))^2 \quad (8)$$

$$\frac{\partial (I^\downarrow(\mathbf{x}_v^g) - I^\downarrow(\mathbf{x}_v^h))^2}{\partial \phi_{x_1}^i} = 2(I^\downarrow(\mathbf{x}_v^g) - I^\downarrow(\mathbf{x}_v^h)) \frac{\partial (I^\downarrow(\mathbf{x}_v^g) - I^\downarrow(\mathbf{x}_v^h))}{\partial \phi_{x_1}^i} \quad (9)$$

$$\frac{\partial (I^\downarrow(\mathbf{x}_v^g) - I^\downarrow(\mathbf{x}_v^h))}{\partial \phi_{x_1}^i} = \left(\nabla_{x_1} I^\downarrow(\mathbf{x}_v^g) \frac{\partial \mathbf{T}\Phi(\mathbf{x}_v^g)}{\partial \phi_{x_1}^i} - \nabla_{x_1} I^\downarrow(\mathbf{x}_v^h) \frac{\partial \mathbf{T}\Phi(\mathbf{x}_v^h)}{\partial \phi_{x_1}^i} \right) \quad (10)$$

$$\frac{\partial \mathbf{T}\Phi(\mathbf{x}_v^g)}{\partial \phi_{x_1}^i} = B_{a_1}(r_1^g) B_{a_2}(r_2^g) \cdots B_{a_{N+1}}(r_{N+1}^g) \quad (11)$$

with $r_1^g, r_2^g, \dots, r_{N+1}^g$ as defined in equation (4) with respect to the location \mathbf{x}_v^g . We position the images along the temporal direction, reflecting its acquisition time. Alternatively, one could decide to use different criteria to position the images along the temporal direction, such as amount of change. For some applications it may further be useful to allow for deformations along the temporal direction, if there is uncertainty about the exact acquisition time, or different parts of the image were acquired at different time points. For our application, we do not need this flexibility so that we set the update along the temporal direction to zero.

3.4 Implementation Details

We apply a multi-resolution approach to reduce computational cost and increase robustness to local minima. Further, we apply a multi-grid approach, with several grid refinements on each image resolution, to capture large displacements, as well as, fine details. In order to guarantee smoothness and consistency over time, we use a single control point grid of one higher dimension and overlay this onto the $(N + 1)$ D image. For the temporal dimension, we use fewer control points than the number of images to increase the dependency between the images and, at the same time, to further decrease the computational cost. The software is developed using open-source libraries ITK and VTK. The application will be made publicly available in the near future.

3.5 Advantages of STORM for Motion Modeling

After having described the method, we want to point out the advantages of STORM for motion modeling.

1. **Smooth and Temporally Consistent Deformation Field:** Current approaches perform first the registration of the images, and then the combination and regularization of the transformation fields. The problem is that for the regularization step, the image information is no longer considered, and therefore transformation fields may be created that do not match to the underlying images. In contrast, our single B-spline deformation field across all the images intrinsically guarantees a smooth and consistent transformation along the temporal direction in each update step. Similarity and regularization aspects are jointly optimized.
2. **Robustness Against Outliers:** Dynamic imaging is more challenging than its static counterpart, causing more artifacts and noise in the images. Our method is robust to such effects because the $(N + 1)$ D deformation field interlaces neighboring images, with the number of control points n_{N+1} determining to which extent. Images containing outliers can therefore still be correctly aligned, if neighboring images drive the registration.
3. **Handling Large Displacements:** During the registration of time-resolved images, one of the main issues is the registration of distant images. If the displacement between the images is too large, pairwise methods fail to find the mapping. By incorporation temporal information and using groupwise registration, STORM is able to better handle large displacements in the sequence.
4. **Time Interpolation:** In some applications, it is desirable to see the organ at a specific state of the motion [7, 21]. However, due to the low temporal resolution of time-resolved images, this is not always possible. Thanks to the higher dimensional deformation field of STORM, it is feasible to resample images for any time point by applying the inverse mapping.
5. **Tracking Using Deformation Field:** In 4D radiation therapy, it is necessary to extract motion trajectories of a tumor region, in order to perform the pre-operative dosimetry planning [8]. Our approach enables tracking of certain structures by clipping the $(N + 1)$ D deformation field along the temporal dimension.

4 Experiments and Results

In order to evaluate the performance of STORM, we experiment with temporally-related synthetic and medical data sets. In addition to using visual inspection for qualitative evaluation, mean and standard deviation (STD) along the temporal direction are computed for the quantitative validation. Due to space limitations, every second image is shown for each sequence.

4.1 Synthetic Data

Synthetic image sequences are created by applying gradually increasing deformations to a template image to simulate a smooth deformation over time, reflecting a smooth motion. The experiments are conducted to demonstrate the advantages listed in section 3.5. For each experiment a multi-resolution setting with 30 iterations on the highest level that is increased by a factor of 1.2 for each lower level is used together with 8 control points in spatial dimensions and 7 control points in the temporal dimension.

Growing Ring

The input images shown in Fig. 1 are binary images of a continuously growing ring of size 100×100 pixels. The registered images are also shown in the same figure. We observe that all images are correctly aligned to the initially unknown target. The wireframe meshes are presented to visualize the deformations. The direction of the displacements is encoded in the meshes with the hue component. Mean and STD images, calculated along the temporal direction, confirm the correct registration. The mean image clearly

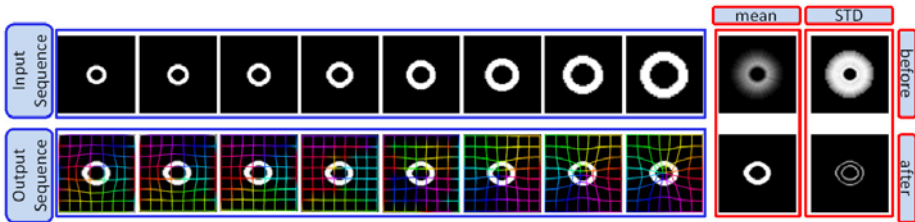


Fig. 1. Input and output sequences together with the statistics images for growing ring experiment

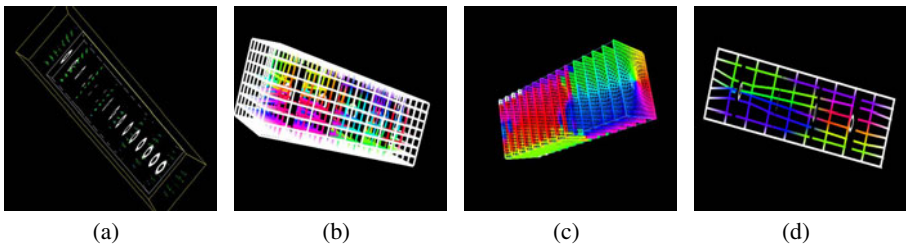


Fig. 2. (a) $(N + 1)$ D image. (b) $(N + 1)$ D wireframe mesh. (c) ND meshes in $(N + 1)$ D. (d) A cut-plane from the $(N + 1)$ D mesh along the temporal direction.

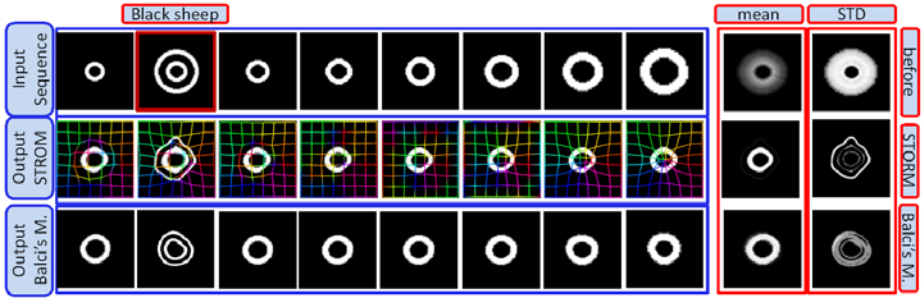


Fig. 3. Input and output sequences together with the statistics images for the black sheep experiment using our method and Balci’s method

shows the ring, and the STD image only shows small deviations on the edges of the ring caused by interpolation artifacts. The $(N + 1)$ D image and the control point grid are shown in Figs. 2(a) and 2(b), respectively. The visualization of the ND meshes in $(N + 1)$ D is shown in Fig. 2(c). The deformation along the temporal direction can be observed by clipping the $(N + 1)$ D deformed mesh along the time-axis as shown in Fig. 2(d). The smoothness and consistency of the deformation along the temporal dimension in this figure, as well as, the smoothness of the change of colors in Fig. 2(c) demonstrates the advantages of our methods described in 3.5.1 and 3.5.5.

Black Sheep

In this experiment, the same growing ring sequence is used with an outlier image introduced to create a ‘black sheep’ in the sequence. The input images are shown in Fig. 3. The registration result that we expect, having the smoothness of motion in mind, is the alignment of the inner ring with the others, while treating the outer ring as a separate structure. We compare our method to the groupwise registration method proposed by Balci et al. [2]. The results for the two methods are shown in Fig. 3. Looking at the final alignment in Balci’s case, it is clear that it does not fulfill the expectations since the outer ring is aligned with the others instead of the inner, real one. In our case, however, the outer ring is identified correctly and the inner one is aligned with the others. These findings are also supported by the STD and mean images shown in the same figure. The reason for the failure of Balci’s method and the success of our method is the consideration of the temporal relationship during the registration, making the method more robust against outliers, as noted in 3.5.2.

First vs. Last

In this case, STORM is tested on a data set with large displacements between the first and last image to model large motion deformations. A randomly created B-spline deformation field is asymmetrically applied to a template image of size 100×100 pixels. The goal of this experiment is to show the advantage mentioned in 3.5.3 and the insufficiency of pairwise registration in such cases. In Fig. 4, the input sequence is shown

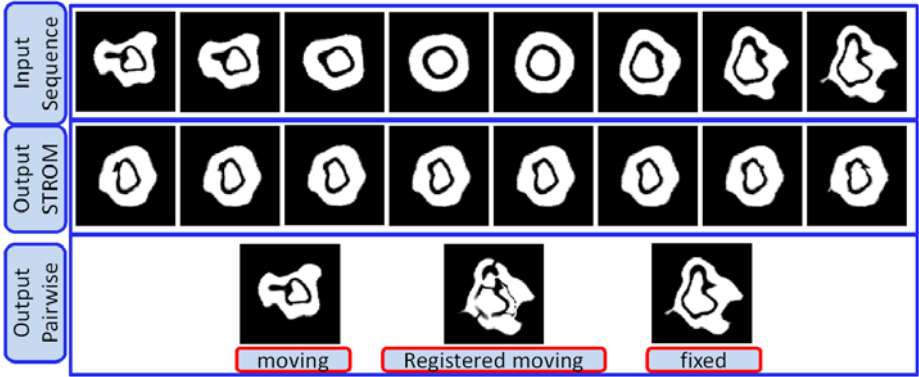


Fig. 4. Input and output sequences together with the statistics images for first vs. last experiment. Last row: Results for experiment with pairwise registration method.

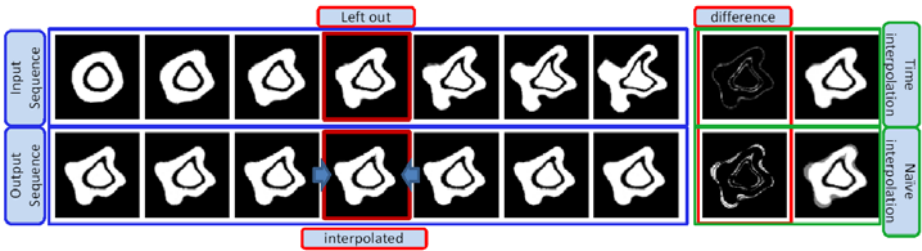


Fig. 5. Illustration of interpolation in time using STORM. Time interpolation and naïve interpolation are compared by showing the difference images to the original image.

where the middle image is used for creating the sequence. The results in the same figure show that the images are aligned up to some small structures which are difficult for many registration algorithms using FFD's. For the pairwise registration, after trying different implementations of demons algorithm available in ITK, we decided to use the diffeomorphic demons proposed by Vercauteren et al. [24], since it achieved the best performance. After registering the first image (moving) to the last image (fixed) in the sequence, we show the deformed moving image in Fig. 4. The dissimilarity between the registration result and the fixed image shows that the large deformation is difficult to handle by pairwise methods.

Time Interpolation

In this experiment, we want to illustrate the advantage mentioned in 3.5.4; the interpolation of images along the temporal dimension. For this we use the sequence in Fig. 5 where we omit the indicated image and try to reconstruct it. First, the registration of the reduced sequence is performed. Second, the temporally adjacent images to the reconstructed time point are selected. Third, a linear intensity interpolation between the neighboring, deformed images is performed - with more emphasis to the temporally closer image. Fourth, the inverse transformation is applied to the interpolated image [18], creating the reconstruction. For validation, we compare the reconstructed

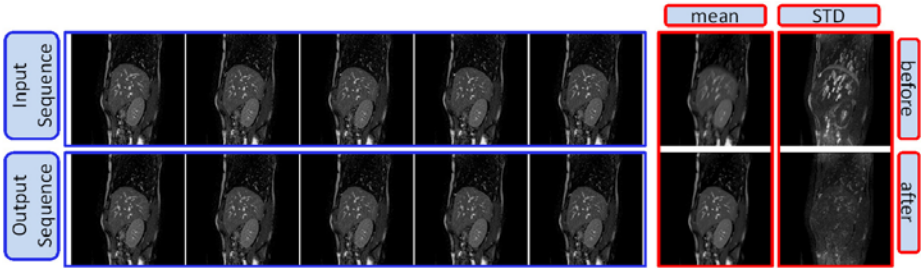


Fig. 6. Experiment with MRI sequence. Input, output sequences and mean, STD images before and after the registration.

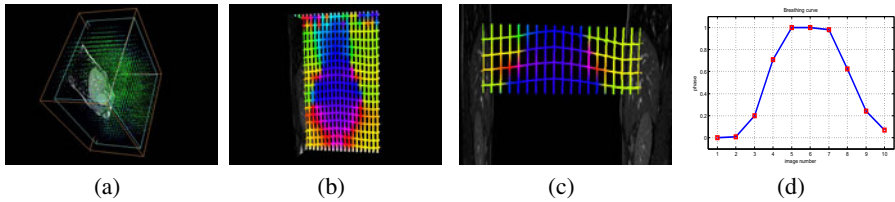


Fig. 7. (a) $(N + 1)$ D deformation field. (b) A temporal slice from $(N + 1)$ D mesh. (c) A closer look into the mesh in (b). (d) The breathing curve for the data set. Note the similarity between the curves in (c) and the curve in (d). The color coded deformation on the mesh shows the direction of motion along the time-axis. Individual lines can be seen as the local motion paths.

image to the original image in the sequence by calculating the difference image. Further, we perform a naïve reconstruction by intensity interpolation of the input images. The results clearly show the good performance of the proposed time interpolation.

4.2 Medical Data

Two experiments have been conducted. First, a time-resolved image sequence consisting of 10 MR images is used. The images are so called “navigator slices” that are used for 4D MRI reconstruction in [26]. The images of size 255x255 pixels and temporal resolution 2.7 Hz are acquired from the liver using the sagittal plane with Philips 1.5T Achieva during one respiratory cycle. The second experiment is done with a CT sequence extracted from the POPI-model [23] which consists of 10 volumes corresponding to phases of a respiratory cycle¹. The images are of size 482x141 pixels and correspond to the sagittal planes in the posterior thoracic region. The same registration settings are used with an additional multi-level grid scheme.

MRI Sequence

An excerpt from the 10 images in the sequence is shown in Fig. 6. The first and the last images in the sequence correspond to the exhale state of the respiration. We perform registration on the sequence using STORM, with the results being presented in

¹ The data was obtained from the the Léon Bérard Cancer Center & CREATIS lab, Lyon, France. <http://www.creatis.insa-lyon.fr/rio/popi-model>

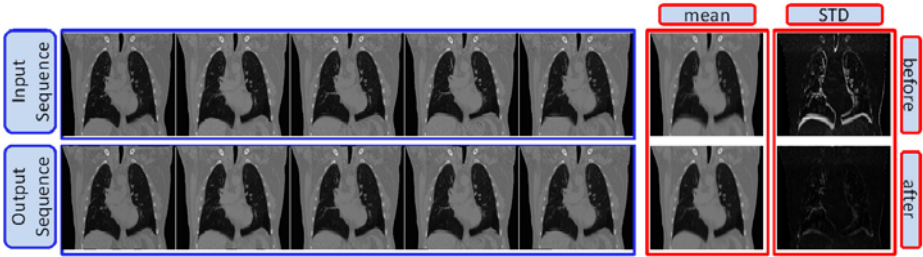


Fig. 8. Experiment with CT sequence. Input, output sequences and mean, STD images before and after the registration.

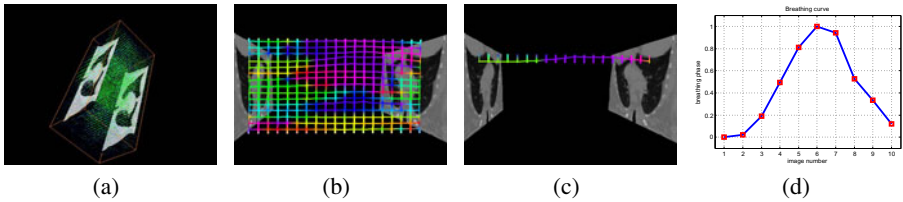


Fig. 9. (a) $(N + 1)$ D deformation field. (b) A temporal slice from $(N + 1)$ D mesh. (c) A closer look into the mesh in (b). (d) The breathing curve for the data set. Note the similarity between the curves in (c) and the curve in (d). The color coded deformation on the mesh shows the direction of motion along the time-axis. Individual lines can be seen as the local motion paths.

the same figure. The visual inspection and also the calculated statistics images show the good performance of our method for aligning the sequence, and therefore its applicability for registering real medical images. We present the $(N + 1)$ D deformation field and a slice along the temporal dimension from the $(N + 1)$ D mesh in Figs. 7(a) and 7(b). Furthermore, we recover the breathing curve for this sequence using the image-based gating technique proposed in [28], see Fig. 7(d). The similarity between the breathing curve and the deformation field in temporal direction in Fig. 7(c) suggests the correspondence between the motion track recovered by both methods.

CT Sequence

We perform registration using 10 images. The results are presented in Fig. 8. The mean and the STD images in the same figure show the very good performance of STORM on the thoracic images. The high STD around the upper edge of the liver is due to the image construction artifacts, which is quite visible in the input images. We again present images of $(N + 1)$ D deformation field and mesh. Furthermore, as done for the MRI sequence, we recover the breathing curve, see Fig. 9(d). The breathing signal from manifold learning is once again similar to the recovered motion trajectories of STORM.

5 Discussion and Conclusion

In this work, we have proposed a novel approach to the registration of time-resolved image sequences by taking advantage of groupwise registration. Unlike the traditional

methods, a single higher dimensional B-splines based transformation was used. This provides the possibility to have a deformation field also in between the images. Thanks to this feature, we were able to interpolate images in the temporal direction, which is necessary in many applications. Recently proposed APE was adapted to the registration of time-resolved images sequences. For optimization, we used stochastic gradient descent optimizer with a line search.

Experiments on synthetic images revealed STORM's advantages in different scenarios over the classical pairwise and groupwise registration methods. It was shown that STORM is more robust against outliers and large deformations within the sequence. The results showed that the use of temporal information ensures the consistency and smoothness in the deformation field along the temporal direction. Moreover, the integration of groupwise registration helped to handle large deformations within the sequence. We performed time interpolation by reconstructing images at any time point in the sequence, which is especially important for applications like radiotherapy treatment planning or 4D image reconstruction. Experiments on CT and MRI sequences showed the good performance of STORM on real medical images.

Acknowledgment. This work was partly funded by the European Commission.

References

1. Bai, W., Brady, S.: Spatio-temporal image registration for respiratory motion correction in pet imaging. In: ISBI, pp. 426–429. IEEE, Los Alamitos (2009)
2. Balci, S., Golland, P., Wells, W.: Non-rigid Groupwise Registration using B-Spline Deformation Model. In: Workshop on Open-Source and Open-Data for 10th MICCAI, pp. 105–121. Citeseer (2007)
3. Bystrov, D., Vik, T., Schulz, H., Klinder, T., Schmidt, S.: Local motion analysis in 4D lung CT using fast groupwise registration. In: Proceedings of the 16th IEEE International Conference on Image Processing, pp. 1729–1732. IEEE Press, Los Alamitos (2009)
4. Castillo, E., Castillo, R., Martinez, J., Shenoy, M., Guerrero, T.: Four-dimensional deformable image registration using trajectory modeling. *PMB* 55, 305 (2010)
5. Colgan, R., McClelland, J., McQuaid, D., Evans, P., Hawkes, D., Brock, J., Landau, D., Webb, S.: Planning lung radiotherapy using 4D CT data and a motion model. *Physics in Medicine and Biology* 53, 5815 (2008)
6. Crum, W., Hartkens, T., Hill, D.: Non-rigid image registration: theory and practice. *British Journal of Radiology* 77(Special Issue 2), S140 (2004)
7. Ehrhardt, J., Werner, R., Frenzel, T., Säring, D., Lu, W., Low, D., Handels, H.: Reconstruction of 4D-CT data sets acquired during free breathing for the analysis of respiratory motion. In: Proc. of SPIE, vol. 6144, pp. 365–372 (2006)
8. Flampouri, S., Jiang, S., Sharp, G., Wolfgang, J., Patel, A., Choi, N.: Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4D-CT data and Monte Carlo simulations. *Physics in Medicine and Biology* 51, 2763 (2006)
9. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage* 23, 151–160 (2004)
10. King, A., Buerger, C., Schaeffter, T.: Cardiac Respiratory Motion Modelling by Simultaneous Registration and Modelling from Dynamic MRI Images. In: WBIR, pp. 222–233 (2010)
11. Klein, G., Huesman, R.: Four-dimensional processing of deformable cardiac PET data. *Medical Image Analysis* 6(1), 29–46 (2002)

12. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. *IEEE Trans on Pattern Analysis and Machine Intelligence* 28(2), 236–250 (2006)
13. Ledesma-Carbayo, M., Kybic, J., Desco, M., Santos, A., Sühling, M., Hunziker, P., Unser, M.: Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. *IEEE Transactions on Medical Imaging* 24(9), 1113 (2005)
14. McClelland, J., Blackall, J., Tarte, S., Chandler, A., Hughes, S., Ahmad, S., Landau, D., Hawkes, D.: A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. *Medical Physics* 33, 3348 (2006)
15. Perperidis, D., Mohiaddin, R., Rueckert, D.: Spatio-temporal free-form registration of cardiac MR image sequences. *Medical Image Analysis* 9(5), 441–456 (2005)
16. Peyrat, J., Delingette, H., Sermesant, M., Xu, C., Ayache, N.: Registration of 4D Cardiac CT Sequences Under Trajectory Constraints With Multichannel Diffeomorphic Demons. *IEEE T-MI* (2010)
17. Rohlfing, T., Maurer Jr., C., O’Dell, W., Zhong, J.: Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Medical Physics* 31, 427 (2004)
18. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic registration using B-splines. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) *MICCAI 2006*. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
19. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18(8) (1999)
20. Sarrut, D., Boldea, V., Miguet, S., Ginestet, C.: Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. *Medical Physics* 33, 605 (2006)
21. Schreibmann, E., Chen, G., Xing, L.: Image interpolation in 4D CT using a BSpline deformable registration model. *IJROBP* 64(5), 1537–1550 (2006)
22. Sundar, H., Litt, H., Shen, D.: Estimating myocardial motion by 4D image warping. *Pattern Recognition* 42(11), 2514–2526 (2009)
23. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: The popi-model, a point-validated pixel-based breathing thorax model. In: *ICCR, Toronto, Canada* (2007)
24. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons using ITK’s finite difference solver hierarchy. *The Insight Journal* (2008)
25. Viola, P., Wells III, W.: Alignment by maximization of mutual information. *International Journal of Computer Vision* 24(2), 137–154 (1997)
26. Von Siebenthal, M., Szekely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, P.: 4D MR imaging of respiratory organ motion and its variability. *PMB* 52, 1547 (2007)
27. Wachinger, C., Navab, N.: Similarity Metrics and Efficient Optimization for Simultaneous Registration. In: *CVPR* (2009)
28. Wachinger, C., Yigitsoy, M., Navab, N.: Manifold learning for image-based breathing gating with application to 4D ultrasound. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 26–33. Springer, Heidelberg (2010)
29. Wu, G., Wang, Q., Jia, H., Shen, D.: Registration of longitudinal image sequences with implicit template and spatial-temporal heuristics. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6362, pp. 618–625. Springer, Heidelberg (2010)
30. Zöllei, L., Learned-Miller, E., Grimson, E., Wells, W.: Efficient population registration of 3D data. In: *Computer Vision for Biomedical Image Applications*, pp. 291–301 (2005)

Fast Brain Matching with Spectral Correspondence

Herve Lombaert^{1,2}, Leo Grady², Jonathan R. Polimeni³, and Farida Cheriet¹

¹ Ecole Polytechnique de Montreal, Canada

² Siemens Corporate Research, Princeton, NJ

³ Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School, Massachusetts General Hospital, Charlestown, MA

Abstract. Brain matching is an important problem in neuroimaging studies. Current surface-based methods for cortex matching and atlas-ing, although quite accurate, can require long computation times. Here we propose an approach based on spectral correspondence, where spectra of graphs derived from the surface model meshes are matched. Cerebral cortex matching problems can thus benefit from the tremendous speed advantage of spectral methods, which are able to calculate a cortex matching in seconds rather than hours. Moreover, spectral methods are extended in order to use additional information that can improve matching. Additional information, such as sulcal depth, surface curvature, and cortical thickness can be represented in a flexible way into graph node weights (rather than only into graph edge weights) and as extra embedded coordinates. In control experiments, cortex matching becomes almost perfect when using additional information. With real data from 12 subjects, the results of 288 correspondence maps are 88% equivalent to (and strongly correlated with) the correspondences computed with FreeSurfer, a leading computational tool used for cerebral cortex matching. Our fast and flexible spectral correspondence method could open new possibilities for brain studies that involve different types of information and that were previously limited by the computational burden.

1 Introduction

The human cerebral cortex is composed of many distinct brain areas whose locations relative to the folding pattern are highly stereotyped. In many neuroimaging studies, finding corresponding locations between two individuals allows data to be pooled across subjects and enables the investigation of functional and anatomical differences between individuals. Early attempts at computing correspondences relied on the extrinsic geometry seen on brain volumetric images. In 1967, Talairach *et al.* [19] introduced an early version of a brain atlas in the form of a 3D stereotaxic coordinate system. Despite its popularity, this method matched volumetric brain data using 3D Euclidean distances, which ignored geometric variabilities in the folding pattern. Techniques based on high-dimensional

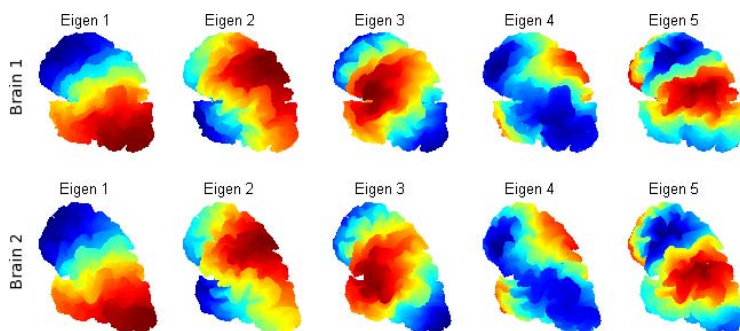


Fig. 1. Eigendecomposition of the graph Laplacian (using (2)) showing the first six eigenvectors (or six spectral components, one per column) of two brain surfaces (top and bottom rows). The color indicates a spectral coordinate for each point.

deformations allow for the alignment of volumetric brain image data. However, the lack of an explicit model for the brain surface often creates misaligned cortical areas [1]. Later, it was demonstrated that *surface-based alignment* [6,5,4,22,20], which operates by directly registering surface models of the cerebral cortex, significantly outperforms volume-based approaches [6,5]. The success of these surface-based techniques depends on the stability of the folding pattern across subjects. Some cortical areas are indeed consistently in a fixed position relative to the folding pattern [6,11], whereas other areas float around and do not seem well correlated to the folding pattern. These areas, however, may be correlated with measurable anatomical features other than the explicit cortical geometry. One successful method for computing brain surface correspondences was introduced by Fischl *et al.* [6]. It inflates each cerebral hemisphere surface to a sphere by a process that minimizes metric distortions, thus preserving local distances and areas. The cortical surface model is then mapped to the sphere using a non-rigid deformation driven by geometric features of the original folding pattern. This is the method used by FreeSurfer, a leading and widely used tool for brain surface reconstruction, matching, and atlas. Despite its accuracy, FreeSurfer suffers from a severe computational burden, which causes it to be very slow—it can take hours to compute a correspondence map between two cerebral cortices.

In order to alleviate this severe computational burden, we introduce a different approach for brain surface matching based on spectral correspondence. Spectral methods [3] present the tremendous advantage of being extremely fast—on the order of seconds. Correspondences are found on a graph *spectrum*, which is essentially the eigendecomposition of the graph Laplacian of an underlying shape model (illustrated on Fig. 1). Spectral methods have already been used in many fields, including in computer vision with the segmentation and registration of shapes in images [15], and recently in medical applications with the analysis of brain shape features [14,16] and with the smoothing of cortical surfaces [2]. Umeyama [21] and later Scott and Longuet-Higgins [17], pioneered the use of spectral methods for the correspondence problem. Shapiro and Brady [18]

compared ordered eigenvectors of a proximity matrix to find correspondences. Their work served as a basis for future spectral correspondence methods. Variants includes the use of different proximity matrices using different kernels, the use of the adjacency matrix, or different normalized Laplacian matrices. Mateus *et al.* [12] proposed an original unsupervised spectral method with an alternative to eigenvalue ordering based on eigenvectors histograms and refining the eigenvectors alignment with a probabilistic point matching. Jain and Zhang [10] tackle the eigenvectors alignment with a nonrigid deformation based on thin plate splines.

Previous spectral correspondence methods employ solely geometric information by weighting the graph edges with the distances between connected pairs of vertices. However, in order to use certain quantities (like sulcal depth, surface curvature, or cortical thickness), we must modify the spectral correspondence approach to incorporate information beyond edge length. To our knowledge, we are the first to present the use of *node weighting* in a spectral correspondence method. Additional information can indeed be incorporated into the Laplace operator, which implicitly contains metric information about nodes and edges. Moreover, additional information can be used as extra embedded coordinates when aligning the eigenvectors. This added level of flexibility makes our method a good candidate for brain studies involving various types of information with a large number of subjects.

After detailing our approach in the next section, we show in a control experiment that additional information can dramatically improve the performance of a spectral method. Using data from 12 subjects, we validate our method by comparing the computed correspondences with those generated by FreeSurfer [6]. We show that our method produces results, in a fraction of time required by FreeSurfer, that approach the accuracy of FreeSurfer. We believe that this large gain in processing speed would open the doors to new brain studies that were previously limited by the computational burden of the cortex matching calculation. Therefore, this method has the potential to be a significant tool for use in neuroscience.

2 Method

The proposed algorithm finds correspondences by comparing cortex representations, called *spectra* (illustrated on Fig. 1). The spectrum of a brain surface mesh is independent of its extrinsic geometry. In order to adapt a spectral method to brain surface matching, we must solve for several issues. First, we show how additional information (sulcal depth [6], surface curvature, and cortical thickness) can be used in spectral methods as weightings of the *graph nodes*. Second, we improve the ordering of the spectral components by finding an optimal permutation of the underlying eigenvectors. And third, we align the spectra in a multidimensional space using a nonrigid point transformation method. Our algorithm is summarized in Figure 2.

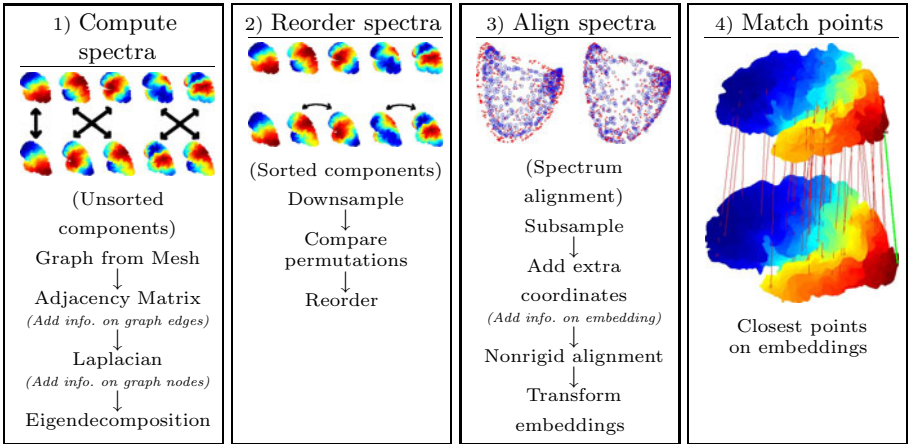


Fig. 2. Algorithm summary: *First*, we build two graphs and set the graph edges (the adjacency matrix) and on the graph nodes (the Laplacian matrix). The eigendecomposition of the graph’s Laplacian reveals the spectral components. *Second*, we reorder the components by finding the optimal permutation of components. *Third*, we deform the spectral embeddings. *Finally*, matching points are found with closest points in both spectral representations.

2.1 Spectral Correspondence

Given a shape defined by its collection of nodes $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ and a neighborhood system \mathcal{N} connecting a subset of these nodes (e.g., a mesh with vertices and faces), it is possible to construct its corresponding graph $\mathcal{G}_{\mathcal{X}}$. Here, we define the adjacency matrix W in terms of *affinity* weights (see [8]), which are derived from a distance metric $\text{dist}(x_i, x_j)$ between two neighboring vertices (x_i, x_j) (ε is a small penalizing term):

$$W_{ij} = \begin{cases} (\text{dist}(i, j) + \varepsilon)^{-1}, & \text{if } i \in \mathcal{N}(j), \text{ and } i \neq j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

It was shown in [8] that the general Laplacian operator on a graph takes the form $\tilde{L} = GL = G(D - W)$, where D is a diagonal matrix defined as $D_{ii} = \sum_j W_{ij}$ and G is the diagonal matrix of node weights. Typically in spectral correspondence, G is set to $G = D^{-1}$. However, we propose here to replace the default assignment $G = D^{-1}$ with *any* meaningful node weighting. Therefore, data associated with a mesh, such as distances or other additional information, can be incorporated in a graph on either its edges (in [1]), or its nodes by manipulating G .

We assume that the sulcal depth at each point, $\{s_1, s_2, \dots, s_n\}$, the Gaussian curvature at each point, $\{\kappa_1, \kappa_2, \dots, \kappa_n\}$, or the cortical thickness at each point, $\{t_1, t_2, \dots, t_n\}$, are pointwise characteristics and should be primarily defined on the graph nodes rather than on edges. We use the exponential of the positive diagonal matrices $\exp(S) = \exp(\text{diag}(s_1, s_2, \dots, s_n))$, and $\exp(K) =$

$\exp(\text{diag}(\kappa_1, \kappa_2, \dots, \kappa_n))$, and $\exp(T) = \exp(\text{diag}(t_1, t_2, \dots, t_n))$. We propose to incorporate additional information in the weighting of the nodes by defining the graph Laplacian as:

$$\tilde{L} = GL, \text{ where } G = D^{-1} (c_s \exp(S) + c_\kappa \exp(K) + c_t \exp(T))^{-1}, \quad (2)$$

where c_s , c_κ , and c_t are weighting factors. We use for instance $c_s = \text{mean}\{D_{ii}\}_{i=1\dots n} / \text{mean}\{\exp(s_i)\}_{i=1\dots n}$. The right eigenvectors of the Laplacian comprise the graph spectrum $\hat{X} = \{\hat{X}^{(1)}, \hat{X}^{(2)}, \dots, \hat{X}^{(n)}\}$. Figure 1 shows an example of spectral components for two brain hemispheres where each column depicts a different spectral component. Each eigenvector $\hat{X}^{(u)}$ represents a different (weighted) harmonic on a mesh surface that represents an intrinsic geometric property. The values $\hat{x}_i^{(u)}$, $i \in [1 \dots n]$, give the *spectral coordinates* for each point x_i . Eigenvectors associated with the lower non-zero eigenvalues (e.g., $\hat{X}^{(2)}, \hat{X}^{(3)}$) represent coarse (low-frequency) intrinsic geometric properties of the shape, the first of them $\hat{X}^{(2)}$ is called the *Fiedler vector*, while eigenvectors associated with higher eigenvalues (e.g., $\hat{X}^{(n-1)}, \hat{X}^{(n)}$) represent fine (high-frequency) geometric properties. The core idea of our method is to match two meshes \mathcal{X} and \mathcal{Y} by comparing their corresponding spectra \hat{X} and \hat{Y} rather than directly comparing the meshes themselves.

2.2 Ordering the Spectra

Each point of the brain surface mesh is represented with K spectral components associated with the K smallest eigenvalues, i.e., the embedded representations are $\hat{X}^K = [\hat{X}^{(2)}, \dots, \hat{X}^{(K+1)}]$ and $\hat{Y}^K = [\hat{Y}^{(2)}, \dots, \hat{Y}^{(K+1)}]$. Unfortunately, the spectral coordinates of the two meshes may not be directly comparable as a result of two phenomena. First, there exists a *sign ambiguity* when computing eigenvectors (i.e., if $Ax = \lambda x$ then $A(-x) = \lambda(-x)$), requiring a sign check of each eigenvector in the two meshes. Additionally, as a result of greater algebraic multiplicity of an eigenvalue, it may be possible that the *ordering* of the lowest eigenvectors will change, e.g., if two eigenvectors correspond to the same eigenvalue in both meshes, then the solver may compute these eigenvectors in one order for the first mesh and in the opposite order for the second mesh. For large meshes, this is a recurrent problem and the eigenvectors must be reordered. Since different brains do not present major discrepancies or major articulated deformations between individuals, the eigenvectors may be reordered by comparing their values at all pairs of closest points between the two brain hemispheres.

To speed up the reordering, all eigenvectors are first subsampled by selecting randomly a few points (we use 500 points in our experiments). Their spectral coordinates are normalized between 0 and 1 and denoted as $\hat{x}^{(i)}$. A spatial integration of all differences within pairs of closest points provides a similarity measure, i.e., if the eigenvectors $\hat{x}^{(i)}$ and $\hat{y}^{(j)}$ correspond to each other in both meshes, for all closest Cartesian points $\{(x_i, y_{i'})\}_{i=1\dots n}$, the difference of their associated spectral coordinates are computed. All the differences of potentially corresponding eigenvectors, $\hat{x}^{(u)}$ and $\hat{y}^{(v)}$, are gathered in a dissimilarity matrix, $C(\hat{x}^{(u)}, \hat{y}^{(v)}) = \sum_{i=1}^N (\hat{x}_i^{(u)} - \hat{y}_{i'}^{(v)})^2$, where $y_{i'} \in \mathcal{Y}$ is closest to $x_i \in \mathcal{X}$.

The Hungarian algorithm may be used to find an optimal permutation of eigenvectors $\hat{y}^{(v)}$ and, in order to remove the sign ambiguity, the minimal dissimilarity between the comparison of $\hat{x}^{(u)}$ and $\hat{y}^{(v)}$, and $\hat{x}^{(u)}$ and $-\hat{y}^{(v)}$ is used. The cost matrix used in the Hungarian algorithm is $Q(u, v) = \min\{C(\hat{x}^{(u)}, \hat{y}^{(v)}), C(\hat{x}^{(u)}, -\hat{y}^{(v)})\}$. After permutation, any eigenvector $\hat{x}^{(u)}$ corresponds with $\hat{y}^{(u)}$ and has a permutation cost $C^{(u)}$.

2.3 Alignment of Spectra

After reordering and handling the sign ambiguity, the eigenvectors of the two meshes may be assumed to have the same ordering in both embeddings (i.e., $\hat{x}^{(u)}$ corresponds with $\hat{y}^{(v)}$). However, the embedded representations, \hat{X}^K and \hat{Y}^K , need to be aligned (as illustrated in the third box of Fig. 2, both spectra have slight differences) so that closest points in these embedded representations would reveal corresponding points in both shapes (i.e., if \hat{y}_j^K is the closest point to \hat{x}_i^K , then x_i corresponds with y_j).

In order to perform this alignment, each eigenvector $\hat{x}^{(u)}$ is first weighted with $\exp(-(C^{(u)}\lambda_{\hat{x}^{(u)}})^2/2\sigma^2)$, where $C^{(u)}$ is the permutation cost, $\lambda_{\hat{x}^{(u)}}$ is its associated eigenvalue, and σ is a penalizing factor, we use $\sigma = \text{mean}\{C^{(u)}\lambda_{\hat{x}^{(u)}}\}_{u=1\dots K}$. Low-frequency eigenvectors, associated with coarser geometric properties (i.e., small eigenvalues $\lambda_{\hat{x}^{(u)}}$), will thus have more importance than the high-frequency eigenvectors associated with finer details, and pairs of corresponding eigenvectors will have more importance if they have strong similarities (i.e., low permutation costs $C^{(u)}$).

To further constrain the alignment of the mesh representations, we add extra information as additional coordinates to the embedded representation. Specifically, we concatenate with our first K spectral components $\hat{X}^K = [\hat{X}^{(2)}, \dots, \hat{X}^{(K+1)}]$ the extra coordinates, $\exp(\mathbf{S})$, $\exp(\mathbf{K})$, and $\exp(\mathbf{T})$. These extra components are also normalized to be comparable with the spectral coordinates. The embedded representations to be aligned are thus:

$$\begin{aligned}\tilde{X} &= [\hat{X}^K, c_s \exp(\mathbf{S}_X), c_\kappa \exp(\mathbf{K}_X), c_t \exp(\mathbf{T}_X)], \text{ and,} \\ \tilde{Y} &= [\hat{Y}^K, c_s \exp(\mathbf{S}_Y), c_\kappa \exp(\mathbf{K}_Y), c_t \exp(\mathbf{T}_Y)].\end{aligned}\quad (3)$$

The alignment of the spectral components can be viewed as a nonrigid registration, $\tilde{X}^K = \Phi(\tilde{Y}^K)$. The third box of Fig. 2 shows the alignment challenge where the first three spectral components ($\hat{X}^{(2)}$, $\hat{X}^{(3)}$, $\hat{X}^{(4)}$) are used as 3D (x, y, z) coordinates for visualization purposes. The Robust Point Matching with a thin plate spline-based transformation is often used for 2D or 3D registration. However, with this approach, the final registration depends on the number and choice of the control points. We apply the Coherent Point Drift method [13] which is fast and demonstrates excellent performance in this application. To increase speed in our algorithm, we subsample \tilde{X} and \tilde{Y} by taking randomly a few points (we used 500 points). The Coherent Point Drift method finds a continuous transformation Φ that can be applied on all points of \tilde{Y} . After aligning both embedded representations (i.e., $\tilde{X} = \Phi(\tilde{Y})$), it is possible to directly compare them, i.e.

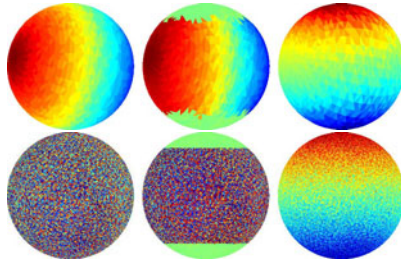


Fig. 3. Effect of node weighting. *Top row*, three cases showing the Fiedler vector on spheres of same orientation. *Left*, with no weighting: the vector is aligned with the sphere axis. *Middle*, with *edge* weighting: the top and bottom sections of the sphere are masked. *Right*, with *node* weighting: the vector is reorientated vertically. *Bottom row*: Same three experiments ran on 100 randomly orientated spheres. Their Fiedler vectors are accumulated. The reorientation of the Fiedler vector via node weighting (*bottom-left*) is clearer.

two points which are closest in the embedded representations, \tilde{x}_i and \tilde{y}_i , are treated as corresponding points in the meshes \mathcal{X} and \mathcal{Y} . The fourth box of Fig. 2 illustrates a few pairs of corresponding points. All pairs of points connected by the red lines have the closest embedded coordinates.

3 Results

Our methodology introduces several new concepts for spectral methods and shows how these methods may be customized for the application of cerebral cortex matching. We first show in an intuitive experiment the effect of node weighting on a simple mesh. Second, we measure the accuracy of spectral methods on a controlled experiment with a known ground truth. Third, we analyze the accuracy of our method against FreeSurfer. For this comparison we used 24 cerebral hemispheres from 12 subjects, and based our comparison on 288 matches using different combinations of additional information. Each brain surface mesh has been extracted using FreeSurfer from T_1 -weighted magnetic resonance images.

3.1 Node Weighting

We believe that we are the first to utilize node weights in a spectral correspondence approach. Consequently, we briefly devote some space to give an intuition about the behavior of these node weights in the context of spectral correspondence. To demonstrate the differences in weighting the edges and the nodes, we choose to show the Fiedler vector on a spherical mesh with an asymmetric vertex distribution. The concentration of vertices at the poles guides the spectral eigen-decomposition (i.e., the Fiedler vector is aligned with the sphere poles as shown in top-left sphere of Fig. 3). The accumulation of Fiedler vector on randomly orientated spheres should yield a uniform distribution (as shown in the bottom-left sphere of Fig. 3). We show that encoding weights on nodes can influence

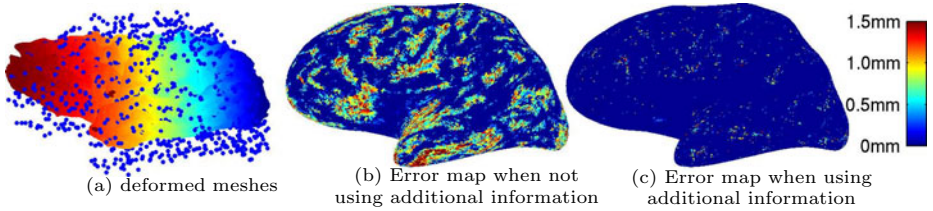


Fig. 4. Comparison with ground truth. *Left:* The deformed mesh (colored with its Fiedler vector) overlaid with the original vertex positions illustrated by the blue dots. *Middle:* The mean distance error is 0.44 mm with a standard spectral correspondence method (i.e., without additional information). *Right:* With additional information (sulcal depth, surface curvature, and cortical thickness), the error drops to 0.05 mm.

the orientation of the Fiedler vector. For instance, to reorient the Fiedler vector vertically, we use a texture data where the top section and the bottom section of a sphere (on a world z axis) are heavily weighted. For all points $\{x_i\}_{i=1\dots n}$, their weights are $\Theta = [\theta_1, \theta_2, \dots, \theta_n]$ where $\theta_i = 1000$ if $|x_i^{(z)}| > 0.4$ (i.e., a large weighting at the top and bottom of the sphere along the z axis) and $\theta_i = 1$ if $|x_i^{(z)}| \leq 0.4$ (i.e., a low weighting around the middle section).

In previous approaches to spectral correspondence, weights are encoded on graph edges. The texture weight Θ is added to the edge weights (1): $w_{ij} = (\text{dist}(x_i, x_j) + \varepsilon)^{-1} \times (|\theta_i - \theta_j| + \varepsilon)^{-1}$. The resulting weighting will highlight the texture boundaries isolating three distinct regions: the top, the middle, and the bottom section. This creates a multiplicity of three in the zero eigenvalues, and as shown in the top-middle sphere of Fig. 3, the Fiedler vector covers the largest section.

In our method, we propose to weight nodes in a graph in addition to weighting the edges. In order to compare with the previous experiment, we do not incorporate the texture Θ on graph edges. It is used on graph nodes (2): $G = D^{-1} \text{diag}(\Theta)^{-1}$. After the spectral decomposition, the multiplicity of the zero eigenvalue is 1 (i.e., there is one distinct object), and the Fiedler vector is aligned with the texture (i.e., with the world z axis). To verify this alignment, we repeated this experiment with 100 spheres orientated randomly and we accumulated the values of the Fiedler vectors (bottom row of Fig. 3). The principal axis of the accumulated values, $(-0.04, 0.12, 0.99)$, is indeed vertical when weighting the graph nodes (bottom-right of Fig. 3). The accumulation of the Fiedler vectors does not show an apparent principal axis when weighting only the graph edges (bottom-middle of Fig. 3), or when the texture data is not used (bottom-left of Fig. 3).

These three experiments show that weighting the graph nodes provides a new way for incorporating additional pointwise information and behaves differently than weighting the graph edges. Spectral methods can thus benefit from this idea, not only in cortex matching, but in various applications. The next experiment shows how node weighting improves matching in spectral methods.

3.2 Matching Deformed Brain Surfaces

Every individual has a unique folding pattern in the cerebral cortex, however there are many large-scale similarities. Before considering inter-subject cortex matching, we study the intra-subject case. We do so with a controlled, simulated deformation to analyze how additional information improves the accuracy of our method, and how the additional information should be incorporated. For our experiment, we match one brain hemisphere with a deformed version of itself. The vertex indexing remains the same in the deformed version, therefore the true matching is known (i.e., for all i , vertex p_i in the first mesh should match the vertex q_i in the second mesh). We severely deform one of the cortical hemisphere surface models with the transformation $q^{(z)} = (1 + \alpha)p^{(z)}$ (a compression in the z -axis controlled by α) and $q^{(x)} = p^{(x)} + \beta r^2 / \max(r^2)$ with $r^2 = p^{(x)^2} + p^{(y)^2}$ (a radial distortion controlled by β). This simulates a deformation due to a drastic change in the head shape. The deformation however preserves the same mesh topology (i.e., with no discontinuity and no intersecting faces). Fig. 4 illustrates the position of the original hemisphere with the blue dots and the deformed hemisphere with the colored mesh. We quantify the accuracy of our matching by measuring the mean distance between all points and their corresponding matched points. That is, for all points p_i in mesh 1 matching q_j in mesh 2, we average the distance: $\text{mean}(\text{dist}(p_i, p_j))$. When no additional information is used, as it is the case in most state-of-the-art spectral methods, we find an average error distance of 0.44 mm as shown in the first error map of Fig. 4. Most errors are actually located on the sulci extrema.

Additional information can be incorporated as *node* weighting by using (2); as *edge* weighting by similarly adding additional term to (1) such as $w_{ij} = 1/(d(i, j) + \epsilon) \exp(-(s_i - s_j)^2/2\sigma_s^2) \exp(-(\kappa_i - \kappa_j)^2/2\sigma_\kappa^2) \exp(-(t_i - t_j)^2/2\sigma_t^2)$, where $\sigma_{s,\kappa,t}$ are penalizing factors; or as additional coordinates in the alignment process by using (3). Three sources of additional information (sulcal depth, surface curvature, and cortical thickness) can thus be used in three different ways in our method. That is 512 possible combinations ($2^{3 \times 3}$). We iterate through all of them and found that adding information as additional coordinate has the strongest impact on the accuracy of our matching. Adding the sulcal depth as the only additional feature yields an error of 0.16 mm; adding only the surface

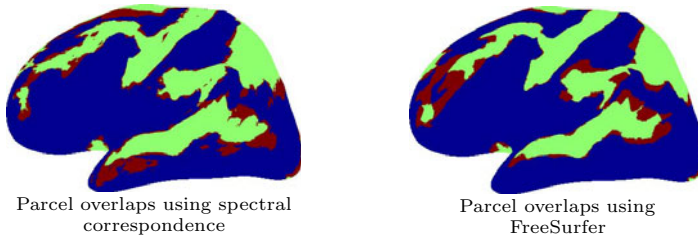


Fig. 5. In *green*, the areas where the projected sulcal regions of two cortices overlap, and in *red* the projection mismatches. (*Left brain*) Correspondences were computed in 124 seconds using our method, while (*right brain*) FreeSurfer required several hours.

curvature yields an error of 0.35 mm; and the cortical thickness yields an error of 0.14 mm. Adding single or multiple source of additional information on only the graph nodes does not yield significant improvement (0.44 mm), nor does representing this additional information on only the graph edges (0.44 mm). However, by adding all three features and using all of them on the graph nodes and on the graph edges, the error drops to 0.06 mm. Our best combination of additional information was obtained when using all three features and when using sulcal depth and cortical thickness on the graph nodes, yielding an error of 0.05 mm. The error map in the right of Fig. 4 shows an almost perfect matching with our best-performing combination.

3.3 Validation

Brain surface matching is an ambiguous problem. Indeed, sulci morphology and topology differ from one individual to another. There is no ground truth available for perfect brain surface matching. However, FreeSurfer [6] has been demonstrated to provide highly accurate cortical matchings that closely align cortical areas across subjects [9] and therefore provides a reliable benchmark for our comparison. The delineations of 81 sulcal regions are available for 24 hemispheres (12 subjects). These sulcal regions were obtained using an automatic parcellation of the cortex [7] and are considered as our gold standard. Although folding parcellations are not expected to align between subjects in all cases (except for the primary folds), they do provide means to compare the two methods. We use correspondence maps generated by FreeSurfer and by our method to project the parcellation areas onto different brain hemispheres and we measure their overlaps (illustrated on Figure 5). To process a mesh of 135,000 vertices, FreeSurfer has a varying processing time which is always on the order of several hours, while the time required by our method is just on the order of a few minutes. To process all our 288 possible pairs of left and right brain hemispheres, our method required on average 124 seconds on a 2.8 GHz Intel Pentium 4 using unoptimized Matlab code (with meshes of 20,000 vertices, our method performed in 19 seconds). The code could benefit further from parallel programming and the use of GPU. The total time was 9 hours on a single computer, a substantial advantage compared to the several weeks required by FreeSurfer to process all 288 cortex matchings in series. Each overlap ratio is defined by the ratio of set intersection to set union. Figure 6 shows the overlap ratios for the largest sulcal parcellations using our method (our best setting is shown in cyan) and FreeSurfer (red). Our results are consistent across all 81 sulcal regions (i.e., whenever FreeSurfer outputs a higher overlap ratio, our method also consistently outputs a higher overlap ratio). Specifically, our results are correlated to FreeSurfer's overlaps with a correlation coefficient of $\rho = 0.816$. When comparing larger regions [4]

¹ Sulcal regions: 9 (*G frontal middle*), 10 (*G frontal middle*), 18 (*G occipit temp med Lingual part*), 23 (*G parietal inferior Supramarginal part*), 24 (*G parietal superior*), 26 (*G precentral*), 41 (*Medial wall*), 42 (*Pole occipital*), 45 (*S central*), 47 (*S cingulate Main part and Intracingulate*), 59 (*S intraparietal and Parietal transverse*), 80 (*S temporal superior*).

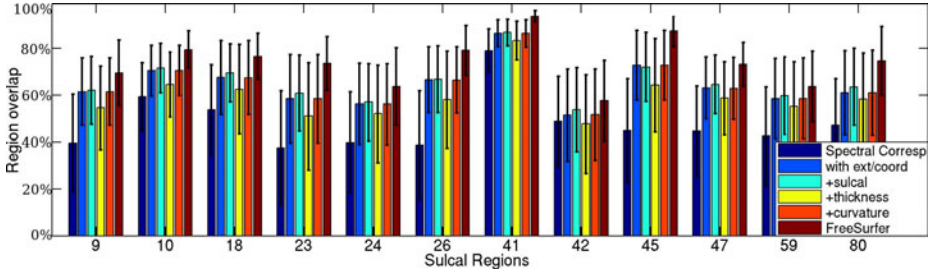


Fig. 6. Overlap ratios of different sulcal regions in several settings. (*Dark blue*) No additional information (65% of FreeSurfer’s performance), (*blue*) all additional information (sulcal depth, cortical thickness and cortical curvature) as extra coordinates (87%), (*cyan*) all additional information as extra coordinates and *sulcal depth* on graph nodes and graph edges (88%), (*yellow*) all additional information as extra coordinates and *cortical thickness* on graph nodes and graph edges (78%), (*orange*) all additional information as extra coordinates and *cortical curvature* on graph nodes and graph edges (87%) and (*red*) by FreeSurfer (requiring weeks of computations). Our method only requires 9 hours and is strongly correlated with FreeSurfer (correlation coefficient of $\rho = 0.816$). The error bars show the standard deviation of each overlap ratio.

(illustrated on Fig. 6) parcels 9, 10, 18, 23, 24, 26, 41, 42, 45, 47, 59, and 80, as defined in 7), FreeSurfer’s overlap ratios are on average 74%. In its best setting (using sulcal depth as additional information), our method gives 88% that of FreeSurfer’s overlap ratios.

3.4 Combination of Additional Information

Besides information on sulcal depth, we had access to information on cortical thickness and on surface curvature. The cortical thickness is another clinically relevant anatomical measure, which is calculated by FreeSurfer from anatomical MRI images. The sulcal curvature is estimated with the Gaussian curvature ($\kappa_1\kappa_2$) of the mesh. We first analyze the performance of our method using five configurations of different combinations of additional features. For each configuration, we ran our method on the 288 pairs of brain hemispheres (totaling 1440 matchings). The results are summarized in Figure 6. The *first* configuration uses no additional information ($G = D^{-1}$ in (2)). In that configuration, the average overlap ratio on the largest parcels is only 48% (in comparison, FreeSurfer performs at 74%). In the *second* configuration, we use sulcal depth, cortical thickness, and cortical curvature as extra coordinates in the spectral alignment (using (3)). The average overlap ratio increases to 64% (or 87% of FreeSurfer’s performance), a 34% increase from the previous configuration. As shown in the previous experiments, using additional information as extra coordinates does increase the accuracy. In the *third* configuration, we also use all additional information as extra coordinates, and we add sulcal depth information on graph nodes ($G = D^{-1} \exp(S)^{-1}$ in (2)) and on graph edges (in (1), $w_{ij} = 1/(d(i, j) + \epsilon) \exp(-(s_i - s_j)^2/2\sigma_s^2)$ where σ_s is a

regularization term). This configuration is actually the best one in which our method performed. The average overlap ratio is 66% (or 88% of FreeSurfer's performance). This suggests that sulcal depth provides crucial information when matching brain hemispheres, as has been previously suggested by Fischl *et al.* [6]. The *fourth* configuration adds cortical thickness as additional information on graph nodes ($G = D^{-1} \exp(T)^{-1}$ in (2)) and on graph edges (in (11), $w_{ij} = 1/(d(i, j) + \epsilon) \exp(-(t_i - t_j)^2/2\sigma_t^2)$). Using cortical thickness alone actually worsen the overlap ratio to 58% (or 78% of FreeSurfer's performance). This suggests that cortical thickness may be a contradictory information in our spectral correspondence method. The *fifth* configuration uses cortical curvature as additional information on graph nodes ($G = D^{-1} \exp(K)^{-1}$ in (2)) and on graph edges (in (11), $w_{ij} = 1/(d(i, j) + \epsilon) \exp(-(\kappa_i - \kappa_j)^2/2\sigma_\kappa^2)$). Cortical curvature shows to be also a significant additional information as it increases the average overlap ratio to 64% (or 87% of FreeSurfer's performance). It is important to note that there is no perfect configuration of additional information. Our experiment showed that certain configurations perform better on particular parcellations rather than on others. The right configuration of additional information thus depends on which sulcal region of the brain should be matched. That said, our experiment suggests that sulcal depth and cortical curvature are significant additional information that improve our matching method.

4 Conclusion

Cerebral cortex matching is an important topic that facilitates basic computational study in neuroscience. Current, surface-based matching methods can be quite accurate, but very slow. We have proposed a new cortex matching algorithm based on spectral correspondence operating at speeds of several orders of magnitude faster than current methods. Furthermore, we extended spectral methods in order to use additional information as weights in graph nodes and as extra embedded coordinates with little or no computational expense. This added flexibility makes our spectral correspondence method a good candidate for brain studies involving many additional information. Our current best configuration of additional information were found when using sulcal depth, surface curvature, and cortical thickness, as extra embedded coordinates and sulcal depth on graph nodes and graph edges. Our brain surface matching method far outperforms the accuracy of the more commonly used volumetric methods and approaches FreeSurfer's level of accuracy when aligning sulcal regions (88% of FreeSurfer's performance). The vast increase in speed and the added flexibility when using additional information gives new perspectives to previously computationally prohibitive experiments. The contribution new features incorporated to help improve the matching (e.g., anatomical or functional features extracted from various data sources) can be tested. Quick parameter sweeps can be performed to isolate the best parameter value sets. These computationally intensive experiments can help us to understand what features are consistently correlated with brain areas across individuals and what their role are during the development of the cortical folding pattern. Currently, the correspondences

found with the pairs of closest spectral neighbors. New schemes, such as the Relaxation Labeling as proposed in [23], will be tested and might improve accuracy. Future work will be to test different weighting functions (both based on nodes and edges), to incorporate more brain information (e.g., vascular density, MRI intensity), to evaluate the performance using additional cortical areas, and to test hypotheses about the relative importance of these features.

Acknowledgments

The authors would like to specially thank Gareth Funka-Lea and the financial support of the National Science and Environment Research Council (NSERC).

References

1. Amunts, K., Malikovic, A., Mohlberg, H., Schormann, T., Zilles, K.: Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable? *NeuroImage* 11(1) (2000)
2. Anqi, Q., Bitouk, D., Miller, M.: Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace-Beltrami operator. *Trans. Med. Im.* 25(10) (2006)
3. Chung, F.: *Spectral Graph Theory* (CBMS Conf. in Math., No. 92). AMS, Providence (1997)
4. Drury, H.A., Van Essen, D.C., Joshi, S.C., Miller, M.I.: Analysis and comparison of areal partitioning schemes using 2-D fluid deformations. *NeuroImage* 3 (1996)
5. Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T.T., Mohlberg, H., Amunts, K., Zilles, K.: Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex* 18(8) (2007)
6. Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M.: High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping* 8(4) (1999)
7. Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M.: Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14(1) (2004)
8. Grady, L., Polimeni, J.R.: *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. Springer, Heidelberg (2010)
9. Hinds, O.P., Rajendran, N., Polimeni, J.R., Augustinack, J.C., Wiggins, G., Wald, L.L., Rosas, D.H., Potthast, A., Schwartz, E.L., Fischl, B.: Accurate prediction of V1 location from cortical folds in a surface coordinate system. *NeuroImage* 39(4) (2008)
10. Jain, V., Zhang, H.: Robust 3D shape correspondence in the spectral domain. In: *Int. Conf. on Shape Mod. and App.* (2006)
11. Lohmann, G., von Cramon, D.Y., Colchester, A.C.: Deep sulcal landmarks provide an organizing framework for human cortical folding. *Cereb Cortex* 18(6) (2008), bhm174 [pii] 10.1093/cercor/bhm174 [doi]
12. Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., Boyer, E.: Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration. In: *CVPR* (2008)

13. Myronenko, A., Song, X.: Point-set registration: Coherent point drift. PAMI (2009)
14. Niethammer, M., Reuter, M., Wolter, F.-E., Bouix, S., Peinecke, N., Koo, M.-S., Shenton, M.E.: Global Medical Shape Analysis Using the Laplace-Beltrami Spectrum. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 850–857. Springer, Heidelberg (2007)
15. Reuter, M.: Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions. *Int. Journal Comp. Vis.* (2009)
16. Reuter, M., Wolter, F.E., Shenton, M., Niethammer, M.: Laplace-Beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *Comput. Aided Des.* 41(10) (2009)
17. Scott, G.L., Longuet-Higgins, H.C.: An algorithm for associating the features of two images. *Proc. Bio. Sc.* 244(1309) (1991)
18. Shapiro, L.S., Brady, J.M.: Feature-based correspondence: an eigenvector approach. *Image Vis. Comp.* 10(5) (1992)
19. Talairach, J., Szikla, G., Tournoux, P., Prosalenti, A., Bordas-Ferrier, M., Covello, L., Jacob, M., Mempel, E.: Atlas d'anatomie stereotaxique du telencephale, Masson, Paris (1967)
20. Thompson, P., Toga, A.W.: A surface-based technique for warping three-dimensional images of the brain. *Trans. on Med. Im.* 15(4) (1996)
21. Umeyama, S.: An eigendecomposition approach to weighted graph matching problems. PAMI 10(5) (1988)
22. Van Essen, D.C., Drury, H.A.: Structural and functional analyses of human cerebral cortex using a surface-based atlas. *J. Neurosci.* 17(18) (1997)
23. Zheng, Y., Doermann, D.: Robust point matching for nonrigid shapes by preserving local neighborhood structures. PAMI 28(4) (April 2006)

Landmark Matching Based Automatic Retinal Image Registration with Linear Programming and Self-similarities

Yuanjie Zheng¹, Allan A. Hunter III², Jue Wu¹, Hongzhi Wang¹,
Jianbin Gao³, Maureen G. Maguire², and James C. Gee¹

¹ PICSL, Department of Radiology, University of Pennsylvania, Philadelphia, PA

² Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA

³ School of Computer Science & Engineering, University of Electronic Science and Technology of China

Abstract. In this paper, we address the problem of landmark matching based retinal image registration. Two major contributions render our registration algorithm distinguished from many previous methods. One is a novel landmark-matching formulation which enables not only a joint estimation of the correspondences and transformation model but also the optimization with linear programming. The other contribution lies in the introduction of a reinforced self-similarities descriptor in characterizing the local appearance of landmarks. Theoretical analysis and a series of preliminary experimental results show both the effectiveness of our optimization scheme and the high differentiating ability of our features.

1 Introduction

Retinal images are widely used in diagnosing and monitoring the progress of a variety of eye diseases, such as diabetic retinopathy, age-related macular degeneration, and glaucoma. As a process of establishing spatial correspondences between two retinal images, retinal image registration is fundamental to applications of as diverse as detecting locations of scars and burns, tracking the progress of diseases, mosaicing to provide a complete view of the retina, and creating tools to assist ophthalmologists during laser surgery or other procedures [4].

However, automatic retinal image registration has remained a difficult problem due to several challenges. First, the transformation between two retinal images to be aligned can be very complicated. Retina is a curved surface and retinal images are projections of this surface from a wide range of view points with an un-calibrated camera. Second, there exist large homogeneous areas in retinal images. Nonvascular surface occupies an overwhelming area of the retina, which is textureless for healthy retinas while exhibits a variety of pathologies for unhealthy retinas. Finally, large appearance variations may be observed between the images to be registered. It is because the pathologies can appear and disappear over time, the effects of disease and poor image quality can obscure the vasculature [16], and images may be captured under vastly different illumination

conditions or acquired from different modalities. To solve this hard problem of retinal image registration, there are two keys to a successful algorithm: selection of an optimization strategy to more efficiently estimate the transformation model no matter the model is simple or complicated, and extraction of reliable features robust to image variations.

Landmark matching [4,17,5,16] is more attractive in retinal image registration compared with pixel-wise registration (a brief review to previous registration techniques is included in Sec. 2). It is because that the large homogenous areas and appearance variations in retinal images make the pixel-wise registration very hard and that landmarks are also used in practice by ophthalmologist [4]. Disregarding certain encouraging results produced by some previous methods [4,16,5], two major weaknesses were recognized in our research from many previous works. *First*, optimization strategies for estimating the transformation model are less efficient. Most previous methods perform an iterative optimization process by repeating the estimations of correspondences and transformation model, which requires a good initialization and results in an inferior local estimation. As discovered by us and also mentioned in [4], joint estimation of the correspondences and an appropriately predefined transformation model can result in a superior estimation. Unfortunately, no effective optimization approach has been described yet to accomplish this. *Second*, features describing landmarks are not always distinguishable. Most previous features (e.g. [5]) were extracted from a binary vascular image in which a lot of information of the local image pattern is lost. Moreover, landmarks are not uniquely identified by the widely used features characterized with vessel orientations [4,17,16] because branching angles at bifurcations can be similar across the retina.

This paper addresses the problem of landmark matching based automatic retinal image registration and bears two major contributions. One is a novel landmark-matching formulation proposed to accomplish a joint estimation of the correspondences and transformation model. This new formulation enables the matching problem to be solved by a very efficient combinatorial optimization technique - linear programming (LP) [11]. The other contribution is the introduction of a reinforced local self-similarities descriptor [14] in characterizing the local appearance of landmark. This descriptor captures the geometric layout in nearby image locations of the repetitions of local intensity-pattern. The features from this descriptor are invariant not only to local affine deformations but also large rotations of image. Theoretical analysis and a series of preliminary experimental results show that this joint estimation strategy using the LP techniques is very effective with the ability to produce superior accuracy and robustness, and that the proposed features bear a high differentiating ability.

2 Previous Work

Image registration has been extensively involved in a wide variety of practical tasks in the fields of retinal image analysis, medical image analysis, computer vision, etc. There exist a huge number of related publications. These

previous methods can basically be categorized as pixel/voxel-wise based, landmark/feature based, or their combinations. Detailed review to all of them is beyond the scope of this paper. We below only provide a brief review to the specific methods of retinal image registration and to the optimization techniques of the landmark matching based image registration.

For retinal image registration, both landmark matching based methods [4,17,5,16] and pixel-wise strategies [10,12] were explored. However, landmark matching has attracted more interests due to several reasons as explained in Sec. 1. In most previous works, landmarks were placed at cross points or bifurcations of vessel and characterized by for example coordinates, orientations or widths of the surrounding vessels. However, it is well known that these features are not robust and not always well distinguishable in practice, introducing severe matching ambiguities. The inter-image motion was described by various models such as weak affine [17], affine [4,10,12], bilinear/projective [12] and quadratic [4]. Optimization techniques include iterative closest point (ICP) algorithm [5], dual-bootstrap ICP [16], Hough Transform [17], and traditional global optimization techniques such as simulated annealing and genetic algorithm [12].

In those more general fields of medical image analysis and computer vision, landmark matching based image registration has been treated as a fundamental task [7,2]. To obtain a high matching accuracy, optimization plays a vital role. There are basically three classes of optimization techniques previously used. The first class assumes the correspondences to be pre-known and only resolves the transformation model [15,3,13]. The second class does not explicitly assume any specific transformation model but solely solves the correspondences with various optimization techniques such as the Hungarian method [1] and the quadratic programming [2]. The third class handles the joint estimation of the correspondences and transformation model. However, for simplicity, an iterative process by repeating the estimations of each of them was more widely used, e.g. the EM like technique in [6] and the iteratively updating process in [18]. With this iterative strategy, only a local optimization can be attained and the accuracy is highly conditioned on the initialization. In contrast, recently explored joint estimation with combinatorial optimization techniques [9] is free from the iterative process and can simultaneously solve the correspondences and transformation model. These methods are free from initialization and able to produce a global optimization. Unfortunately, only the simple rigid spatial transformation has been investigated [9].

3 Landmark Matching with Linear Programming

We provide a new formulation for the landmark matching problem. It bears at least two benefits compared with previous work. First, correspondences and the transformation model can be solved simultaneously. This joint estimation can result in superior accuracies because the correspondences and transformation model can assist each other in the optimization process compared with estimating them independently. The other benefit is that the new formulation enables

the optimization solved by linear programming (LP) - an important class of combinatorial optimization techniques. LP is attracting more and more interests in computer vision and medical image analysis due to its ability to obtain a globally optimal solution and the availability of more and more efficient and free solvers online. However, few work with LP can handle the joint estimation.

3.1 Problem Formulation

Suppose we have one reference-landmark set $\Omega^r = \{\xi_i^r, i = 1, 2, \dots, N_r\}$ and one floating-landmark set $\Omega^m = \{\xi_i^m, i = 1, 2, \dots, N_m\}$ where N_r and N_m represent the number of reference-landmarks and floating-landmarks, respectively. Taking the reference-landmark i as an example, ξ_i^r is expressed by the features (denoted by vertical vector \mathbf{v}_i^r) extracted to describe the visual appearance of this point and its coordinates $\chi_i^r = [x_i^r \ y_i^r]^T$, i.e. $\xi_i^r = [\mathbf{v}_i^{rT} \ \chi_i^{rT}]^T$ where T means transpose. For later usages, we provide as well the vector $\chi_i^{r'} = [x_i^r \ y_i^r \ 1]^T$ to denote the corresponding homogeneous coordinates, vector $\chi_i^{r''} = [x_i^{r2} \ x_i^r y_i^r \ y_i^{r2} \ x_i^r \ y_i^r \ 1]^T$ to represent the 12-parameter transformation model in [4], matrix $\chi^r = [\chi_1^r \ \chi_2^r \ \dots \ \chi_{N_r}^r]^T$, matrix $\chi'^r = [\chi_1^{r'} \ \chi_2^{r'} \ \dots \ \chi_{N_r}^{r'}]^T$, and matrix $\chi''^r = [\chi_1^{r''} \ \chi_2^{r''} \ \dots \ \chi_{N_r}^{r''}]^T$. We omit the definitions of similar symbols for the floating-landmarks for brevity.

Point matching is defined as estimating a mapping σ which indicates that a floating-landmark i_m corresponds to a reference-landmark i_r , i.e. $i_r = \sigma(i_m)$. Mapping σ can be represented by a correspondence/assignment matrix (denoted by E) [9] or a transformation model (denoted by T). T is usually represented by a parametric model in order for the robustness to noise/outliers and for reducing the involved computational burden.

Our point matching scheme is formulated as jointly estimating the correspondences and transformation model by maximizing the feature matching quality and the transformation model compliance quality. Feature matching quality represents how similar the features \mathbf{v}^m of each floating-landmark are to \mathbf{v}^r of its corresponding reference-landmark. Transformation model compliance quality evaluates how well the coordinates \mathbf{x}^m of each floating point and \mathbf{x}^r of its corresponding reference-landmark comply with the estimated transformation model.

Correspondence Matrix. We first define the correspondence matrix E as a binary matrix and then relax it later. E is of size $N_m \times N_r$, for which each row contains exactly one 1. If a floating-landmark i_m is matched to the reference-landmark i_r , then $E_{i_m, i_r} = 1$ and other elements of the i_m th row all equal to 0. These can be stated alternatively as the following two constraints:

$$E \in \{0, 1\}^{N_m \times N_r}, \tag{1}$$

$$E\mathbf{1} = \mathbf{1}, \tag{2}$$

where $\mathbf{1}$ is a vertical vector of ones.

Discreteness of the values taken by the elements of E in Eq. (1) introduces hardships into designing an efficient optimization algorithm for landmark matching. Many papers [6,9] tried to relax it to a continuous value within $[0, 1]$. This softassign strategy [6] can make the resulting energy function better behaved. This relaxation can be guaranteed by Eq. (2) together with the below constraint

$$E \geq 0. \quad (3)$$

Unfortunately, we found through our experiments that the softassign strategy can also cause ambiguities in the matching process. Imagine that one floating-landmark is matched to five reference-landmarks and the matching likelihood value equals to 0.2 for each of them. We then have no idea to decide the assignment. To deal with this dilemma, we propose to penalize the to-centroid spatial deviations of the matched reference-landmarks to enforce a spatial concentration.

To define the to-centroid spatial deviations, we first locate the centroid of the matched reference-landmarks by computing its coordinates with

$$\bar{\chi}^r = E\chi^r. \quad (4)$$

We then use $\bar{\chi}_x^r$ and $\bar{\chi}_y^r$ to denote the vertical vectors concatenated by the x -axis and y -axis coordinates of all centroids, respectively, and denote $\bar{\chi}^r = [\bar{\chi}_x^r \ \bar{\chi}_y^r]$. Then, the to-centroid deviation matrix is written as

$$D = \sqrt{(\bar{\chi}_x^r \mathbf{1}^T - \mathbf{1}(\chi_x^r)^T)^2 + (\bar{\chi}_y^r \mathbf{1}^T - \mathbf{1}(\chi_y^r)^T)^2} \quad (5)$$

where lengths of the $\mathbf{1}$ s are set appropriately to get a matrix D in size $N_m \times N_r$.

Penalizing the to-centroid deviations amounts to minimizing the function

$$\mathcal{O}_c = tr(D^T E), \quad (6)$$

where $tr(\cdot)$ denotes the trace of matrix. Minimization of the objective function in Eq. (6) with the constraints in Eq. (2) and Eq. (3) and other criteria explained below will result in a correspondence matrix E taking a continuous values in $[0, 1]$ but with a significant bias towards 0 or 1.

Feature Matching Quality. Feature matching quality measures how similar the visual appearance of each floating-landmark is to the matched reference-landmark(s). We use the reinforced self-similarities (to be proposed in Sec. 4) to describe the landmarks, which is invariant to local affine deformation, radially increasing non-rigid deformation, and rotation. We use the negative of the correlation between two series of features as the matching cost and then obtain a feature matching cost matrix C in size $N_m \times N_r$ for which the element $C_{i,j}$ denotes the cost matching the i th floating-landmark to the j th reference-landmark. Note that C can be computed beforehand.

Maximization of feature matching quality is then expressed as the minimization of the below objective function:

$$\mathcal{O}_f = tr(C^T E). \quad (7)$$

Transformation. Most transformation models can be incorporated in our landmark matching scheme, such as the widely used general models as diverse as affine, elastic, quadratic, deformable represented by Thin-plate Spline (TPS) or others, and some specific models designed for retinal imaging, e.g. the model in [4]. In the followings, we provide a discussion on the model in [4] and the TPS based deformable model as examples.

We propose to apply the constraints of the predefined transformation model on the location relations of each floating-landmark and the corresponding centroid of its matched reference-landmarks. As an example, the 12-parameter transformation model in [4] can be enforced with the following equation

$$\bar{\chi}^r = \chi'^m \Theta^T, \quad (8)$$

where $\bar{\chi}^r$ and χ'^m are defined previously and Θ is a 2×6 matrix. In Eq. (8), only the elements of Θ are unknowns. This transformation model is obtained in [4] for retinal imaging by assuming a quadratic surface for the retinal, a rigid transformation between two viewpoints, and a weak-perspective camera projection model. All involved parameters are combined in Θ .

When the transformation between the two landmark point sets is deformable, the TPS model [6] can be employed, as expressed by

$$\bar{\chi}^r = \chi^m + \chi'^m A^T + \Phi \Delta^m \quad (9)$$

where A is a 2×3 matrix containing the parameters of an affine transformation, Φ is a $N_m \times N_m$ symmetric matrix containing the information about the floating-landmark set's internal structural relationships and its elements can be pre-computed as $\Phi_{i,j} = \|\chi_i^m - \chi_j^m\|^2 \log \|\chi_i^m - \chi_j^m\|$, and Δ^m is an $N_m \times 2$ matrix for which each row denotes the non-affine deformation of the corresponding floating-landmark. In Eq. (9), A and Δ^m are unknowns and need to be estimated.

Note that Eq. (8) and Eq. (9) are all linear to the unknowns, which leads to a linear formulation which can be solved with the LP techniques.

3.2 Optimization

The optimization scheme for our point matching task can be formulated as a constrained minimization problem of an integrated objective function:

$$\mathcal{O} = \mathcal{O}_f + \lambda \mathcal{O}_c = \text{tr}((C + \lambda D)^T E) \quad (10)$$

where \mathcal{O}_f and \mathcal{O}_c are defined in Eq. (7) and Eq. (6), respectively, and λ (e.g. 0.5) is an adjusting parameter to control the ‘‘softness’’ of the estimated correspondence matrix E . For this optimization scheme, constraints are expressed in Eq. (2), Eq. (3), Eq. (8) or Eq. (9), unknowns are the correspondence matrix E and the transformation parameters: Θ in Eq. (8), or A and Δ^m in Eq. (9).

One benefit of our optimization scheme is that it can be resolved with linear programming (LP) [11]. It is obvious because the objective function and the constraints involved in the optimization scheme are all linear to the unknowns if

removing \mathcal{O}_c from Eq. (10) (i.e. setting $\lambda = 0$). With \mathcal{O}_c , the objective function is not linear any more due to the nonlinearity of \mathcal{O}_c to E in Eq. (6). Fortunately, \mathcal{O}_c can be linearized by computing D with Eq. (5) while fixing E in Eq. (4). Hence, our optimization scheme beomes an iterative process. At each iteration, the estimation of E in previous iteration is used to compute D . For the first iteration, E can be estimated by removing \mathcal{O}_c from Eq. (10). *This means that our landmark matching technique does not require any manual initialization.*

Formulation as Linear Programming. We provide expressions in a canonical form for our LP based optimization scheme by taking the transformation in Eq. (8) as an example. Other models can be handled with a similar way.

We first define some additional symbols by treating all elements in the correspondence matrix E , matrix Θ in Eq. (8), matching cost matrix C , and the to-centroid deviation matrix D as a single column and denoting them with vertical vectors \mathbf{e} , θ , \mathbf{c} , and \mathbf{d} , respectively.

We then create a binary permutation matrix P_e in size $N_m \times (N_m \cdot N_r)$, such that $P_e \mathbf{e} = E \mathbf{1}$. Then, we have the reformulation of Eq. (2):

$$P_e \mathbf{e} = \mathbf{1}. \tag{11}$$

We further divide Eq. (8) into two parts: x -mapping and y -mapping, and represent each of them as a linear function of Θ . The x -mapping is written as

$$\bar{\chi}_x^r = \chi_x^m + \chi'^m \Theta_x^T \tag{12}$$

where $\bar{\chi}_x^r$ is the first column corresponding to the x -coordinates of $\bar{\chi}^r$, χ_x^m is the first column of χ^m , and Θ_x is the first row of Θ and in charge of x -mapping. It is easy to further create a permutation matrix P_θ^x in size of 6×12 , such that $\Theta_x^T = P_\theta^x \theta$. Therefore, Eq. (12) can be rewritten as

$$\chi'^m P_\theta^x \theta = \bar{\chi}_x^r - \chi_x^m. \tag{13}$$

With $Q_\theta^x = \chi'^m P_\theta^x$ and $b_\theta^x = \bar{\chi}_x^r - \chi_x^m$, Eq. (13) can be rewritten concisely as

$$Q_\theta^x \theta = b_\theta^x. \tag{14}$$

Similarly y -mapping is expressed as

$$Q_\theta^y \theta = b_\theta^y. \tag{15}$$

We are now ready to provide the formulation of linear programming:

$$\min \begin{bmatrix} \mathbf{c} + \lambda \mathbf{d} \\ \mathbf{0} \end{bmatrix}^T \begin{bmatrix} \mathbf{e} \\ \theta \end{bmatrix} \tag{16}$$

subject to

$$\begin{bmatrix} P_e & \mathbf{0} \\ \mathbf{0} & Q_\theta^x \\ \mathbf{0} & Q_\theta^y \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \theta \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ b_\theta^x \\ b_\theta^y \end{bmatrix} \tag{17}$$

and

$$-\mathbf{e} \leq 0. \quad (18)$$

In this formulation, \mathbf{e} and θ are unknowns.

There are different ways to solve a LP problem and various solvers are available online. We chose the interior point solver of GLPK¹ for its efficiency.

Algorithm Overview. The optimization of our landmark matching approach is briefed as the following five steps:

1. Solving the LP problem specified by Eq. (16), Eq. (17), and Eq. (18) without considering the to-centroid deviations (i.e. set $\lambda = 0$ in Eq. (16)).
2. Updating the to-centroid deviation matrix using Eq. (5) with current estimation of the correspondence matrix.
3. Increasing λ by $\delta\lambda$ (e.g. 0.2).
4. Solving the LP problem specified by Eq. (16), Eq. (17), and Eq. (18) with considering the to-centroid deviation matrix.
5. Go to step 2 until the stopping criterion is satisfied.

The high accuracy of our algorithm comes from its at least five unique characteristics. First, step 1 conveniently provides automatic initialization. Second, steps 1 and 4 provide optimal solution with the LP techniques. Third, the correspondences and transformation model are estimated simultaneously in step 4 and the transformation model can help to guide the estimation of correspondences. Fourth, gradually increasing λ in step 4 approaches a binary correspondence matrix with less ambiguity in the assignment. Finally, the transformation model help to resist noise and outliers.

Our algorithm can produce an accurate estimation by repeating steps 2-5 for at most 3 times. Through our experiments, it can finish the optimization process within 1.5 seconds to match about 180 landmarks using a Dell PC with 2.39 GHz Intel Core 2 CPU. At the end, each floating-landmark is assigned the reference-landmark with the largest estimated correspondence likelihood value.

4 Landmark Detection and Feature Extraction

Landmark detection and feature extraction are also two important procedures of our retinal image registration approach. For landmark detection, the vascular-landmark detection technique in [4] is employed. Several steps are involved in this detection process: specification of seed points of vessel using 1D edge detection along a series of evenly spaced rows and columns, tracing out vascular structures with an exploratory procedure, and detection of landmarks by locating intersections of traces where three or more traces meet.

To extract features characterizing the local appearance of each landmark, we introduce a reinforced self-similarities (SS) descriptor. The SS descriptor was

¹ <http://gnuwin32.sourceforge.net/packages/glpk.htm>

recently proposed in [14] to measure similarity between two visual entities of images or videos. SS has been shown to bear high differentiating ability. Compared with SS, most other widely known descriptors (e.g. SIFT [3]) assume the existence of a common underlying visual property between matching images, e.g. intensities or responses of certain filters. SS relaxes this assumption and tries to capture the geometric layouts of local intensity-pattern’s repetitions in nearby image locations. In other words, SS describes the local internal layouts of the self-similarities instead of the intensity pattern itself. This is very important for a robust descriptor. For example, landmarks of retinal images from different modalities can have very different local intensity patterns but the geometric layouts of pattern’s repetitions in the vicinity of the landmark are very similar. As a result, SS can be used in both intra-modality and inter-modality retinal image registrations. We omit the details of computing SS for brevity. They can be found in [14].

The proposed reinforced SS descriptor is based on the original SS descriptor but tailored to be invariant to large image rotations. Although the original SS features are robust to local affine deformation and radially increasing non-rigid deformation [14], they become fragile when a large rotation happens between the matching images. It is because both the order of the entries in the SS descriptor vector and the partitioned bins [14] can change with rotation. Invariance to large rotations is particularly demanding for the retinal images considering the large range of view points involved in the retinal imaging process [5]. This rotation invariance can be simply accomplished by computing Hu’s seven moment invariants [8] of the outputs of the original SS descriptors. Therefore, the reinforced SS results in a feature vector with seven entries for each landmark point.

5 Results and Discussions

We performed a series of experiments to evaluate our retinal image registration algorithm both quantitatively and qualitatively. The transformation of each experimental image pair was formed synthetically by ourselves or introduced in the imaging process.

5.1 Synthetic Transformation

In order to quantitatively compare our method with state-of-the-art, we chose a digital fundus retinal image (as shown in Fig. 1 (a)), warped it with 50 synthetic affine transformations, and then ran our LP based method with the features in [4] (denoted by “LP”), our LP based method with the reinforced SS features introduced in this paper (denoted by “LP-SS”), the hierarchical method in [4] (denoted by “Hierarchical”) and the hybrid method in [5] (denoted by “Hybrid”). These synthetic affine warps were created by randomly choosing a rotation angle in $[0\ 2\pi]$, translation in $[-8\ 8]$ pixels for both directions, scale in $[0.3\ 3.5]$, and shear in $[-1.8\ 1.8]$ for both directions. Note that for each pair of images, the ground-truth position in the warped image of each pixel in the original image is known. To measure the registration accuracy, we employed the mean square

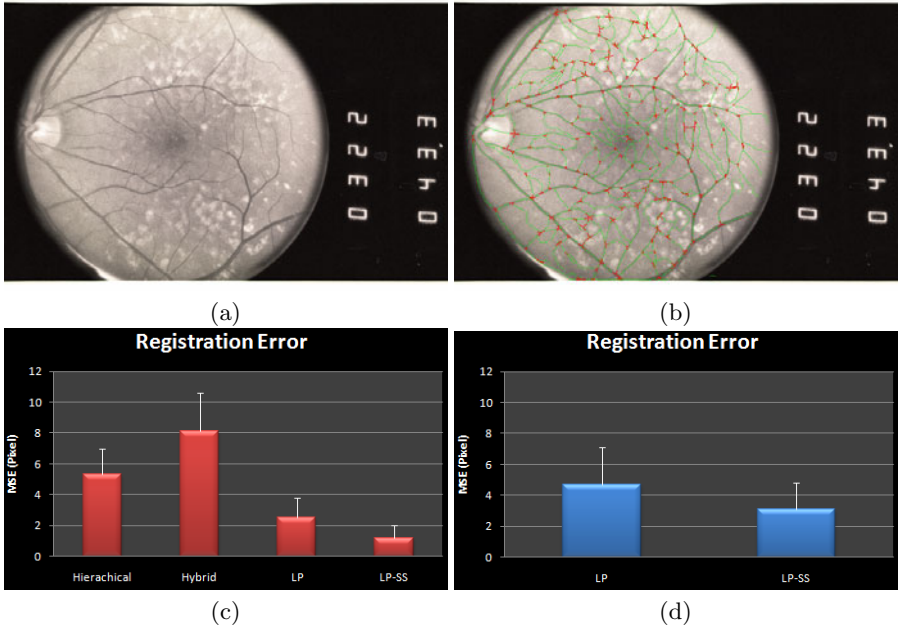


Fig. 1. (a): A digital fundus retinal image. (b): Bifurcation/crossover landmark points (red) and vessel’s centerline pixels (green) detected by the technique in [4]. (c) and (d): Registration error bars over 50 synthetic affine transformations and 50 synthetic nonlinear deformations, respectively.

error (MSE) between the ground-truth position and the estimation over all pixels in the original image. As shown by the error bars in Fig. 1 (c), we can see that our LP based algorithms produced higher accuracies, showing the effectiveness of our optimization technique, and that the reinforced SS features outperformed the features in [4], showing the superior differentiating ability of our features.

We also tested our algorithm on 50 nonlinear deformation fields simulated with the TPS model. We first manually specified 15 bifurcation/crossover points in the image and treated them as the control points of the TPS model. We then synthesized these fields by randomly setting the affine transformation with the same way as above and randomly displacing control points in a range $[-4, 4]$ pixel for both directions. The registration errors of our LP algorithm with the features in [4] and the reinforced SS features are shown in Fig. 1 (d). We can see that both cases performed well on the nonlinearly deformed images while the reinforced SS features produced higher accuracies than the features in [4].

5.2 Real Transformation

We chose 19 pairs of retinal images and tested on them with our LP based algorithm using the reinforced SS features and the hierarchical approach in [4]. The centerline error described in [4] was used as the error measure for validation.

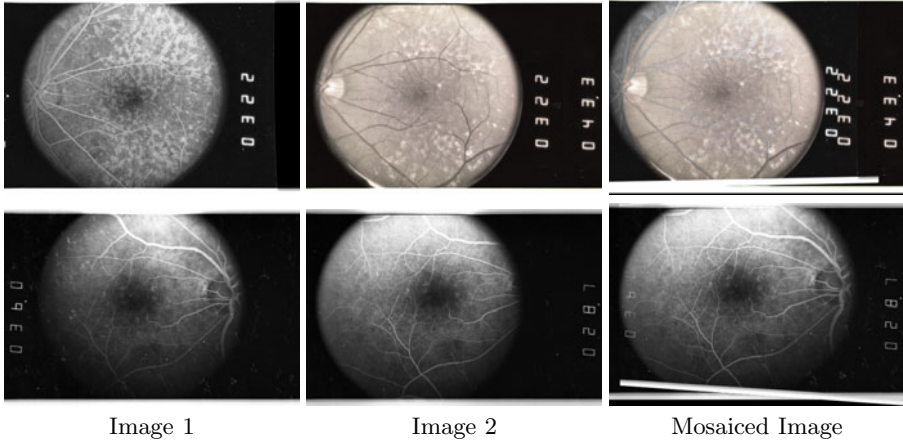


Fig. 2. Two example pairwise image mosaics produced by our approach

The hierarchical approach in [4] resulted in a mean/std error value of 4.3/3.8, and in contrast, our method achieved a mean/std error value of 1.2/0.7, showing the obvious performance improvements of our method.

Two examples of pairwise image mosaic formed by our retinal registration algorithm are shown in Fig. 2. Disregarding the image modality difference (fluorescein angiogram image vs. digital fundus image) and image changes caused by drusen in the above pair and the severe bias field artifact in the below pair, the mosaicing accuracy of our method is perceptually very high.

6 Conclusion and Future Work

We proposed a new landmark matching based retinal image registration scheme which bears at least two superior properties. First of all, a better optimization can be achieved. It originates from the proposed novel landmark matching formulation which enables the linear programming techniques in solving the matching problem and enables a joint estimation of correspondences and transformation model. Moreover, a lower matching ambiguity can be obtained. It benefits from the proposed reinforced self-similarities (SS) descriptor. The features from this descriptor were shown to yield a higher differentiating ability in both intra-modality and inter-modality registrations through our experiments.

Our future work would include more validation experiments, extension of our method to the 3D case, and applications to other images in various modalities (e.g. MR brain images and CT lung images). The extension to a 3D case can simply be achieved by extending the reinforced SS descriptor to 3D and including a 3D transformation model.

Acknowledgement

The authors gratefully acknowledge NIH support of this work via grants P30 EY001583 and P30 NS045839.

References

1. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE TPAMI* 24, 509–522 (2001)
2. Berg, A.C., Berg, T.L., Malik, J.: Shape matching and object recognition using low distortion correspondence. In: *CVPR*, pp. 26–33 (2005)
3. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. *IJCV* 74, 59–73 (2007)
4. Can, A., Stewart, C.V., Roysam, B., Tanenbaum, H.L.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE TPAMI* 24, 347–364 (2002)
5. Chanwimaluang, T., Fan, G., Fransen, S.R.: Hybrid retinal image registration. *IEEE Transactions on Information Technology in Biomedicine* 10, 130–142 (2006)
6. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding* 89, 114–141 (2003)
7. Gholipour, A., Kehtarnavaz, N., Briggs, R., Devous, M., Gopinath, K.: Brain functional localization: a survey of image registration techniques. *IEEE TMI* 26, 427–451 (2007)
8. Hu, M.K.: Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 179–187 (1962)
9. Jiang, H., Yu, S.X.: Linear solution to scale and rotation invariant object matching. In: *CVPR* (2009)
10. Kolar, R., Kubecka, L., Jan, J.: Registration and fusion of the autofluorescent and infrared retinal images. *IJBI* 2008, 513478 (2008)
11. Matousek, J., Gartner, B.: *Understanding and Using Linear Programming*. Springer, Heidelberg (2007)
12. Matsopoulos, G.K., Mouravliansky, N.A., Delibasis, K.K., Nikita, K.S.: Automatic retinal image registration scheme using global optimization techniques. *IEEE TMI* 3, 47–60 (1999)
13. Rohr, K., Stiehl, H.S., Sprengel, R., Buzug, T.M., Weese, J., Kuhn, M.H.: Landmark-based elastic registration using approximating thin-plate splines. *IEEE TMI* 20, 526–534 (2001)
14. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: *CVPR* (June 2007)
15. Shen, D., Davatzikos, C.: Hammer: Hierarchical attribute matching mechanism for elastic registration. *IEEE TMI* 21, 1421–1439 (2002)
16. Stewart, C.V., ling Tsai, C., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE TMI* 22, 1379–1394 (2003)
17. Zana, F., Klein, J.C.: A multimodal registration algorithm of eye fundus images using vessels by hough transform. *IEEE TMI* 18, 419–428 (1999)
18. Zheng, Y., Doermann, D.: Robust point matching for nonrigid shapes by preserving local neighborhood structures. *IEEE TPAMI* 28, 643–649 (2006)

Reconstruction of 4D-CT from a Single Free-Breathing 3D-CT by Spatial-Temporal Image Registration

Guorong Wu¹, Qian Wang^{1,2}, Jun Lian³, and Dinggang Shen¹

¹ Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{grwu, dgshen}@med.unc.edu

² Department of Computer Science, University of North Carolina at Chapel Hill
qianwang@cs.unc.edu

³ Department of Radiation Physics, University of North Carolina at Chapel Hill
Jun_lian@med.unc.edu

Abstract. In the radiation therapy of lung cancer, a free-breathing 3D-CT image is usually acquired in the treatment day for image-guided patient setup, by registering with the free-breathing 3D-CT image acquired in the planning day. In this way, the optimal dose plan computed in the planning day can be transferred onto the treatment day for cancer radiotherapy. However, patient setup based on the simple registration of the free-breathing 3D-CT images of the planning and the treatment days may mislead the radiotherapy, since the free-breathing 3D-CT is actually the mixed-phase image, with different slices often acquired from different respiratory phases. Moreover, a 4D-CT that is generally acquired in the planning day for improvement of dose planning is often ignored for guiding patient setup in the treatment day. To overcome these limitations, we present a novel two-step method to reconstruct the 4D-CT from a single free-breathing 3D-CT of the treatment day, by utilizing the 4D-CT model built in the planning day. Specifically, in the first step, we proposed a new spatial-temporal registration algorithm to align all phase images of the 4D-CT acquired in the planning day, for building a 4D-CT model with temporal correspondences established among all respiratory phases. In the second step, we first determine the optimal phase for each slice of the free-breathing (mixed-phase) 3D-CT of the treatment day by comparing with the 4D-CT of the planning day and thus obtain a sequence of partial 3D-CT images for the treatment day, each with only the incomplete image information in certain slices; and then we reconstruct a complete 4D-CT for the treatment day by warping the 4D-CT of the planning day (with complete information) to the sequence of partial 3D-CT images of the treatment day, under the guidance of the 4D-CT model built in the planning day. We have comprehensively evaluated our 4D-CT model building algorithm on a public lung image database, achieving the best registration accuracy over all other state-of-the-art methods. Also, we have validated our proposed 4D-CT reconstruction algorithm on the simulated free-breathing data, obtaining very promising 4D-CT reconstruction results.

Keywords: 4D-CT, spatial-temporal registration, radiation therapy, lung cancer.

1 Introduction

A major goal of image-guided radiation therapy is to deliver the optimal radiation dose to the tumor area, while minimizing the dose to the nearby normal structures. However, patient motions, such as respiratory and muscular motion, can cause significant artifacts during 3D-CT imaging. Recently, 4D-CT imaging, with respiration phase as the fourth dimension, has been developed and becomes more and more popular since it can capture 3D-CT in each respiration phase. Importantly, it has been shown that 4D-CT imaging helps reduce the artifacts from patient motion and thus leads to a more accurate delineation of tumor and other critical structures.

For treating lung cancer with radiation therapy, generally a 4D-CT and a free-breathing 3D-CT are acquired in the planning day. With these images, physician can manually delineate the outlines of lung, heart, tumor, and other critical structures from the free-breathing 3D-CT, or sometimes also from the inhalation and exhalation images of the 4D-CT. With these manual delineations, the optimal dose plan can be computed to maximize dose in cancer and minimize dose in nearby normal structures. In the treatment day, a free-breathing 3D-CT image is typically acquired for image-guided patient setup. Specifically, the free-breathing 3D-CT acquired in the planning day is aligned with the free-breathing 3D-CT acquired in the treatment day, i.e., through the registration of their segmented bones. Thus, the dose plan defined in the planning day can be transferred onto the treatment day for cancer radiotherapy.

However, there are two major limitations in the current lung cancer treatment pipeline. First, the 4D information available in the 4D-CT of the planning day is not well utilized for transferring of the dose plan onto the treatment day. Since the 4D-CT takes much longer scanning time and some patients have the difficulty in breath holding, only a free-breathing 3D-CT is usually acquired in the treatment day. Therefore, the dose plan in the treatment day is simply determined by registering its free-breathing 3D-CT with that in the planning day, without considering the useful information available in the 4D-CT of the planning day.

Second, it is difficult to register two free-breathing 3D-CT images, and their inaccurate registration could immediately undermine the dose plan in the treatment day. The registration challenge mainly comes from the inconsistent image information included in the two free-breathing 3D-CT images, since their corresponding slices could be acquired from the completely different respiratory phases in the planning and treatment days. Also, since the free-breathing 3D-CT is actually a mixed-phase image that never exists in the reality, it could provide misleading information to the observers. For example, as demonstrated in [1, 2], a moving spherical object could be acquired with different shapes under the free-breathing condition, depending on which phase the scanner starts to acquire the images.

In this paper, we propose a novel two-step method to overcome these two limitations by reconstructing a new 4D-CT from a single free-breathing 3D-CT for the treatment day, with the guidance from the 4D-CT model built in the planning day. Thus, the dose plan in the treatment day can be better determined based on the new reconstructed 4D-CT in the treatment day, rather than the free-breathing 3D-CT.

In the first step of our method (planning day), we develop a new spatial-temporal registration algorithm to simultaneously register all phase images of the 4D-CT acquired in the planning day, for building a 4D-CT model. Specifically, rather than

equally calculating the correspondence for each image point, we propose to hierarchically select key points in each phase image and then let them drive the initial registration by robust feature matching. Also, by taking advantage of key points, we further determine their temporal correspondences across all phase images of the 4D-CT by defining a set of ‘artificial’ temporal fibers to connect them along the respiratory phases. After registering all phase images, a 4D-CT model can be built to guide the reconstruction of a new 4D-CT in the treatment day for the next step.

In the second step of our method (treatment day), a new 4D-CT in the treatment day will be reconstructed from a free-breathing 3D-CT image by mapping the 4D-CT of the planning day to the treatment day. Specifically, we first determine an optimal phase for each slice of the free-breathing (mixed-phase) 3D-CT of the treatment day by comparing with various slices in the 4D-CT of the planning day and thus obtain a sequence of partial 3D-CT images in the treatment day, with each new 3D-CT image having only the partial image information in certain slices. Then, we reconstruct a new 4D-CT for the treatment day by warping the 4D-CT of the planning day to the sequence of partial 3D-CT images of the treatment day, with the guidance from the 4D-CT model built in the planning day.

Each of the two above-described steps has been evaluated in our experiments. Specifically, our spatial-temporal registration algorithm in the first step has been tested on a dataset with ten lung 4D-CT images [3], each having a set of manually-delineated landmarks in both selected and all phase images. Our algorithm achieves the best registration accuracy, compared to the other start-of-the-art deformable registration algorithms, including diffeomorphic Demons [4] and a recently-published B-splines based 4D registration algorithm [5]. Our 4D-CT reconstruction algorithm in the second step has been evaluated on a simulated free-breathing dataset, indicating its excellent performance in reconstructing 4D-CT from a single free-breathing 3D-CT.

2 Method

The goal of our method is to reconstruct a 4D-CT $\mathbf{Q} = \{Q_s | s = 1, \dots, N\}$ with N phases for the treatment day, based on a single free-breathing 3D-CT image I taken in the treatment day, and also to build a 4D-CT model from the planning 4D-CT $\mathbf{P} = \{P_s | s = 1, \dots, N\}$ to guide the described reconstruction. Our whole method consists of two steps: (1) build a 4D-CT model by estimating deformation fields of all phase images in the planning day to a reference space R , i.e., $\mathbf{F} = \{f_s(x) | x \in \Omega_{R \rightarrow P_s}, s = 1, \dots, N\}$; and (2) reconstruct the 4D-CT \mathbf{Q} for the treatment day by utilizing the 4D-CT model built in the planning day.

Fig. 1 gives an overview of our method, and also explains the difficulty in registration of the free-breathing 3D-CT images. For simplicity and easy explanation, we use an example of face expression scanning, instead of lung scanning, and also assume only two phases, smiling and angry phases, in the whole cycle of face expression, as shown in Fig. 1(a). It can be observed that the shapes of mouth and eyebrows are key to distinguish between the smiling and angry phases. Assume that a subject’s face turns from smile to anger during the image acquisition, which can be used to simulate the effect of respiration in the lung imaging. Thus, the top half face (i.e., the eyes and

eyebrows) can be scanned with smiling expression, while the bottom half face (i.e., the mouth) can be scanned with the angry expression. As the result, a mixed-expression face is obtained in Fig. 1(d); note that this half smiling and half angry expression may never happen in the reality. To simulate the possible shape changes during acquisition, we also simulate some local distortions around eyes (i.e., deformed from ellipse to circle) and mouth (i.e., deformed from close to open), as shown in red and blue in Fig. 1(d), respectively. It is worth noting that the direct registration of this mixed-expression face with the smile and angry phase images (in the planning day) will not provide a reasonable 4D-CT reconstruction result as shown in Fig. 1(e), since the interlaced phases in the mixed-expression image could provide misleading information.

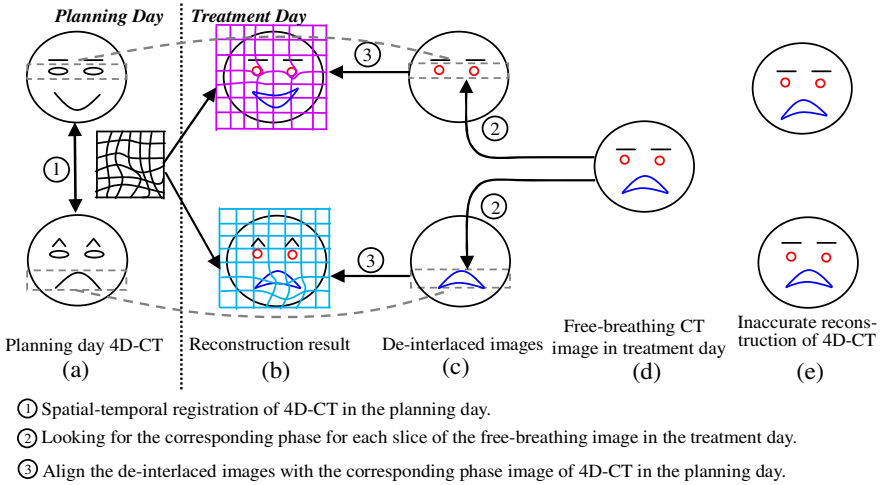


Fig. 1. The overview of our proposed method for estimating the 4D-CT images in the treatment day (b) from a free-breathing image in the treatment day (d), by using the 4D-CT model built in the planning day (a). The failed case of directly registering the free-breathing image to the 4D-CT is shown in (e). To avoid this potential problem, our method first de-interlaces the free-breathing image into the respective phase images (c), and then reconstructs a new 4D-CT in the treatment day (b) by warping the planning-day 4D-CT onto the de-interlaced images, with the guidance of the 4D-CT model built in the planning day.

To overcome this difficulty, we first build a 4D-CT model in the planning day by a robust spatial-temporal registration algorithm which we will explain later. As shown in Fig. 2 (c), we first determine the phase for each slice of the free-breathing image by looking for the best matched slice in the corresponding phase image of the 4D-CT. In the meanwhile, the neighboring slices in the free-breathing image of the treatment day are required to be identified with similar phases. Thus, the whole free-breathing image can be de-interlaced into several 3D-CT images, denoted as $\mathbf{D} = \{D_s | s = 1, \dots, N\}$, while the image information in each D_s is not complete (see Fig. 1(c)), i.e., mouth and eyes are missing in the smile and angry phases, respectively. After that, the spatial deformation fields $\mathbf{G} = \{g_s(x) | x \in \Omega_{P_s \rightarrow D_s}, s = 1, \dots, N\}$ between the 4D-CT in the

planning day and the image sequence $\{D_s\}$ can be estimated in two ways. (1) For the slices with image content in D_s , their correspondences can be determined with P_s directly. (2) For other missing slices in D_s , their correspondences can be compensated from other phase images according to the 4D-CT model built in the planning day. For the face example, the de-interlaced results of the free-breathing image and the final reconstructed 4D-CT in the treatment day are shown in Figs. 1 (c) and (b), respectively. It can be seen that the final reconstruction results are reasonable (i.e., the smiling face is still the smiling face and the angry face is still the angry face) and the local distortions around the mouth and eyebrows are well preserved in each phase.

In the following, we first present our spatial-temporal registration algorithm for building the 4D-CT model in the planning day. Then, we detail an algorithm for reconstruction of a new 4D-CT from a free-breathing image in the treatment day.

4D-CT Model in the Planning Day. Given the 4D-CT image \mathbf{P} in the planning day, the 4D-CT model can be built by aligning all phase images P_s to a reference image R (i.e., the maximum inhale phase image used in this paper) by estimating the deformation fields $\{f_s\}$. Bending energy [6], denoted as $L_s(f_s)$, is used in this paper as a smoothness term for regularization of deformation field f_s , thus TPS can be used to efficiently compute the optimal f_s with minimal bending energy.

Different from other intensity-based registration methods [4, 5, 7], we propose to use the attribute vector as the morphological signature of each image point for reliable correspondence detection. Therefore, each lung CT image is first segmented into lung, bone, and body regions by [8]. Furthermore, we use the multi-scale Hessian filter [9] to delineate the tiny vessels within lung, in order to improve the registration for the vessels. After that, the attribute vector $\vec{a}(x)$ is calculated at each location x , to include image intensity, Canny edge response, and geometric moments of structure labels (i.e., lung, bone, body, and vessel). By setting the threshold on these attribute vectors and gradually relaxing the selection criteria, a set of key points $\mathbf{X}_s = \{x_s^j | j = 1, \dots, M_s\}$ (where M_s is the overall number of key points in each P_s) can be hierarchically selected in P_s . These key points are generally locating at salient areas in the CT image, i.e., the green points in Fig. 2. Also, we can compute the attribute vectors for the reference image R and thus obtain the key points $\mathbf{Y} = \{y^i | i = 1, \dots, W\}$ in the reference space (i.e., the red points in Fig. 2).

Inspired by [10], the shape of each phase image P_s and the reference image R can be represented by its key points \mathbf{X}_s and \mathbf{Y} . Each key point x_s^j in one phase image P_s can be considered as a random variable drawn from a Gaussian mixture model (GMM) with its centers as the deformed shape $f_s(\mathbf{Y})$ in P_s . Next, we introduce the hidden variables $m_s^{i,j} \in \{1, \dots, W\}$ (where W is the number of reference key points) to indicate the spatial assignment of a specific GMM center y^i in generating x_s^j . Therefore, the probability density function of x_s^j can be given by:

$$P(x_s^j | m_s^{i,j}, f_s(y^i)) = \mathcal{N}(x_s^j; \mu, \Sigma) \propto \exp\left\{-\frac{\|x_s^j - f_s(y^i)\|^2}{\sigma^2}\right\} \quad (1)$$

where $\mu = f_s(y^i)$ and $\Sigma = \sigma^2 \cdot Id_{W \times W}$ denote the mean and the covariance matrix of the normal distribution $\mathcal{N}(x_s^j; \mu, \Sigma)$, respectively. $Id_{W \times W}$ denotes the $W \times W$ identity matrix. For sake of simplicity, the isotropic covariance matrix Σ is assumed here.

The appearance information of key points x_s^j and y^i are represented by their corresponding attribute vectors $\vec{a}(x_s^j)$ and $\vec{a}(y^i)$, which embed rich anatomical information from the neighborhood. Given the association $m_s^{i,j}$ between x_s^j and y^i , the attribute vector $\vec{a}(x_s^j)$ can also be regarded as the random variable generated from a GMM with its centers as the collection of attributes $\{\vec{a}(f_s(y^i))\}$. Then, given f_s, y^i , and the corresponding assignment $m_s^{i,j}$, the conditional probability of attribute vector $\vec{a}(x_s^j)$ is defined as:

$$P(\vec{a}(x_s^j) | x_s^j, m_s^{i,j}, f_s(y^i)) = \mathcal{N}(\vec{a}(x_s^j); \vec{a}(f_s(y^i)), \tau^2) \propto \exp\left\{-\frac{\|\vec{a}(x_s^j) - \vec{a}(f_s(y^i))\|^2}{\tau^2}\right\} \quad (2)$$

where we assume elements in attribute vector are independent and have the same variance τ .

Given the assignment $m_s^{i,j}$ and f_s , the conditional probability of $(x_s^j, \vec{a}(x_s^j))$ can be formulated as:

$$P(x_s^j, \vec{a}(x_s^j) | m_s^{i,j}, f_s(y^i)) = e^{-\eta(x_s^j, f_s(y^i))}, \quad \eta(x_s^j, f_s(y^i)) = \frac{\|x_s^j - f_s(y^i)\|^2}{\sigma^2} + \frac{\|\vec{a}(x_s^j) - \vec{a}(f_s(y^i))\|^2}{\tau^2} \quad (3)$$

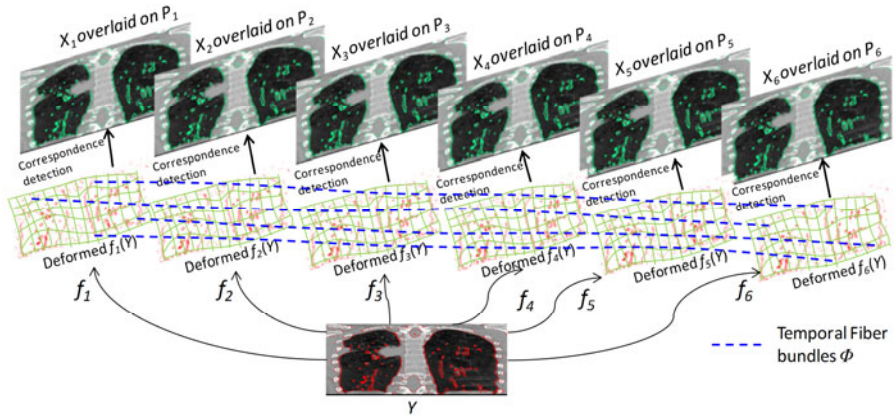


Fig. 2. The schematic illustration of our spatial-temporal registration on 4D-CT. The deformation fields $\{f_s\}$ are estimated based on the correspondences between key points Y (red points) and X_s (green points) by robust feature matching w.r.t. their shape and appearance. Meanwhile, the temporal continuity is preserved by performing kernel regression along the temporal fiber bundles Φ (see the blue dash curves).

In our registration method, we achieve spatial-temporal correspondence in the following way: (1) The deformation field f_s is steered by the correspondences between Y and X_s , and other non-key points will only follow the deformation of the nearby key points; (2) The temporal continuity constraint is performed based on the deformed position $f_s(y^i)$ in all phase images. Thus, for each y^i , its deformed position in the domain of each P_s along the respiratory phase s , denoted as $\varphi_i = \{f_s(y^i) | s = 1, \dots, N\}$, can be regarded as a temporal fiber (see the blue dash curves in Fig. 2),

which can be used to enforce the temporal continuity along the fiber bundle $\Phi = \{\varphi_i | i = 1, \dots, W\}$. For each fiber φ_i , a kernel regression [11] is used to preserve its temporal continuity. Accordingly, the temporal fibers after kernel regression $\hat{\varphi}_i = \{\hat{f}_s(y^i) | s = 1, \dots, N\}$ can be obtained by optimizing the following energy function L_T :

$$\arg \min_{f_s} L_T(\hat{f}_s(y^i)) = \sum_{s'=1}^N [f_s(y^i) - f_{s'}(y^i)]^2 \frac{1}{\sigma_s} \psi\left(\frac{s-s'}{\sigma_s}\right) \quad (4)$$

where ψ is the regression function. A Gaussian kernel with bandwidth σ_s is used in this paper.

Considering $h = (\{f_s\}, \{m_s^{i,j}\})$ as the model parameters and $v = (\{x_s^j\}, \{\tilde{a}(x_s^j)\})$ as the observations, the joint probability $P(h, v)$ is given by:

$$P(h, v) = \left[\prod_{s=1}^N \prod_{j=1}^{M_s} \pi_s^{i,j} \cdot P(x_s^j, \tilde{a}(x_s^j) | m_s^{i,j}, f_s(y^i)) \right] \cdot \left[\prod_{s=1}^N P(f_s) \cdot \prod_{i=1}^W P(\hat{f}_s(y^i)) \right], \quad (5)$$

where $\pi_s^{i,j} = P(m_s^{i,j} = i)$, and the probability functions of spatial transformation field $P(f_s)$ and temporal fiber bundles $P(\hat{f}_s(y^i))$ are, respectively, defined as:

$$P(f_s) = e^{-\frac{\lambda_1}{2} \|L_s(f_s)\|^2}, P(\hat{f}_s(y^i)) = e^{-\frac{\lambda_2}{2} \|L_T(\hat{f}_s(y^i))\|^2}, \quad (6)$$

where λ_1 and λ_2 are the scales used to control the smoothness of the spatial transformation field f_s^P and the temporal fiber bundles Φ , respectively. Note that L_s is the bending energy [6] on spatial deformation field and L_T , defined in Eq. 4, is the residual energy after kernel regression.

To find the optimal $\{f_s\}$ and $\{m_s^{i,j}\}$ in Eq. 5, we resort to the “free energy with annealing” method in [12] to minimize the distribution distance between $p(h|v)$ and $p(h, v)$, which bounds for maximizing the likelihood of the observation $P(v)$. By omitting some constant terms and taking logarithm to Eq. 5, we obtain the free energy for our registration algorithm as:

$$E = \sum_{s=1}^N \left\{ \sum_{j=1}^{M_s} [\pi_s^{i,j} \cdot \eta(x_s^j, y^i) + r \cdot \pi_s^{i,j} \log(\pi_s^{i,j})] + \lambda_1 L_s(f_s) \right\} + \lambda_2 \sum_{i=1}^W L_T(\hat{f}_s(y^i)) \quad (7)$$

where r serves as the role of temperature in the annealing system [13] for robust correspondence detection. We will clarify it in Eq. 8.

The optimization of E in Eq. 7 is performed in the “divide-and-conquer” way. In the first step (*correspondence detection*), the probability of spatial assignment $\pi_s^{i,j}$ can be immediately computed by only minimizing E w.r.t. $\pi_s^{i,j}$:

$$\pi_s^{i,j} = \exp\left(-\frac{\eta(x_s^j, y^i)}{r}\right) / \sum_{j=1}^{M_s} \exp\left(-\frac{\eta(x_s^j, y^i)}{r}\right). \quad (8)$$

It is obvious that $\pi_s^{i,j}$ is penalized in the exponential way according to the discrepancy degree $\eta(x_s^j, y^i)$ of shape and appearance in Eq. 3. The temperature r controls the penalty from mild to aggressive as it gradually decreases with the development of registration.

Once $\pi_s^{i,j}$ has been obtained, the correspondence of each y^i w.r.t. the shape of X_s is the weighed mean location of all x_s^j s, i.e., $f_s(y^i) = \sum_{j=1}^{M_s} \pi_s^{i,j} \cdot x_s^j$, by discarding all unnecessary terms with $f_s(y^i)$ in Eq. 7. In the second step (*dense deformation interpolation*), the temporal fibers Φ can be updated according to the latest correspondences $f_s(y^i)$ in each phase. Then the dense deformation field f_s is obtained by

TPS (to minimize the bending energy $L_s(f_s)$) based on the sparse correspondences $\{\hat{f}_s(y^i)\}$, which optimizes $L_T(f_s)$ in Eq. 4 after kernel smoothing on each fiber φ_i . By repeating these two steps, the deformation of the reference image R to all P_s can be estimated and the temporal consistency can be well preserved by implicit temporal fibers.

In the end of registration, the deformation fields $\{f_s\}$ towards the reference image are obtained, along with the implicit temporal fibers Φ , which will facilitate the reconstruction of 4D-CT in the treatment day, as detailed below.

Reconstruction of 4D-CT in the Treatment Day. The overview of reconstruction of the 4D-CT in the treatment day is summarized in Fig. 3. There are two steps, i.e., phase de-interlace and the spatial-temporal registration, involved for reconstructing the 4D-CT in the treatment day. These two steps are explained below.

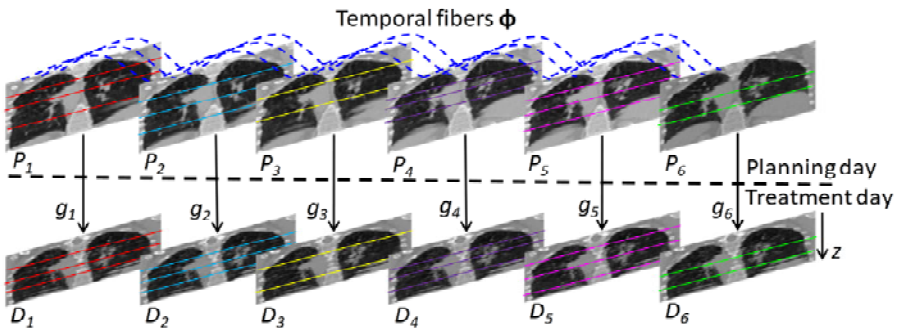


Fig. 3. The scheme for reconstruction of 4D-CT in the treatment day. The de-interlaced results at each phase are designated by the lines with different colors. Note that the free-breathing 3D-CT is repeated at each phase for easy description. The reconstruction of 4D-CT in the treatment day is obtained by estimating the deformation field g_s between the de-interlaced image sequence D and the planning-day image sequence P , with the guidance of temporal heuristics obtained in the planning day.

To determine the phase at each slice of the free-breathing image I , all phase images $P_s (s = 1, \dots, N)$ in the planning day will roughly deform to the free-breathing image I in the treatment day, by alignment of bones since the bone will almost not move with lung. In this way, for each slice $I(z)$ (where z is from the superior to inferior of human body), we can simply determine its corresponding phase by the two criteria: (1) the image similarity between $I(z)$ and its corresponding $P_s(z)$ should be as high as possible; (2) the detected phases for the neighboring slices of image I should be as similar as possible. We call this procedure as phase de-interlace. After that, several 3D volumes $D_s (s = 1, \dots, N)$ which are associated with the particular phase images P_s are obtained, although many slices in each D_s are empty. The lines in the bottom of Fig. 3, with different colors denoting for different detected phases, designate the de-interlaced results in each phase, along with the matched corresponding planning-day slices shown in the top of Fig. 3.

Here, we consider $\{P_s\}$ as the model image sequence with complete information and $\{D_s\}$ as the subject image sequence with partial information after de-interlace. As shown in the Fig. 3, the deformation field g_s from each P_s to the de-interlaced result D_s will be estimated for mapping the planning-day 4D-CT to the treatment-day space. As we will explain next that the estimation of only the deformation field g_s between P_s and D_s is not enough to obtain good registration, the temporal heuristics in the planning day is also used to compensate the missing information in D_s .

Following the spatial-temporal registration method we have proposed above, the estimation of each g_s from D_s to P_s can be performed independently by regarding D_s as the reference image R and only the underlying phase s (i.e., $N = 1$) will be considered in optimizing the energy function (Eq. 7) in registration. Similarly, the kernel regression term (the second term in Eq. 7) is also removed since N in Eq. 7 is now set to 1. Specifically, the attribute vector for each point in D_s will be calculated for feature matching. Then, the key points in D_s , i.e., $\mathbf{U}_s = \{u_s^k | k = 1, \dots, K_s\}$, are hierarchically selected to establish the correspondences with P_s . Each key point x_s^j in P_s has the contribution in determining the correspondence for u_s^k . The contribution is computed by Eq. 8, which is the exponential function of the geometric distance between u_s^k and x_s^j , as well as the difference between attribute vectors $\vec{a}(u_s^k)$ and $\vec{a}(x_s^j)$. Finally, the destination of u_s^k in P_s is the weighed mean location of all x_s^j s w.r.t. the previously calculated contributions.

Recall that only a few slices in the free-breathing image are associated with the particular D_s after de-interlacing. Therefore, no correspondence can be detected at the slice-missing areas due to the lack of key points. However, since we have roughly warp the planning-day 4D-CT to the treatment day, the detected correspondences on one D_s can be propagated to other de-interlaced images $D_{s'} (s' \neq s)$ by following the trajectory of temporal fibers established in the planning day. In this way, the motion in the slice-missing area of one D_s can be compensated from the regions in other phases where the correspondences have been identified.

Considering all the key points \mathbf{U}_s s selected in all D_s s as the set of source points in the treatment-day space, TPS can be performed to interpolate the dense deformation field g_s one by one. Finally, the 4D-CT in the treatment day can be reconstructed by mapping each P_s according to the estimated dense deformation field g_s .

3 Experiments

To demonstrate the performance of our proposed method, we first evaluate the registration accuracy in building 4D-CT model on DIR-lab data [3], with comparison to the state-of-the-art diffeomorphic Demons algorithm [4] and B-splines based 4D registration algorithm [5]. Then, we validate our algorithm in the reconstruction of 4D-CT for the treatment day from the simulated free-breathing 3D-CT.

Evaluation of Registration Performance on DIR-lab Dataset. There are 10 cases in DIR-lab dataset, each having a 4D-CT with six phases (from phase 1 to phase 6). For each case, 300 landmarks between the maximum inhale and the maximum exhale phases are available by repeated manual delineations, and also the correspondences of 75 landmarks for each phase are provided. Thus, we can evaluate the registration

accuracy at two fixed phases and also in the entire 4D-CT. The registration accuracy is computed by the Euclidean distance between the reference landmark positions and the landmark positions propagated from the first time point to all other time points by the registration algorithm. It is worth noting that our registration is always performed on the entire 4D-CT, regardless in the evaluation of 300 landmarks in inhale and exhale phases or 75 landmarks in all 6 phases.

The registration results by Demons, B-splines based 4D registration algorithm, and our algorithm on 300 landmarks between maximum inhale and maximum exhale phases are shown in the left part of Table 1. Note that we show the results only for the first 5 cases, since the authors in [5] reported their results only for these 5 cases. It can be observed that our method achieves the lowest mean registration errors over the other two registration methods, both based on simple intensity matching. The mean and standard deviation on 75 landmark points over all six phases by the B-spline based 4D registration algorithm and our registration algorithm (which are both 4D registration methods) are shown in the right part of Table 1. In all cases, our method achieves both lower mean registration error and smaller standard deviation.

Table 1. The mean and standard deviation of registration error (mm) on 300 landmark points between maximum inhale and exhale phases, and on 75 landmark points in all six phases

#	300 landmark points				75 landmark points		
	Initial	Demons	Bspline4D	Our method	Initial	Bspline4D	Our Method
1	3.89(2.78)	2.91(2.34)	1.02 (0.50)	0.80(0.70)	2.18(2.54)	0.92 (0.66)	0.62(0.41)
2	4.34(3.90)	4.09(3.67)	1.06 (0.56)	0.72(0.64)	3.78(3.69)	1.00 (0.62)	0.60(0.41)
3	6.94(4.05)	4.21(3.98)	1.19 (0.66)	1.13(0.87)	5.05(3.81)	1.14 (0.61)	0.76(0.47)
4	9.83(4.85)	4.81(4.26)	1.57 (1.20)	1.23(0.93)	6.69(4.72)	1.40 (1.02)	0.90(0.59)
5	7.48(5.50)	5.15(4.85)	1.73 (1.49)	1.28(1.28)	5.22(4.61)	1.50 (1.31)	0.96(0.79)

We also evaluate the temporal continuity of our spatial-temporal registration algorithm, with comparison to the Demons algorithm, since we do not have the registered landmark positions by the B-spline based 4D registration method. Two typical trajectories from the maximum inhale to the maximum exhale phases are displayed in Fig. 4 (a)-(b) by using red, green, and blue curves to denote the reference landmark positions, the estimated landmark positions by Demons, and the estimated landmark positions by our algorithm, respectively. The estimated trajectories by our algorithm

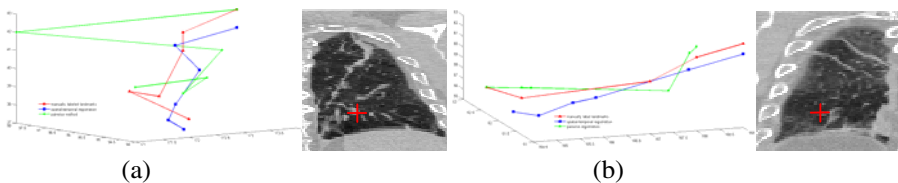


Fig. 4. Trajectories of two landmarks (selected from 75 landmarks) are shown in (a) and (b), where red, green, and blue curves denote the reference landmark positions, the estimated landmark positions by Demons, and the estimated landmark positions by our algorithm. The locations of these two reference landmarks are also shown, by the two red crosses near to the diaphragm.

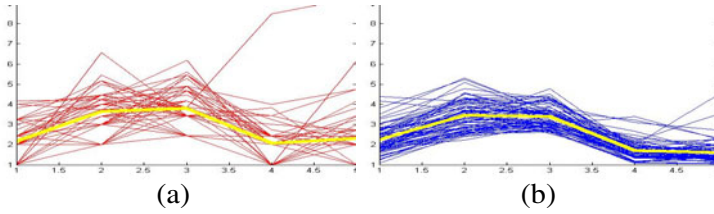


Fig. 5. The estimated velocity magnitudes from the maximum inhale to the maximum exhale, by Demons (a) and our algorithm (b), respectively. Here, the yellow curve is the mean velocity magnitude along phases.

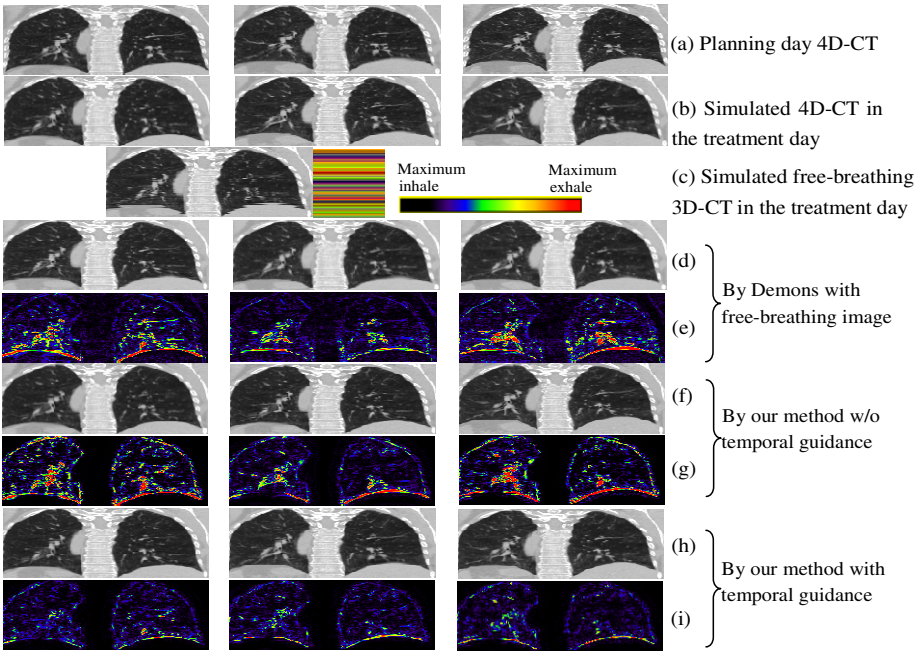


Fig. 6. The performance of 4D-CT reconstruction from a simulated free-breathing 3D-CT. A planning-day 4D-CT is first obtained, with its maximum inhale, middle, and maximum exhale phases shown in (a). The simulated treatment-day 4D-CT is displayed in (b), along with a simulated free-breathing 3D-CT of the treatment day shown in (c). The registration result between the planning-day 4D-CT and the free-breathing 3D-CT and its difference w.r.t. the ground-truth (b) are shown in (d)-(e), respectively. The reconstruction results without/with temporal guidance are displayed in (f) and (h), along with their corresponding differences w.r.t. the ground-truth (b) shown in (g) and (i), respectively.

(blue curves) is not only closer to the reference trajectories (red curves), but also much smoother than those produced by Demons (green curves). These two reference landmarks are selected near to the diaphragm (see red crosses in Fig. 4), where the motion is generally larger during respiration.

Figs. 5 (a) and (b) show the velocity magnitudes between consecutive phases for all 75 trajectories, obtained by Demons and our algorithm, respectively. The yellow curves denote the mean velocity magnitude along phases. Since we use the temporal fibers to constrain the continuity along phases, it can be observed that the velocity magnitude is much more continuous by our algorithm.

Evaluation of 4D-CT Reconstruction in the Treatment Day. To validate the performance of our reconstruction method, we simulate the free-breathing image for the treatment day as follows. We first obtain a planning-day 4D-CT as shown in Fig. 6(a), with the inhale, middle, and exhale phases displayed from left to right. We then simulate the deformation field by randomizing the parameters of B-spline control points, and apply this simulated deformation field to all phase images of the planning-day 4D-CT, for obtaining a ground-truth 4D-CT in the treatment day as shown in Fig. 6(b). Finally, we generate a free-breathing 3D-CT (Fig. 6(c)) by extracting the image slice from the corresponding phase image of the ground-truth treatment-day 4D-CT according to the phase table given for each slice in the right of Fig. 6(c). Accordingly, we can now estimate the treatment-day 4D-CT from this free-breathing 3D-CT by (1) directly registering the planning-day 4D-CT with the free-breathing 3D-CT, (2) reconstruction without temporal guidance from the planning-day 4D-CT model, and (3) reconstruction with temporal guidance from the planning-day 4D-CT model, with the respective results reported in Figs. 6 (d), (f) and (h). Also, their difference images w.r.t. the ground-truth treatment-day 4D-CT images (Fig. 6(b)) are given in Figs. 6 (e), (g) and (i), respectively. It is obvious that the reconstruction results have been significantly improved by utilizing the temporal model in our algorithm.

4 Conclusion

We have proposed a novel two-step method to reconstruct a new 4D-CT from a single free-breathing 3D-CT in the treatment day, for possible improvement of cancer radiotherapy which is currently using the registration of free-breathing 3D-CT images for patient setup. In our method, the 4D-CT model is first built in the planning day by robust spatial-temporal registration. Then, with the guidance of this 4D-CT model, the reconstruction of a 4D-CT in the treatment day is accomplished by warping the planning-day 4D-CT (with complete information) to the sequence of partial 3D-CT images of the treatment day (with missing image content in certain slices), obtained from the free-breathing 3D-CT after the phase de-interlace. We have extensively evaluated each step of our method with both real and simulated datasets, and obtained very promising results in building 4D-CT model and reconstructing new 4D-CT from a single free-breathing 3D-CT.

References

1. Chen, G.T.Y., Kung, J.H., Beaudette, K.P.: Artifacts in computed tomography scanning of moving objects. *Seminars in Radiation Oncology* 14, 19–26 (2004)
2. Vedam, S.S., Keall, P.J., Kini, V.R., Mostafavi, H., Shukla, H.P., Mohan, R.: Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Physics in Medicine and Biology* 48, 45–62 (2003)

3. Castillo, R., Castillo, E., Guerra, R., Johnson, V.E., McPhail, T., Garg, A.K., Guerrero, T.: A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine and Biology* 54, 1849–1870 (2009)
4. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage* 45, S61–S72 (2009)
5. Metz, C.T., Klein, S., Schaap, M., van Walsum, T., Niessen, W.J.: Nonrigid registration of dynamic medical imaging data using $nD + t$ B-splines and a groupwise optimization approach. *Medical Image Analysis* 15, 238–249 (2011)
6. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 567–585 (1989)
7. Christensen, G.E., Song, J.H., Lu, W., Naqa, I.E., Low, D.A.: Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. *Med. Phys.* 34, 2155–2163 (2007)
8. Armato III, S.G., Sensakovic, W.F.: Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis. *Academic Radiology* 11, 1011–1021 (2004)
9. Frangi, A., Niessen, W., Vincken, K., Viergever, M.: 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)
10. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.M.: Unsupervised Learning of an Atlas from Unlabeled Point-Sets. *IEEE Computer Society* 26, 160–172 (2004)
11. Wand, M.P., Jones, M.C.: Kernel Smoothing. *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York (1995)
12. Frey, B.J., Jovic, N.: A comparison of algorithms for inference and learning in probabilistic graphical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1392–1416 (2005)
13. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding* 89, 114–141 (2003)

Probabilistic Elastography: Estimating Lung Elasticity

Petter Risholm¹, James Ross², George R. Washko², and William M. Wells¹

¹ Surgical Planning Lab, Department of Radiology

² Pulmonary and Critical Division, Department of Medicine,
Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts
pettri@bwh.harvard.edu

Abstract. We formulate registration-based elastography in a probabilistic framework and apply it to study lung elasticity in the presence of emphysematous and fibrotic tissue. The elasticity calculations are based on a Finite Element discretization of a linear elastic biomechanical model. We marginalize over the boundary conditions (deformation) of the biomechanical model to determine the posterior distribution over elasticity parameters. Image similarity is included in the likelihood, an elastic prior is included to constrain the boundary conditions, while a Markov model is used to spatially smooth the inhomogeneous elasticity. We use a Markov Chain Monte Carlo (MCMC) technique to characterize the posterior distribution over elasticity from which we extract the most probable elasticity as well as the uncertainty of this estimate. Even though registration-based lung elastography with inhomogeneous elasticity is challenging due to the problem's highly underdetermined nature and the sparse image information available in lung CT, we show promising preliminary results on estimating lung elasticity contrast in the presence of emphysematous and fibrotic tissue.

1 Introduction

Lung function is closely related to the mechanical properties of lung parenchyma which is in turn determined by the orientation and proportion of the elastin and collagen fibers in the tissue matrix. The disparate properties of these tissues provide both the natural lung elastic recoil necessary for passive exhalation as well as the structural support for the pulmonary vasculature and airways. Diseases of the lung that alter lung function often do so by altering amounts of elastin and collagen [12].

In clinical practice, lung function tests are used to assess the severity of chronic lung diseases such as chronic obstructive pulmonary disease (COPD) and pulmonary fibrosis. While these tests are simple and safe to do, they reflect aggregate lung function and are unable to differentiate the relative amounts of emphysema or scarring that may be found in the lung of a smoker. In these cases, clinicians often turn to CT to assess both the extent and distribution of lung disease. A common method employed for the detection and quantification of emphysema is based upon lung densitometry, i.e. voxels with values less

than e.g. -950 HU are indicative of emphysematous tissue. A clear limitation to this observation is that the detection of abnormal tissue is not followed by an assessment of its mechanical properties.

In [5], image registration was proposed as a way of monitoring emphysema progression in lung CT. They align a baseline CT scan with a follow-up scan and relate intensity differences between the two images, corrected for intensity differences due to local tissue expansion and compression, to the disease progression. Yin et al. [13] propose a similar approach to handle differing intensity levels due to compression/expansion when matching inspiration to expiration images of the lung, however, their results do not significantly improve over using a standard sum of squared difference similarity measure.

Elastography is a non-invasive method to determine local elastic properties, e.g. Young's modulus. One approach to elastography imaging is to apply a low frequency oscillation to the tissue and measure the resulting deformations with ultrasound (US). The elasticity of the material can be extracted from the observed movement of the slowly oscillating tissue, e.g. to distinguish healthy from abnormal tissue. In Magnetic Resonance Elastography (MRE), a synchronized cyclic shear wave displacement is encoded into the phase of the MR signal. Because of the air in the lung, it has been difficult to perform elastography imaging of the lung. However, McGee et al. [6] recently showed it technically feasible to do MRE of the lung using hyperpolarized ^3He . They present results on estimation of the shear modulus of an ex vivo porcine lung in 2D. The main challenges with using this method for estimation of elasticity in COPD patients are signal to noise ratio and relatively long acquisition times.

Another approach to elastography is taken by Miga et al. [7] who introduce modality independent registration-based elastography to estimate tissue stiffness. Given two, potentially multi-modal, images of an object undergoing an elastic deformation, they minimize a modality independent similarity measure with regards to the elastic parameters. The method uses a Finite Element (FE)-discretization of the solution domain and the linear elastic energy, and it requires as input the boundary conditions of the biomechanical model, i.e. the deformation of the boundary of the FE-mesh. In Ou et al. [8], the same group applies the method to study elasticity of breast lesions and shows that the estimates can be quite sensitive to the boundary conditions. In some cases, the linear elastic assumption, i.e. geometric linearity of the strain tensor and linear stress-strain response of the tissue, might not be reasonable. Gokhale et al. [4] use a hyperelastic model of soft tissue and minimize a cost function representing the difference between the measured and predicted displacement fields. However, their results are preliminary – they only apply the method to 2D synthetic images.

Among the most sophisticated current methods for registering intra-patient images are those based on biomechanical modeling of tissue deformation where specific material parameters that reflect the underlying tissue properties can be included. However, reported material parameters for many tissue types are estimated from in-vitro samples and quite divergent. Furthermore, it is unknown how medication and tissue pathologies might affect the material parameters.

We have recently shown the feasibility of determining the posterior distribution over registration parameters restricted by a linear elastic biomechanical model discretized by a FE method [9,10]. To avoid using point-estimates for the material parameters, we showed in [9] how the material parameters can be marginalized out of the posterior by modeling them as random variables with broad priors. The method was applied in the realm of intra-operative registration of brain images with a very coarse spatial distribution of the elastic parameters: a three compartment model (cerebrospinal fluid, white matter, gray matter).

In this paper we estimate the posterior distribution over the inhomogeneous elastic modulus without any prior information regarding the underlying tissue labels. To reduce the sensitivity of the approach to boundary conditions (deformation), we propose to marginalize over the boundary condition of the biomechanical model. An MCMC method, where marginalization over parameters is easily achieved, is designed to characterize the posterior distribution. We show how important application dependent estimates, such as the most probable elasticity parameters and their corresponding uncertainty, can be derived from our results. Preliminary lung elastography experiments are carried out on both synthetic and clinical lung inspiration/expiration CT data which shows the feasibility of the method.

2 Methods

The deformation \mathbf{u} of an elastic object Ω , e.g. the lung, caused by a static loading, can be explained by a biomechanical model. It is common to separate the deformation into the boundary deformation \mathbf{u}_b which occurs on $\partial\Omega$ and the internal deformation \mathbf{u}_i in $\Omega \setminus \partial\Omega$. The parameters of a linear elastic biomechanical model are usually defined in terms of the Young's modulus \mathbf{E} (stiffness) and Poisson's ratio ν (compressibility). Suppose we are given the Dirichlet (deformation) boundary condition \mathbf{u}_b , then the permissible elastic deformation \mathbf{u} is parameterized by the unknown inhomogeneous elastic parameters. In this paper we label the unknown elastic parameters $\boldsymbol{\theta}$ and note that they can be \mathbf{E} , ν , or both depending on the situation. In registration-based elastography, we maximize the similarity between two images of Ω acquired under different static loading by adjusting the elastic properties of the biomechanical model.

2.1 Probabilistic Elastography

We treat \mathbf{u}_b and $\boldsymbol{\theta}$ as unknown random variables and pose registration-based elastography in a Bayesian framework where we marginalize over \mathbf{u}_b to get at the posterior distribution over $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta} \mid I_i, I_e) \propto \int_{\mathbf{u}_b} p(I_i, I_e \mid \boldsymbol{\theta}, \mathbf{u}_b) p(\mathbf{u}_b \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) . \quad (1)$$

Our likelihood model estimates the similarity between two images, $I_i(\mathbf{x})$ (inspiration image) and $I_e(\mathbf{x})$ (expiration image) for $\mathbf{x} \in \Omega$, given the elastic parameters

$\theta(\mathbf{x})$, $\mathbf{x} \in \Omega$, and the boundary conditions $\mathbf{u}_b(\mathbf{x})$, $\mathbf{x} \in \partial\Omega$. A prior on \mathbf{u}_b puts low probability on configurations with high elastic energies, and a spatial prior on the elastic parameters is included to constrain them to be relatively smoothly varying. We employ a Markov Chain Monte Carlo (MCMC) sampling approach to characterize the posterior distribution. An advantage of MCMC is that marginalization over random variables is easily carried out, and from the posterior distribution we can estimate the most probable estimate and assess the uncertainty of this estimate. In the following sections we describe the main components of the proposed elastography method.

Biomechanical Model. The linear elastic energy of an elastic body Ω with elastic material parameters θ and deformed by \mathbf{u} is defined as:

$$E_{elastic}(\mathbf{u}, \theta) = \int_{\Omega} \epsilon(\mathbf{u})^T \sigma(\mathbf{u}, \theta) d\mathbf{x} , \tag{2}$$

where ϵ is the *engineering strain* vector and σ is the *engineering stress* vector [2]. We discretize Eq. (2) with a Finite Element (FE) model of the displacement field. The volumetric solid Ω is discretized into a tetrahedral FE mesh which consists of nodes $\mathbf{n}_b = \{\mathbf{n}_b^1, \dots, \mathbf{n}_b^B\}$ on $\partial\Omega$, nodes $\mathbf{n}_i = \{\mathbf{n}_i^1 \dots \mathbf{n}_i^I\}$ on $\Omega \setminus \partial\Omega$ and P tetrahedral finite elements Ω_e such that $\cap_{e=1 \dots P} \Omega_e = \emptyset$ and $\cup_{e=1 \dots P} \Omega_e = \Omega$. The elastic parameters are defined per element, however, the space of unknown elastic parameters can be reduced by, e.g., grouping together elements or defining another lower dimensional material space and interpolating per element from this space. We discretize the elastic parameters per FE node and use the FE machinery to linearly interpolate these parameters to each element. Consequently, since \mathbf{u}_i can be computed by minimizing Eq. (2) given \mathbf{u}_b and θ , the unknowns in our biomechanical model are the B boundary deformation vectors \mathbf{u}_b and the $Q = B + I$ material parameters θ .

We construct a prior on the boundary deformation parameters \mathbf{u}_b by including the elastic energy in a Boltzmann’s distribution with a temperature T_e :

$$p(\mathbf{u}_b \mid \theta) = \frac{1}{Z} \exp(-E_{elastic}/T_e) . \tag{3}$$

Similarity Model. We incorporate an intensity-based similarity measure to estimate the likelihood of observing the two images given the elastic parameters and the boundary deformation. From the elastic parameters and the boundary deformation, we can find the internal deformation \mathbf{u}_i by minimizing Eq. (2). Any similarity measure that can be formulated as an energy function can be included in the framework, but in this paper we restrict ourselves to the Sum of Squared Differences (SSD) $E_{SSD} = \int_{\Omega} (I_i(\mathbf{x}) - I_e(\mathbf{x} + \mathbf{u}(\mathbf{x})))^2 d\mathbf{x}$. The energy function is converted into a probability by way of the Boltzmann distribution:

$$p(I_i, I_e \mid \theta, \mathbf{u}_b) = \frac{1}{Z} \exp(-E_{SSD}/T_s) , \tag{4}$$

where Z_s is a normalizing constant and T_s is the temperature of the distribution. This is equivalent to modeling the similarity distribution with independent zero-mean Gaussian distributions with a variance of $T_s/2$ for each voxel.

Spatial Prior on Elasticity. Registration based elastography is an underdetermined problem which benefits from including prior knowledge into the estimation process. Furthermore, in areas where there is a lack of image information, the method will favor soft tissue because that will reduce the elastic energy. Our framework includes an elasticity prior which can be used to regularize elasticity at neighboring nodes. We incorporate a prior distribution on the elastic parameters in the form of a Markov model which penalizes the difference in the elastic parameters of neighboring nodes:

$$p(\boldsymbol{\theta}) = \frac{1}{Z} \prod_c \exp(-E_c(\boldsymbol{\theta}_c)/T_c), \quad (5)$$

where T_c is the temperature and $c = \{n_i, n_j\}$ denotes a clique where n_i and n_j are nodes in the FE-mesh connected by an element. The clique energy is defined as $E_c(\boldsymbol{\theta}_c) = (\boldsymbol{\theta}(n_i) - \boldsymbol{\theta}(n_j))^2$. Many organs have relatively homogeneous tissue characteristics (e.g. liver and prostate), while others might have more heterogeneous tissue characteristics (e.g. an unhealthy lung which exhibits emphysema and fibrotic tissue), so the “temperature” T_c can be set according to the prior knowledge we have about tissue homogeneity.

2.2 Sampling the Posterior Distribution

We propose to fully characterize the posterior distribution over the elastic parameters by generating samples from the posterior. Unfortunately, because of the complexity of the posterior, it is not possible to draw samples directly from it. Instead, we use a Markov Chain Monte Carlo (MCMC) (Metropolis-Hastings) method to draw samples from the posterior [3]. It constructs a Markov chain that after a sufficient *burn-in* period has the posterior distribution as its stationary distribution. The necessary burn-in period is dependent on the initial parameters where a poor choice of starting values can require a long burn-in period. Finding optimal starting points for MCMC estimation is currently an area of much research, but the general advice is to initialize the chain with estimates as close to the center of the distribution’s mode as possible [3].

The marginalization of the boundary deformation is easily achieved with a MCMC method by sampling $(\boldsymbol{\theta}^*, \mathbf{u}_b^*) \sim p(\boldsymbol{\theta}, \mathbf{u}_b \mid I_i, I_e)$ and “discarding” the \mathbf{u}_b^* samples. In practice, we start with an initial estimate of the parameters, $\boldsymbol{\tau}^0 = (\boldsymbol{\theta}^0, \mathbf{u}_b^0)$, and generate candidate samples $\hat{\boldsymbol{\tau}}$ from a proposal distribution $q(\boldsymbol{\tau} \mid \boldsymbol{\tau}^{t-1})$ which only depends on the previous sample $\boldsymbol{\tau}^{t-1}$. Candidate samples are accepted with probability

$$A(\hat{\boldsymbol{\tau}}, \boldsymbol{\tau}^t) = \min\left(1, \frac{p(\hat{\boldsymbol{\tau}} \mid I_i, I_e)}{p(\boldsymbol{\tau}^t \mid I_i, I_e)}\right). \quad (6)$$

If a sample is accepted we set $\boldsymbol{\tau}^{t+1} = \hat{\boldsymbol{\tau}}$, otherwise $\boldsymbol{\tau}_{t+1} = \boldsymbol{\tau}^t$.

To reduce the autocorrelation between subsequent samples from the chain and to reduce the memory/storage footprint, *thinning* of the chain is common procedure, i.e. to discard every k -th sample. After burn-in and thinning we have a set of statistically independent samples $\{\boldsymbol{\tau}^0, \dots, \boldsymbol{\tau}^N\}$ from the posterior distribution.

Proposal Distributions. Because the two parameter sets \mathbf{u}_b and $\boldsymbol{\theta}$ are different in nature, we split the parameters into two blocks from which proposal samples are generated independently: $q(\mathbf{u}_b, \boldsymbol{\theta} | \mathbf{u}_b^{t-1}, \boldsymbol{\theta}^{t-1}) = q(\mathbf{u}_b | \mathbf{u}_b^{t-1}) q(\boldsymbol{\theta} | \boldsymbol{\theta}^{t-1})$. A proposal boundary deformation is generated from a univariate normal distribution $\hat{\mathbf{u}}_b \sim q(\mathbf{u}_b | \mathbf{u}_b^{t-1}) = \mathcal{N}(\mathbf{u}_b; \mathbf{u}_b^{t-1}, \sigma_d^2)$ centered on the previous estimate \mathbf{u}_b^{t-1} and with σ_d^2 variance. The elastic parameters are sampled from a multivariate Gaussian $\hat{\boldsymbol{\theta}} \sim q(\boldsymbol{\theta} | \boldsymbol{\theta}^{t-1}) = \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\theta}^{t-1}, \sigma_e^2 \boldsymbol{\Sigma})$ centered on $\boldsymbol{\theta}^{t-1}$ and with covariance $\sigma_e^2 \boldsymbol{\Sigma}$. We assume that the material at node i is likely to vary closely with node j if they are in proximity of each other. We compute $\Sigma_{i,j}$, the element in the covariance matrix which relates the FE-nodes i and j , by a distance function $\Sigma_{i,j} = \exp(-\text{dist}(\mathbf{n}_i, \mathbf{n}_j)/D)$ where D is a constant.

Preventing Folding of Finite Elements. Folding of the tetrahedral elements should be prevented in our sampling scheme because it constitutes non-physical deformations. The boundary of the mesh consists of nodes and a set of triangles that span these nodes. If we can prevent the sampler from drawing samples which folds the boundary triangles, the minimization of Eq. (2) with respect to the elastic parameters will almost guarantee that the resulting deformation is free from folded elements. One exception may occur, however, when one side of the mesh is “pushed” through the opposite side of the mesh. We assume that the mesh is close to convex, that the two mesh sides are relatively far apart, and that there is enough image intensity information between two sides to restrict such a “collapse” of the mesh.

After sampling \mathbf{u}_b , our method searches for folded boundary triangles. When a folded triangle is detected, the deformation of the nodes that span this triangle, as well as any neighboring triangles, are resampled, and this procedure is reiterated until no triangles are folded.

2.3 Summarizing the Posterior Distribution

The local marginal distribution over each random variable contains important information regarding the most probable value and its uncertainty, while the full posterior distribution is a high-dimensional complex object from which complex covariance information can be extracted. Even though we marginalize over the boundary deformations, we still retain the \mathbf{u}_b samples so we can summarize both marginal statistics on $\boldsymbol{\theta}$ and \mathbf{u}_b as well as joint statistics.

A local marginal distribution over one component of $\boldsymbol{\theta}$ or \mathbf{u}_b is a 1D probability distribution. The mode of this distribution is the most probable estimate and the dispersion of the distribution can be interpreted as the uncertainty of the most probable estimate. We use the Inter Quartile Range (IQR), i.e. the difference between the third and first quartiles, to estimate the statistical dispersion of the samples.

3 Results

We have tested the proposed elastography method on both synthetic and clinical CT lung images, and in this section we report on these experiments and the

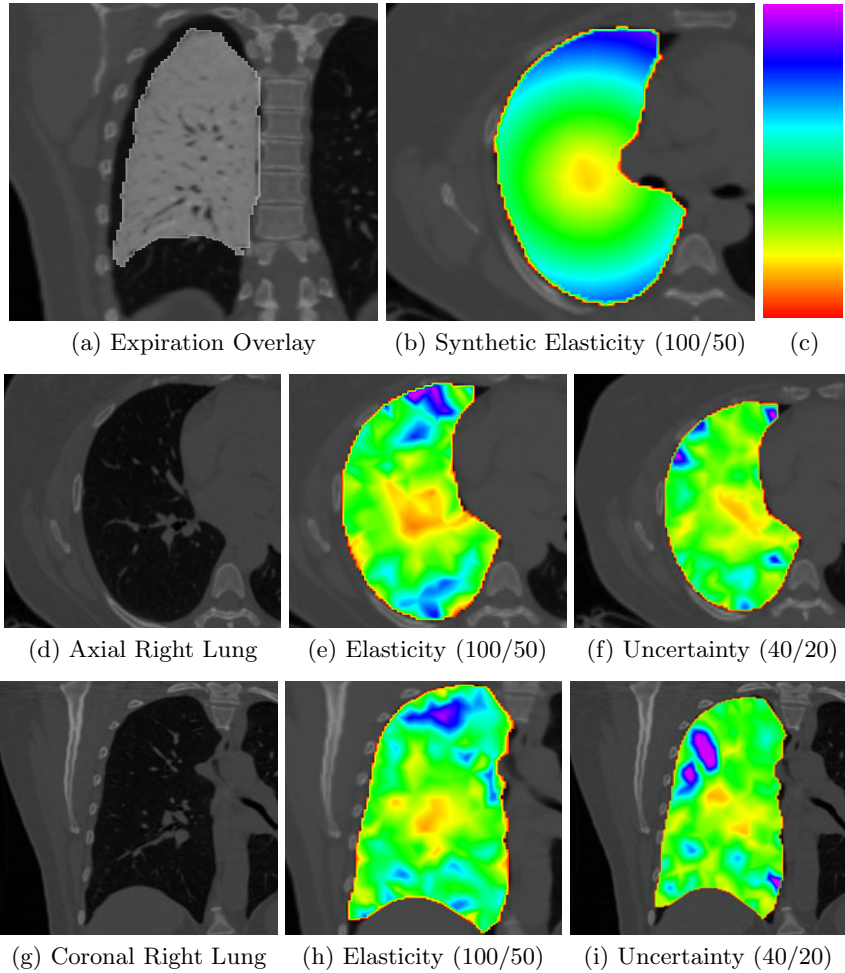


Fig. 1. Synthetic data. All values in parentheses describe the window level setting we use to highlight tissue stiffness where the lower value of the window is colored red (softest) and the upper value is colored purple (stiffest) according to the rainbow color scheme included in (c). (a) Synthetic expiration lung with an inverted gray colormap overlaid on the inspiration volume. Notice the large deformations due to expiration. (b) The synthetic elasticity map corresponding to the slice in (d). (d) Axial slice of the inspiration lung volume. (e) Estimated elasticity. (f) The IQR uncertainty map. (g) Coronal slice of the inspiration lung volume. (h) Estimated elasticity. (i) The IQR uncertainty map. Notice that these results are consistent with the results in Fig. 2 where the general stiffness can be estimated, but with large outliers.

corresponding results. Based on results from [1], we assume that the lung is somewhat compressible with a Poisson's ratio of 0.4 and restrict ourselves to estimate an inhomogeneous Young's modulus ($\theta = \mathbf{E}$).

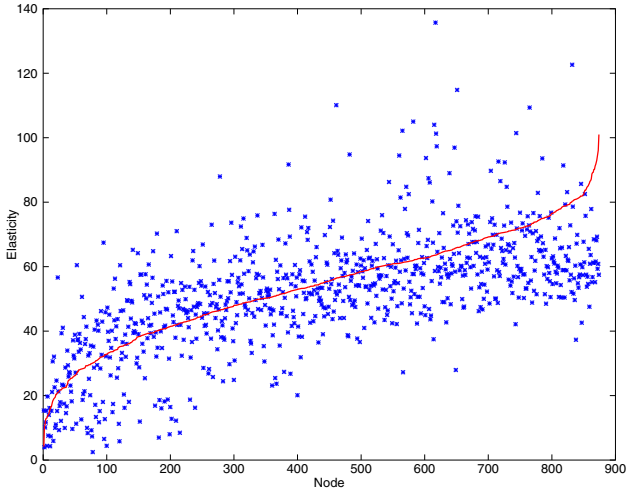


Fig. 2. This figure includes quantitative results on estimating the stiffness (Young’s modulus) for the synthetic case (qualitative results are included in Figure 1). The nodes in the plot have been sorted according to ground truth stiffness and plotted in red while the corresponding estimated stiffness is plotted in blue. It is clear from the plot that the method is sensitive to major differences in stiffness, but the estimates are scattered around the ground truth with a relatively high variance. It is also evident that the method does a better job of estimating the stiffness in the middle to the left of the graph than on the far right. The mean of the difference between the estimated values and the ground truth was 2.6 with a standard deviation of 14.5.

3.1 Synthetic Experiments

We constructed a synthetic dataset with known ground truth elasticity values and tested the ability of the method to recover them. Our inspiration dataset of size $512 \times 512 \times 419$ and spacing (0.95, 0.95, 0.65) mm was acquired from a healthy patient. To reduce the computation time, we downsampled the images by a factor of 2 in each dimension and applied a Gaussian smoothing filter with $\sigma^2 = 0.5$. Slices can be seen in Fig. 1 (d) and (g). A boundary deformation was constructed based on observed deformations between inspiration and expiration datasets. From the geometric center of the right lung, we let the elasticity (Young’s modulus) values increase continuously by a factor of 100 up to the part of the lung at the furthest distance from the center. Using the synthetic boundary deformation and the inhomogeneous elasticity, we could construct the full deformation \mathbf{u} by minimizing Eq. (2) with regards to the deformation. Using the synthetic deformation, we deformed the inspiration lung into a synthetic “expiration” lung. Figure 1 (b) shows a slice of the synthetic elastography map and (a) shows the deformed “expiration” lung overlaid on the full inspiration dataset. From the inspiration label map we created a tetrahedral FE mesh with 874 nodes and 3603 elements. For the estimation we used parameters $T_s = 0.5$, $T_e = 10^7$, $T_c = 300$, $D = 10\text{mm}$, $\sigma_d^2 = 0.005$ and $\sigma_e = 0.1$, where

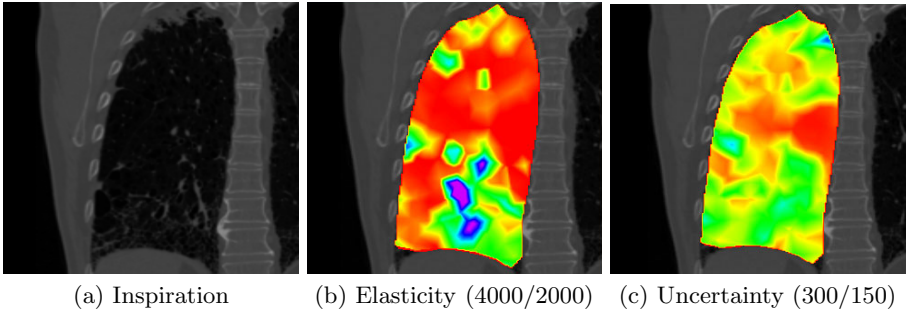


Fig. 3. Fibrotic tissue. The elasticity and uncertainty maps are color coded using the rainbow colormap in Fig. 1 (red is soft, purple is stiff) (c) and the window level in parentheses. (a) A coronal slice of the inspiration dataset where the upper and lower part is identified as fibrotic by a physician. (b) Notice how the fibrotic areas in the lower and upper lung appear stiffer than the surrounding tissue, and also that the cyst in the lower left of the image is estimated as “soft” which could mean that it is ventilated and therefore collapses during expiration. (c) The IQR of the estimated elasticity map. The uncertainty is somewhat larger in the stiffer fibrotic area.

the temperatures are set so that the likelihood is the predominant term in the posterior. The initial Young’s modulus was set to 50 and the boundary deformation was initialized to the ground truth. The sampler generated approximately 7 samples per second. We generated approximately 100 000 samples, discarded the first 50 000 to avoid burn-in effects, and used a thinning factor of 10 to end up with 5 000 independent samples. We computed the mode and IQR and show qualitative results in Fig. 1 and quantitative results in Fig. 2. The method is sensitive to stiffness, but the estimates are scattered around the ground truth value with a relatively high variance which can be explained by the fact that lung CT in some areas contain little image contrast (image information) which may make it difficult for the method to locally estimate a reliable stiffness value. In Fig. 1 we show that the method is able to detect major aspects of the stiffness patterns (the synthetic tissue stiffness increasingly goes from soft in the middle of the lung to stiff on the lung boundary).

3.2 Clinical Experiments

We also tested the method on a clinical dataset which exhibits both fibrotic and emphysematous lung tissue. The original inspiration and expiration CT images have size 512x512x703 and spacing (0.75, 0.75, 0.5) mm, however, to reduce the computation time we downsampled the images by a factor of 2 in each dimension and applied a Gaussian smoothing filter with $\sigma^2 = 0.5$. Labelmasks of both the inspiration and expiration lungs were created using the segmentation method in [11]. From the inspiration labelmask we created a tetrahedral FE mesh with 867 nodes and 3670 elements. The elasticity (Young’s modulus) estimates were homogeneously initialized to 1000. An initial deformation can be found e.g. by

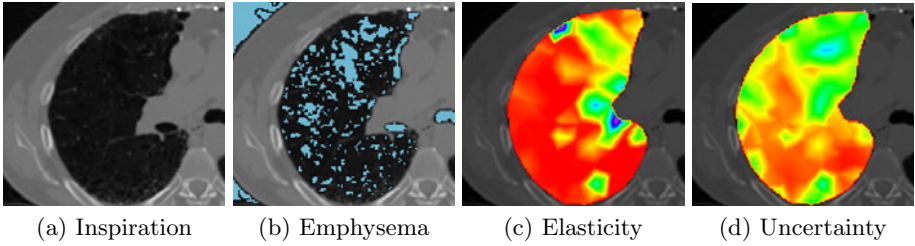


Fig. 4. Emphysematous tissue. The elasticity and uncertainty maps are color coded using the rainbow colormap in Fig. 3(c) and the window levels (4000/2000) and (300/150) for figures (c) and (d) respectively. (a) Axial slice of the inspiration CT dataset. (b) Thresholded image at -950Hu to identify emphysematous tissue and overlaid on the axial slice in (a). (c) Estimated elasticity map. Notice how the elasticity in the lower part of the image is relatively homogeneous, while in the upper area where the emphysema is most predominant it appears stiffer. The reason may be that this emphysematous region is not ventilated and the air filled emphysematous regions do not collapse during expiration. (d) Uncertainty map. We can see that because of the emphysematous regions and consequently the lack of image information, the uncertainty is larger in the area of emphysema compared to non-emphysematous regions.

using a registration algorithm [13], or as we did, by minimizing the distance between \mathbf{u}_b and the expiration lung labelmap by way of a distance field generated from the expiration labelmap.

The sampler generated approximately 4 samples per second while using the following parameters $T_c = 10000$, $T_e = 10^7$, $T_s = 3.0$, $\sigma_d^2 = 0.005$, $\sigma_e = 5$ and $D = 10\text{mm}$. Because we wanted to detect local changes in stiffness in unhealthy lungs we used a high temperature for the T_c and T_e , while the similarity temperature was set relatively low to make it the predominant distribution in the posterior. Approximately 200 000 samples were generated, but after discarding the first half as “burn-in” samples and with a thinning factor of 10, we ended up with 10 000 statistically independent samples. Figure 3 shows elasticity results from parts of the lung where there is considerable fibrosis, while Fig. 4 shows elasticity results from an emphysematous part of the lung.

4 Discussion

This paper has introduced a probabilistic elastography method which can estimate inhomogeneous tissue elasticity given two images of the same object under different external loading. The synthetic results indicate that the method can distinguish between major differences in tissue characteristics. Because expiration causes large deformations and compression of lung tissue, we can expect intensity differences between perfectly aligned inspiration and expiration images because of interpolation artifacts and the fact that compressed tissue in the expiration dataset will appear brighter than the corresponding image in the inspiration image. Lung images also contain regions of relatively homogeneous tissue and sparse image information/contrast which makes it difficult to match the images

without prior information. Both the intensity differences and the lack of image information may explain the outliers we see in the synthetic experiments.

It is generally accepted that lung tissue can be modeled as an elastic material. Our method is very computationally intensive, and consequently we used a linear elastic FE-model with tetrahedral elements to limit computation times to approximately 14 hours. Linear elasticity is applicable when we expect infinitesimal strains or small deformations. However, lung deformations due to inspiration/expiration are large, thus the estimated elasticity can only be assumed to be approximate. Better elastography results may be achieved with lung images where the volume change is smaller. Furthermore, because no external or internal forces are used in the biomechanical calculations, we can only estimate elasticity contrast and not the real elasticity. Another limitation of the simple elastic model is that it assumes an isotropic material model which cannot be assumed to be accurate for the highly anisotropic blood vessels in the lung. Also, because some of the blood vessels are quite large and are connected to the exterior of the lung, they may induce external forces on the internal lung tissue. Consequently, the model may be inaccurate around large vessels.

In future work we will increase the number of FE elements to better discriminate between emphysematous, fibrotic and healthy tissue. However, because of the added degrees of freedom, additional prior information needs to be included e.g. through densitometry to identify regions with emphysema and fibrosis or by having additional correspondence information such as vessel or bronchial branches. FE-methods are known to be sensitive to boundary conditions. The covariance between elasticity and boundary estimates can be extracted from the posterior distribution and used to investigate this sensitivity. Even though registration-based elastography is a challenging under-determined problem, we have shown that it is feasible to estimate lung elasticity with a registration based approach. We believe that with better modeling, e.g. through the inclusion of more prior information and finer resolution meshes, the clinical impact of registration based elastography for studying lung function can be substantial, especially since traditional ultrasound and MR elastography is difficult because of the air filled lungs. However, further validation is required, e.g. by comparing our results with histology samples or MRE. A potential high-impact clinical application would be to study the mechanical nature of smoking related lung disease.

Acknowledgments. This work was supported by NIH grants P41 RR019703, P41 RR13218, K23 HL089353-03 U01 HL089897, U01 HL089856 and an award from the Parker B. Francis Fellowship.

References

1. Al-Mayah, A., Moseley, J., Velec, M., Brock, K.K.: Sliding characteristic and material compressibility of human lung: parametric study and verification. *Med. Phys.* 36(10), 4625–4633 (2009)
2. Bro-nielsen, M.: Finite element modeling in surgery simulation. *Proceedings of the IEEE* 86, 490–503 (1998)

3. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis, 2nd edn. Chapman & Hall/CRC (July 2003)
4. Gokhale, N.H., Barbone, P.E., Oberai, A.A.: Solution of the nonlinear elasticity imaging inverse problem: the compressible case. *Inverse Problems* 24(4) (2008)
5. Gorbunova, V., Lo, P., Ashraf, H., Dirksen, A., Nielsen, M., de Bruijne, M.: Weight preserving image registration for monitoring disease progression in lung CT. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 863–870. Springer, Heidelberg (2008)
6. McGee, K., Hubmayr, R., Ehman, R.: Mr elastography of the lung with hyperpolarized ^3He . *Magn. Reson. Med.* 59, 14–8 (2008)
7. Miga, M.I.: A new approach to elastography using mutual information and finite elements. *Physics in Medicine and Biology* 48(4), 467 (2003)
8. Ou, J.J., Ong, R.E., Yankeelov, T.E., Miga, M.I.: Evaluation of 3d modality-independent elastography for breast imaging: a simulation study. *Physics in Medicine and Biology* 53(1), 147 (2008)
9. Risholm, P., Pieper, S., Samset, E., Wells III, W.M.: Summarizing and visualizing uncertainty in non-rigid registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6362, pp. 554–561. Springer, Heidelberg (2010)
10. Risholm, P., Samset, E., Wells III, W.: Bayesian estimation of deformation and elastic parameters in non-rigid registration. In: Fischer, B., Dawant, B.M., Lorenz, C. (eds.) WBIR 2010. LNCS, vol. 6204, pp. 104–115. Springer, Heidelberg (2010)
11. Ross, J.C., Estépar, R.S.J., Díaz, A., Westin, C.-F., Kikinis, R., Silverman, E.K., Washko, G.R.: Lung extraction, lobe segmentation and hierarchical region assessment for quantitative analysis on high resolution computed tomography images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 690–698. Springer, Heidelberg (2009)
12. Webb, W.R.: Thin-section ct of the secondary pulmonary lobule: anatomy and the image – the 2004 fleischner lecture. *Radiology* 239(2), 322–338 (2006)
13. Yin, Y., Hoffman, E.A., Lin, C.L.: Mass preserving nonrigid registration of CT lung images using cubic B-spline. *Medical Physics* 36(9), 4213–4222 (2009)

Rotation Invariant Completion Fields for Mapping Diffusion MRI Connectivity

Parya MomayyezSiahkal and Kaleem Siddiqi

School of Computer Science & Centre for Intelligent Machines, McGill University

Abstract. Partial differential equations have been successfully used for fibre tractography and for mapping connectivity indices in the brain. However, the current implementation of methods which require 3D orientation to be tracked can suffer from serious shortcomings when invariance to 3D rotation is desired. In this paper we focus on the 3D stochastic completion field and introduce a new methodology to solve the underlying PDE in a manner that achieves rotation invariance. The key idea is to use spherical harmonics to solve the Fokker-Planck equation representing the evolution of the probability density function of a 3D directional random walk. We validate the new approach by presenting improved connectivity indices on synthetic data, on the MICCAI 2009 Fibre Cup phantom and on a biological phantom comprised of two rat spinal chords in a crossing configuration.

Keywords: Diffusion MRI, Completion fields, Connectivity, Spherical harmonics, Fokker-Planck equation.

1 Introduction

Partial differential equations (PDEs) have been used extensively for white matter tractography in diffusion magnetic resonance imaging (Diffusion MRI) [10], [2], [8], [9]. These methods offer the advantage that a theoretically well-grounded measure of connectivity between two regions can be found via a direct computation of a connectivity map, represented by the steady state solution to the PDE. Additionally, the minimum cost pathways can be easily found via a post-processing step. A further advantage is that they can incorporate relevant information from all directions, not just the maxima of the fibre orientation distributions (FODs), into the framework.

Among the PDE-based tractography/connectivity methods, some try to find the steady state solution of variations of the diffusion equation in an anisotropic medium. O'Donnell et al. [10] work directly with the diffusion equation in an anisotropic medium, using 3-dimensional tensors as conductivity tensors. The degree of connectivity is approximated by the steady state flow along any path. Batchelor et al. [2] modify this algorithm by adding a convection term to the PDE, integrating a measure of anisotropy into the concentration flow. In [8], Hageman et al. extend the probabilistic first-order PDE methods using a model based on fluid-mechanics. A fluid flow is simulated through a diffusion tensor

field such that local anisotropy values are encompassed in the PDE via a viscosity term. In a related approach, Fletcher et al. propose a front evolution scheme to compute the minimal cost of connecting two regions of interest [6]. Pichon et al. [11] have generalized PDE methods to the case of high angular resolution diffusion imaging (HARDI) by solving the Hamilton-Jacobi-Bellman equation for minimum cost curves. Motivated by Williams and Jacobs' 2D completion field model in computer vision [14], Momayyez et al. [9] propose a somewhat different computational model, where the probability density of water molecules in a 3D directional random walk is described by the underlying Fokker-Planck equation. Their approach distinguishes itself from the other PDE-based methods by developing a computational model that is fundamentally linked to the underlying physical process, i.e., the diffusion of water molecules in an anisotropic medium.

Although the algorithm suggested in [9] is numerically stable and offers a convenient way to incorporate local diffusion data by setting diffusion and decay parameters based on the FOD computed at each voxel, the uniform sampling of θ and ϕ (the colatitude and azimuth in spherical coordinates) leads to rotation variance. Rotation variance implies that the computed connectivity indices depend on the orientation of the diffusion weighted image (DWI) in a given global coordinate system, which is clearly unacceptable. The underlying uniform sampling of θ and ϕ introduces singularities at $\theta = 0, \pi$ and also contributes to a nonuniform sampling of the spherical surface such that the density of samples on the sphere varies with the distance from the poles. One might be tempted to simply use a near-uniform sampling scheme on a spherical shell, but this complicates the numerical method since the notion of the nearest neighbour in θ and ϕ coordinates, which plays a role in the discretized PDE, is no longer well-defined.

In the present article we introduce a new methodology for achieving rotation invariance in the solution of the PDE underlying the stochastic completion field algorithm. The key idea is to use a spherical harmonics (SPH) basis to provide a rotation invariant numerical estimation of the evolution of the probability density function associated with a 3D directional random walk. While the development in this paper is tailored to a specific PDE, the basic idea of solving PDEs on domains with a spherical topology has been of significant interest in the fields of astronomy, geophysics, and nuclear physics [7] and can find applications in a wide variety of medical imaging and computer vision applications where 3D orientation data is involved. We show that the proposed framework can be formulated in a manner that is independent of the sampling scheme used for ϕ and θ . Therefore, using a near-uniform sampling on the spherical shell together with the SPH-based algorithm achieves the desired rotation invariance. A second contribution of this article is the modification of the 3D directional random walk such that the Fokker-Planck equation has two additional drift terms, which improve the coherence of the simulated trajectories by better exploiting local fiber orientations. We validate this new approach by comparing the obtained connectivity indices with those computed by the method of [9] on synthetic data. Further results are presented on a rat spinal cord phantom [4] and on the MICCAI 2009 Fibre Cup phantom [12].

2 Background: 3D Stochastic Completion Fields

In the 3D stochastic completion field model introduced in [9] a 3D directional random walk is used to model the Brownian motion of water molecules. The state of a particle (x, y, z, θ, ϕ) in $\mathbb{R}^3 \times S^2$ is updated according to the following set of differential equations:

$$\begin{aligned} dx &= \sin \theta \cos \phi dt; & dy &= \sin \theta \sin \phi dt; & dz &= \cos \theta dt \\ d\theta &= \sigma_\theta dB(\theta); & d\phi &= \sigma_\phi dB(\phi); & dB(\theta), dB(\phi) &\sim N(0, dt). \end{aligned} \quad (1)$$

Under this model particles tend to travel in a straight line along the direction of their current orientation followed by a slight change in the orientation at each step. The deviations in orientation are controlled by the Brownian motion terms $B(\phi)$ and $B(\theta)$ and are proportional to the diffusion parameters σ_ϕ^2 and σ_θ^2 . Additionally, an average lifetime ζ is associated with each particle, thus favouring shorter paths over longer ones. Under this model, the probability of passing through a particular state while bridging the gap between a source and a sink region is given by the product of a *stochastic source field* and a *stochastic sink field* [14], [9]. Furthermore, the Fokker-Planck equation describing the time evolution of the probability density of a particle in the course of its 3D directional random walk is given as:

$$\frac{\partial P}{\partial t} = -\sin \theta \cos \phi \frac{\partial P}{\partial x} - \sin \theta \sin \phi \frac{\partial P}{\partial y} - \cos \theta \frac{\partial P}{\partial z} + \frac{\sigma_\phi^2}{2} \frac{\partial^2 P}{\partial \phi^2} + \frac{\sigma_\theta^2}{2} \frac{\partial^2 P}{\partial \theta^2} - \frac{1}{\zeta} P. \quad (2)$$

The results presented in [9] demonstrate the ability of a Lax-Wendroff scheme for discretizing the advection terms coupled with a Crank-Nicholson scheme for the diffusion terms to achieve numerically stable connectivity measures between two seed regions. However, this implementation suffers from the serious limitation of having singularities at the poles due to the uniform sampling of ϕ and θ , and of being rotationally variant due to the subsequent nonuniform density of spherical samples. In the following we overcome this limitation by using a SPH formulation of the PDE in the spherical domain to achieve rotation invariance, which is the main contribution of this article. This strategy can be applied in other medical imaging settings where 3D orientations have to be tracked and updated in a PDE.

We also suggest two modifications of the above model to yield improved performance. The first one has to do with the use of the Lax-Wendroff scheme for the advective part of Equation 2. Using the Lax method, e.g., for $\frac{\partial P}{\partial t} = -\sin \theta \cos \phi \frac{\partial P}{\partial x}$, a change in the probability value of a voxel (x, y, z) at time t is propagated to voxels $(x+1, y, z)$ and $(x-1, y, z)$ at time $t+1$. In fact, if $\sin \theta \cos \phi > 0$, then only the value of the voxel $(x+1, y, z)$ must be updated. A similar effect occurs for negative values of $\sin \theta \cos \phi$. In the current implementation, we use an upwind differencing method for the advection terms to resolve this issue. The second modification is the addition of two extra angular advection steps to the algorithm, as described in the following section, to exploit the local fibre orientations at each voxel more carefully. This in turn reduces the penalty for highly curved tracts, when the underlying data supports such pathways.

3 Angular Constraints on 3D Directional Random Walk

In a directional random walk, as described in [9], the stochastic component of the particle’s motion is constrained to the orientation change which is entirely Brownian. Adhering to this original definition, the probability of aligning the particle’s orientation with an FOD maximum is inversely related to the exponential of the angular difference between the incoming direction and the maxima. Due to this characteristic a lower probability value will be associated with curved fibre tracts compared to straight ones. While this can be a desirable property in situations where one attempts to bridge the gap between two edge segments with no other informative features in between, one would want to remove this bias when the underlying data supports larger orientation changes.

To resolve this issue and to remove the bias towards straight fibre tracts, we propose a variation of the 3D directional random walk where a particle’s state is updated according to the following steps:

1. The particle moves in 3D in the direction of its current orientation:

$$dx = \sin \theta \cos \phi dt; \quad dy = \sin \theta \sin \phi dt; \quad dz = \cos \theta dt. \tag{3}$$

2. The particle’s orientation is changed according to two deviations in the osculating and the binormal planes incorporating the stochastic motion of the random walk. Unlike the original directional random walk, a drift term is added to the orientation changes $d\theta$ and $d\phi$ at each step proportional to μ_θ and μ_ϕ which are set based on the angular difference between the current orientation and the FOD maximum. The diffusion terms σ_ϕ^2 and σ_θ^2 , on the other hand, govern the amount of diffusion allowed at each step contributing to the Brownian component of the directional random walk as before.

$$d\theta = \mu_\theta dt + \sigma_\theta dB(\theta); \quad d\phi = \mu_\phi dt + \sigma_\phi dB(\phi); \quad dB(\theta), dB(\phi) \sim N(0, dt). \tag{4}$$

The stochastic differential equations given in Eq. [4] push the incoming direction towards the most likely orientation at each voxel (when available) through the associated nonzero drift terms. This encourages the particles to follow the closest FOD maxima in the course of their directional random walk, while the nonzero diffusion coefficients allow the particles to follow other directions as well to compensate for noise, data and other model-dependent inaccuracies.

Modified Fokker-Planck Equation. The introduction of the drift terms into the orientation changes at each step leads to the modification of the underlying Fokker-Planck equation by the addition of extra angular advection terms to the PDE:

$$\begin{aligned} \frac{\partial P}{\partial t} = & -\sin \theta \cos \phi \frac{\partial P}{\partial x} - \sin \theta \sin \phi \frac{\partial P}{\partial y} - \cos \theta \frac{\partial P}{\partial z} \\ & - \mu_\phi \frac{\partial P}{\partial \phi} - \mu_\theta \frac{\partial P}{\partial \theta} + \frac{\sigma_\phi^2}{2} \frac{\partial^2 P}{\partial \phi^2} + \frac{\sigma_\theta^2}{2} \frac{\partial^2 P}{\partial \theta^2} - \frac{1}{\zeta} P. \end{aligned} \tag{5}$$

In the rest of this paper, the Fokker-Planck equation will refer to Equation [5].

4 Rotation Invariant Computation

In [9], the solution to the diffusion terms of Eq. 5 is computed using the implicit Crank-Nicholson numerical scheme. The numerical estimation is based upon uniform sampling of ϕ and θ on the spherical surface, which suffers from the inherent problem of rotation variance. Rotation invariant computation of the Fokker-Planck equation requires a uniform sampling of the spherical surface, but this makes the application of finite differencing methods in the spherical domain challenging. Whereas there is no known point set on a spherical surface which can be realized as the analog of uniform sampling in Euclidean space, our goal is to minimize the error associated with the global properties of the sampling points. To accomplish this, we use a quasi-uniform point set obtained by an electrostatic repulsion approach on the sphere, known as the minimum energy point distribution (ME) [4]. A numerical solution based on the SPH expansion of the probability function on the spherical shell is consequently devised. This SPH-based formulation is not linked to any particular sampling scheme, providing the appropriate framework for employing a quasi-uniform spherical point set.

4.1 Spherical Harmonics Based Formulation

Spherical harmonics (SPH), normally denoted by Y_l^m , are the spherical analog of the Fourier transform basis defined for complex functions on the unit sphere. These functions form a complete orthonormal system of the space of square integrable functions on the sphere $L^2(\mathbf{S}^2)$ and are given as:

$$Y_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) \exp(im\phi), \tag{6}$$

where P_l^m is an associated Legendre polynomial[4]. SPHs provide a natural framework to work with functions living on the sphere and as such, they have been used widely in different application areas dealing with data on sphere. Many algorithms have also been introduced in the diffusion-MRI community which have adapted spherical harmonics basis to model the diffusion signal [13], [5].

Forming an orthonormal basis on the spherical shell, any spherical function can be written as $P(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_l^m Y_l^m(\theta, \phi)$ where the expansion is usually truncated at some order L . Given N data points on the sphere $\mathbf{P} = \{P_1, \dots, P_N\}$ and choosing an approximation order of L , a linear least-squares scheme can be used to solve for the unknown SPH coefficients $\mathbf{c} = \{c_0^0, c_1^{-1}, c_1^0, c_1^1, \dots, c_L^L\}$ [1]. The linear least-squares approach seeks to solve the following linear system of equations using the pseudo-inverse of the SPH matrix:

$$\begin{bmatrix} Y_0^0(\theta_1, \phi_1) & Y_1^{-1}(\theta_1, \phi_1) & \dots & Y_L^L(\theta_1, \phi_1) \\ \vdots & \vdots & \ddots & \vdots \\ Y_0^0(\theta_N, \phi_N) & Y_1^{-1}(\theta_N, \phi_N) & \dots & Y_L^L(\theta_N, \phi_N) \end{bmatrix} \times \begin{bmatrix} c_0^0 \\ \vdots \\ c_L^L \end{bmatrix} = \begin{bmatrix} P_1 \\ \vdots \\ P_N \end{bmatrix}, \tag{7}$$

$$\mathbf{Yc} = \mathbf{P}, \quad \mathbf{c} = \mathbf{Y}^+\mathbf{P} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{P}.$$

¹ In this work, we use the real form of the spherical harmonics since the function to be expanded is a real-valued probability density function.

The SPH expansion together with the above equation is next used to develop an algorithm to solve the spherical part of the PDE given in Eq. 5:

$$\begin{aligned}
P(t, \theta, \phi) &= \sum_{l=0}^L \sum_{m=-l}^l c_l^m(t) Y_l^m(\theta, \phi) \\
\frac{\partial P}{\partial t} &= -\mu_\phi \frac{\partial P}{\partial \phi} - \mu_\theta \frac{\partial P}{\partial \theta} + \frac{\sigma_\phi^2}{2} \frac{\partial^2 P}{\partial \phi^2} + \frac{\sigma_\theta^2}{2} \frac{\partial^2 P}{\partial \theta^2} \\
&= \sum_{l=0}^L \sum_{m=-l}^l c_l^m(t) \left(-\mu_\phi(\theta, \phi) \frac{\partial Y_l^m(\theta, \phi)}{\partial \phi} - \mu_\theta(\theta, \phi) \frac{\partial Y_l^m(\theta, \phi)}{\partial \theta} \right. \\
&\quad \left. + \frac{\sigma_\phi^2(\theta, \phi)}{2} \frac{\partial^2 Y_l^m(\theta, \phi)}{\partial \phi^2} + \frac{\sigma_\theta^2(\theta, \phi)}{2} \frac{\partial^2 Y_l^m(\theta, \phi)}{\partial \theta^2} \right).
\end{aligned} \tag{8}$$

Note that in Eq. 8 the x -, y -, z - dependence of P is implicit since the spherical PDE is solved locally for each voxel independent of the neighbouring voxels. The partial derivatives of Y_l^m with respect to θ and ϕ at a point can be expressed as the combination of SPHs of different phase factors m (but the same order l) at the given point. Using this property, Eq. 8 can be written simultaneously for all the sample points in matrix form leading to a linear system of ordinary differential equations (ODEs):

$$\begin{aligned}
\frac{\partial \mathbf{P}}{\partial t} &= (-\mathbf{M}_\phi \mathbf{Y}_\phi - \mathbf{M}_\theta \mathbf{Y}_\theta + \Sigma_\phi \mathbf{Y}_{\phi^2} + \Sigma_\theta \mathbf{Y}_{\theta^2}) \mathbf{c} \\
&= (-\mathbf{M}_\phi \mathbf{Y}_\phi - \mathbf{M}_\theta \mathbf{Y}_\theta + \Sigma_\phi \mathbf{Y}_{\phi^2} + \Sigma_\theta \mathbf{Y}_{\theta^2}) \mathbf{Y}^+ \mathbf{P} = \mathbf{D} \mathbf{P},
\end{aligned} \tag{9}$$

where:

$$\begin{aligned}
\mathbf{M}_\phi &= \mathbf{diag}[\mu_\phi(\theta_1, \phi_1), \dots, \mu_\phi(\theta_N, \phi_N)] \quad \mathbf{M}_\theta = \mathbf{diag}[\mu_\theta(\theta_1, \phi_1), \dots, \mu_\theta(\theta_N, \phi_N)], \\
\Sigma_\phi &= \mathbf{diag}[\sigma_\phi(\theta_1, \phi_1), \dots, \sigma_\phi(\theta_N, \phi_N)] \quad \Sigma_\theta = \mathbf{diag}[\sigma_\theta(\theta_1, \phi_1), \dots, \sigma_\theta(\theta_N, \phi_N)], \\
\mathbf{Y}_\phi &= \begin{bmatrix} \frac{\partial Y_0^0(\theta_1, \phi_1)}{\partial \phi} & \dots & \frac{\partial Y_L^L(\theta_1, \phi_1)}{\partial \phi} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_0^0(\theta_N, \phi_N)}{\partial \phi} & \dots & \frac{\partial Y_L^L(\theta_N, \phi_N)}{\partial \phi} \end{bmatrix} \quad \mathbf{Y}_\theta = \begin{bmatrix} \frac{\partial Y_0^0(\theta_1, \phi_1)}{\partial \theta} & \dots & \frac{\partial Y_L^L(\theta_1, \phi_1)}{\partial \theta} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_0^0(\theta_N, \phi_N)}{\partial \theta} & \dots & \frac{\partial Y_L^L(\theta_N, \phi_N)}{\partial \theta} \end{bmatrix}, \\
\mathbf{Y}_{\phi^2} &= \begin{bmatrix} \frac{\partial^2 Y_0^0(\theta_1, \phi_1)}{\partial \phi^2} & \dots & \frac{\partial^2 Y_L^L(\theta_1, \phi_1)}{\partial \phi^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 Y_0^0(\theta_N, \phi_N)}{\partial \phi^2} & \dots & \frac{\partial^2 Y_L^L(\theta_N, \phi_N)}{\partial \phi^2} \end{bmatrix} \quad \mathbf{Y}_{\theta^2} = \begin{bmatrix} \frac{\partial^2 Y_0^0(\theta_1, \phi_1)}{\partial \theta^2} & \dots & \frac{\partial^2 Y_L^L(\theta_1, \phi_1)}{\partial \theta^2} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 Y_0^0(\theta_N, \phi_N)}{\partial \theta^2} & \dots & \frac{\partial^2 Y_L^L(\theta_N, \phi_N)}{\partial \theta^2} \end{bmatrix}.
\end{aligned} \tag{10}$$

It is important to emphasize that the spherical harmonic formulation still allows for local specification of the angular diffusion $\sigma_\phi^2, \sigma_\theta^2$ and advection μ_ϕ, μ_θ coefficients. These parameters are selected locally based on the diffusion-MRI data available throughout the volume. In order to avoid too large or too small time steps, an adaptive step size numerical algorithm, the Runge-Kutta-Fehlberg (RKF45) method, is used to advance the solution of Eq. 9 in time.

5 The Complete Computational Model

A summary of the overall computational method is provided in the following set of equations where, similar to [9], a time splitting approach has been employed. As mentioned in Section 1, we replace the Lax-Wendroff scheme in [9] by an upwind differencing scheme to solve the PDE along the x -, y -, and z - coordinates. The SPH formulation is also used to solve the PDE in the spherical domain:

$$\begin{aligned}
 P_{x,y,z,\theta,\phi}^{t+\frac{1}{5}} &= P_{x,y,z,\theta,\phi}^t - \sin \theta \cos \phi \begin{cases} P_{x,y,z,\theta,\phi}^t - P_{x-1,y,z,\theta,\phi}^t & \text{if } \sin \theta \cos \phi > 0 \\ P_{x+1,y,z,\theta,\phi}^t - P_{x,y,z,\theta,\phi}^t & \text{if } \sin \theta \cos \phi < 0 \end{cases}, \\
 P_{x,y,z,\theta,\phi}^{t+\frac{2}{5}} &= P_{x,y,z,\theta,\phi}^{t+\frac{1}{5}} - \sin \theta \sin \phi \begin{cases} P_{x,y,z,\theta,\phi}^{t+\frac{1}{5}} - P_{x,y-1,z,\theta,\phi}^{t+\frac{1}{5}} & \text{if } \sin \theta \sin \phi > 0 \\ P_{x,y+1,z,\theta,\phi}^{t+\frac{1}{5}} - P_{x,y,z,\theta,\phi}^{t+\frac{1}{5}} & \text{if } \sin \theta \sin \phi < 0 \end{cases}, \\
 P_{x,y,z,\theta,\phi}^{t+\frac{3}{5}} &= P_{x,y,z,\theta,\phi}^{t+\frac{2}{5}} - \cos \theta \begin{cases} P_{x,y,z,\theta,\phi}^{t+\frac{2}{5}} - P_{x,y,z-1,\theta,\phi}^{t+\frac{2}{5}} & \text{if } \cos \theta > 0 \\ P_{x,y,z+1,\theta,\phi}^{t+\frac{2}{5}} - P_{x,y,z,\theta,\phi}^{t+\frac{2}{5}} & \text{if } \cos \theta < 0 \end{cases}, \\
 \mathbf{P}_{x,y,z,\theta,\phi}^{t+\frac{4}{5}} &= \mathbf{D}\mathbf{P}_{x,y,z,\theta,\phi}^{t+\frac{3}{5}}, \\
 P_{x,y,z,\theta,\phi}^{t+1} &= e^{-\frac{1}{\zeta}} P_{x,y,z,\theta,\phi}^{t+\frac{4}{5}}.
 \end{aligned} \tag{11}$$

Both the source field and the sink field are computed by time integration of the above probability density function representing a particle’s state in the course of its directional random walk:

$$P'(x, y, z, \theta, \phi) = \int_0^\infty P^t(x, y, z, \theta, \phi) dt \approx \sum_0^{t'} P^t(x, y, z, \theta, \phi). \tag{12}$$

The angular advection and diffusion coefficients $\mu_\theta, \mu_\phi, \sigma_\theta, \sigma_\phi$ and the life time coefficient ζ are set locally for each state using the fibre orientation distribution (FOD) calculated at each voxel. For all orientations within a certain angular difference from the FOD maxima, μ_θ and μ_ϕ are equal to the angular difference between the closest FOD maximum and the current orientation. σ_θ and σ_ϕ are set to some fixed value to take the effect of the underlying noise and inaccuracies into account. Finally, the particle’s life time ζ for each state is also set based on the angular difference, i.e., ζ is chosen to be very small for orientations far from the FOD maxima, and it diminishes linearly with increasing angular difference for the rest of the orientations.

6 Experimental Results

In this section, we evaluate the performance of our algorithm by running a series of validation experiments. Qualitative and quantitative validation of our method is provided using synthetic multi-tensor data, a biological phantom and the MICCAI 2009 Fibre Cup phantom made of small-diameter acrylic fibres. We also provide comparative results between the connectivity measures provided by the 3D completion field algorithm and a standard probabilistic fibre tractography algorithm, similar to that of [3].

6.1 Validation on Synthetic Data

To investigate the rotation invariant property of our new computational model, we generate synthetic data using the multi-tensor model of [11]. For our experiments, a quasi-uniform ME point set of 100 directions on the hemisphere is chosen for diffusion encoding directions. The synthetic data is created by placing diffusion tensors (DTs) along the fibre pathways where the major eigenvector of a DT aligns with the tangent vector of the curves. Partial volume averaging is also considered for voxels with more than one curve passing through them. Background voxels are filled with isotropic tensors with eigenvalues equal to $600 * 10^{-6} \frac{mm^2}{s}$. The eigenvalues for the anisotropic voxels are chosen based on the desired fractional anisotropy. Finally, we simulate the effect of noise by adding complex Gaussian noise to the raw signal.

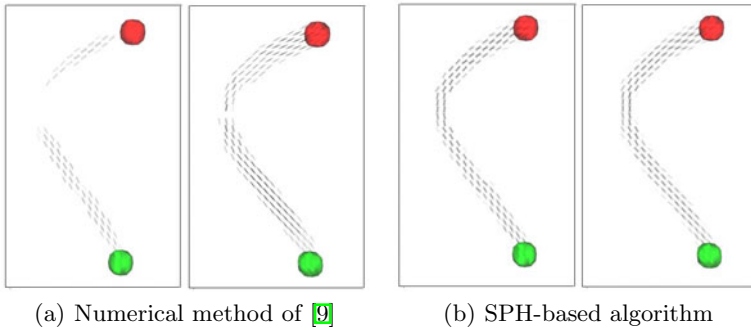


Fig. 1. Visualizing the effect of rotation on the performance of the SPH-based algorithm and the numerical method of [9]

Figure 1 shows the probability distributions obtained for a synthetic curved tract in two perpendicular spatial orientations, using the numerical method given in [9] and the new rotation invariant method [1]. It is clear that the SPH formulation recovers the underlying connection more accurately. More significant though is the difference between the probability distributions obtained by the algorithm of [9] for the two spatial orientations. The algorithm even fails to completely recover the connectivity pattern for one of the spatial orientations (Fig. 1(a) left). Conversely, the probability distributions obtained by the SPH-based algorithm clearly demonstrate the robustness of the algorithm to rotation in space. The same conclusion can be drawn from the connectivity measures summarized in Table 1. The small difference still observed for the rotation invariant method is due to the errors introduced by the time splitting approach, which updates the probability distribution function for x -, y - and z - coordinates in separate steps. The effect is more severe for our synthetically generated data with clear

² For the purpose of visualization, all results are projected onto the 2D plane. The reader is also encouraged to zoom-in on the figures in the electronic version for a better view of the orientations and the transparency values.

jumps from voxel to voxel. We expect this effect to be milder for less coarse real data. The connectivity measures, as described in [9], are computed by taking the average of the probability values of the voxels contained in the source and sink region pairs.

Table 1. Connectivity indices obtained for two spatial orientations of the same synthetic fibre tract

Algorithm	1st Configuration	2nd Configuration
numerical method of [9]	0.0332511	0.271475
SPH-based Algorithm	0.154295	0.197969

6.2 Validation on Phantom Data

Fiber Cup Phantom Data. The new SPH-based algorithm was also tested on the MICCAI 2009 Fibre cup phantom [12], with known ground truth. For our experiments, we used the DWIs with a spatial resolution of $3 \times 3 \times 3 \text{mm}^3$ and a b -value of 1500 s/mm^2 .

Figure 2 provides a probabilistic view of the connectivity patterns obtained for different sink and source region pairs, using our SPH-based algorithm. We have used the same pairs of sink and source regions as in [9], where the first four are associated with true fibre tracts while the last two are not. The connectivity measures provided by the algorithm are also summarized in Table 2 and are compared to the measures obtained by the numerical method of [9] and those of a probabilistic tractography approach similar to [3]. The probabilistic method uses the uncertainty computed from a residual bootstrapping approach to run many iterations of a streamline tractography technique, where the weakest link along the reconstructed fibre tracts represents the connectivity strength from the seed region to every voxel.

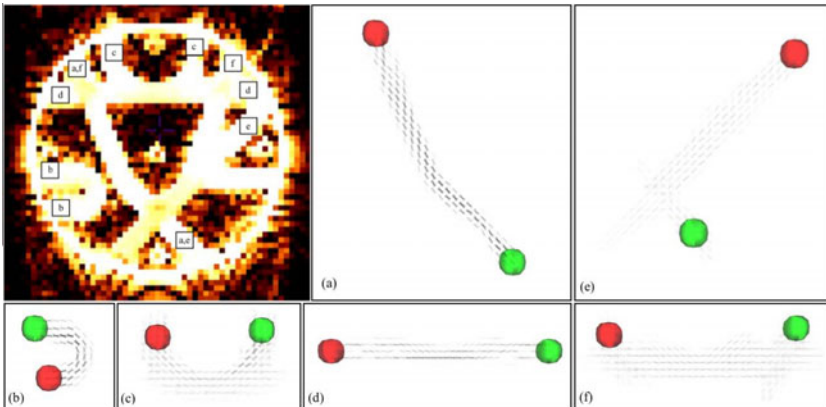


Fig. 2. Results of the SPH-based model. The darker the vectors, the higher the associated probabilities. Top-Left: Mean diffusivity image and the ROIs used.

Similar to [9], one major observation is the failure of the weakest link measure to provide a consistent connectivity index for true connections. While the algorithm correctly assigns zero connectivity to the connections (e) and (f), no connection is found for the straight connection of (d). Moreover, the connectivity measures for the other three true connections extend over a wide range of values, making any kind of inference based on these measures inaccurate. More significant though is the improvement gained by using our new SPH-based algorithm over the numerical method of [9]. It should be noted that since this phantom is almost planar, the results obtained mainly show the effect of adding the angular advection terms to the Fokker-Planck equation. While the previous method assigns almost similar measures to the cases (c) and (f), where the first is a true connection and the second is not, the connectivity measure provided by the new algorithm is very low for both cases (e) and (f) and much higher for the other cases which represent true connections. The fact that the connectivity indices of cases (c) and (d) are an order of magnitude lower than those of cases (a) and (b), is mainly due to the complex configuration of fibre tracts at the top part of the phantom, causing the connectivity measure to drop as expected.

Table 2. Connectivity indices obtained for six different seed region pairs

Algorithm	(a)	(b)	(c)	(d)	(e)	(f)
SPH CF	0.327592	0.258567	0.0614615	0.0947138	0.000047583	0.000594305
Num. Mth of [9]	0.17152	0.22445	0.04716	0.24342	0.00551	0.07559
Probabilistic	0.00041	0.239901	0.00792	0	0	0

Biological Phantom Data. We also evaluated the performance of our algorithm on a physical phantom with known connectivity built from excised rat spinal cords [4]. The cords were scanned to acquire 90 DWIs as well as 10 T2-weighted images. Unlike the physical phantom used in the experiments of the previous section, the biological phantom is not planar and thus this experiment better explores the 3D performance of our algorithm.

Figure 3 (left) shows the anisotropy image of the phantom and the seed regions used. The table on the right summarizes the connectivity measures obtained between different sink and source regions for our method and the probabilistic tractography algorithm. Based on the ground truth, it is known that ROI (a) is connected to ROI (b) and (c) is connected to (d). The biological phantom has a complex crossing section and low anisotropy values, which make tracking within the phantom a difficult task. However, from the connectivity measures provided by the SPH-based algorithm, it is clear that our algorithm does a better job in distinguishing true connections from false connections when compared to the probabilistic algorithm. The probabilistic tractography approach fails to reconstruct the pathway from ROI (c) to (d). In fact, either starting from ROI (c) or (d), the algorithm cannot pass through the crossing region, providing no connectivity for any of the (cd), (ac) and (bd) ROI pairs. In contrast, the SPH-based completion field algorithm provides us with connectivity measures that are an order of magnitude higher for the true connections.



Algorithm	SPH-Based CF	Probabilistic
ab	0.353533	0.201719
cd	0.174609	0
ac	0.057683	0
bd	0.028613	0

Fig. 3. Anisotropy image and connectivity indices obtained for the biological phantom

7 Conclusion

This paper introduces a mathematical formulation to achieve rotation invariant solutions for partial differential equations where 3D orientation data is involved. The method has been applied to the specific problem of diffusion MRI connectivity measurements using 3D stochastic completion fields. In addition, to make the application of 3D stochastic completion fields more appropriate for brain connectivity measurement in the presence of dense orientation information, additional angular advection terms are incorporated into the 3D directional random walk, leading to a modified Fokker-Planck equation. Using synthetic diffusion MRI data, we have demonstrated that our SPH-based method is more robust to rotation in space. Further experiments on two physical and biological phantoms show that the new formulation outperforms some existing tractography algorithms, a standard probabilistic tractography method and the rotation variant implementation of the completion field, by providing more reliable connectivity measures across the phantoms and by distinguishing true connections from false ones. In future, we plan to further explore the performance of our algorithm by performing experiments on human brain data and also by setting the diffusion parameters based on measurements of the underlying noise and data uncertainty.

Acknowledgments. We are grateful to Jennifer Campbell and Bruce Pike for many helpful discussions and for providing the biological phantom data for our experiments. This work was supported by grants from NSERC and FQRNT.

References

- Alexander, D.C., Barker, G.J., Arridge, S.R.: Detection and modeling of non-gaussian apparent diffusion coefficient profiles in human brain data. *Mag. Reson. in Med.* 48, 331–340 (2002)
- Batchelor, P.G., Hill, D.L.G., Calamante, F., Atkinson, D.: Study of connectivity in the brain using the full diffusion tensor from MRI. In: Insana, M.F., Leahy, R.M. (eds.) *IPMI 2001*. LNCS, vol. 2082, pp. 121–133. Springer, Heidelberg (2001)
- Berman, J.I., Chung, S., Mukherjee, P., Hess, C.P., Han, E.T., Henry, R.G.: Probabilistic streamline Q-ball tractography using the residual bootstrap. *NeuroImage* 39, 215–222 (2008)
- Campbell, J.S.W., Siddiqi, K., Rymar, V.V., Sadikot, A.F., Pike, G.B.: Flow-based fiber tracking with diffusion tensor and Q-ball data: validation and comparison to principal diffusion direction techniques. *NeuroImage* 27, 725–736 (2005)

5. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-ball imaging. *Mag. Reson. in Med.* 58, 497–510 (2007)
6. Fletcher, P.T., Tao, R., Jeong, W.-K., Whitaker, R.T.: A volumetric approach to quantifying region-to-region white matter connectivity in diffusion tensor MRI. In: Karssemeijer, N., Lelieveldt, B. (eds.) *IPMI 2007*. LNCS, vol. 4584, pp. 346–358. Springer, Heidelberg (2007)
7. Fornberg, B., Merrill, D.: Comparison of finite difference and pseudo-spectral methods for convective flow over a sphere. *Geoph. Res. Letters* 24, 3245–3248 (1997)
8. Hageman, N.S., Toga, A.W., Narr, K.L., Shattuck, D.W.: A diffusion tensor imaging tractography algorithm based on Navier-Stokes fluid mechanics. *IEEE Transactions on Medical Imaging* 28(3), 348–360 (2009)
9. MomayyezSiahkal, P., Siddiqi, K.: Probabilistic anatomical connectivity using completion fields. In: Jiang, T., Navab, N., Plum, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6361, pp. 566–573. Springer, Heidelberg (2010)
10. O’Donnell, L., Haker, S., Westin, C.-F.: New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic pDEs and geodesics in a tensor-warped space. In: Dohi, T., Kikinis, R. (eds.) *MICCAI 2002*. LNCS, vol. 2488, pp. 459–466. Springer, Heidelberg (2002)
11. Pichon, E., Westin, C.-F., Tannenbaum, A.R.: A hamilton-jacobi-bellman approach to high angular resolution diffusion tractography. In: Duncan, J.S., Gerig, G. (eds.) *MICCAI 2005*. LNCS, vol. 3749, pp. 180–187. Springer, Heidelberg (2005)
12. Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., Mangin, J.: New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models. *Mag. Reson. in Med.* 60, 1276–1283 (2008)
13. Tournier, J.D., Yeh, C., Calamante, F., Cho, H., Connelly, A., Lin, P.: Resolving crossing fibres using constrained spherical deconvolution: validation using diffusion-weighted imaging phantom data. *NeuroImage* 41, 617–625 (2008)
14. Williams, L.R., Jacobs, D.W.: Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation* 9, 837–858 (1997)

A Polynomial Approach for Maxima Extraction and Its Application to Tractography in HARDI*

Aurobrata Ghosh¹, Demian Wassermann², and Rachid Deriche¹

¹ Athéna Project Team, INRIA Sophia Antipolis-Méditerranée, France
{Aurobrata.Ghosh,Rachid.Deriche}@inria.fr

² Department of Radiology, Harvard Medical School & Brigham and Women's Hospital, Boston, USA

Abstract. A number of non-parametrically represented High Angular Resolution Diffusion Imaging (HARDI) spherical diffusion functions have been proposed to infer more and more accurately the heterogeneous and complex tissue microarchitecture of the cerebral white-matter. These spherical functions overcome the limitation of Diffusion Tensor Imaging (DTI) at discerning crossing, merging and fanning axonal fiber bundle configurations inside a voxel. Tractography graphically reconstructs the axonal connectivity of the cerebral white-matter *in vivo* and *non-invasively*, by integrating along the direction indicated by the local geometry of the spherical diffusion functions. Tractography is acutely sensitive to the local geometry and its correct estimation. In this paper we first propose a polynomial approach for analytically bracketing and numerically refining with high precision all the maxima, or fiber directions, of any spherical diffusion function represented non-parametrically. This permits an accurate inference of the fiber layout from the spherical diffusion function. Then we propose an extension of the deterministic Streamline tractography to HARDI diffusion functions that clearly discern fiber crossings. We also extend the Tensorline algorithm to these HARDI functions, to improve on the extended Streamline tractography. We illustrate our proposed methods using the Solid Angle diffusion Orientation Distribution Function (ODF-SA). We present results on multi-tensor synthetic data, and real *in vivo* data of the cerebral white-matter that show markedly improved tractography results.

1 Introduction

Diffusion MRI (dMRI) is a state-of-the-art method for studying the complex micro-architecture of the cerebral white matter *in vivo* and *non-invasively*. dMRI is sensitive to and measures the diffusion of water molecules. The complex geometry of the underlying tissue can be inferred by fundamentally assuming that the diffusion of water molecules is relatively less hindered parallel to coherent microstructures, such as axonal fiber bundles, than perpendicular to these structures. Hence the geometry or shape of the reconstructed diffusion function is an

* This work was partially supported by the ANR project NucleiPark and the France-Parkinson Association.

indicator of the microarchitecture of the tissue. The orientation information, which indicates axon fiber bundle directions, is of greater importance than the radial part, therefore, naturally the diffusion function is represented as a spherical function. However, it is well known that the current spatial resolution of dMRI at typically 2 mm^3 [9] is coarse compared to the true scale of an axon, which is of the order of $1 \text{ }\mu\text{m}$. Therefore, the spherical diffusion function (SDF) represents at best the average or dominant fiber direction of the underlying tissue locally, and is affected by partial voluming effects.

Diffusion Tensor Imaging (DTI) [4] is the most popular and well-utilized dMRI reconstruction protocol. DTI is known to work well in regions with a single fiber configuration, but to become ambiguous in regions where fiber bundles cross, merge, fan or kiss. Since about a third of dMRI voxels contain such complex fiber configurations [6], SDFs with richer geometries and multiple peaks capable of discerning crossings, have been of central importance in recent dMRI research. Diffusion acquisition protocols have also evolved to overcome this limitation of DTI with High Angular Resolution Diffusion Imaging (HARDI). A number of non-parametrically represented HARDI SDFs have been recently proposed that can indicate fiber crossings. e.g. the Orientation Distribution Function (ODF) [18], the Fiber Orientation Density (FOD) [17], the Persistent Angular Structure (PAS) [2], and the Diffusion Orientation Transform (DOT) [16].

Tractography graphically reconstructs the connectivity of the cerebral white-matter by integrating along the SDF's geometry locally. It is a modern tool that is unique in the sense that it permits an indirect dissection visualization of the brain in vivo and non-invasively [8]. The underpinnings of tractography are also based on the fundamental assumption of dMRI, i.e. that the diffusion of water molecules is hindered to a greater extent perpendicular to coherent fiber bundle structures than parallel to these. Therefore, following the geometry of the local diffusion function and integrating along reveals the continuous dominant structure of the fiber bundle. However, tractography is acutely sensitive to the local geometry, and its correct estimation is crucial.

Deterministic tractography has seen considerable success in researching neurological disorders [9]. Classically it was defined on DTI [5,21]. However, since DTI is ambiguous in regions with fiber crossings, the trend in recent years has been to extend tractography to complex SDFs that describe fiber directions more accurately [20,12]. Probabilistic tractography was proposed to address the reliability of deterministic tractography which remains sensitive to a number of parameters. Probabilistic tractography measures the likelihood of two regions being connected. Given the capabilities of dMRI schemes of date, due to partial voluming, noise, etc., probabilistic tractography provides a more complete statement. However, there exist state-of-the-art probabilistic schemes that rely on deterministic tracking to compute likelihood measures [14]. Therefore, deterministic tractography is an important problem.

In this paper we propose a deterministic tractography scheme based on a polynomial approach for accurately extracting the maxima of any non-parametrically represented SDF. Our paper has two main contributions. First, the polynomial

maxima extraction we use can analytically bracket all the maxima of any spherical (diffusion) function (SDF). This is neither a heuristic approach, like a finite difference mesh search, nor a local search approach, like optimization. It guarantees that all the maxima are located analytically, and then refined numerically to any degree of precision based on known numerical schemes. This ensures that given a SDF we can accurately quantify its local geometry for tracking, and that no maxima are overlooked. Computationally, we find in our experiments that the tracking time using our approach is comparable to the discrete mesh search proposed in [12]. Our maxima extraction can be considered the limiting case of the mesh search in [12] with a complete mathematical framework like in an optimization approach. Essentially it operates on the continuous SDF and locates *all* its extrema, i.e. $\nabla(\text{SDF}) = \mathbf{0}$, while not being dependent on initialization.

The second contribution of our paper is in our extension of the classical DTI Streamline tractography [5]. We adapt the Streamline tractography to the multiple maxima that can be discerned by complex SDFs in regions with fiber crossings, allowing us to trace through such regions with greater accuracy. We also extend the well known Tensorline tractography [21] to complex SDFs, to smooth out local kinks that can make the fiber tracks unnaturally “wiggly” in plain Streamline tractography due to acquisition noise and partial voluming which make the estimated SDF field spatially irregular. This is important, since kinks can violate the curvature threshold and stop the tracking algorithm.

We illustrate our method using Solid Angle diffusion ODFs (ODF-SA) [1]. The ODF-SA is a non-parametric SDF represented in the Spherical Harmonic (SH) basis and is a good generic SDF for applying our maxima extraction method. We first experiment on synthetic data generated from a multi-tensor model, then on in vivo human cerebral data [3] where we show marked improvements in detecting lateral radiations of the Corpus Callosum using our modified Tensorline tractography.

2 Materials and Methods

We first describe in detail our polynomial maxima extraction method. Then we describe the extensions of the Streamline and the Tensorline algorithms to complex SDFs.

Maxima Extraction: A Polynomial Approach. The SDF of DTI is the diffusion ellipsoid parametrized by the 3×3 symmetric tensor. Its geometry can be easily quantified from the eigen-decomposition of the tensor, e.g. the major eigenvector indicates the locally dominant fiber direction. However, quantifying the geometry of more complex SDFs with multiple peaks isn’t evident. We propose here a polynomial framework for extracting the maxima of such SDFs.

Our approach can be broken down into three steps. First, we rewrite a non-parametric SDF described in the SH basis, which is the most popular basis used in the dMRI, in an equivalent homogeneous polynomial basis constrained to the sphere. Second, we formulate a constrained polynomial optimization problem for identifying all the stationary points or extrema of the SDF. We then solve it

using a novel polynomial system solver instead of a local optimization approach. We can do so since the optimization problem can be shown to be a root finding problem for a system of homogeneous polynomials. This approach allows us to analytically bracket all the real roots of the polynomial system without depending on an initial solution. We refine the roots numerically to a high precision to accurately quantify all the extrema of the SDF. Finally in the third step, we categorize the extrema into maxima, minima and saddle-points of the SDF and thus extract all the maxima of the SDF.

The SH basis is an ideal formulation for describing spherical functions, e.g. SDFs with rich geometries, non parametrically, since they form a complex complete orthonormal basis for square integrable functions on the unit sphere. A modified, real, and symmetric SH basis is popularly used in dMRI to describe SDFs [11], since the diffusion function is real and assumed to be symmetric:

$$Y_j(\theta, \phi) = \begin{cases} \sqrt{2} \cdot \text{Re}(Y_l^{|m|}(\theta, \phi)) & \text{if } m < 0 \\ Y_l^m(\theta, \phi) & \text{if } m = 0 \\ \sqrt{2} \cdot \text{Im}(Y_l^m(\theta, \phi)) & \text{if } m > 0 \end{cases}, \quad (1)$$

where $j = (l^2 + l + 2)/2 + m$, $\theta \in [0, \pi]$, $\phi \in [0, 2\pi]$, and $Y_l^m(\theta, \phi)$ are the SHs.

The space of truncated SHs of rank- d is linearly bijective to the space of homogeneous polynomials (HP) of degree- d constrained to the sphere [10]. If a spherical function is described by a HP: $P(\mathbf{x} = [x_1, x_2, x_3]^T) = \sum_{i_1+i_2+i_3=d} A_{i_1 i_2 i_3} x_1^{i_1} x_2^{i_2} x_3^{i_3}$, where ($\|\mathbf{x}\|_2 = 1$) or $x_i = x_i(\theta, \phi)$, $i = 1..3$, then the SH coefficients of this spherical function can be computed from the spherical harmonic transform (SHT):

$$c_j = \sum_{i_1+i_2+i_3=d} A_{i_1 i_2 i_3} \int_{\Omega} x_1^{i_1}(\theta, \phi) x_2^{i_2}(\theta, \phi) x_3^{i_3}(\theta, \phi) \cdot Y_j(\theta, \phi) d\Omega. \quad (2)$$

Since the coefficients $A_{i_1 i_2 i_3}$ of the HP are outside the integration in the SHT Eq.2 can be rewritten as a linear transformation $\mathbf{C} = \mathbf{M}\mathbf{A}$ by reordering the indices i_1, i_2, i_3 suitably, where \mathbf{C} are the coefficients in the SH basis, \mathbf{A} the coefficients in the HP basis, and \mathbf{M} is the transformation matrix. When \mathbf{C} and \mathbf{A} have the same dimension, \mathbf{M} becomes an invertible square matrix, and \mathbf{A} can be computed from \mathbf{C} . In other words a SDF described in a truncated SH basis of rank- d can be equivalently rewritten in a constrained HP basis of degree- d .

Finding the maxima of the SDF can be formulated as a constrained optimization (maximization) problem: $\max_{\mathbf{x}} P(\mathbf{x})$ subject to $\|\mathbf{x}\|_2^2 - 1 = 0$. Using Lagrange Multipliers, it can be rewritten as an unconstrained functional: $F(\mathbf{x}, \lambda) = P(\mathbf{x}) - \lambda(\|\mathbf{x}\|_2^2 - 1)$. From optimization theory, the maxima \mathbf{x}^* (and its corresponding λ^*) of the SDF would have to satisfy $\nabla F(\mathbf{x}^*, \lambda^*) = 0$:

$$\frac{\partial F(\mathbf{x}^*, \lambda^*)}{\partial x_1} = \frac{\partial F(\mathbf{x}^*, \lambda^*)}{\partial x_2} = \frac{\partial F(\mathbf{x}^*, \lambda^*)}{\partial x_3} = \|\mathbf{x}^*\|_2^2 - 1 = 0. \quad (3)$$

Eq.3 is a system of HPs, $\{Q_k(\mathbf{X})\}$, and implies that $\mathbf{X}^* = (\mathbf{x}^*, \lambda^*)$ would be a root of this system. Since the SDF is a real function, only the real roots of this

system are of interest. However, $\nabla F(\tilde{\mathbf{x}}, \tilde{\lambda}) = 0$ identifies all the stationary points or extrema of the SDF. Therefore, once the real roots of the SDF are identified, they would have to be categorized into the maxima, minima, and saddle-points.

To find the real roots of Eq 3 we use the subdivision solver for polynomial systems proposed in [15]. The solver works as follows. The polynomials are converted from the monomial basis to the Bernstein basis, since the latter has intuitive geometric interpretations:

$$Q(\mathbf{X}) = \sum_{i_1=0}^{d_1} \sum_{i_2=0}^{d_2} \sum_{i_3=0}^{d_3} \sum_{i_4=0}^{d_4} B_{i_1, i_2, i_3, i_4} B_{d_1}^{i_1}(x_1; a_1, b_1) B_{d_2}^{i_2}(x_2; a_2, b_2) \cdot B_{d_3}^{i_3}(x_3; a_3, b_3) B_{d_4}^{i_4}(x_4 = \lambda; a_4, b_4), \quad (4)$$

where $B_d^i(x; a, b) = \binom{d}{i} \frac{1}{(b-a)^d} (x-a)^i (b-x)^{d-i}$ form the Bernstein basis on $[a, b]$ along any dimension, and the domain of $Q(\mathbf{X})$ is $[a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \times [a_4, b_4]$. Since the SDF is defined on the unit sphere, the initial domain of all the $Q(\mathbf{X})$'s can be $[-1, 1] \times [-1, 1] \times [-1, 1] \times [-K, K]$, for very large K . The coefficients are converted to the Bernstein basis using exact arithmetic to avoid loss of precision. This operation is required once.

Along every dimension, De Casteljau's algorithm efficiently subdivides the Bernstein representation of any $Q(\mathbf{X})$ into two sub-representations in two sub-domains of the initial domain. This allows to subdivide the entire domain into sub-domains efficiently by subdividing along all four dimensions to locate the roots. Moreover, along every dimension Descartes' theorem states that *the number of real roots of $q(x) = \sum b_i B_d^i(x; a, b)$ in $]a, b[$ is bounded by the number of sign changes of $\{b_i\}$ and is equal modulo 2*, where $q(x)$ is the projection of $Q(\mathbf{x})$ along that dimension. This implies an *exclusion* test that is negative when there are no real roots in the interval and is positive when there are one or more real roots in the interval. Therefore, making it possible to analytically identify and subdivide intervals along every dimension to bracket the real roots. Once a root has been bracketed or isolated any standard numerical one dimensional root-finder of choice can be used to refine the real root with high precision.

From the Bernstein coefficients in Eq 4 it is easy to sandwich the projection $q(x)$ of $Q(\mathbf{X})$ along any dimension j by $m_j(q; x_j) = \sum_{i_j}^{d_j} \min_{(0 \leq i_k \leq d_k, k \neq j)} B_{i_1, i_2, i_3, i_4} B_{d_j}^{i_j}(x_j; a_j, b_j)$ and by $M_j(q; x_j) = \sum_{i_j}^{d_j} \max_{(0 \leq i_k \leq d_k, k \neq j)} B_{i_1, i_2, i_3, i_4} B_{d_j}^{i_j}(x_j; a_j, b_j)$, such that the roots of $q(x_j)$ (along dimension j) are sandwiched by the roots of m_j and M_j [15]. Thus if m_j and M_j don't have any roots in an interval $[a_j, b_j]$, as indicated by Descartes' theorem applied to m_j and M_j , then $q(x_j)$ has no root in that interval, as implied by the exclusion test. Furthermore, if any of the projections $q_k(x_j)$ of the system of polynomials $Q_k(\mathbf{X})$ has no roots in the interval $[a_j, b_j]$, then the system $Q_k(\mathbf{X})$ has no real roots in the sub-domain overlapping this interval, and the sub-domain can be discarded.

The powerful exclusion test allows to analytically subdivide and reject sub-domains of the initial domain in such a fashion that the rejected sub-domains are guaranteed to not contain any real roots of Eq 3. Once the intervals that were not rejected along every dimension are small enough, they are numerically refined to

locate the roots along those dimensions in those intervals. However, theoretically, these roots (along given dimensions) are only roots of the projections $q_k(x_j)$ and may not be a root $\tilde{\mathbf{X}}$ of $Q_k(\mathbf{X})$. An additionally test would be required to reject solutions to the above subdivision process that may not be real roots of the polynomial system. In practice, however, we have never come across such solutions, but simply testing for $\nabla F(\tilde{\mathbf{X}}) = 0$ would provide such a test.

Once all the extrema of the SDF have been quantified precisely, these would have to be categorized into maxima, minima and saddle-points to identify the maxima of the SDF. We use the Bordered Hessian test [13] to do this. In unconstrained multi-dimensional optimization the extremum, $\tilde{\mathbf{x}}$, can be categorized by evaluating the Hessian, $H_{\mathbf{x}}F(\tilde{\mathbf{x}})$. If $H_{\mathbf{x}}F(\tilde{\mathbf{x}})$ is positive definite, then $\tilde{\mathbf{x}}$ is a local minimum, if $H_{\mathbf{x}}F(\tilde{\mathbf{x}})$ is negative definite, then $\tilde{\mathbf{x}}$ is a local maximum, and if $H_{\mathbf{x}}F(\tilde{\mathbf{x}})$ has eigenvalues with mixed signs, then $\tilde{\mathbf{x}}$ is a saddle-point.

In constrained optimization, which is our case, the Hessian test is extended to the Bordered Hessian test to account for the constraints – the Hessian has to be positive/negative definite at $\tilde{\mathbf{x}}$, while the gradient of the constraints has to be zero (at $\tilde{\mathbf{x}}$), for $\tilde{\mathbf{x}}$ to be a minimum/maximum [13]. Given an n dimensional functional $P(\mathbf{x})$ to be maximized, subject to a set of m constraints $\mathbf{g}(\mathbf{x})$, the Bordered Hessian of the corresponding Lagrangian functional $F(\mathbf{x}, \boldsymbol{\lambda}) = P(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x})$, is defined as:

$$\overline{H}F(\mathbf{x}, \boldsymbol{\lambda}) = \begin{bmatrix} 0_{m \times m} & \nabla \mathbf{g}(\mathbf{x})_{m \times n} \\ \nabla \mathbf{g}(\mathbf{x})_{n \times m}^T & H_{\mathbf{x}}F(\mathbf{x}, \boldsymbol{\lambda})_{n \times n} \end{bmatrix}_{(m+n) \times (m+n)}, \tag{5}$$

where the Hessian of the Lagrangian functional is bordered by the Jacobian of the constraints $\nabla \mathbf{g}(\mathbf{x})$, and padded by a corner-block of zeros. In our case $n = 3$ and $m = 1$, therefore, the Bordered Hessian is a 4×4 matrix. The Bordered Hessian is rank deficient and cannot satisfy the definiteness conditions of the Hessian test. However, an extremum, $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}, \tilde{\boldsymbol{\lambda}})$, of the constrained optimization can be categorized using the following alternating sign tests:

$$\begin{aligned} (-1)^m |\overline{H}_r F(\tilde{\mathbf{X}})| &> 0 \text{ strict minimum,} \\ (-1)^r |\overline{H}_r F(\tilde{\mathbf{X}})| &> 0 \text{ strict maximum,} \\ r &= m + 1, \dots, n, \end{aligned} \tag{6}$$

$$\text{where : } \overline{H}_r F = \begin{bmatrix} 0_{m \times m} & \nabla \mathbf{g}_{m \times r} \\ \nabla \mathbf{g}_{r \times m}^T & H_{\mathbf{x}}F_{r \times r} \end{bmatrix}_{(m+r) \times (m+r)}, \tag{7}$$

where $\nabla \mathbf{g}_{m \times r}$ denotes the first r columns of the Jacobian $\nabla \mathbf{g}(\mathbf{x})$ and $H_{\mathbf{x}}F_{r \times r}$ denotes the principal $r \times r$ sub-matrix of the Hessian $H_{\mathbf{x}}F_{n \times n}$. There are other ways of categorizing the extremum $\tilde{\mathbf{X}}$ into a maximum, or a minimum, e.g. using differential geometry as proposed in [7].

Tractography. The continuous version of Streamline tractography [5] defined for DTI, considers a fiber tract as a 3D space curve parametrized by its arc-length, $\mathbf{r}(s)$, and describes it by its Frenet equation:

$$\frac{d\mathbf{r}(s)}{ds} = \mathbf{t}(s) = \epsilon_1(\mathbf{r}(s)), \quad \mathbf{r}(0) = \mathbf{r}_0 \quad (8)$$

where $\mathbf{t}(s)$ the tangent vector to $\mathbf{r}(s)$ at s is equal to the unit major eigenvector $\epsilon_1(\mathbf{r}(s))$ of the diffusion tensor at $\mathbf{r}(s)$. This implies that fiber tracts are locally tangent to the major eigenvector of the diffusion tensor.

Integrating Eq. 8 requires two things – first, a spatially continuous tensor (or SDF) field, and second, a numerical integration scheme. [5] proposed two approaches for estimating a spatially continuous tensor field from a discrete DTI tensor field, namely approximation and interpolation. They also proposed the Euler’s method, the 2nd order Runge-Kutta method, and the adaptive 4th order Runge Kutta method as numerical integration schemes. Finally, for stopping they proposed four criteria – the tracts are within the image volume, the tracts are in regions with FA value higher than a threshold, the curvature of a tract is smaller than a threshold, and that a tract is better aligned with the major eigenvector in the next spatial location than any of the two other eigenvectors.

We adapt this tractography algorithm to SDFs with multiple maxima by modifying Eq. 8 to:

$$\frac{d\mathbf{r}(s)}{ds} = \eta_{\theta_{min}}(\mathbf{r}(s)), \quad \mathbf{r}(0) = \eta_{max}(0) \quad (9)$$

where $\eta_i(\mathbf{r}(s))$ are all the unit maxima vectors of the SDF extracted by our method at $\mathbf{r}(s)$, η_{max} is the unit maximum vector whose function value is the largest amongst all the η_i , and $\eta_{\theta_{min}}$ is the unit maximum vector in the current step that is most collinear to the unit maximum vector followed by the integration in the previous step. Eq. 9 and the initial condition state that at the starting point we begin integrating along the dominant maximum direction, and at each consecutive step we first extract all the maxima of the SDFs and choose the maximum direction most collinear to the maximum direction from the previous integration step, to move forward in the integration.

Since we require a continuous field of SDFs for integrating Eq. 9, we consider the Euclidean interpolation of the SDFs, which is equivalent to the L2-norm interpolation of the SDFs, since the SHs form an orthonormal basis [12]. For the numerical integration scheme we employ the 2nd order Runge-Kutta method due to its robustness and simplicity with an integration step of length 0.5mm.

For stopping we only use two criteria – the tracts are within the image volume, and tracts aren’t allowed to have high curvatures, or the radius of curvature of tracts should be smaller than 0.87mm. Currently we don’t employ any stopping criteria based on anisotropy indices because indices such as GFA for complex SDFs show similar contrasts to FA, and therefore have low values in regions with fiber crossings.

Since Streamline tractography traces fibers that are always tangent to the maxima of the local SDF, these fibers can be unnaturally “wiggly”. Due to acquisition noise the estimated discrete SDF field is generally spatially irregular. Thus closely following the local maximum can cause the fiber to wriggle. This effect can be important enough to violate the curvature criterion and cause the

tracking algorithm to abort. Furthermore, partial voluming effects can also cause certain SDFs to lack maxima along the fiber direction, especially in voxels with fiber crossings, even when neighbouring SDFs may have maxima well aligned with the fiber direction. This can cause the tracking algorithm to suddenly deviate to another fiber track, violating again the curvature criterion.

DTI Tensorline tractography was proposed to specifically address these issues [21]. ϵ_1 in Eq. 8 was replaced by:

$$\mathbf{v}_{\text{out}} = f\epsilon_1 + (1 - f)((1 - g)\mathbf{v}_{\text{in}} + g\mathbf{D} \cdot \mathbf{v}_{\text{in}}), \quad (10)$$

where f, g are user defined weights, $\mathbf{v}_{\text{in}}, \mathbf{v}_{\text{out}}$ are the incoming and outgoing tangents respectively, \mathbf{D} is the local diffusion tensor with ϵ_1 its unit major eigenvector, and $\mathbf{D} \cdot \mathbf{v}_{\text{in}}$ is the tensor deflection (TEND) term. This smooths away unnatural kinks, and also helps to plough through regions with uncertainty, where \mathbf{D} is oblate or spherical.

With general SDFs, \mathbf{D} doesn't exist, but the SDFs have more complex geometries that are meant to better resolve the angular uncertainties in regions with crossings, implying that the TEND term can be ignored. Therefore, we adapt the Tensorline tractography to general SDFs with multiple maxima by replacing $\eta_{\theta_{\text{min}}}$ in Eq. 9 by:

$$\mathbf{v}_{\text{out}} = f\eta_{\theta_{\text{min}}} + (1 - f)\mathbf{v}_{\text{in}}, \quad (11)$$

with f a user defined weight, and with $\mathbf{v}_{\text{in}}, \mathbf{v}_{\text{out}}$ as defined above. \mathbf{v}_{in} acts like an inertial factor to maintain the fiber's general trend, and Eq. 11 smooths away kinks and helps to navigate regions with SDFs affected by partial voluming.

Thus we have proposed a polynomial based maxima extraction method to correctly extract all the maxima of any SDF described non-parametrically in the SH basis. And we have extended the deterministic Streamline and Tensorline tractography methods to deal with the multiple maxima discerned by complex SDFs in regions with fiber crossings, to be able to track fibers more accurately through such regions.

3 Experiments

We illustrate our approach on ODF-SAs which are the orientation marginal distribution of the diffusion probability density function $\Psi(\theta, \phi) = \int_0^\infty P(r, \theta, \phi)r^2 dr$ [1]. They are SDFs with multiple peaks aligned along directions with high probability of diffusion of water molecules, i.e. fiber directions, and can discern crossing fiber configurations. ODF-SAs are a good example of non-parametrically represented SDFs described in the real and symmetric SH basis. From the Diffusion Weighted Images (DWIs), we estimated rank-4 ODF-SAs and then tracked fibers from specific seed regions using our proposed maxima extraction method and extensions of deterministic Streamline and Tensorline tractographies. We experimented on a synthetic dataset, and on an in vivo human cerebral dataset.

Synthetic Dataset: To conduct controlled experiments with known ground truths, we used a multi-tensor approach parametrized by a diagonal tensor $\mathbf{D} = [1700,$

$300, 300] \times 10^{-6} \text{mm}^2/\text{s}$ to generate synthetic DWIs [12]. The DWIs were generated along 81 directions for a b -value of $3000 \text{s}/\text{mm}^2$. The synthetic data was used to first validate our maxima extraction method with voxels with 2, and 3 fibers crossing. It was then used to generate a synthetic image of overlapping fiber bundles to validate fiber tracking using DTI tracking and ODF-SA tracking.

In vivo Human Cerebral Dataset: The in vivo human cerebral data was acquired with 60 encoding gradient directions, a b -value of $1000 \text{s}/\text{mm}^2$, twice-refocused spin-echo EPI sequence, with $\text{TE} = 100 \text{ms}$, and $\text{TR} = 12 \text{ms}$. It has isotropic voxels of $1.7 \text{mm} \times 1.7 \text{mm} \times 1.7 \text{mm}$, with three repetitions, and was corrected for subject motion. It was acquired on a whole-body 3T Siemens Trio scanner [3]. This dataset was used to compare DTI tractography and ODF-SA tractography. We tracked specifically from seeds with the Corpus Callosum (CC). It is well known that due to the important crossings between the lateral radiations of the CC and Pyramidal Tract (PT) it is difficult to track the lateral radiations of the CC. We used the ODF-SA and our modified tracking algorithms to specifically recover these lateral radiations of the CC [19]. We also tracked the PT and the Superior Longitudinal Fasciculus (SLF) which crosses the lateral radiations of the CC and the PT to validate our tracking method in regions where all 3 fiber bundles cross.

4 Results

Figure 1 shows the results on the synthetic multitensor data test. In Fig. 1a and Fig. 1b we present the results of our maxima extraction. These voxels represent 2 and 3 fiber-crossings. All the extrema were extracted and categorized: maxima in thick yellow, minima in green, and saddle-points in thin blue. In the case of the 2 fiber configuration, even the unimportant maxima along the z -axis was correctly detected.

Figure 1c,d,e, present a synthetic dataset image with 2 fiber bundles crossing. Fig. 1c shows the estimated ODF-SAs with the seed region highlighted in grey in the background. Fig. 1d is the result of DTI Tensorline tracking with $f = 0.3$ and $g = 0$ (Eq. 10). These weights imply that the TEND term was ignored and only the inertia term \mathbf{v}_{in} played a role. These weights were chosen to make the DTI Tensorline tracking comparable to the ODF-SA Tensorline tracking. The ODF-SA Tensorline tracking with $f = 0.3$ (Eq. 11) is visible in Fig. 1e. A closer inspection at the junction areas clearly indicate that DTI tracking has no crossings, and the fibers bend to accommodate for partial voluming, whereas in ODF-SA tracking the fibers cross each other as expected.

We also compared the computation time of our maxima extraction method to that of the discrete mesh search method proposed in [12], where a 16th order icosahedral tessellation of the sphere with 1281 samples of the SDF on a hemisphere was used. Comparing neighbouring directions in this tessellation revealed an angular resolution limit of $\sim 4^\circ$. Our maxima extraction can be considered as the limiting case of the discrete mesh search approach. Choosing the ODF-SA Tensorline algorithm and only changing the maxima extraction method, we

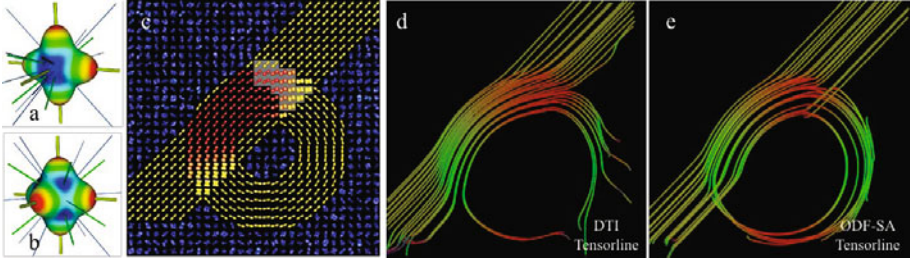


Fig. 1. (a,b): Extrema extraction & Categorization on synthetic data. Glyph-colour: red: high anisotropy, blue: low anisotropy Lines: thick-yellow: Maxima, thin-blue: Saddle points, green: Minima. a) 2-fibers b) 3-fibers. (c,d,e) Fiber tracking on synthetic dataset image with 2 fiber bundles crossing. c) ODF-SA glyphs with seeds highlighted as grey in the background. d) DTI Tensorline tracking. e) ODF Tensorline tracking. DTI tracking never shows crossings, whereas ODF-SA tracking clearly shows crossings.

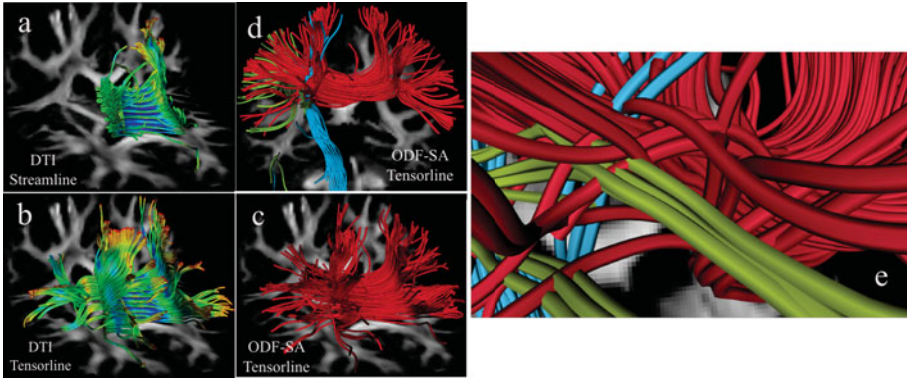


Fig. 2. Tractography from the Corpus Callosum (CC). a) DTI Streamline cannot recover lateral radiations of CC. b) DTI Tensorline improves on (a), some lateral radiations are recovered. c) ODF-SA Tensorline recovers a rich distribution of the CC. d) ODF-SA Tensorline tracking from seeds in the CC (red: left-right), Pyramidal Tract (PT: superio-inferior: blue) and Superior Longitudinal Fasciculus (SLF: antero-posterior: green). e) Zoom: Tracking visible through regions where all 3 fiber bundles cross.

found the computation times to be highly comparable. Our method clocked 17mins 34secs while the discrete mesh search registered 15mins 45secs.

Results of the in vivo human cerebral dataset are presented in Figure-2. Neither the estimated DTI field nor the ODF-SA field were regularized prior to tracking. Tracking was seeded in regions of the CC. It is well known that fibers radiating laterally from the CC cross the PT [19], and are hard to track. This is clearly seen in Fig-2a which presents the DTI Streamline tractography result. This is improved by using DTI Tensorline tractography (parameters as above), as seen in Fig-2b. However, only a select few lateral radiations are found. A

much richer distribution of fibers from the CC are tracked using the ODF-SA Tensorline tractography (parameters as above), as seen in Fig-2c. Fig-2d highlights ODF-SA Tensorline tractography by presenting fibers from three bundles that are known to cross. In red is the CC (left-right), in blue is the PT (superio-inferior), and in green is the SLF (anterio-posterior). The close up in Fig-2e clearly shows fibers traced through a region where all three bundles cross.

5 Conclusion

In this paper, we first proposed a polynomial based approach for correctly extracting the maxima of any non-parametrically represented SDF in the SH basis. Recently a number of such HARDI SDFs have been proposed to infer accurately the micro-architecture of the underlying cerebral tissue. We illustrated our method on the ODF-SA. Our approach guarantees to analytically bracket all the maxima, and to numerically refine them to any precision, to accurately quantify the SDF's geometry. It can be seen as the limiting case of a discrete mesh search, with the complete mathematical framework of optimization. However, it differs from optimization since it's not a local search and isn't dependent on an initial solution. Further, all the maxima are found together.

Maxima extraction is sensitive to both the estimation of the SDFs, which can be affected by signal noise, and the localization of the SDFs' maxima after estimation, which can be affected by the maxima extraction process. Our method is independent of the SDFs' estimation process – it computes on the SDFs' SH coefficients after these have been estimated from the signal and provides an error free localization of the correct maxima. However, when the SDFs estimation isn't robust, the maxima can be affected by signal noise.

As an application of our maxima extraction method, we also proposed an extension of the DTI Streamline and Tensorline algorithms. We tested on synthetic and in vivo human cerebral datasets. Using the ODF-SA Tensorline algorithms, we showed that we were able to dramatically improve deterministic tractography results. We were able to recover a rich distribution of the lateral radiations of the CC [19], which are hard to track since they cross the PT. Further, by tracking from seeds in the CC, PT, and SLF, we were able to correctly trace through regions where all three fiber bundles cross.

References

1. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the orientation distribution function in single and multiple shell Q-ball imaging within constant solid angle. *Magnetic Resonance in Medicine* 64(2), 554–566 (2010)
2. Alexander, D.C.: Maximum Entropy Spherical Deconvolution for Diffusion MRI. In: Christensen, G.E., Sonka, M. (eds.) *IPMI 2005*. LNCS, vol. 3565, pp. 76–87. Springer, Heidelberg (2005)
3. Anwander, A., Tittgemeyer, M., von Cramon, D.Y., Friederici, A.D., Knosche, T.R.: Connectivity-based parcellation of broca's area. *Cerebral Cortex* 17(4), 816–825 (2007)

4. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. *Biophysical Journal* 66(1), 259–267 (1994)
5. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine* 44(4), 625–632 (2000), <http://view.ncbi.nlm.nih.gov/pubmed/11025519>
6. Behrens, T.E.J., Johansen-Berg, H., Jbabdi, S., Rushworth, M.F.S., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations. what can we gain? *NeuroImage* 34(1), 144–155 (2007)
7. Bloy, L., Verma, R.: On computing the underlying fiber directions from the diffusion orientation distribution function. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 1–8. Springer, Heidelberg (2008)
8. Catani, M., Howard, R.J., Pajevic, S., Jones, D.K.: Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage* 17, 77–94 (2002)
9. Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., Thompson, A.: Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. *Lancet. Neurol.* 7(8), 715–727 (2008)
10. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications. *Magnetic Resonance in Medicine* 56, 395–410 (2006)
11. Descoteaux, M., Angelino, E., Fitzgibbons, S.: R.Deriche: A Fast and Robust ODF Estimation Algorithm in Q-Ball Imaging. In: Third IEEE ISBI: from Nano to Macro, Arlington, Virginia, USA, pp. 81–84 (April 2006)
12. Descoteaux, M., Deriche, R., Knosche, T.R., Anwander, A.: Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *IEEE Transactions on Medical Imaging* 28(2), 269–286 (2009)
13. Im, E.: Hessian sufficiency for bordered Hessian. *Res. Lett. Inf. Math. Sci.* 8, 189–196 (2005)
14. Jeurissen, B., Leemans, A., Jones, D.K., Tournier, J.D., Sijbers, J.: Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. *Human Brain Mapping* (2010)
15. Mourrain, B., Pavone, J.-P.: Subdivision methods for solving polynomial equations. *Journal of Symbolic Computation* 44, 292–306 (2009)
16. Ozarslan, E., Shepherd, T., Vemuri, B., Blackband, S., Mareci, T.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (dot). *NeuroImage* 31(3), 1086–1103 (2006)
17. Tournier, J.D., Calamante, F., Gadian, D., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23, 1176–1185 (2004)
18. Tuch, D.: Q-ball imaging. *Magnetic Resonance in Medicine* 52(6), 1358–1372 (2004)
19. Türe, U., Yasargil, M., Friedman, A., Al-Mefty, O.: Fiber dissection technique: lateral aspect of the brain. *Neurosurgery* 47(2), 417 (2000)
20. 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. *Magn. Reson. Med.* 54(6), 1377–1386 (2005)
21. Weinstein, D.M., Kindlmann, G.L., Lundberg, E.C.: Tensorlines: Advection-diffusion based propagation through diffusion tensor fields. In: Visualization Conference. IEEE, Los Alamitos (1999)

A Generative Approach for Image-Based Modeling of Tumor Growth

Bjoern H. Menze^{1,2}, Koen Van Leemput^{1,3,4},
Antti Honkela⁵, Ender Konukoglu⁶, Marc-André Weber⁷,
Nicholas Ayache², and Polina Golland¹

¹ Computer Science and Artificial Intelligence Laboratory,
Massachusetts Institute of Technology, USA

² Asclepios Research Project, INRIA Sophia-Antipolis, France

³ Dept. of Radiology, Massachusetts General Hospital, Harvard Medical School, USA

⁴ Department of Information and Computer Science, Aalto University, Finland

⁵ Helsinki Institute for Information Technology HIIT, University of Helsinki, Finland

⁶ Machine Learning and Perception Group, Microsoft Research, Cambridge, UK

⁷ Department of Diagnostic Radiology, Heidelberg University Hospital, Germany

Abstract. Extensive imaging is routinely used in brain tumor patients to monitor the state of the disease and to evaluate therapeutic options. A large number of multi-modal and multi-temporal image volumes is acquired in standard clinical cases, requiring new approaches for comprehensive integration of information from different image sources and different time points. In this work we propose a joint generative model of tumor growth and of image observation that naturally handles multi-modal and longitudinal data. We use the model for analyzing imaging data in patients with glioma. The tumor growth model is based on a reaction-diffusion framework. Model personalization relies only on a forward model for the growth process and on image likelihood. We take advantage of an adaptive sparse grid approximation for efficient inference via Markov Chain Monte Carlo sampling. The approach can be used for integrating information from different multi-modal imaging protocols and can easily be adapted to other tumor growth models.

1 Introduction

Processes related to tumor growth can be modeled at different scales ranging from signaling at sub-cellular level, via multi-cellular processes determining metabolic properties of the tumor, to the gross bio-mechanical behavior of tumor-affected tissue at a macroscopic scale. Tumor models in medical image analysis rely almost exclusively on information from morphologic images and consequently focus on the macroscopic phenomena of tumor evolution. An important class of macroscopic tumor models is based on the reaction-diffusion equations [5,24,6,11,4,23,13,16,10]. In this paper, we propose an efficient Bayesian framework for image-based inference in this type of tumor growth models.

Tumor models are used to describe effects a lesion has on surrounding healthy tissue. Modeling this effect helps to improve inter-subject image and atlas reg-

istration [10,17,25,26,2] and to improve tissue segmentation in the presence of the lesion [12,8]. As the tumor location it is not consistent across patients, tumor shape is often employed for constructing spatial priors. For example, a large number of simulated tumors can be used to learn characteristic tumor-induced deformations [21,2]. For data of individual patients growth models also help to estimate deformation fields and tissue perturbations [10,26]. Most of these methods take a static approach to modeling tumors for *removing* the tumor-induced effects from the analysis of the subject's brain itself.

Other biophysical tumor models describe explicitly the evolution of the lesion. (For a short review see [20].) For glioma most such models follow a reaction-diffusion framework [5]. They assume an infiltrative growth of the tumor cells, may consider differences in cell diffusion in white and gray matter [24], and model locally anisotropic migration patterns by integrating information from diffusion tensor images (DTI) [6]. Some models also include the mechanical effect of the lesion on surrounding structures by modeling the interaction between the tumor mass and the invaded tissue [6,14], or couple the growth dynamics with phenomenological image models for Gadolinium uptake and changes in DTI [23]. The primary application for these image-based dynamical tumor models is in simulating tumor growth, either to analyze macroscopic patterns of disease progression [1], or to generate realistically appearing tumor images to be used for validation of tumor segmentation methods [23].

Many dynamical tumor models could serve as a framework for integrating complex information from different image modalities and longitudinal data sets. But unlike the static whole brain models, the image-based personalization remains challenging for most forward simulators. As a consequence, some studies only provide qualitative measures for analyzing lesions and their progression [7], employ strict assumptions on the relationship between the tumor cell density and image observation [24], or focus on a theoretical treatment of the inverse problem under the assumption of having appropriate observables available at a later stage [14]. The PDE-constrained optimization approach in [14] relates the tumor growth model to landmark-based registration using a reaction-diffusion-advection model. This approach is similar to the whole brain modeling [10,26,21], but depends critically on the tumor-tissue interaction model. Alternatively, a traveling wave approximation of the reaction-diffusion model can be used to estimate model parameters from the temporal evolution of the tumor front [16]. Unfortunately, this approach only provides the speed of growth. Furthermore, all of the previous methods focus on point-estimates of the process. This presents significant challenges in presence of noisy observations and uncertainty in evolution models.

In this paper, we develop a novel approach for personalizing tumor models, for integrating multi-modal data, and for quantifying uncertainty. We formulate a generative model that includes a biophysical tumor growth model and statistical observation models for different image modalities. We devise an efficient inference scheme based on the forward model and applicable in high-dimensional observation settings. We demonstrate the resulting method in experiments with

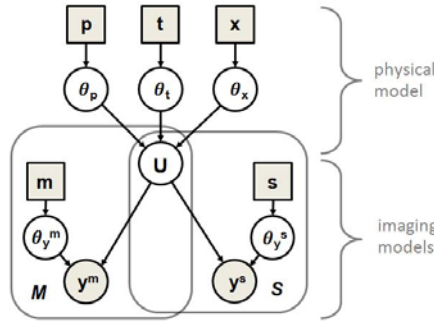


Fig. 1. Joint process and observation model. The physical model defines the state variable \mathbf{u} as a function of model parameters θ_p and initial conditions θ_x and θ_t . The imaging models relate \mathbf{u} to the image observations \mathbf{y}^m and \mathbf{y}^s , representing functional maps and tumor segmentations, respectively. Variables are represented by circles, parameters are shown as squares. We estimate the model parameters θ_p to characterize the state of disease visible from images \mathbf{y} .

both synthetic and clinical patient data. We envision that a joint tumor and image modeling will close the gap between functional image interpretation and disease modeling, and will provide new directions for therapy optimization.

2 Tumor Growth Model

Our approach includes a physical process model that describes the progression of the disease through the evolution of the tumor cell concentration \mathbf{u} and a probabilistic imaging model that captures the relationship between the latent state \mathbf{u} and the image observations \mathbf{y} . Figure 1 presents the full graphical model described in this section.

2.1 Physical Process Model

We let $\mathbf{u} = (u_1, \dots, u_I)^T$ be the latent state variables where $u_i \in [0, 1]$ is the tumor cell concentration in voxel i ($1 \leq i \leq I$). We model the temporal evolution of the tumor as an inhomogeneous anisotropic diffusion governed by the Fisher-Kolmogorov equation [5]:

$$\frac{\partial \mathbf{u}}{\partial t} = \nabla_x (\mathbf{D} \nabla_x \mathbf{u}) + \rho \cdot \mathbf{u} (1 - \mathbf{u}), \tag{1}$$

where ∇_x represents the spatial derivative operator. The equation describes the growth of a glioma as an infiltrative process with diffusivity tensor \mathbf{D} and self-limiting growth with proliferation rate ρ . We further assume that $\mathbf{D} = D \cdot \hat{\mathbf{D}}$, where $\hat{\mathbf{D}}$ is a patient-specific diffusion tensor, observed via DTI, and D is a global diffusivity parameter [6]. We apply Neumann border conditions, and assume that tumor cells migrate in the white and gray matter only, with higher

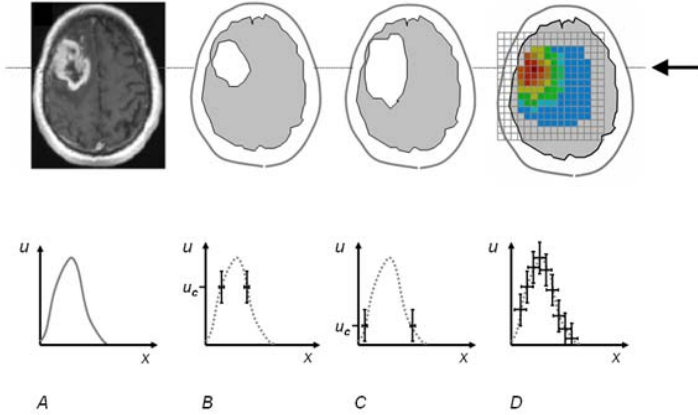


Fig. 2. Example relationships between image information \mathbf{y} (top) and tumor cell density \mathbf{u} (bottom). *A*: actual T_1 gad image and the schematic tumor cell distributions along the line indicated by the arrow in the top row. *B*, *C*: exemplary modalities where only the visible tumor segmentation \mathbf{y}^s can be correlated with specific values u_c of the state variable \mathbf{u} . *D*: exemplary modality where all entries of the functional map \mathbf{y}^m can be correlated with the latent state variable \mathbf{u} .

diffusivity in the white matter ($D_w \gg D_g$). To personalize the model, we construct estimates of the model parameters $\theta_p = \{D, \rho\}$ individually for every patient.

By integrating Eq. (1) over time – starting from the initial state $\mathbf{u}(t = 0)$ when the first image of the tumor was acquired – we obtain a 4D functional model $U(x, y, z, t)$ that describes the evolution $\mathbf{u}(t)$ at any time point t . We choose a parametric representation for the initial state by placing a seed u^{init} in a voxel located at $\theta_x = \{x, y, z\}$ and by growing the tumor for the time interval θ_t . We form the deterministic tumor evolution model:

$$p(\mathbf{u}(t)|\theta_x, \theta_t, \theta_p) = \delta(U(x, y, z, t; \theta_x, \theta_t, \theta_p) - \mathbf{u}(x, y, z, t)), \tag{2}$$

where δ is Dirac’s delta indicating that \mathbf{u} is an exact solution of the functional $U(x, y, z, t)$. We choose to use a deterministic model without a stochastic component in Eq. (1) or noise term in Eq. (2) as it represents a stronger constraint on the growth process. We incorporate non-deterministic behavior into our framework via a probabilistic model of the imaging process, as described below.

2.2 Imaging Models

Observations $\mathbf{y} = (y_1, \dots, y_I)^T$ are acquired at each voxel i at specific times t , representing partial observations of the process. We evaluate every image volume $\mathbf{y}(t)$ with the corresponding state vector $\mathbf{u}(t)$. For notational simplicity we drop the time index in the imaging models below. As illustrated in Fig. 2, we assume

that there are two different types of observations that can be related to \mathbf{u} : continuous *functional maps* and binary *tumor segmentations* which are the most common observation.

Functional maps. Functional maps \mathbf{y}^m contain continuous values $y_i^m \in \mathbb{R}$. They represent imaging modalities with functional parameters that can be directly correlated with the latent physiological variable \mathbf{u} using a (possibly nonlinear) forward observational operator F . Formally,

$$p(\mathbf{y}^m | \mathbf{u}, \theta_y^m) = \prod_i p(y_i^m | \mathbf{u}, \theta_y^m) = \prod_i \mathcal{N}(y_i^m; \mu_i, \sigma_m^2), \tag{3}$$

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2 . Eq. 3 represents a noisy observation of

$$\mu_i = F(b^c \cdot \mathbf{u})_i. \tag{4}$$

We let $\theta_y^m = \{b^c, \sigma_m^2\}$ be the parameters of this model, where b^c is the coefficient in a linear model relating \mathbf{u} and \mathbf{y}^m in a first order approximation, relevant for example for magnetic resonance spectroscopic images (MRSI). A function $F(\cdot)$ could be, for example, a known nonlinear smoothing and subsampling operator, modeling the point spread function of the imaging process and different spatial resolution of \mathbf{y}^m and \mathbf{u} . Examples for \mathbf{y}^m include metabolic maps from MRSI [9], or apparent diffusion coefficient maps from DTI.

Binary segmentations. Tumor segmentations contain discrete labels $\mathbf{y}^s, y_i^s \in \{0, 1\}$. Since tumor outlines are commonly associated with constant tumor infiltration, we model observation \mathbf{y}^s for a given tumor cell concentration \mathbf{u} as a Bernoulli random variable:

$$p(\mathbf{y}^s | \mathbf{u}, \theta_y^s) = \prod_i p(y_i^s | \mathbf{u}, \theta_y^s) = \prod_i \alpha_i^{y_i^s} \cdot (1 - \alpha_i)^{1-y_i^s}, \tag{5}$$

where α_i is the probability of observing characteristic tumor-induced changes:

$$\alpha_i = .5 + .5 \cdot \text{sign}(u_i - u^c) \left(1 - e^{-\frac{(u_i - u^c)^2}{\sigma_s^2}} \right), \tag{6}$$

which is a double logistic sigmoid function with parameters $\theta_y^s = \{u^c, \sigma_s^2\}$. We essentially assume that the tumor cell infiltration is invisible below the threshold u^c but modifies the image signal in a predictable way after surpassing u^c [24]. Parameter σ_s^2 transforms the hard threshold into a smoother decision, also reflecting the uncertainty in the threshold u^c .

2.3 Joint Model

Combining the deterministic tumor growth model in Eq. (2) with the image observation models in Eqs. (3)–(5) and letting $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_k]$ denote the collection of all k image observations acquired at N time points t_n , we obtain

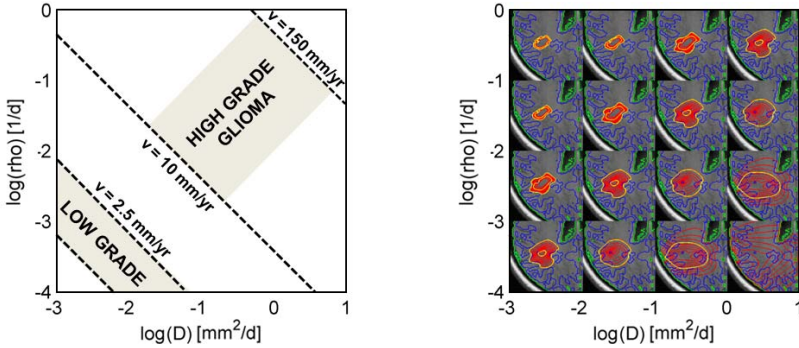


Fig. 3. Parameter space of the tumor model with tumor cell diffusivity D and proliferation rate ρ . Left: Parameterization for high- and low-grade glioma [122]. The speed of growth v can be obtained from times series of images [16]. It provides complementary information to tumor shape. Right: Shapes of equally large tumors for different parameterizations. Isolines of $\mathbf{u} = .05$ and $\mathbf{u} = .6$ are shown in yellow. We use the information from both time series and tumor shape jointly to characterize infiltrative and proliferative behavior of the tumor.

$$\begin{aligned}
 p(\mathbf{Y}|\theta_x, \theta_t, \theta_p, \theta_y) &= \prod_{t_n} p(\mathbf{y}(t_n)|\theta_x, \theta_t, \theta_p, \theta_y) \\
 &= \prod_{t_n} \int_{\mathbf{u}(t_n)} p(\mathbf{y}(t_n)|\mathbf{u}(t_n), \theta_y) p(\mathbf{u}(t_n)|\theta_x, \theta_t, \theta_p) d\mathbf{u}(t_n) \\
 &= \prod_{t_n} \int_{\mathbf{u}(t_n)} \prod_k p(\mathbf{y}_k(t_n)|\mathbf{u}(t_n), \theta_y) p(\mathbf{u}(t_n)|\theta_x, \theta_t, \theta_p) d\mathbf{u}(t_n), \tag{7}
 \end{aligned}$$

assuming that all observations are conditionally independent given the latent state \mathbf{u} . We adapt a factored prior $p(\theta_x, \theta_t, \theta_p, \theta_y) = p(\theta_x)p(\theta_t)p(\theta_p)p(\theta_y)$ and choose uniform distributions for all parameters. Similar to [1], we choose the range for the evolution parameters θ_p as illustrated in Fig. 3. We use experimental evidence [24,9] to set the range of the imaging parameters θ_y and use the life expectancy of a healthy person to set the range for the temporal parameter θ_t . We assume that the tumor started growing from a location within the hyperintense areas in T₂ MRI at $t = 0$ to set a prior on the spatial parameter θ_x . We obtain the joint posterior distribution of the parameters using Bayes’ rule:

$$p(\theta_x, \theta_t, \theta_p, \theta_y|\mathbf{Y}) \propto p(\mathbf{Y}|\theta_x, \theta_t, \theta_p, \theta_y)p(\theta_x, \theta_t, \theta_p, \theta_y), \tag{8}$$

from which we compute the posterior distribution of θ_p via marginalization

$$p(\theta_p|\mathbf{Y}) = \int_{\theta_x} \int_{\theta_t} \int_{\theta_y} p(\theta_x, \theta_t, \theta_p, \theta_y|\mathbf{Y}) d\theta_x d\theta_t d\theta_y. \tag{9}$$

$p(\theta_p|\mathbf{Y}) = p(D, \rho|\mathbf{Y})$ is the desired quantity of diagnostic interest in our application.

3 Inference

The posterior distribution of $\theta_p = \{\rho, D\}$ allows us to summarize a large set of image observations \mathbf{y} through diffusivity D and cell doubling rate ρ (Fig. 3). We aim at visualizing $p(\theta_p|\mathbf{Y})$ by drawing samples from it using an efficient Markov Chain Monte Carlo (MCMC) sampling strategy. MCMC constructs a random walk through the parameter space which converges towards the solution. The method avoids intractable integrals in Eq. (7). Unfortunately, the repeated evaluation of Eq. (9) for this walk requires the costly forward integration of the physical model in Eq. (11) for any sample from θ_p . To overcome this challenge, we separate the integration over the model parameters in Eq. (9): we integrate over parameters θ_t and θ_y which can be calculated sequentially at any step when propagating the forward model, and rely on MCMC only to sample from θ_x and θ_p .

Sequential integration of θ_t and θ_y . Given the parameters of the tumor growth model θ_p and a starting point θ_x , we propagate the physical model in Eq. (11) based the deterministic model U in Eq. (2). During the forward integration of Eq. (11), the integral

$$p(\theta_p, \theta_x|\mathbf{Y}) = \int_{\theta_y} \int_{\theta_t} p(\theta_x, \theta_t, \theta_p, \theta_y|\mathbf{Y})d\theta_t d\theta_y \tag{10}$$

can be calculated efficiently in a single sweep, by evaluating the image likelihood $p(\mathbf{y}|\mathbf{u}, \theta_y)$ at any step forward in time, and subsequently integrating over all steps. This procedure represents an integration over θ_t on a regularly spaced grid. Similarly, we integrate over the parameters θ_y^m and θ_y^s on a regularly spaced grid. This step can be performed very quickly once $\mathbf{u}(t_n)$ has been calculated for the different observations $\mathbf{y}(t_n)$. This procedure approximates Eq. (10) by a sum:

$$p(\theta_p, \theta_x|\mathbf{Y}) \approx \sum_{\theta_t^{(g)}} \sum_{\theta_y^{(g)}} p(\theta_x, \theta_t^{(g)}, \theta_p, \theta_y^{(g)}|\mathbf{Y}). \tag{11}$$

Efficient sampling of θ_x and θ_p . Even with the proposed sequential integration scheme, evaluating $p(\theta_p, \theta_x|\mathbf{Y})$ repeatedly for different values $\{\theta_p, \theta_x\}$ during MCMC sampling remains prohibitively time consuming. We therefore rely on a representation of $p(\theta_p, \theta_x|\mathbf{Y})$ that can be sampled more efficiently. In particular, we approximate $p(\theta_p, \theta_x|\mathbf{Y})$ through $\tilde{p}(\theta_p, \theta_x|\mathbf{Y})$ using a so-called *sparse grid* basis [3,15] in the first step. This approximation allows us to sample the space of $p(\theta_p, \theta_x|\mathbf{Y})$ in a *structured* way, with less redundant computations, and systematically increasing the accuracy of the approximation with every sample. More specifically, we interpolate between G pre-computed sampling points:

$$\tilde{p}(\theta_p, \theta_x|\mathbf{Y}) = \sum_{g=1}^G p(\theta_p^{(g)}, \theta_x^{(g)}|\mathbf{Y}) \cdot \Phi_s(\theta_p, \theta_x), \tag{12}$$

relying on a hierarchical linear basis function Φ_G that spans the space of θ_p and θ_x . This approximation is constructed from a small number of evaluations of $p(\theta_p, \theta_x | \mathbf{Y})$, sampled at specific grid locations $\{\theta_p^{(g)}, \theta_x^{(g)}\}$ within the range of $p(\theta_x)$ and $p(\theta_p)$. The interpolated posterior distribution can be evaluated at little computational cost. In the second step, we use the pre-computed approximation $\tilde{p}(\theta_p, \theta_x | \mathbf{Y})$ when evaluating a large number of samples from $p(\theta_p, \theta_x | \mathbf{Y})$ via MCMC. Efficiently sampling from $\tilde{p}(\theta_p, \theta_x | \mathbf{Y})$ also allows us to tune the sampling algorithm more easily in the presence of local minima. In the final step, we construct

$$\tilde{p}(\theta_p | \mathbf{Y}) = \int_{\theta_x} \tilde{p}(\theta_p, \theta_x | \mathbf{Y}) d\theta_x \quad (13)$$

by aggregating values of θ_p from our samples θ_x, θ_p . In order to minimize the number of necessary samples from $p(\theta_p, \theta_x | \mathbf{Y})$, i.e., the number of forward integrations of the physical model, we choose the sparse grid collocation from [18,27], but with a depth first search strategy for local refinement. To sample efficiently from Eq. (12) we use the Delayed Rejection Adaptive Metropolis (DRAM) variant of the Metropolis-Hastings MCMC sampler [11].

4 Experiments

We evaluate our method on a series of images acquired for monitoring patients with low-grade glioma. The low-grade glioma is typically only visible as hyperintense lesion in T₂/FLAIR images. Patients are monitored for the occurrence of T₁gad enhancements indicating the transition to high-grade disease and to be treated immediately. However, clinical evidence suggests that the dynamics of the tumor growth can predict this transition at a much earlier stage [22,4] (Fig. 3, left). We hypothesize that model parameters θ_p may provide more accurate information about the state of disease. Here we evaluate our approach under this clinical objective on synthetic ground truth and real clinical data. We also compare standard MCMC sampling with the proposed adaptive sparse grid MCMC sampling.

Implementation. We implement the tumor growth model in Eq. (1) by employing the preconditioned conjugate gradient descent to solve the state equations [19]. We sequentially evaluate Eq. (10) while integrating Eq. (1). In every forward step we evaluate the image likelihood, integrating on a grid over $u_c(\text{T}_2) = .01 \dots .2$ and $u_c(\text{T}_1\text{gad}) = .6 \dots .8$ with $\sigma_s^2 = .05 \dots .100$ and for metabolic maps for $b^c(\text{Choline}) = 1 \dots .8$ and $b^c(\text{NAA}) = -8 \dots .1$ with $\sigma_m^2 = .5 \dots .2$. We impose general physiological constraints such as a maximum time of growth of 50 years and a maximal sustainable tumor volume of 150 cm³.

We perform integration by sampling in a five dimensional space with $\{D, \rho\} = \theta_p$ and $\{x, y, z\} = \theta_x$ and always initialize the sampling in the center of the valid parameter space. We evaluate the samples via MCMC both with and without

sparse grid representation. In the first case (*direct MCMC*) we perform a random walk with 2000 samples. We use the last 1500 samples for evaluating the statistic of the parameters. Similar to [11], we use 100 samples to adapt the proposal distribution, and propagate up to two steps into suboptimal directions of the sampling space. We obtain rejection rates of 40%-60% with approximately 3000-5000 integrations of Eq. (10). In the second case (*sparse grid MCMC*) we evaluate Eq. (10) about the same number of times, i.e., for 3000 basis vectors, but at positions in the parameter space defined by an adaptive sparse grid with “maximum norm” basis [15]. We then perform MCMC sampling under the same settings as for the *direct MCMC*, but use the approximated posterior, interpolated from the sparse grid basis, as the function to be evaluated for each sample. The *direct* sampling takes about 6-12 hours for each test case on a standard personal computer. The *sparse grid* sampling takes a similar amount of computing, but can be easily parallelized, resulting in an algorithm that is 8-10 times faster. Subsequent MCMC sampling using the sparse grid interpolation is accomplished within minutes.

Data. In the first experiment, we model a synthetic low-grade glioma. We use DTI and tissue segmentations from a healthy segment of a patient data set to evolve a synthetic tumor with $D = 10^{-2}$ and $\rho = 10^{-2.3}$ for $\theta_t = 2000$ days, starting from a tumor seed point u^{init} with 10% infiltration in a 1 mm^3 volume. We assume $D = D_w = 10^3 D_g$ [11,16]. We model a second observation by propagating the model for another 180 days. For both observations we model the segmentations from T_2 /FLAIR images using a threshold of $u_c(T_2) = .095$. We set $u_c(T_{1gad}) = .695$. This is higher than the maximum value of \mathbf{u} at both time points and our synthetic T_{1gad} images do not show contrast agent-enhanced regions. We model metabolic maps from MRSI by smoothing \mathbf{u} with a 1 cm wide Gaussian kernel and subsampling the volume on a grid with 1 cm voxels and model two metabolite maps with coefficients $b^c(\text{Cho}) = 1.8$ and $b^c(\text{NAA}) = -4.6$. In the second experiment, we model a developing high-grade tumor. We use the same setting as above, but with $D = 10^{-0.2}$ and $\rho = 10^{-1.3}$. The tumor is evolved for 250 days for the first observation, and another 90 days for the second. Figure 4 (left) shows the second time point. All observations are modeled as above; T_{1gad} images do not show enhancements in this experiment either, and the tumor still appears as a low-grade glioma.

In addition to the two test cases with ground truth for D and ρ , we evaluate our model on two clinical data sets. The first data set comprises six sets of images acquired approximately every 3 months over 1.5 years. The second data set comprises four sets of images, acquired every 3-6 months in a similar time span. Available are in both cases DTI, T_2 /FLAIR and T_{1gad} images. T_2 /FLAIR images show a visible progression of the tumor front, while T_{1gad} is free of Gadolinium enhancements. The lesion is segmented manually in three 2D slices intersecting with the tumor center in the first case; it is manually segmented in all slices for the second case. To ensure the diffusion tensor to be free of tumor-induced effects, we use the mirrored DTI of the contra-lateral, disease-free hemisphere to evolve the tumor. The second data set was provided by the

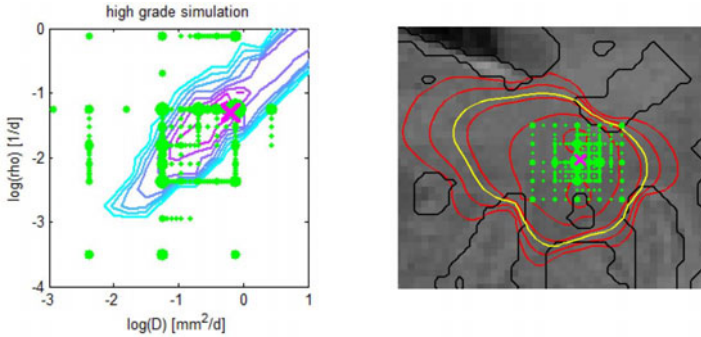


Fig. 4. Adaptive sampling of the parameter space for the synthetic high-grade data set. Left: Sampling points (green) for the model parameters $\theta_p = \{D, \rho\}$. The contour map shows the posterior distribution $p(\theta_p | \mathbf{Y})$ evaluated on a regular grid (cyan–low, pink–high). Right: Sampling points (green) for the $\{x, y\}$ coordinates of the tumor seed θ_x . The Fig. also shows isolines of tumor cell density (red), the predicted extensions of the T₂ hyper-intense area (yellow) and tissue boundaries (black). In both figures the size of the green sampling points indicates how often the indicated parameter was evaluated under different combinations. The ground truth is indicated by the pink cross. Most adaptively chosen sampling points are close to the ground truth.

authors of [16], who reported the speed of tumor growth in the white matter of $v_w = 2\sqrt{D_w\rho} = 2\sqrt{.2 \cdot .008}$ mm/d for this patient.

Results. Fig. 4 shows the adaptive sparse grid for the parameters of the growth model θ_p (left) and the tumor seed point θ_x (right). Green dots represent sampling points used to construct the approximation of the posterior in Eq. (12). The right image shows the central slice of the synthetic *high grade* data set. Indicated by dots are $\{x, y\}$ positions of the location parameters θ_x . Dot size indicates the number of model evaluations with varying $\{z, D, \rho\}$ for the given $\{x, y\}$. We find most of the adaptively chosen locations to be close to the real location of the tumor (pink cross). This also holds true for the parameter space spanned by diffusivity D and proliferation ρ in the left image. Here the location close to the actual parameter (pink cross) is the one evaluated for the highest number of different seed points. The contour map shows the likelihood evaluated on a regular grid with the true seed point θ_x and is in good agreement with the adaptively sampled locations on the sparse grid. The most likely region is a diagonal line in D - ρ space, similar to areas of constant shape in Fig. 3 (right) and orthogonal to regions of constant speed of growth in Fig. 3 (left). This indicates that information about the shape of the tumor is effectively used to infer model parameters in our proposed approach.

Fig. 5 reports estimates of the model parameters D and ρ for all four test cases and both the direct sampling (blue) and the sparse grid sampling (green). For the first two data sets (A, B), the sparse grid sampling is in good agreement with the ground truth (pink cross). The mean value of the direct sampling

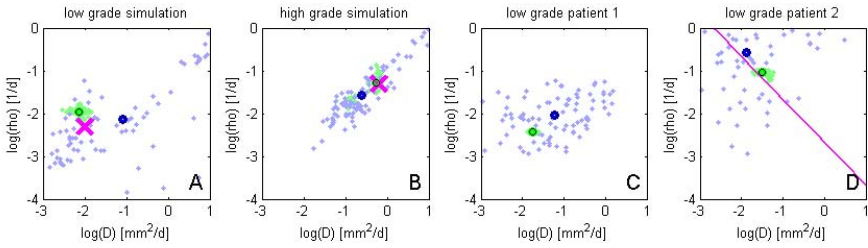


Fig. 5. MCMC sampling results in the space spanned by model parameters D and ρ , for the four experiments. Green samples are obtained from the sparse grid interpolation Eq. (12), blue-purple samples come from the direct sampling in Eq. (10). Black circles indicate means of the two distributions. Ground truth for A and B are indicated by the pink cross. In D the previously estimated speed of growth [16] is shown by the pink line. The sparse grid sampling approximation performs better than the direct MCMC (A - B). Estimates correlate well with results from [16], but provide a more accurate characterization of the state of disease (D).

(blue/black circle) is relatively close to ground truth and the mean of the sparse grid sampling (green/black circle). We find the sparse grid results to be much more compact with noticeably less variation. This observation is also true for the patient data (C , D). We used the current number of 3000 samples for sparse grid MCMC only to allow for a direct comparison of both approaches. Already an interpolation with a few hundred samples led to results similar to those in Fig. 3, demonstrating the benefits of the sparse sampling approach for estimating parameters of the tumor model.

Results in Fig. 5 can also be interpreted in terms of the diagnostic task. In subfigure C , results suggest a “classical” low-grade tumor, well in agreement with the expected parameters for low-grade patients (Fig. 3, left). In subfigure D , we observe a more infiltrative and rapidly growing tumor. This is in good agreement with the clinical observation of the tumor expanding several centimeters per year and showing a star-shaped pattern, often associated with poor prognosis. The data in D had already been evaluated in [16] and the authors had estimated the speed of the tumor front shown as the red pink diagonal line in D . We observe a good agreement of our results with this earlier estimate, but obtain the additional information that the expansion of the tumor is not due to a fast infiltration, but a rapid doubling of tumor cells with a direr prognosis for this particular patient.

5 Conclusions

The proposed tumor modeling approach links the physical process model with the statistical image observation model and enables efficient estimation of model parameters in a Bayesian framework, potentially also to be used with model components at different scales and with different sources of information. Preliminary results are promising. The approach can readily be adapted to incorporate

information from other functional imaging modalities, and to model mass effect and tissue translations, or tissue alterations resulting from therapy.

Acknowledgements. This work was supported by the German Academy of Sciences Leopoldina (Fellowship Programme LPDS 2009-10), the Academy of Finland (133611), INRIA CompuTumor, NIH NIBIB NIMIC U54-EB005149, NIH NCRR NAC P41- RR13218, NIH NINDS R01-NS051826, NIH R01-NS052585, NIH R01-EB006758, NIH R01-EB009051, NIH P41-RR014075 and the NSF CAREER Award 0642971.

References

1. Alvord, E.C., Swanson, K.R.: Using mathematical modeling to predict survival of low-grade gliomas. *Ann. Neurol.* 61, 496–497 (2007)
2. Cuadra, B.B., Pollo, C., Bardera, A., Cuisenaire, O., Thiran, J.P.: Atlas-based segmentation of pathological brain MR images using a model of lesion growth. *IEEE TMI* 23, 1301–1314 (2004)
3. Bungartz, H.-J., Griebel, M.: Sparse grids. *Acta Numerica* 13, 147–269 (2004)
4. Caseiras, G.B., Ciccarelli, O., Altmann, D.R., Benton, C.E., Tozer, D.J., Tofts, P.S., Yousry, T.A., Rees, J., Waldman, A.D., Jäger, H.R.: Low-grade gliomas: six-month tumor growth predicts patient outcome better than admission tumor volume, relative cerebral blood volume, and apparent diffusion coefficient. *Radiology* 253, 505–512 (2009)
5. Chaplain, M.A.J., Stuart, A.M.: A mathematical model for the diffusion of tumour angiogenesis factor into the surrounding host tissue. *J. Math. Appl. Med. Biol.* 8, 191–220 (1991)
6. Clatz, O., Sermesant, M., Bondiau, P.-Y., Delingette, H., Warfield, S.K., Malandain, G., Ayache, N.: Realistic simulation of the 3-D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. *IEEE TMI* 24, 1334–1346 (2005)
7. Cobzas, D., Mosayebi, P., Murtha, A., Jagersand, M.: Tumor invasion margin on the riemannian space of brain fibers. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009*. LNCS, vol. 5762, pp. 531–539. Springer, Heidelberg (2009)
8. 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 TMI* 9, 629–640 (2008)
9. Ganslandt, O., Stadlbauer, A., Fahlbusch, R., Kamada, K., Buslei, R., Blumcke, I., Moser, E., Nimsky, C.: ^1H -MRSI integrated into image-guided surgery: correlation to standard MR imaging and tumor cell density. *Neurosurg.* 56, 291–298 (2005)
10. Gooya, A., Biros, G., Davatzikos, C.: Deformable registration of glioma images using EM algorithm and diffusion reaction modeling. *IEEE TMI* 30, 375–390 (2011)
11. Haario, H., Laine, M., Mira, A., Saksman, E.: DRAM: Efficient adaptive MCMC. *Statistics and Computing* 16, 339–354 (2006)
12. Hamamci, A., Unal, G., Kucuk, N., Engin, K.: Cellular automata segmentation of brain tumors on post contrast MR images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6363, pp. 137–146. Springer, Heidelberg (2010)

13. Hirsch, S., Szczerba, D., Lloyd, B., Bajka, M., Kuster, N., Székely, G.: A mechanochemical model of a solid tumor for therapy outcome predictions. In: Allen, G., Nabrzyski, J., Seidel, E., van Albada, G.D., Dongarra, J., Sloot, P.M.A. (eds.) ICCS 2009. LNCS, vol. 5544, pp. 715–724. Springer, Heidelberg (2009)
14. Hogeia, 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)
15. Klimke, A., Wohlmuth, B.: Piecewise multilinear hierarchical sparse grid interpolation in MATLAB. *ACM Trans. Math. Software* 31, 1–20 (2005)
16. Konukoglu, E., Clatz, O., Menze, B.H., Weber, M.-A., Stieltjes, B., Mandonnet, E., Delingette, H., Ayache, N.: Image guided personalization of reaction-diffusion type tumor growth models using modified anisotropic Eikonal equations. *IEEE TMI* 29, 77–95 (2010)
17. Kyriacou, S.K., Davatzikos, C., Zinreich, S.J., Bryan, R.N.: Nonlinear elastic registration of brain images with tumor pathology using a biomechanical model. *IEEE TMI* 18, 580–592 (1999)
18. Ma, X., Zabarab, N.: An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. *Inverse Problems* 25 (2009)
19. McCorquodale, P., Colella, P., Johansen, H.: A Cartesian grid embedded boundary method for the heat equation in irregular domains. *J. Comp. Phys.* 173(2), 620–635 (2001)
20. Menze, B.H., Stretton, E., Konukoglu, E., Ayache, N.: Image-based modeling of tumor growth in patients with glioma. In: Garbe, C.S., Rannacher, R., Platt, U., Wagner, T. (eds.) *Optimal Control in Image Processing*. Springer, Heidelberg (2011)
21. Mohamed, A., Zacharakib, E.I., Shena, D., Davatzikos, C.: Deformable registration of brain tumor images via a statistical model of tumor-induced deformation. *MedIA* 10, 752–763 (2006)
22. Pallud, J., Mandonnet, E., Duffau, H., Galanaud, D., Taillandier, L., Capelle, L.: Prognostic value of initial magnetic resonance imaging growth rates for World Health Organization grade II gliomas. *Ann. Neurol.* 60, 380–383 (2006)
23. Prastawa, M., Bullitt, E., Gerig, G.: Simulation of brain tumors in MR images for evaluation of segmentation efficacy. *MedIA* 13, 297–311 (2009)
24. 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)
25. Zacharaki, E., Shen, D., Mohamed, A., Davatzikos, C.: Registration of brain images with tumors: Towards the construction of statistical atlases for therapy planning. In: *Proc. ISBI* (2006)
26. Zacharaki, E.I., Hogeia, C.S., Shen, D., Biros, G., Davatzikos, C.: Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. *Neuroimage* 46, 762–774 (2009)
27. Zarabab, N.: Solving stochastic inverse problems: A sparse grid collocation approach. In: Biegler, L. (ed.) *Large-Scale Inverse Problems and Quantification of Uncertainty*. Wiley, Chichester (2011)

An Event-Based Disease Progression Model and Its Application to Familial Alzheimer's Disease

Hubert M. Fonteijn¹, Matthew J. Clarkson², Marc Modat², Josephine Barnes³, Manja Lehmann³, Sebastien Ourselin², Nick C. Fox³, and Daniel C. Alexander¹

¹ Centre for Medical Image Computing, Department of Computer Science, University College London, UK

h.fonteijn@cs.ucl.ac.uk

² Centre for Medical Image Computing, Department of Medical Physics and Bioengineering, University College London, UK

³ Dementia Research Centre, Institute of Neurology, University College London, UK

Abstract. This study introduces a novel event-based model for disease progression. The model describes disease progression as a series of events. An event can consist of a significant change in symptoms or in tissue. We construct a forward model that relates heterogeneous measurements from a whole cohort of patients and controls to the event sequence and fit the model with a Bayesian estimation framework. The model does not rely on *a priori* classification of patients and therefore has the potential to describe disease progression in much greater detail than previous approaches. We demonstrate our model on serial T1 MRI data from a familial Alzheimer's disease cohort. We show progression of neuronal atrophy on a much finer level than previous studies, while confirming progression patterns from pathological studies, and integrate clinical events into the model.

Keywords: Disease progression, Alzheimer's disease, atrophy, computational model.

1 Introduction

Every disease is uniquely characterized by a progression of symptoms and pathology, whose understanding is vital for accurate diagnosis and treatment planning. Modeling patterns of disease progression is a key aim of medical science, because it enables basic understanding of the disease and provides staging systems that assist treatment. Examples of staging systems that are based on disease progression can be found in diseases such as cancer and HIV. For instance, the Ann Arbor system [2] classifies lymphomas into four stages which progress from a localized tumor in a single lymph node to a widespread involvement of lymph nodes and organs.

Disease progression occurs at various levels, ranging from the symptoms a patient experiences to cellular and biochemical changes. For instance, Alzheimer's disease (AD) is characterized by a progressive deterioration of cognitive abilities,

which is caused by severe pathology at the cellular level. This pathology consists of extracellular Amyloid plaques and intracellular neurofibrillary tangles (NFT), which are thought to cause local neuronal atrophy. Braak and Braak [1] show that the spread of NFT's follows a very specific pattern, spreading from memory related areas such as the hippocampus towards areas in the medial temporal lobe, the parietal cortex and the prefrontal cortex. Only in the last stage of AD are primary cortices affected. Interestingly, there is now ample evidence that neuronal atrophy can be measured *in vivo* with T1-weighted Magnetic Resonance Imaging (MRI). Scahill et al. [16] use non-linear registration methods to measure atrophy and show that the spread of atrophy follows a very similar pattern to the spread of NFT's. Thompson et al. [18] and Dickerson et al. [4] show a very similar effect when studying differences in cortical thickness.

Current models of disease progression are relatively crude. These models use symptomatic staging to divide the patients into a small number of groups, which typically characterize disease state as "presymptomatic", "mild", "moderate" or "severe". They then assess the differences in biomarkers among those groups. Symptomatic staging is an invaluable tool to explain disease progression to patients and their family. However, it relies on imprecise clinical data and therefore has limited temporal resolution, which limits its power to discriminate and stage diseases. Furthermore, symptomatic staging often suffers from high inter-rater variability. Current analysis methods therefore do not deliver the promise of *in vivo* measurements to develop models of disease progression that match or even surpass the level of temporal detail of pathology-based models.

In this study we take a direct computational approach to disease progression modeling. We model the progression pattern as a sequence of discrete events. Similar models arise in palaeontology, where researchers fit time lines of the emergence and extinction of species to the fossil occurrence of species in different geological time periods [14]. In our model, the events are changes in patient state, such as the onset of a new symptom ("patient reports memory loss to GP") or the first measurement of a tissue pathology ("lumbar puncture shows raised beta amyloid"). The aim is to find the ordering of the events that is most consistent with the measurements from a cohort of patients.

The event-based computational modeling approach has major advantages over current approaches to modeling disease progression: First, unlike all previous approaches, it does not rely on *a priori* staging of patients using clinical symptoms, but extracts the event ordering directly from the data. This enables the model to express a much more detailed time line of disease progression (see figures 1 and 2) limited only by the number of events considered rather than crude clinical staging. The approach also naturally captures uncertainty and variance in the ordering (figure 2) reflecting measurement error and heterogeneity over the population or cohort. Furthermore, the more detailed time line provides a finer grained assessment of new patients, as it is straightforward to take a new patient's data and identify their position along the model time line. It is also a more natural description of the progression than those obtained indirectly through clinical staging. Second, model fitting naturally exploits longitudinal or

cross-sectional data sets or, as we demonstrate here, heterogeneous data sets with different numbers of time points for each patient. Third, the approach combines information from different sources naturally. Here, for example, we demonstrate orderings of events including regional atrophy events informed by imaging as well as clinical events informed by patients' scores in behavioural tests.

We demonstrate this model on serial T1 MRI data from familial Alzheimer's disease (fAD) patients. fAD is a rare autosomal-dominantly inherited variant of Alzheimer's disease which causes early onset. Because fAD mutation carriers will develop AD at a predictable time it is feasible to enroll them in studies before they develop any clinical symptoms, thereby providing a unique opportunity to study changes in biomarkers in the very earliest stages of AD. The event-based disease progression model shows the disease development in much finer detail than previously seen while confirming existing knowledge about the pattern and progression of atrophy in fAD.

2 Theory

2.1 Event-Based Model

The event-based disease progression model consists of a set of events E_1, \dots, E_N and an ordering $S = (s(1), \dots, s(N))$, which is a permutation of the integers $1, \dots, N$ determining the event ordering $E_{s(1)}, \dots, E_{s(N)}$. The set of events are specified a priori and we estimate S from a data set X . The full data set X contains a set of measurements X_j , $j = 1, \dots, J$, from each of the J patients and X_l , $l = 1, \dots, L$, from each of L controls. In this study, each measurement informs just one event in one patient, and each event has a corresponding measurement. Thus each $X_j = [x_{1j}, x_{2j}, \dots, x_{Nj}]^T$ contains exactly N corresponding measurements: x_{ij} informs as to whether E_i in patient j has occurred. However, the formulation adapts naturally to allow multiple measurements to inform single events, single measurements to inform multiple events, and the set of measurements to vary among patients.

To fit the model to the data, we need to evaluate the likelihood $p(X | S)$ of the model given the data. We start by fitting simple models for the probability distribution $p(X_i | E_i)$ on the measurement X_i , the i^{th} measurement in all patients, given that E_i has occurred and, similarly, $p(X_i | \neg E_i)$ on x_i given that E_i has not occurred. We provide more detail about this procedure in section [2.2](#).

If patient j is at position n in the progression model, events $E_{s(1)}, \dots, E_{s(n)}$ have occurred, while events $E_{s(n+1)}, \dots, E_{s(N)}$ have not, and we can write the likelihood of that patient's data given S :

$$p(X_j | S, k) = \prod_{i=1}^k p(X_{s(i)j} | E_{s(i)}) \prod_{i=k+1}^N p(X_{s(i)j} | \neg E_{s(i)}), \quad (1)$$

where we assume individual measurements are independent. We integrate out the hidden variable k to obtain

$$p(X_j | S) = \sum_{k=1}^N p(k)p(X_j | S, k), \tag{2}$$

where $p(k)$ is the prior probability of being at position k in the ordering; we assume uniform priors here. Next, we assume independence of measurements among patients to obtain

$$p(X | S) = \prod_{j=1}^J p(X_j | S). \tag{3}$$

We use Bayes theorem to obtain the posterior distribution $p(S | X)$:

$$p(S | X) = \frac{p(S)p(X | S)}{p(X)}. \tag{4}$$

The normalization constant $p(X)$ is analytically intractable. We therefore use a Markov Chain Monte Carlo algorithm [9] to sample from $p(S | X)$. We use flat priors on the ordering S , on the assumption that *a priori* all orderings are equally likely. We provide more details about this algorithm in section 2.3.

2.2 The Likelihood of the Data Given Events

The probability distribution on a measurement x associated with one particular event E has two key components:

$$p(x) = p(x | E)p(E) + p(x | \neg E)p(\neg E), \tag{5}$$

which are the distributions on the measurement, conditional on whether E has or has not occurred. However, evaluating equation 1 requires separate models for both $p(x | E)$ and $p(x | \neg E)$ in order to evaluate equation 5. To obtain models for these two conditional distributions for a particular event E_i , we start by fitting a mixture of Gaussians to the set of measurements $Y_i = [x_{ij} | j = 1, \dots, J] \cup [x_{il} | l = 1, \dots, L]$, associated with event E_i from all patients and controls. To fit the mixture of Gaussians, we use the Expectation Maximization algorithm [3] with the number of components set to 1, 2, ..., 5 and pick the model that minimizes the Bayesian Information Criterion [11]. The resulting mixture provides a crude model for $p(x)$ in equation 5 for this particular i . Next, we associate each measurement x_i with the Gaussian component that maximizes its likelihood. As a model for $p(x | \neg E)$, we take the single Gaussian component that is associated to the largest proportion of measurements from controls. The model for $p(x | E)$ is the mixture of the remaining components.

Clinical events require slightly different treatment, because, in the algorithm, they are binary rather than informed by a measurement: either the patient was classified as AD from the cognitive tests or they were not. We simply assign fixed high and low likelihood values to $p(x | E)$ and $p(x | \neg E)$, respectively if the event has occurred, or vice versa if not. We choose fixed values at the extremes of the

range of values we observe for the atrophy events. However, the precise choice of settings of the fixed values has very little effect on the outcome.

2.3 The MCMC Algorithm

The MCMC algorithm starts by proposing a random candidate sequence. At each iteration t , a perturbation of the current model S_t swaps the positions of two randomly chosen events, in the next step $S_{t+1} = S'$ with probability $\min(a, 1)$ where $a = p(X | S')/p(X | S_t)$ is the likelihood ratio; otherwise $S_{t+1} = S_t$. MCMC algorithms depend on accurate initialization to achieve good mixing properties within realistic time spans. We therefore initialize the MCMC algorithm with the maximum likelihood solution S_{ML} for $p(X | S)$, which we find with a greedy ascent algorithm. The greedy ascent algorithm uses the same perturbation rule as the MCMC algorithm, but sets $a = 1$ for $p(X | S') > p(X | S_t)$ or $a = 0$ otherwise. We run the algorithm for 2,000 iterations, which generally is enough to reach a local maximum. We repeat the algorithm 10 times from different initialization points to ensure we reach the global maximum. Next, we run the MCMC procedure for 200,000 iterations, of which we discard 100,000 burnin iterations and retain the remaining 100,000 samples.

3 Methods

In this study, we use two types of event: clinical events and atrophy events. Clinical events include the progression to a later clinical status, such as from presymptomatic AD to Mild Cognitive Impairment (MCI). Criteria for progression are outlined in section 3.1. Atrophy events correspond to the occurrence of significant regional atrophy within a patient and are derived from serial T1 MRI data. Section 3.2 provides details about the MRI acquisitions. Section 3.3 outlines the preprocessing steps which we use to compute regional atrophy values from the MR images. To fit an ordering of all these events, we then run the MCMC algorithm using the procedures as outlined in section 2.

3.1 Patients

The fAD data we use has been previously analyzed in [15]. Briefly, nine carriers of autosomal mutations associated to Alzheimer's Disease were recruited from the Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery. 25 Age-matched and sex-matched controls (two to three controls per mutation carrier) were also recruited from spouses, relatives and healthy volunteers. All participants gave written informed consent as approved by the local ethics committee. All mutation carriers underwent comprehensive clinical and neuropsychological assessments, including the mini-mental state examination (MMSE) [7] and volumetric MRI scans, at each visit (41 visits: three to eight per patient, interval between visits ≈ 12 month). Each control patient had two MRI scans (except two participants who had four scans each) adding up to

54 scans in total. The clinical status of each mutation carrier was classified at each time point as: 1) familial Alzheimer's disease, if the patient fulfilled National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders (NINCDS-ADRDA) diagnostic criteria for probable Alzheimer's disease; 2) MCI, if the patient fulfilled MCI criteria; and 3) presymptomatic, if participants fell short of both NINCDS-ADRDA and MCI criteria.

3.2 Imaging

Imaging on the fAD cohort was undertaken with two scanners: in the period between 1991 and 2000 a 1.5 T GE Signa MRI scanner (General Electric Medical Systems, Waukesha, WI, USA) was used. T1 images were acquired using a Spoiled Gradient Echo technique with the following parameters: (256 x 128 matrix, field of view 24 x 24 cm, TR/TE/NEX/FA=35 ms/5 ms/1/35°) yielding 124 contiguous 15 mm thick slices. Between 2000 and 2005 scans were acquired on a different 1.5 T GE Signa scanner. On this scanner T1 images were acquired using an inversion recovery fast spoiled gradient echo sequence with the following parameters: 256 x 256 matrix, field of view 24 x 18 cm, TR/TE/TI/NEX/FA=14 ms/5.4 ms/650 ms/1/15°, yielding 124 contiguous 1.5 mm thick slices.

3.3 Preprocessing

The preprocessing pipeline extracts anatomically defined regions, determines voxel-wise atrophy using non-rigid registration methods and combines both steps to compute regional atrophy values. We use Freesurfer to segment the brain into cortical and subcortical structures. Freesurfer starts with a gray matter/white matter segmentation. It then uses a labeling procedure on each subject's data which is based on a training set of manually segmented scans [5,6]. Thus an anatomical segmentation is achieved which is dependent on each subject's individual anatomy and which has been shown to be statistically equivalent to manual segmentation [6]. In the fAD data, we supplement Freesurfer's segmentation with subject-specific hippocampus masks, which were manually traced by [15]. We calculate voxel-wise atrophy values using a free form deformation registration method [12]. This method first uses a rigid registration to realign the repeat scans to the baseline scan and then uses a set of cubic B-splines to locally deform the repeat scan to match the baseline scan. In each voxel, the change that is required by this matching procedure is quantified by the determinant of the Jacobian of the deformation field. We calculate regional atrophy by calculating the mean Jacobian value for each segmented region.

4 Results

Figure 1 shows the average event sequence for the fAD cohort. This model is based on clinical events and regional atrophy values. We reduce the number of

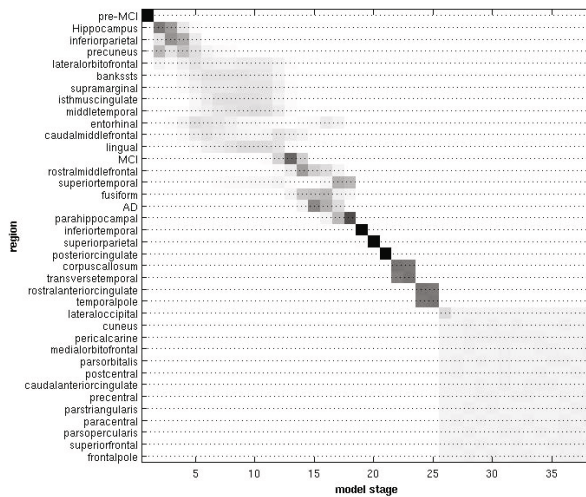


Fig. 2. Positional variance of the event sequence for the fAD cohort. For each region, the frequency with which a region visits a position in the event sequence during the MCMC procedure is plotted. The regions are ordered according to position in the event sequence, averaged over all MCMC iterations.

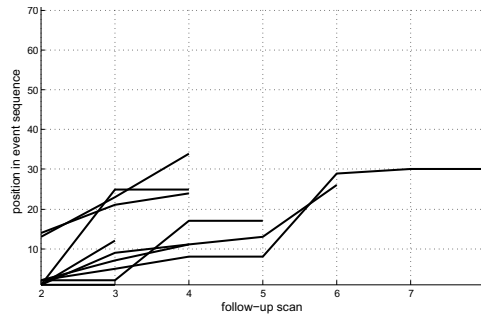


Fig. 3. Classification results for all follow-up scans in all patients. We use the event-based disease progression model to compute each scan's stage in the model. Each separate line represents one individual patient. Within patients, later follow-up scans correspond to later stages in the model.

the separation of many events close together in the ordering is weak, but events further apart show clear separations. Moreover, we see some separation of consecutive clusters of regions whose ordering rarely overlaps. Finally, the bottom right corner of this figure shows a block of events with equal probability for all positions within that block. These events do not show any evidence that the corresponding event has occurred and their position in the ordering is therefore spread over all positions after the last event, in which there is evidence that the corresponding event has occurred.

The event-based disease progression model develops a time line of disease progression that can be used to position each within the event sequence. As a simple validation of the model, we exploit the fact that we have multiple time points for each patient; recall that we ignore this information when fitting the model. We determine the position of each patient by maximizing the likelihood of their data, given the maximum a posteriori estimate of the event sequence from the MCMC algorithm, with respect to the hidden variable k in equation 2. The test for the validity of the event-based disease progression model is whether later sets of measurements from the same subject are positioned later in the event sequence than earlier measurements. Figure 3 shows that this is consistently the case for the event-based disease progression model of the fAD cohort.

5 Discussion

In this study, we introduce a new event-based disease progression model and estimation procedure. This model represents a radical departure from previous approaches to study disease progression: instead of relying on classification of patients according to symptomatic criteria, the event-based disease progression model utilizes all aspects of the available data to develop a time line of disease progression with maximum detail. We apply the model to MRI data from a fAD cohort and find excellent agreement with existing models of disease progression. These results highlight the strength of our approach in several ways: First, the disease progression model does not rely on *a priori* classification of patients and therefore makes less assumptions about the data than previous studies. Nevertheless, it reconstructs a time line of disease progression which agrees well with those studies. Second, our model dramatically improves the temporal resolution of time line of disease progression, when compared to previous studies. Third, the fAD data set contains only a very limited number of patients. It is therefore remarkable that we are able to reconstruct such a detailed time line of disease progression from this data. Finally, we develop a measure of positional variance which indicates the confidence we have about the position of each event in the event sequence. Previously, time lines of disease progression have been developed in an ad hoc fashion by reviewing data from many studies and such time lines therefore did not contain any formal measure of confidence in the sequence of events. To the best of our knowledge, this study is the first to introduce a model that explicitly fits such time lines and that quantifies their uncertainty.

We have omitted several further experiments for brevity, which confirm the robustness of our approach. We only show results from regional atrophy measures that are averaged across hemispheres. We have however also fitted the event-based disease progression model to the complete data set and have found a high correlation between the positions of regions in both hemispheres. We therefore conclude that hemispheric averaging is valid for this disease. We have also tested our model on regional atrophy that was computed with a different non-linear registration algorithm [8]. We found broad agreement of event sequences between both registration methods. Future work will test the algorithm on parcellations other than the Freesurfer parcellation.

In many neurodegenerative diseases, including AD [13] and HD [17], large-scale multi-center studies are now underway to investigate the use of a range of measurements, such as structural MRI and CSF measurements, as biomarkers for disease progression. Our model scales up well to larger data sets and would therefore be an excellent candidate to analyze these data sets. It will be interesting to investigate how the increased number of subjects in these data sets will affect the positional variance: the increase in statistical power will lead to lower variance in the event sequence, but the between-subject variability in disease progression will be better sampled, leading potentially to a higher variance in the event sequence. One drawback of such larger data sets is that they are likely to contain misclassifications: for instance, patients can be included in an AD study on the basis of AD-like symptoms, but can in fact have a different disease. We avoid this issue in this study, because in the fAD cohort each patient is selected based on genetic characteristics that are typical for fAD. Future studies will have to investigate the robustness of our model to the misclassifications inherent in current large-scale studies.

We have assumed that event measurements are independent, even if they are repeated measures within the same subject. It would be interesting to investigate whether incorporating such information for (partially) longitudinal data decreases the variance in the event sequence. Likewise, we have not exploited the considerable correlation we observe between different event measurements in both patients and in controls. In patients, these correlations could be used in a data reduction step, in which events with very similar measurement patterns are grouped together. This would considerably reduce the number of stages and thereby the complexity of the model. In controls, such correlations could be used to develop more sensitive algorithms for event detection. Instead of only relying on the distribution of single event measures in controls to determine whether an event has happened, the conditional distribution of one event measure, given several other (correlated) event measures could be used to predict the event measure in patients and to detect deviations from the predicted value.

The event-based disease progression model can readily be applied to other diseases, neurodegenerative or otherwise. This will be especially useful in cases where a well-established model for disease progression is lacking. Future work will develop the event-based disease progression model further, by for instance including the possibility for event onsets to be simultaneous and incorporating a variance around the event onsets. Another important improvement to the model will be the incorporation of a mixture of event sequences, which will allow the classification of diseases into sub-types. Mixtures of event orderings have already been studied by Manila and Meek [10], but this approach will have to be adapted considerably to be used in the context of disease progression. Other future work will further integrate more measurement types, such as functional and anatomical connectivity measures, CSF measurements and others into the currently developed model for disease progression. This will eventually lead to a rich and detailed model of disease progression in AD and other diseases.

Acknowledgments

We thank Martin Rossor, Basil Ridha and Natalie Ryan for their contribution to the clinical aspects of this study. EPSRC grants EP/D506468/01 and EP/E007748 supported HMF and DCA on this project.

References

1. Braak, H., Braak, E.: Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica* 82(4), 239–259 (1991)
2. Carbone, P., Kaplan, H., Musshoff, K., Smithers, D., Tubiana, M.: Report of the committee on Hodgkin's disease staging classification. *Cancer Research* 31(11), 1860 (1971)
3. Dempster, A., Laird, N., Rubin, D., et al.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1), 1–38 (1977)
4. Dickerson, B., Bakkour, A., Salat, D., Feczko, E., Pacheco, J., Greve, D., Grodstein, F., Wright, C., Blacker, D., Rosas, H., et al.: The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex* 19(3), 497 (2009)
5. Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al.: Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron* 33(3), 341–355 (2002)
6. Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., et al.: Automatically parcellating the human cerebral cortex. *Cerebral Cortex* 14(1), 11 (2004)
7. Folstein, M.F., Folstein, S.E., McHugh, P.R.: Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 12(3), 189 (1975)
8. Freeborough, P., Fox, N.: Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *Journal of Computer Assisted Tomography* 22(5), 838 (1998)
9. Gilks, W.R., Richardson, S., Spiegelhalter, D.J.: Markov chain Monte Carlo in practice. Chapman & Hall/CRC (1996)
10. Mannila, H., Meek, C.: Global partial orders from sequential data. In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 168. ACM, New York (2000)
11. McLachlan, G., Peel, D.: Finite mixture models. Wiley Interscience, Hoboken (2000)
12. Modat, M., Ridgway, G., Taylor, Z., Lehmann, M., Barnes, J., Hawkes, D., Fox, N., Ourselin, S.: Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine* 98(3), 278–284 (2010)
13. Mueller, S., Weiner, M., Thal, L., Petersen, R., Jack, C., Jagust, W., Trojanowski, J., Toga, A., Beckett, L.: Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's and Dementia: The Journal of the Alzheimer's Association* 1(1), 55–66 (2005)

14. Puolamaki, K., Fortelius, M., Mannila, H.: Seriation in paleontological data using Markov chain Monte Carlo methods. *PLoS Computational Biology* 2(2), e6 (2006)
15. Ridha, B.H., Barnes, J., Bartlett, J.W., Godbolt, A., Pepple, T., Rossor, M.N., Fox, N.C.: Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *The Lancet Neurology* 5(10), 828–834 (2006)
16. Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C.: Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences of the United States of America* 99(7), 4703 (2002)
17. Tabrizi, S., Langbehn, D., Leavitt, B., Roos, R., Durr, A., Craufurd, D., Kennard, C., Hicks, S., Fox, N., Scahill, R., et al.: Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *The Lancet Neurology* 8(9), 791–801 (2009)
18. Thompson, P., Mega, M., Woods, R., Zoumalan, C., Lindshield, C., Blanton, R., Moussai, J., Holmes, C., Cummings, J., Toga, A.: Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex* 11(1), 1 (2001)

The Ideal Observer Objective Assessment Metric for Magnetic Resonance Imaging Application to Signal Detection Tasks

Christian G. Graff and Kyle J. Myers

Division of Imaging and Applied Mathematics, Center for Devices and Radiological Health, U. S. Food and Drug Administration, Silver Spring MD USA
`christian.graff@fda.hhs.gov`

Abstract. The ideal Bayesian observer is a mathematical construct which makes optimal use of all statistical information about the object and imaging system to perform a task. Its performance serves as an upper bound on any observer's task performance. In this paper a methodology based on the ideal observer for assessing magnetic resonance (MR) acquisition sequences and reconstruction algorithms is developed. The ideal observer in the context of MR imaging is defined and expressions for ideal observer performance metrics are derived. Comparisons are made between the raw-data ideal observer and image-based ideal observer to elucidate the effect of image reconstruction on task performance. Lesion detection tasks are studied in detail via analytical expressions and simulations. The effect of imaging sequence parameters on lesion detectability is shown and the advantages of this methodology over image quality metrics currently in use in MR imaging is demonstrated.

1 Introduction

The assessment of MR sequences and reconstruction algorithms is dominated by pixel-based SNR (defined as the mean of a region of interest divided by the noise standard deviation), resolution considerations and qualitative discussions of image artifacts [5]. These measures of image quality are easy to determine, however they are only weakly correlated with the ability of an observer to perform a clinical task. Since medical images are created to perform a task, it is more meaningful to base the assessment of image quality on the ability of a reader (human or computer algorithm) to perform a clearly-defined task [13].

Many tasks performed via MR involve detecting abnormalities, i.e., edema in the brain, liver or breast lesions, etc. which may be regarded as 2-class classification problems, with the task being to correctly classify the patient as normal or abnormal. The mechanism that allows these tasks to be performed is class-dependent MR signal contrast brought about by differences in the chemical properties of the object. These differences manifest themselves as spatial and temporal changes in MR tissue properties.

The observer performing this task is typically a human, though increasingly algorithms in the form of computer-aided diagnosis tools are gaining attention.

An important tool for assessing image quality is the ideal observer (IO), a mathematical construct which makes use of all available statistical information to make classification decisions. It is optimal in the sense that it minimizes Bayes risk, or equivalently maximizes receiver operating characteristic (ROC) performance [9,10]. Ideal observer performance provides an objective measure of the suitability of the acquired data to the performance of the task and an upper bound on performance of all other observers (both human and computer-based). The use of the IO to assess MR images has been suggested in the past [7,10] and model observers for MR reconstructions have been studied more recently [8]. In this work MR IO theory is rigorously developed and the relationship between practical MR sequence parameters and IO performance is studied in detail.

In the following section the MR imaging operator is defined and the IO and its performance metrics are developed for an arbitrary classification task. In subsequent sections the focus turns to the particular task of small lesion detection, with analytical results for a signal-known-exactly (SKE) task using an echo-planar acquisition and numerical results for a signal-known-statistically (SKS) task, where the lesion location is not known a priori. The final section gives conclusions about the potential of this methodology to improve MR image quality assessment.

2 Fundamentals

The MR Imaging Operator. A general digital imaging system collects a finite set of data, represented by vector \mathbf{g} from a continuous object $\mathbf{f}(\mathbf{r}, t)$. In many imaging modalities the object is assumed to have spatial, but no temporal variability, however for MR imaging, temporal variability plays a vital role. The system operator which maps the continuous object to discrete data space is represented by \mathcal{H} , and the signal equation is given by

$$\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n}, \quad (1)$$

where \mathbf{n} represents system noise.

An MR system does not image the physical object directly, but rather the magnetic spins that precess within it, excited by the magnetic field. Several physical parameters characterize the behavior of these spins: the proton density $\rho(\mathbf{r}, t)$, the spin-lattice relaxation rate $R_1(\mathbf{r}, t)$, and two spin-spin relaxation rates $R_2(\mathbf{r}, t)$ and $R_2^\dagger(\mathbf{r}, t)$. Thus the 'object' being imaged may be represented by the vector field

$$\mathbf{f}(\mathbf{r}, t) = \{\rho(\mathbf{r}, t), R_1(\mathbf{r}, t), R_2(\mathbf{r}, t), R_2^\dagger(\mathbf{r}, t)\}. \quad (2)$$

The spatio-temporal distribution of the spins,

$$M_0(\mathbf{r}, t) \triangleq M_0[\mathbf{f}(\mathbf{r}, t)], \quad (3)$$

created by a non-linear interaction between the imaging system \mathcal{H} and the object \mathbf{f} , as governed by the Bloch equations [5], creates the signal measured in an MR experiment.

MR data are acquired with an imaging sequence which consists of a series of rf pulses that excite spins into the plane transverse to the main field and magnetic field gradients $\mathbf{G}(t)$ that control what data gets collected when. Referring to (II), a complex-valued element of vector \mathbf{g} may be written as

$$\begin{aligned} \mathbf{g}_{bp} &= \int_{\infty} \mathbf{C}_b(\mathbf{r})M_0(\mathbf{r}, t_p)e^{-2\pi i\mathbf{k}_p \cdot \mathbf{r}} d\mathbf{r} + \mathbf{n}_{bp} \\ &\triangleq \boldsymbol{\mu}_{bp} + \mathbf{n}_{bp}, \end{aligned} \tag{4}$$

where $b = 1 \dots B$ indexes the receive coil in a multi-coil acquisition, $p = 1 \dots P$ indexes Fourier domain frequency, \mathbf{C}_b represents the coil sensitivity, t_p is the time relative to the rf excitation when \mathbf{g}_{bp} is acquired, \mathbf{n}_{bp} is noise and \mathbf{k}_p is the frequency component measured at t_p , controlled by the field gradients, as

$$\mathbf{k}_p = \int_0^{t_p} \mathbf{G}(t') dt'. \tag{5}$$

The thermally-generated noise \mathbf{n} is modeled by a normal distribution with zero mean and equal variance in every component. Assuming no electromagnetic coupling between coils, it may be assumed independent for each element of \mathbf{n} and thus

$$\text{pr}(\mathbf{g}|\mathbf{f}) = \frac{1}{(2\pi\sigma^2)^{BP}} e^{-(\mathbf{g}-\boldsymbol{\mu})^\dagger(\mathbf{g}-\boldsymbol{\mu})/2\sigma^2}. \tag{6}$$

The IO, with its complete knowledge of the object statistics and imaging system, can deal with raw data directly; however, an IO for reconstructed images is useful for examining the information provided to a human observer and measuring the information loss and corresponding drop in optimal performance caused by the reconstruction. For an inverse fast Fourier transform sum-of-squares reconstruction, the magnitude of the j -th image pixel may be represented by

$$\hat{f}_j = \frac{1}{\text{FOV}^2} \sqrt{\sum_b \left| \sum_p \mathbf{g}_{bp} e^{2\pi i \mathbf{k}_p \cdot \mathbf{r}_j} \right|^2}, \tag{7}$$

where FOV is the field of view and \mathbf{r}_j is the location of pixel j . The non-linearity of the reconstruction changes the noise distribution. It can be shown that for a reconstruction involving B coils, \hat{f}_j will follow a non-central chi distribution with $2B$ degrees of freedom [4] given by

$$\text{pr}_{\hat{f}_j|\mathbf{f}}(x) = \frac{\nu_j}{\sigma^2} e^{-(x^2 + \nu_j^2)/2\sigma^2} \left(\frac{x}{\nu_j}\right)^B I_{B-1}\left(\frac{x\nu_j}{\sigma^2}\right), \tag{8}$$

where I_{B-1} is the modified Bessel function of order $B - 1$ and ν_j is the reconstruction of the mean data, given by

$$\nu_j = \frac{1}{\text{FOV}^2} \sqrt{\sum_b \left| \sum_p \boldsymbol{\mu}_{bp} e^{2\pi i \mathbf{k}_p \cdot \mathbf{r}_j} \right|^2}. \tag{9}$$

The distribution of each pixel given \mathbf{f} is independent due to the independence of \mathbf{n} and the orthogonality of the Fourier kernels. Equations (6) and (8) are required to describe the IO.

The Ideal Observer. For a two-class problem every object is considered to be a member of one of two classes, H_1 or H_2 , which in this work correspond to with and without a lesion. The IO decides which is the correct class for a given \mathbf{f} by calculating the raw-data likelihood ratio

$$A^d(\mathbf{g}) = \frac{\text{pr}(\mathbf{g}|H_2)}{\text{pr}(\mathbf{g}|H_1)}, \tag{10}$$

or the reconstructed-image likelihood ratio

$$A^i(\hat{\mathbf{f}}) = \frac{\text{pr}(\hat{\mathbf{f}}|H_2)}{\text{pr}(\hat{\mathbf{f}}|H_1)} \tag{11}$$

and comparing it to a threshold A_{th} , deciding H_2 is the correct class if $A > A_{th}$. By varying the threshold the observer changes the true-positive fraction TPF (the fraction of objects with a lesion that it correctly identifies) and the false-positive fraction FPF (the fraction of objects without a lesion that it incorrectly determines has a lesion). Considering TPF and FPF for all thresholds produces the receiver operating characteristic (ROC) curve. The area under the curve (AUC) is a summary of the ROC curve and can be used as an objective metric for task performance. AUC is often mapped to detectability, defined as

$$d_A = 2\text{erf}^{-1} [2 \cdot \text{AUC}_A - 1]. \tag{12}$$

For exponential-family noise distributions, as in MR imaging, it is convenient to deal with the log-likelihood

$$\lambda = \log A, \tag{13}$$

which is a monotonic transformation of A , and thus does not affect the summary performance metrics. A common alternative figure of merit to AUC or equivalently d_A , is the IO SNR, defined as

$$\text{SNR}_\lambda = \frac{\langle \lambda|H_2 \rangle - \langle \lambda|H_1 \rangle}{\sqrt{\frac{1}{2}\sigma_{H_2}^2 + \frac{1}{2}\sigma_{H_1}^2}}, \tag{14}$$

where $\langle \cdot \rangle$ denotes expectation and $\sigma_{H_j}^2$ represents the variance of λ for class H_j . Typically SNR_λ is easier to compute than AUC since it only requires knowledge of the first two moments of the likelihood ratio. If the data has a Gaussian distribution under both hypotheses, as in raw MR data,

$$d_A = \text{SNR}_\lambda \tag{15}$$

and thus the calculation of AUC is greatly simplified via (12) and (14). Throughout this work the performance using raw data and A^d is compared with performance using reconstructed images and A^i .

The SKE/BKE Task. The signal known exactly/background known exactly (SKE/BKE) task is useful for analyzing the effect of noise and sequence design on detectability. The IO is assumed to have complete a priori knowledge of the object and thus system noise and the data acquisition parameters (timings, which data are acquired, etc.) are the sole determining factors for IO performance. For this task

$$A_{\text{SKE}}^d(\mathbf{g}) = \frac{\prod_{b,p} \text{pr}(\mathbf{g}_{bp}|\mathbf{f}, H_2)}{\prod_{b,p} \text{pr}(\mathbf{g}_{bp}|\mathbf{f}, H_1)} \tag{16}$$

and by inserting the probability density (6) into the likelihood ratio (10) it is easy to show that (14) is equivalent to

$$\text{SNR}_{\lambda, \text{SKE}}^d = \frac{\|\boldsymbol{\mu}_{|H_2} - \boldsymbol{\mu}_{|H_1}\|}{\sigma}, \tag{17}$$

where $\boldsymbol{\mu}_{|H_j}$ is the mean data vector defined in (4) conditioned on class j . Using (15), the inverse of (12) and (17) $\text{AUC}_{\lambda, \text{SKE}}^d$ can be calculated since the noise is Gaussian for raw MR data.

For the non-Gaussian reconstructed images no simple analytical expression exists for $\text{AUC}_{\text{SKE}}^i$ or $\text{SNR}_{\lambda, \text{SKE}}^i$; however if we define conditional pixel SNR values

$$s_{j|H_1} = \frac{\nu_{j|H_1}}{\sigma^i}, \quad s_{j|H_2} = \frac{\nu_{j|H_2}}{\sigma^i}, \tag{18}$$

where $\sigma^i = \sigma\sqrt{P}/\text{FOV}^2$ is the image noise variance (P is the total number of data points acquired per coil), it is possible to derive relatively simple formulae for the first two moments of λ_{SKE}^i . Inserting (9) into (11) and dropping constant terms which do not affect the IO performance metrics, the first and second order moments are

$$\langle \lambda_{\text{SKE}}^i | H_1 \rangle = - \sum_j F_B^1(s_{j|H_1}, s_{j|H_2}) \tag{19}$$

$$\langle \lambda_{\text{SKE}}^i | H_2 \rangle = \sum_j F_B^1(s_{j|H_2}, s_{j|H_1}) \tag{20}$$

$$\text{var}(\lambda_{\text{SKE}}^i | H_1) = \sum_j [F_B^2(s_{j|H_2}, s_{j|H_1}) - F_B^1(s_{j|H_2}, s_{j|H_1})^2] \tag{21}$$

$$\text{var}(\lambda_{\text{SKE}}^i | H_2) = \sum_j [F_B^2(s_{j|H_1}, s_{j|H_2}) - F_B^1(s_{j|H_1}, s_{j|H_2})^2], \tag{22}$$

where F_B^k is the integral function

$$F_B^k(a, b) = \int_0^\infty \left(\log \left[\frac{I_{B-1}(aw)}{I_{B-1}(bw)} \right] \right)^k b \left(\frac{w}{b} \right)^B e^{-(b^2+w^2)/2} I_{B-1}(bw) dw. \tag{23}$$

This 1-D integral does not have an analytic solution, however the integrand is well-behaved for all parameter values and it is amenable to quadrature integration techniques. Using (19) - (22) an estimate of $\text{SNR}_{\lambda, \text{SKE}}^i$ is easily obtained.

Estimates of AUC_{SKE}^i may be obtained from the approximation $d_A^i \approx SNR_{\lambda, SKE}^i$ and (12) or through other approximations [2]. Alternatively, Monte Carlo techniques may be used.

The SKS/BKE Task. The observer will never have complete a priori knowledge of the object and so it is useful to consider detection tasks for which the IO has only statistical knowledge of a random signal. If it is assumed that the signal is parametrized by a random variable θ with known distribution $pr_{\theta}(\theta)$, the distribution of data g conditioned on H_2 is given by marginalizing over θ ,

$$pr(g|H_2) = \int_{\infty} pr(g|H_2, \theta)pr_{\theta}(\theta)d\theta. \tag{24}$$

The distribution of g under H_1 is unaffected by the signal. This allows us to write

$$A_{SKS}^d(g) = \int_{\infty} A_{SKE}^d(g)pr_{\theta}(\theta)d\theta. \tag{25}$$

Equation (25) along with knowledge of $pr_{\theta}(\theta)$ and the derivations given above for the SKE/BKE task allow us to calculate IO performance metrics for the SKS task using Monte Carlo techniques.

3 Experimental Design and Methods

The task of small lesion detection is analyzed. Early detection of primary and metastatic lesions can greatly improve patient outcome, making this an important imaging task. Assuming an ideal slice selection pulse, consider the object consisting of a small circular lesion embedded in a uniform circular background as shown in Fig. 1. This object is imaged using a Cartesian echo-planar (EPI) acquisition sequence which collects all data during a single RF excitation and is controlled by various timing parameters. The relationship between these parameters and the data acquisition procedure are summarized in Fig. 2. The primary

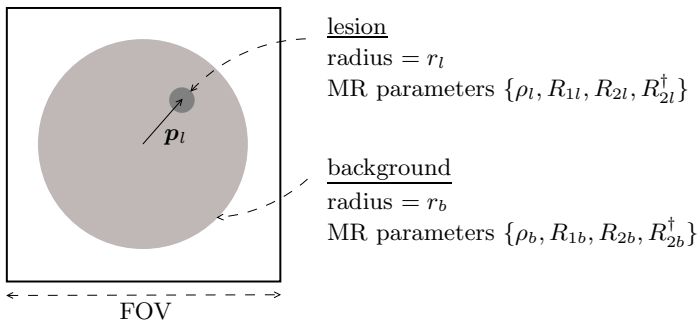


Fig. 1. Small lesion detection phantom

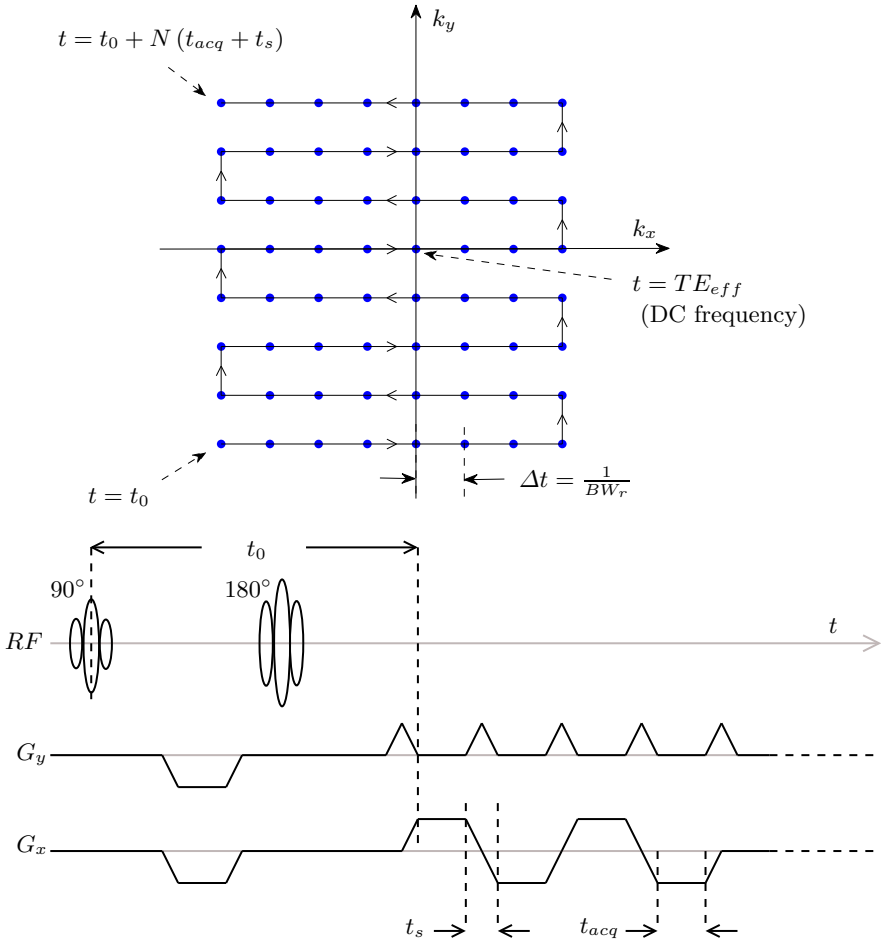


Fig. 2. Cartesian EPI data acquisition scheme (restricted to an 8x8 matrix for illustrative purposes) (top) and corresponding pulse sequence (bottom). Slice selection gradients not depicted. Time t_0 is the delay between RF excitation and the commencement of data collection. The time required to flip gradient polarities is t_s and is proportional to the scanner gradient slew rate. The time to acquire a line of N data points is t_{acq} and is governed by the receiver bandwidth BW_r .

source of contrast in an EPI acquisition is $R_2^* \triangleq R_2 + R_2^\dagger$, which is reflected in the magnetization,

$$M_0(\mathbf{r}, t) = \frac{K\rho(\mathbf{r})\gamma^2\hbar^2 B_0 \sin\alpha}{k_B T} e^{-tR_2^*(\mathbf{r})} \tag{26}$$

$$\triangleq I(\mathbf{r})e^{-tR_2^*(\mathbf{r})}, \tag{27}$$

where K is a scaling constant, γ is the gyromagnetic ratio of hydrogen, \hbar is Planck’s constant, B_0 is the magnetic field strength, k_B is Boltzman’s constant, T is the temperature and α is the flip angle of the rf pulse. It is convenient to define the equilibrium magnetization $I(\mathbf{r})$ as in (27). Note that the time dependence of the components of \mathbf{f} has been dropped, corresponding to a static, non-contrast-enhanced object.

4 Results

The task defined in the previous section is analyzed using the IO defined in Section 2. Assuming $\rho_b = \rho_l$ and spatially invariant coil sensitivities allows for simple analytical expressions for the mean data for both classes. From (17),

$$\text{SNR}_\lambda^d = \frac{Ir_l}{\sigma} C \sqrt{\sum_p |e^{-t_p R_{2l}^*} - e^{-t_p R_{2b}^*}|^2 \frac{|J_1(2\pi r_l |\mathbf{k}_p|)|^2}{|\mathbf{k}_p|^2}}, \tag{28}$$

where $C = \|C\|$ is the magnitude of the coil sensitivity vector, J_1 is the first order Bessel function of the first kind and t_p is when spatial frequency \mathbf{k}_p is measured as defined in Fig. 2. If the receiver bandwidth is high enough to neglect R_2^* decay during the acquisition, we may assume $t_p \approx TE_{eff} \forall t_p$, the time when the DC frequency is measured (see Fig. 2). This assumption permits factoring out the exponential decay terms from the summation to obtain

$$\text{SNR}_\lambda^d \approx \frac{|ICe^{-TE_{eff} R_{2l}^*} - ICe^{-TE_{eff} R_{2b}^*}|}{\sigma} \cdot r_l \cdot \sqrt{\sum_p \frac{|J_1(2\pi r_l |\mathbf{k}_p|)|^2}{|\mathbf{k}_p|^2}}, \tag{29}$$

where the first term is approximately the contrast-to-noise (CNR) of the lesion in the reconstructed image. This leads to the relationship

$$\text{SNR}_\lambda^d \propto \text{CNR} \cdot \sqrt{\text{Lesion Area}} \cdot \sqrt{\sum_p \frac{|J_1(2\pi r_l |\mathbf{k}_p|)|^2}{|\mathbf{k}_p|^2}}. \tag{30}$$

Thus the IO SNR for lesion detection is approximately proportional to the CNR measured in the image domain, the square root of lesion cross-sectional area in the image plane and a term reflecting the relationship between the lesion and the data collection scheme.

Figure 3 compares $\text{AUC}_{\text{SKE}}^d$ for three acquisition sequences with differing receive bandwidths (16, 64 and 256 kHz) but equal scan times via contrast-detail diagrams. Higher receiver bandwidth increases the noise level in the acquired data; however, all three sequences are designed to require the same amount of data collection time, thus the 64 kHz acquisition can achieve double the resolution of the 16 kHz acquisition. Similarly, the 256 kHz acquisition has double the resolution of the 64 kHz acquisition. We can clearly see from Fig. 3 that

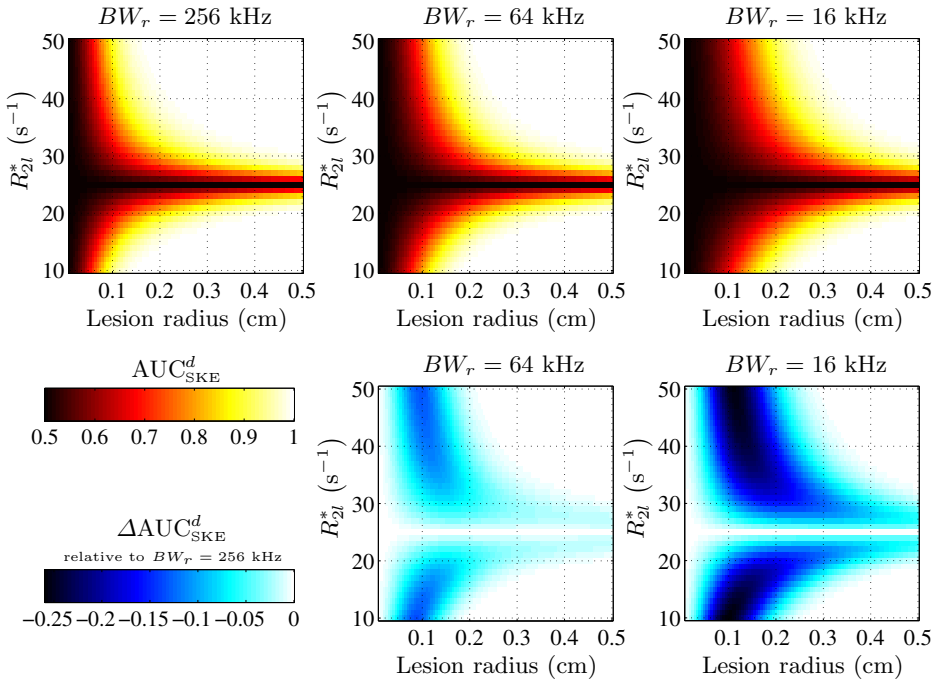


Fig. 3. Comparison of lesion detectability as measured by AUC_{SKE}^d for different choices of BW_r (noise level) and resolution for fixed scan time. Left-most column: $BW_r = 256$ kHz, 128×128 data set (pixel size = 0.1875 cm). Middle column: $BW_r = 64$ kHz, 64×64 data set (pixel size = 0.375 cm). Bottom plot represents drop in AUC_{SKE}^d relative to the 256 kHz case. Right-most column: $BW_r = 16$ kHz, 32×32 data set (pixel size = 0.75 cm). Noise level such that pixel SNR = 20 @ $BW_r = 16$ kHz. Background relaxation rate $R_{2b}^* = 25 \text{ s}^{-1}$.

for fixed scan time, resolution is more vital than noise level, even for very low contrast signals, with the high resolution/high noise sequence giving superior AUC_{SKE}^d for all combinations of lesion size and contrast. A drop of up to 0.25 in AUC_{SKE}^d was observed comparing the 16 kHz to the 256 kHz sequence and 0.1 comparing the 64 kHz to the 256 kHz sequence (lower plots of Fig. 3), primarily for lesions with radius ≈ 1 mm, with smaller lesions generally not detectable by any sequence ($AUC \approx 0.5$) and larger lesions detectable by all sequences ($AUC \approx 1$). As R_{2l}^* approaches the background value of 25 s^{-1} , the contrast between background and lesion is eliminated and AUC drops to 0.5, irrespective of lesion size. The ability of the IO to detect low-contrast lesions smaller than those a radiologists can detect in vivo is due in large part to the simplicity of the task. Everything about the lesion and background is known and the major limiting factor is system noise. We will show that if the task is made more realistic, the IO performance drops significantly.

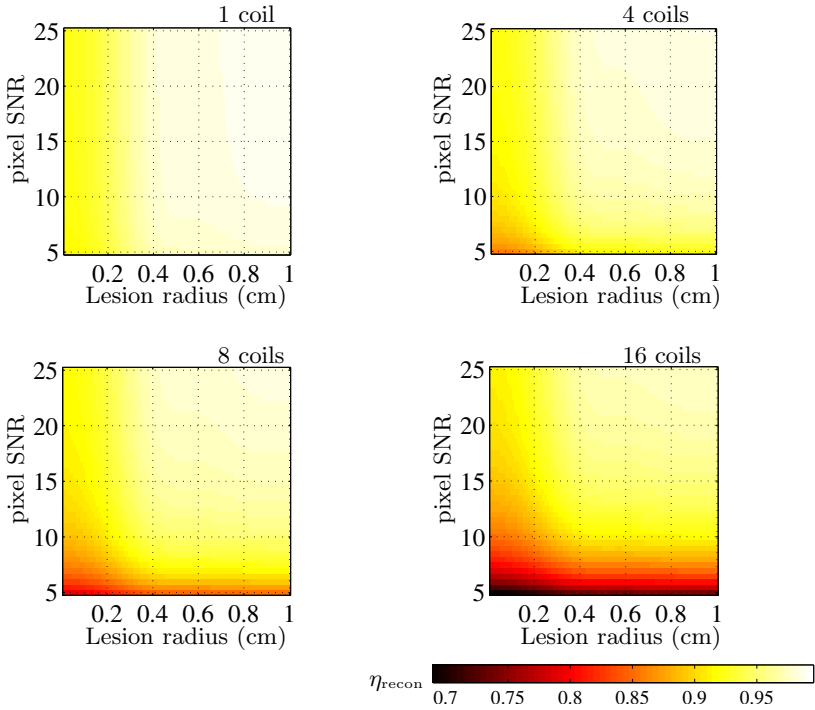


Fig. 4. Reconstruction efficiency η_{recon} as determined by (31) for 1, 4, 8 and 16 receive coils, demonstrating a drop in efficiency for lower pixel SNR and higher number of coil elements. $BW_r = 64$ kHz, 64×64 data set per coil, $R_{2l}^* = 23 \text{ s}^{-1}$ and $R_{2b}^* = 25 \text{ s}^{-1}$.

Figure 4 compares the IO performance for reconstructed images using a reconstruction efficiency score [10] defined as

$$\eta_{\text{recon}} = \frac{(\text{SNR}_{\lambda, \text{SKE}}^i)^2}{(\text{SNR}_{\lambda, \text{SKE}}^d)^2}. \tag{31}$$

Since the IO makes optimal use of all available information, a reconstruction algorithm can never increase the IO performance and thus $0 \leq \eta_{\text{recon}} \leq 1$. We observe a drop in efficiency as the number of coils increases, particularly for higher noise levels. This does not imply that a single-channel receive coil is superior to multi-channel receive coils, but does indicate that some information is lost when combining data from multiple coils to produce a human-interpretable magnitude image. This corroborates the work of Larsson, et. al [6] who studied sum-of-squares reconstructions from multiple-channel coils in terms of maximum-likelihood estimation.

To study the affect of object variability on IO performance we considered the SKS task where lesion location is unknown. The IO does not know the lesion location \mathbf{p}_l a priori, but knows it can occur anywhere in the background support with equal probability. Equation (25) was used to produce samples of $\mathcal{L}_{\text{SKS}}^d$ for

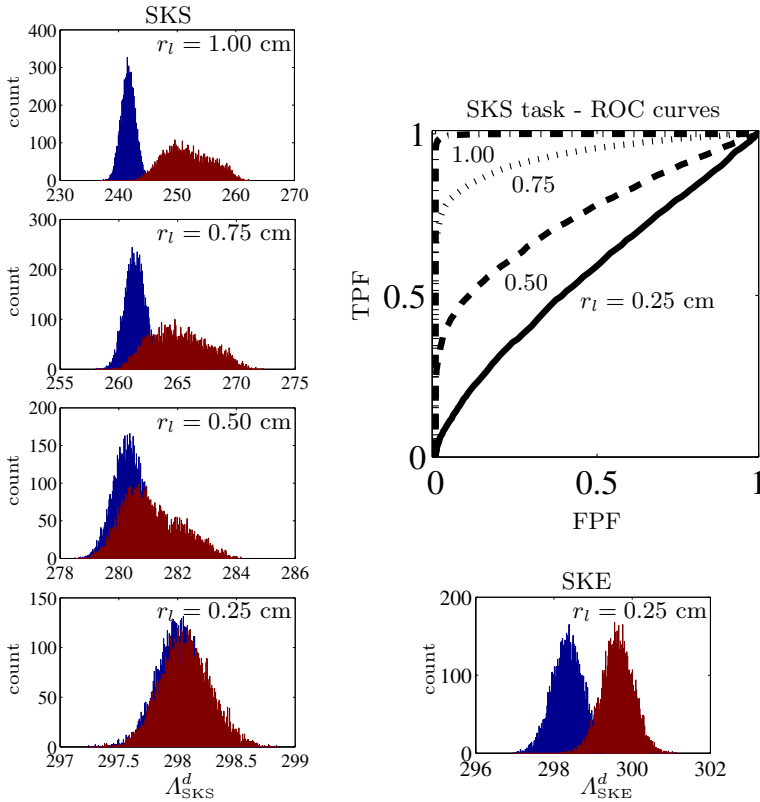


Fig. 5. Histograms of the likelihood ratio for classes H_1 (blue) and H_2 (red) when lesion location is unknown for various lesion sizes (left column) and corresponding SKE histograms for the smallest lesion size (lower right). SKS ROC curves are given (upper-right). SKE curves are not displayed as $AUC_{\text{SKE}}^d \approx 1$. Other imaging parameters: $BW_r = 64$ kHz, $R_{2l}^* = 20$ s $^{-1}$, $R_{2b}^* = 25$ s $^{-1}$, pixel SNR = 20, 64×64 data set, lesion centered at the origin for SKE task.

vectors \mathbf{g} drawn from the H_1 and H_2 data distributions, with 7200 lesion-present and 7200 lesion-absent data vectors per imaging task. Figure 5 gives examples of likelihood ratio histograms and corresponding empirical ROC curves. The increased difficulty of the SKS task is reflected in significantly lower IO performance. For $r_l = 0.25$ cm, $AUC_{\text{SKE}}^d = 0.98$, while $AUC_{\text{SKS}}^d = 0.57$, indicating that if the IO knows the location a priori, it can almost always correctly detect the lesion; if it does not know the location, it does barely better than chance for this size lesion. We note the non-Gaussian distribution of $\text{pr}(\Lambda|H_2)$ for the SKS task, indicating that $\text{SNR}_{\lambda, \text{SKS}}^d$ would not be a good surrogate figure of merit for AUC.

5 Conclusions

This paper describes the IO for MR classification tasks and demonstrates its use by assessing image quality for a lesion detection task. The presented image

quality metrics provide direct task-specific information about MR data and image quality which is not provided by pixel SNR or CNR.

The raw-data IO performance, AUC^d or SNR_{λ}^d is useful for measuring the quality of the acquisition sequence, i.e., the suitability of the data to the required task. The reconstruction efficiency η_{recon} provides a measure of the decrease in IO performance caused by the reconstruction algorithm. The IO does not require a reconstruction step, however human readers do, and the reconstruction efficiency can play a role in evaluating reconstruction algorithms along with human or model-human reader studies. The reconstruction-based IO performance AUC^i , or where appropriate SNR_{λ}^i , provides an upper bound on human reader performance and can be used for overall system optimization, incorporating both sequence design and image reconstruction.

Technical developments in MR imaging are progressing rapidly with recent advances in sequence design, non-Cartesian acquisitions, parallel imaging and under-sampling techniques, iterative reconstructions involving compressed-sensing algorithms, etc., all of which challenge the validity of traditional image quality metrics. The task-based assessment framework provides a way to move towards more meaningful image quality metrics that can quantitatively determine the efficacy of emerging MR techniques. To achieve this goal, this initial work will be extended via Monte Carlo simulation to include more modern data acquisition and reconstruction techniques, as well as more realistic object models with more variability and more sophisticated imaging tasks.

References

1. Barrett, H.H., Myers, K.J.: Foundations of Image Science. Wiley, Chichester (2004)
2. Clarkson, E.W., Barrett, H.H.: Approximations to ideal-observer performance on signal-detection tasks. *Applied Optics* 39, 1783–1793 (2000)
3. Constable, R.T., Skudlarski, P., Gore, J.C.: An ROC approach for evaluating functional brain MR imaging and postprocessing protocols. *Magn. Reson. Med.* 34, 57–64 (1995)
4. Constantinides, C., Atalar, E., McVeigh, E.R.: Signal-to-noise measurements in magnitude images from NMR based arrays. *Magn. Reson. Med.* 38, 852–857 (1997)
5. Haacke, E.M., Brown, R.W., Thompson, M.R., Venkatesan, R.: *Magnetic Resonance Imaging: Principles and Sequence Design*. Wiley-Liss (1999)
6. Larsson, E.G.: SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging. *J. Magn. Reson.* 163, 121–123 (2003)
7. Tapiovaara, M.J., Wagner, R.F.: SNR and noise measurements for medical imaging i. *Phys. Med. Biol.* 38, 71–92 (1993)
8. Tisdall, M.D., Atkins, M.S.: Using human and model performance to compare MRI reconstructions. *IEEE Trans. Med. Imag.* 25, 1510–1517 (2006)
9. Van Trees, H.L.: *Detection, Estimation and Modulation Theory*. Wiley, Chichester (1968)
10. Wagner, R.F., Brown, D.G.: Unified SNR analysis of medical imaging systems. *Phys. Med. Biol.* 30, 489–518 (1985)

3D Shape Analysis for Early Diagnosis of Malignant Lung Nodules

Ayman El-Baz^{1,*}, Matthew Nitzken¹, Fahmi Khalifa¹, Ahmed Elnakib¹,
Georgy Gimel'farb², Robert Falk³, and Mohammed Abo El-Ghar⁴

¹ Bioimaging Laboratory, Bioengineering Department, University of Louisville,
Louisville, KY, USA
Tel.: (502)-852-5092
aselba01@louisville.edu

² Department of Computer Science, University of Auckland, Auckland, New Zealand

³ Department of Radiology, Jewish Hospital, Louisville, KY, USA

⁴ Urology and Nephrology Department, University of Mansoura, Mansoura, Egypt

Abstract. An alternative method for diagnosing malignant lung nodules by their shape rather than conventional growth rate is proposed. The 3D surfaces of the detected lung nodules are delineated by spherical harmonic analysis, which represents a 3D surface of the lung nodule supported by the unit sphere with a linear combination of special basis functions, called spherical harmonics (SHs). The proposed 3D shape analysis is carried out in five steps: (*i*) 3D lung nodule segmentation with a deformable 3D boundary controlled by two probabilistic visual appearance models (the learned prior and the estimated current appearance one); (*ii*) 3D Delaunay triangulation to construct a 3D mesh model of the segmented lung nodule surface; (*iii*) mapping this model to the unit sphere; (*iv*) computing the SHs for the surface, and (*v*) determining the number of the SHs to delineate the lung nodule. We describe the lung nodule shape complexity with a new shape index, the estimated number of the SHs, and use it for the K -nearest classification to distinguish malignant and benign lung nodules. Preliminary experiments on 327 lung nodules (153 malignant and 174 benign) resulted in the 93.6% correct classification (for the 95% confidence interval), showing that the proposed method is a promising supplement to current technologies for the early diagnosis of lung cancer.

Keywords: Shape, Spherical Harmonics, Lung, Diagnosis, Nodules.

1 Introduction

Pulmonary nodules are the most common manifestation of lung cancer and are the principal cause of cancer-related deaths [16]. Fast and accurate classification of the nodules is of major importance for medical computer-aided diagnostic systems (CAD). A nodule is an approximately spherical volume of higher-density

* Corresponding author.

tissue visible in an X-ray lung image. Large malignant nodules (generally defined as greater than 1 cm in diameter) are easily detected with any traditional imaging equipment and are then diagnosed by needle biopsy or bronchoscopy. However, diagnostic options for small malignant nodules are limited due to difficulties in accessing them, especially if they are located deep in the tissue or away from the large airways. Therefore, additional imaging and CAD techniques are needed. The popular direction of detecting small cancerous nodules is to analyze their growth rate over time. This paper introduces a new approach to characterize detected nodules based on their shape.

Related work: A great deal of work has been published regarding the usefulness of morphologic features for discriminating malignant from benign pulmonary nodules on computed tomography (CT) and to a lesser extent, chest radiographs. Several studies have shown a correlation between different nodule shape characteristics and underlying pathology. For example, Furuya et al. [7] analyzed the margin characteristics of 193 pulmonary nodules on high-resolution CT and subjectively classified them as one of several types, including round, lobulated, densely spiculated, ragged, and halo. They found a high level of malignancy among the lobulated (82%), spiculated (97%), ragged (93%), and halo nodules (100%), while 66% of the round nodules proved to be benign.

Automatically extracted features have also been shown to correlate with underlying malignancy. Kawata et al. [10] quantified the surface curvature and degree of surrounding radiating pattern in biopsy-proven benign and malignant nodules, and compared the resulting feature maps. Their results showed good separation of the feature maps between the two categories. Similarly, fractal analysis has been used to quantify the nodule margin characteristics of benign and malignant nodules. Kido et al. [12] used 2D and 3D fractal dimensions to analyze the lung-nodule interface in a series of 117 peripheral pulmonary nodules of various underlying pathology including benign hamartomas, tuberculomas, and pneumonias as well as malignant diagnoses including bronchogenic carcinomas. They noted statistically significant differences between the 2D fractal dimensions of hamartomas and all other nodules, as well as differences between the 3D fractal dimensions pneumonias and tuberculomas and bronchogenic carcinomas. Although none of these studies directly assessed the accuracy of their methods in predicting a diagnosis, they support the notion that nodule shape can potentially be used by automated systems to distinguish benign from malignant nodules.

Several groups have designed CAD systems with the goal of predicting a diagnosis based on features extracted from CT or chest radiographs. In general, they share common schema, first extracting features from the images, then designing and using an automatic classifier to categorize nodules based on these features, and lastly evaluating the performance of the system with receiver operating characteristics (ROC) analysis. The CAD systems differ in the specific extracted features and the type of classifier used, with linear discriminant classifiers (LDC) and neural networks (NN) being the most common. Below, LDCs-based classifier systems will be discussed followed by NN-based classifier systems.

Kawata and colleagues [11] designed a CT-based CAD system that classified pulmonary nodules based on a combination of their curvature index and the relationship of the nodules to their surrounding features. Mori et al. [13] also designed a CAD system using a curvedness index in combination with dynamic contrast-enhanced CT in order to evaluate temporal change as a possible discriminating feature of benign and malignant nodules.

One of the early neural network based CAD systems was developed by Gurney and Swensen [8]. They compared two systems, one using a neural network based classifier and one using a Bayesian classifier. Both systems used a combination of subjectively evaluated clinical and radiological characteristics including border smoothness, spiculation and lobulation.

In summary, the aforementioned existing approaches show the following limitations:

- Most of them classify the lung nodules based on extracted 2D features (e.g., round, lobulated, ragged, and halo, etc.) and they did not take into account the 3D features of lung nodules.
- Most of them did not provide a quantitative measure that has the ability to describe the shape complexity of detected lung nodules.
- Most of the existing features (e.g., curvature, round, etc.) depend on the accuracy of the used nodule segmentation algorithm which make them are difficult for clinical practitioners to use.

This work aims to address these limitations in a way that will make evaluating small lung masses more consistent.

2 Methods

In this work, we propose a novel shape-based approach for the analysis of lung nodules variability between malignant and benign nodules (see Fig. 1). *In this paper we will focus on the steps from 2 to 5 and the first step is shown in detail in [4].*

2.1 Lung Nodules Segmentation

Accurate lung nodules segmentation from 3D LDCT images is a challenging problem because the intensities of the lung nodules and surrounding tissues (e.g., blood vessels, chest, etc.) are not clearly distinguishable. To overcome this problem, we follow our approach introduced in [6], which depends on using a conventional 3D parametric deformable boundary [9], but control its evolution with two probabilistic visual appearance models, namely, a learned lung nodule appearance prior and a current appearance model of the image to be segmented. The prior is a 3D Markov-Gibbs random field (MGRF) model of the lung nodules' intensities with translation- and rotation-invariant pairwise voxel interaction, **being learned analytically based on developing a new maximum likelihood estimator from the training data.** The current appearance is modeled

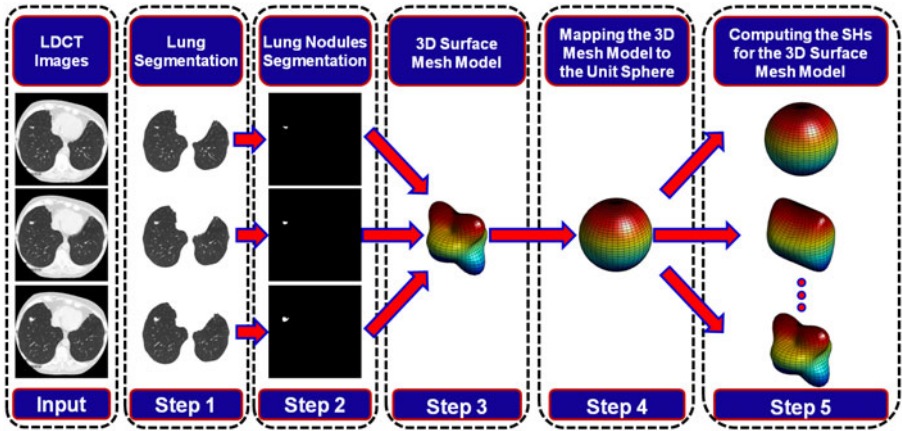


Fig. 1. Proposed shape-based system for early diagnosis of malignant nodules

by a mixed marginal distribution of the voxel intensities in both the lung nodule and surrounding tissues. To extract the voxel-wise model of the current nodule appearance, the mixture is precisely approximated with a linear combination of discrete Gaussians (LCDG) [3] and automatically separated into the lung nodule and background LCDG models. Let (x, y, z) be Cartesian 3D point coordinates. A conventional parametric deformable surface, $\mathbf{B}(\mathbf{P}_1, \dots, \mathbf{P}_K)$, specified by K control vertices $\mathbf{P}_k = (x_k, y_k, z_k)$, evolves in the directions that minimize its energy, E , depending on internal, $\zeta_{\text{int}}(\mathbf{B})$, and external, $\zeta_{\text{ext}}(\mathbf{B})$, forces [9]:

$$E = E_{\text{int}} + E_{\text{ext}} \equiv \int_{\mathbf{B}} (\zeta_{\text{int}}(\mathbf{B}) + \zeta_{\text{ext}}(\mathbf{B})) d\mathbf{B} \quad (1)$$

In this paper, we introduce a new type of external energy that depends on both the learned prior and the current (on-going) appearance model. Let $\mathbf{Q} = \{0, 1, \dots, Q-1\}$ and $\mathbf{L} = \{\text{nl}, \text{bg}\}$ be finite sets of image intensities (gray values) and region labels, respectively. Let a finite 3D arithmetic lattice $\mathbf{R} = [(x, y, z) : x = 0, \dots, X-1; y = 0, \dots, Y-1, z = 1, \dots, Z-1]$ support a 3D image $\mathbf{g} : \mathbf{R} \rightarrow \mathbf{Q}$ and its region map $\mathbf{m} : \mathbf{R} \rightarrow \mathbf{L}$. The label, $m_{x,y,z}$, associates the voxel, $g_{x,y,z}$, with the lung nodule or the background. To reduce the impact of global contrast and offset deviations of intensities due to different sensors, each input 3D image is normalized by mapping its signal range $[q_{\min}, q_{\max}]$ to the maximal range of $[0, 255]$.

To consider the normalized images as samples of a prior MGRF model but exclude any image alignment before the segmentation, we use a generic translation- and rotation-invariant MGRF with only voxel-wise and central-symmetric pairwise voxel interaction. The latter is specified by a set \mathbf{N} of characteristic central-symmetric voxel neighborhoods $\{\mathbf{n}_\nu : \nu \in \mathbf{N}\}$ on \mathbf{R} and a corresponding set \mathbf{V} of Gibbs potentials, one per neighborhood. A central-symmetric neighborhood \mathbf{n}_ν embraces all voxel pairs such that the (x, y, z) -coordinate offsets

between any voxel (x, y, z) and its neighbor (x', y', z') belong to an indexed semi-open interval $[d_{\nu, \min}, d_{\nu, \max})$; $\nu \in \mathbf{N} \subset \{1, 2, 3, \dots\}$ of the inter-voxel distances: $d_{\nu, \min} \leq \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} < d_{\nu, \max}$.

Learning the appearance prior. Let $\mathbf{S} = \{(g_t, m_t) : t = 1, \dots, T\}$ be a training set of 3D images with known region maps. Let $\mathbf{R}_t = \{(x, y, z) : (x, y, z) \in \mathbf{R} \wedge m_{t;x,y,z} = \text{nl}\}$ denote the part of \mathbf{R} supporting lung nodule in the t -th training pair (g_t, m_t) ; $t = 1, \dots, T$. Let $\mathbf{C}_{\nu,t}$ be a family of voxel pairs in \mathbf{R}_t^2 with the co-ordinate offset $(\xi, \eta, \gamma) \in \mathbf{n}_\nu$ in a particular neighborhood. Let $\mathbf{F}_{\text{vx},t}$ and $\mathbf{F}_{\nu,t}$ be empirical marginal probability distributions of voxel intensities and of intensity co-occurrences, respectively, in the training lung nodule from \mathbf{g}_t : $\mathbf{F}_{\text{vx},t} = [f_{\text{vx},t}(q) = \frac{|\mathbf{R}_{t,q}|}{|\mathbf{R}_t|} : q \in \mathbf{Q}]$ and $\mathbf{F}_{\nu,t} = [f_{\nu,t}(q, q') = \frac{|\mathbf{C}_{\nu,t;q,q'}|}{|\mathbf{C}_{\nu,t}|} : (q, q') \in \mathbf{Q}^2]$ where $\mathbf{R}_{t,q} = \{(x, y, z) : (x, y, z) \in \mathbf{R}_t \wedge g_{x,y,z} = q\}$ is a subset of voxels supporting the intensity q and $\mathbf{C}_{\nu,t;q,q'}$ is a subset of the voxel pairs $\mathbf{c}_{\xi,\eta,\gamma}(x, y, z) = ((x, y, z), (x + \xi, y + \eta, z + \gamma)) \in \mathbf{R}_t^2$ supporting the intensity co-occurrence (q, q') in the training lung nodule from \mathbf{g}_t . Let $\mathbf{V}_{\text{vx}} = [V_{\text{vx}}(q) : q \in \mathbf{Q}]$ be a potential function of voxel intensities that describes the voxel-wise interaction. Let $\mathbf{V}_\nu = [V_\nu(q, q') : (q, q') \in \mathbf{Q}^2]$ be a potential function of intensity co-occurrences in the neighboring voxel pairs that describes the pairwise interaction in the neighborhood \mathbf{n}_ν ; $\nu \in \mathbf{N}$. The MGRF model of the t -th training pair is specified by the joint Gibbs probability distribution on the sublattice \mathbf{R}_t :

$$P_t = \frac{1}{Z_t} \exp(|\mathbf{R}_t| (\mathbf{V}_{\text{vx}}^\top \mathbf{F}_{\text{vx},t} + \sum_{\nu \in \mathbf{N}} \rho_{\nu,t} \mathbf{V}_{\nu,t}^\top \mathbf{F}_{\nu,t})) \tag{2}$$

where $\rho_{\nu,t} = |\mathbf{C}_{\nu,t}|/|\mathbf{R}_t|$ is the average cardinality of \mathbf{n}_ν with respect to \mathbf{R}_t .

To identify the MGRF model in Eq. (2), the Gibbs potentials are approximated analytically:

$$V_{\text{vx},\text{nl}}(q) = \log f_{\text{vx},\text{nl}}(q) - \frac{1}{Q} \sum_{\kappa \in \mathbf{Q}} \log f_{\text{vx},\text{nl}}(\kappa) \quad \text{for } q \in \mathbf{Q}; \text{ and} \tag{3}$$

$$V_{\nu,\text{nl}}(q, q') = \lambda \rho_\nu (f_{\nu,\text{nl}}(q, q') - f_{\text{vx},\text{nl}}(q) f_{\text{vx},\text{nl}}(q')) \quad \text{for } (q, q') \in \mathbf{Q}^2 \tag{4}$$

where the common factor λ is also computed analytically.

Modeling the current appearance with LCDG. Non-linear intensity variations in a data acquisition system due to scanner type and scanning parameters affect visual appearance of lung nodules in each data set \mathbf{g} to be segmented. Thus in addition to the learned appearance prior, an on-going lung nodule appearance within a current position of the evolving boundary \mathbf{B} in \mathbf{g} is modeled with its marginal intensity distribution. The whole marginal distribution of the voxel intensities within the boundary is considered as a dynamic mixture of two probability models that characterize the lung nodule and its background, respectively. The mixture is partitioned into these two LCDG models by using the EM-based approach detailed in [3].

Boundary evolution under the appearance models. Let $p_{\text{vx},\text{nl}}(q)$ be the marginal probability of the intensity q in the estimated current LCDG model

for the lung nodule. To guide the boundary evolution, we combine in the external energy term of Eq. (10), both the learned prior and the on-going nodule appearance model as follows:

$$\zeta_{\text{ext}}(\mathbf{P} = (x, y, z)) = -p_{\text{vx,nl}}(g_{x,y,z})\pi_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) \tag{5}$$

Here, $\pi_{\mathbf{P}}(q|\mathbf{S})$ is the prior conditional probability of q , given the fixed current intensities in the characteristic central-symmetric neighborhood of \mathbf{P} for the MGRF prior model of Eq. (2):

$$\pi_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) = \exp(E_{\mathbf{P}}(g_{x,y,z}|\mathbf{S})) / \sum_{q \in \mathbf{Q}} \exp(E_{\mathbf{P}}(q|\mathbf{S}))$$

where $E_{\mathbf{P}}(q|\mathbf{S})$ is the conditional Gibbs energy of pairwise interaction for the voxel \mathbf{P} provided that an intensity q is assigned to the lung nodule while the other current intensities in all its neighboring voxels over the characteristic neighborhoods \mathbf{n}_{ν} ; $\nu \in \mathbf{N}$, remains fixed:

$$E_{\mathbf{P}}(q|\mathbf{S}) = V_{\text{vx,nl}}(q) + \sum_{\nu \in \mathbf{N}} \sum_{(\xi, \eta, \gamma) \in \mathbf{n}_{\nu}} (V_{\nu, \text{nl}}(g_{x-\xi, y-\eta, z-\gamma}, q) + V_{\nu, \text{nl}}(q, g_{x+\xi, y+\eta, z+\gamma}))$$

After the changes in the total energy, $E_{\mathbf{B}}$, of the 3D region $\mathbf{R}_{\mathbf{B}} \subset \mathbf{R}$ inside the evolving boundary \mathbf{B} :

$$E_{\mathbf{B}} = \sum_{\forall \mathbf{P}=(x,y,z) \in \mathbf{R}_{\mathbf{B}}} E_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) \tag{6}$$

stop, the evolution terminates.

2.2 Spherical Harmonics (SHs) Shape Analysis

Spectral SH analysis [1] considers 3D surface data as a linear combination of specific basis functions. In our case, the surface of the segmented lung nodule is approximated first by a triangulated 3D mesh (see Fig. 2) built with an algorithm by Fang and Boas [5]. Secondly, the lung nodule surface for each subject is mapped for the SH decomposition to the unit sphere. We propose a novel mapping approach, called “Attraction-Repulsion” that calls for all mesh nodes to meet two conditions: (i) the unit distance of each node from the lung nodule center as shown in Fig. 3, and (ii) an equal distance of each node from all of its nearest neighbors as shown in Fig. 4.

To detail our Attraction-Repulsion Algorithm (see its summary in Algorithm 1), let τ denote the iteration index, I be the total number of the mesh nodes (in all the experiments below $I = 4896$ nodes), and $\mathbf{P}_{\tau,i}$ be the Cartesian coordinates of the surface node i at iteration τ ; $i = 1, \dots, I$. Let J be the number of the neighbors for a mesh node (see e.g. Fig. 4) and $d_{\tau,ij}$ denote the Euclidean distance between the surface nodes i and j at iteration τ (as shown in Fig. 4(b)), where $i = 1, \dots, I$ and $j = 1, \dots, J$. Let $\Delta_{\tau,ji} = \mathbf{P}_{\tau,j} - \mathbf{P}_{\tau,i}$ denote the displacement between the nodes j and i at iteration τ . Let $C_{A,1}, C_{A,2}, C_R$ be the attraction and repulsion constants, respectively, that control the displacement of each surface node.



Fig. 2. Generating a 3D mesh for the lung nodule surface from a stack of successive segmented 2D LDCT slices

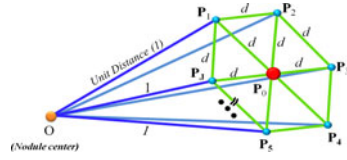


Fig. 3. 3D illustration of the unit distance from all surface nodes to the center of the lung nodule

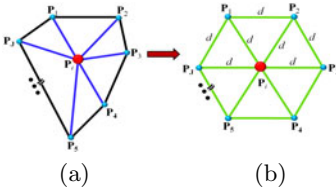


Fig. 4. 2D illustration of the neighbors rearrangement: initial (a) vs. final equidistant locations (b) in all the directions.

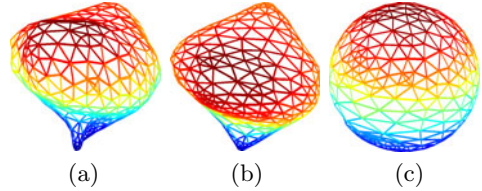


Fig. 5. Lung nodule mesh (a), its smoothed version (b), and the Attraction-Repulsion mapping to the unit sphere (c)

Algorithm 1: Attraction-Repulsion Algorithm

Initialization

- Construct the 3D lung nodule mesh (Fig. 5.a).
- Smooth it by the Laplacian filtering (Fig. 5.b).
- Initialize the mapping of the smoothed mesh to the unit sphere.

Repeat

- **For** $i = 1 \rightarrow I$
 - **Attraction:**
 - * Select a node to process.
 - * Update the node using Eq. (7).
 - **Repulsion:**
 - * Update the node using Eq. (8).
- **End** (all nodes in the mesh are shifted and back-projected onto the unit sphere).

While changes in the node positions occur (Fig. 5.c).

The starting attraction step of the proposed mapping tends to center each node P_i ; $i = 1, \dots, I$, with respect to its neighbors by iteratively adjusting its location:

$$P'_{\tau,i} = P_{\tau,i} + C_{A,1} \sum_{j=1; j \neq i}^J \Delta_{\tau,ji} d_{\tau,ji}^2 + C_{A,2} \frac{\Delta_{\tau,ji}}{d_{\tau,ji}} \tag{7}$$

where the factor $C_{A,2}$ keeps the tightly packed nodes from collision and also pushes the adjusted nodes away from their neighbors if a certain neighbor is much closer than the others.

The subsequent repulsion step inflates the whole mesh by pushing all the nodes outward to become evenly spaced after their final back-projection onto the unit sphere along the rays from the center of the sphere. To ensure that the nodes that have not been shifted will not collide with the altered node, the location of each node $\mathbf{P}_i; i = 1, \dots, I$, is updated before the back-projection as follows:

$$\mathbf{P}_{\tau+1,i}^\circ = \mathbf{P}'_{\tau,i} + \frac{C_R}{2I} \sum_{j=1; j \neq i}^I \left(\frac{\Delta_{\tau,ji}}{|\Delta_{\tau,ji}|^2} \right) \tag{8}$$

where a repulsion constant C_R controls the displacement of each surface node and establishes a balance between the processing time and accuracy (e.g. a smaller C_R values guarantees that the node faces will not become crossed during the iterations at the expense of the increased processing time). All the experiments below were obtained with $0.3 \leq C_R \leq 0.7$.

The original lung nodule mapped to the unit sphere with the proposed Attraction-Repulsion algorithm is approximated by a linear combination of SHs, the lower-order harmonics being sufficient to represent more generic information, while the finer details requiring the higher - order ones. The SHs are generated by the solving an isotropic heat equation for the nodule surface on the unit sphere. Let $\mathcal{S} : \mathbf{M} \rightarrow \mathbf{U}$ denote the mapping of a nodule mesh \mathbf{M} to the unit sphere \mathbf{U} . Each node $\mathbf{P} = (x, y, z) \in \mathbf{M}$ mapped to the spherical position $\mathbf{u} = S(\mathbf{P})$ is represented by the spherical coordinate $\mathbf{u} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta)$; where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$ are the polar and azimuth angles, respectively. The SH $Y_{\alpha\beta}$ of degree α and order β is defined as [2]:

$$Y_{\alpha\beta} = \begin{cases} c_{\alpha\beta} G_\alpha^{|\beta|} \cos \theta \sin(|\beta|\varphi) & -\alpha \leq \beta \leq -1 \\ \frac{c_{\alpha\beta}}{\sqrt{2}} G_\alpha^{|\beta|} \cos \theta & \beta = 0 \\ c_{\alpha\beta} G_\alpha^{|\beta|} \cos \theta \cos(|\beta|\varphi) & 1 \leq \beta \leq \alpha \end{cases} \tag{9}$$

where $c_{\alpha\beta} = \left(\frac{2\alpha+1}{2\pi} \frac{(\alpha-|\beta|)!}{(\alpha+|\beta|)!} \right)^{\frac{1}{2}}$ and $G_\alpha^{|\beta|}$ is the associated Legendre polynomial of degree α and order β . For the fixed α , the polynomials G_α^β are orthogonal over the range $[-1, 1]$. As shown in [2], the Legendre polynomials are effective in calculating SHs, and this is the main motivation behind their use in this work.

Finally, the lung nodule is reconstructed from the SHs of Eq. (9). In the case of the SHs expansion, the standard least-square fitting does not accurately model the 3D shape of the lung nodule and can miss some of the shape details that discriminate between the malignant and benign lung nodules. To circumvent this problem, we used the iterative residual fitting by Shen and Chung [15] that accurately approximates the 3D shape of malignant and benign lung nodules. As demonstrated in Fig. 6, the model accuracy does not significantly change for the benign nodule from the 15 to 60 SHs, while it continues to increase for the malignant nodule.

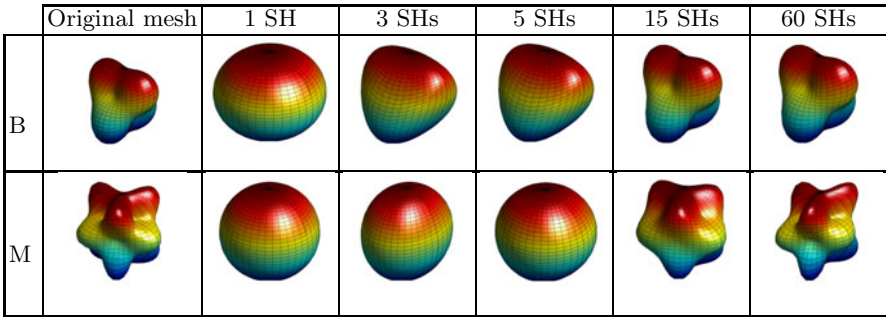


Fig. 6. 3D shape approximation for malignant (M) and benign (B) lung nodules

2.3 Quantitative Lung Nodule Shape Analysis

Our main hypothesis is that the shape of malignant nodules is more complicated (e.g., with spiculation) if it is compared with the shape of benign nodules which is simpler (smoothed shape) as shown in Fig. 6, so that more SHs have to be used for accurate approximation of the shape of the malignant lung nodule. Therefore, the number of the SHs after which there are no significant changes in the approximations can be used as a new shape index quantifying the shape complexity of the detected lung nodules. Due to the unit sphere mapping, the original mesh for each nodule is inherently aligned with the mesh for the approximate shape, and the sum of the Euclidean distances between the corresponding nodes gives the total error between both the mesh models. As shown in Fig. 7, the total error curves for the increasing number \mathcal{K} of the SHs can be statistically analyzed to differentiate between the subjects.

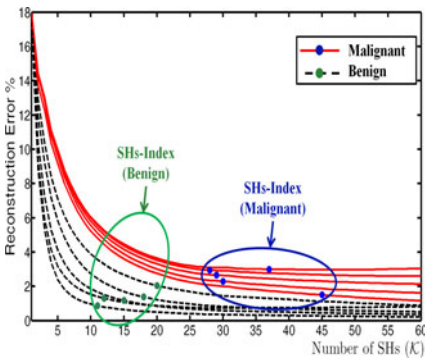


Fig. 7. Estimation of the shape index from the total nodule approximation error for malignant and benign nodules

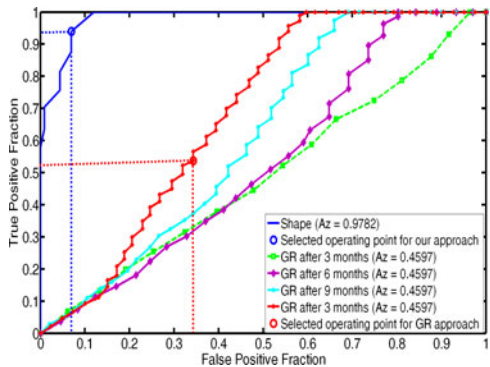


Fig. 8. The ROC curves for the proposed approach and the growth rate based diagnostic approach. Note that 'GR' stands for the growth rate.

3 Experimental Results and Conclusions

To justify the proposed methodology of analyzing the 3D shape of both malignant and benign nodules, the above proposed shape analysis framework was pilot-tested on a database of clinical multislice chest LDCT scans of 327 lung nodules (153 malignant and 174 benign). The CT data sets each have $0.7 \times 0.7 \times 2.0 \text{ mm}^3$ voxels, with nodule diameters ranging from 3 mm to 30 mm. Note that these 327 nodules were diagnosed using a biopsy (our ground truth).

Segmentation results: Figure 9 illustrates results of segmenting pleural attached nodules shown by axial, sagittal, and coronal cross sections. The pixel-wise Gibbs energies in each cross section are higher for the nodules than for any other lung voxels including the attached artery. Therefore, our approach separates accurately the pulmonary nodules from any part of the attached artery. The evolution terminates after 50 iterations because the changes in the total energy become close to zero. The error of our segmentation with respect to the radiologist “ground truth” is 1.86%. In total, our segmentation of the 327 nodules has an error range of 0.29% – 2.17% with a mean error of 0.71%, and a standard error deviation of 1.21%.

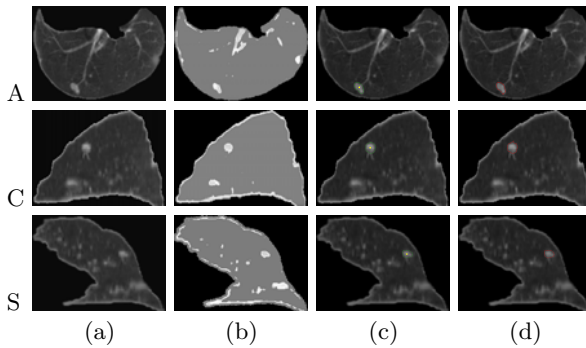


Fig. 9. 3D segmentation of pleural attached nodules; results are projected onto 2D axial (A), coronal (C), and sagittal (S) planes for visualization: 2D profile of the original nodule (a), pixel-wise Gibbs energies (b) for $\nu \leq 11$, our segmentation (c), and (d) the radiologist’s segmentation.

Diagnostic results: The training subset for classification (15 malignant lung nodules and 15 benign lung nodules) was arbitrarily selected among all of the 327 lung nodules. The accuracy of classification based on using K-nearest classifier of both the training and test subjects was evaluated using the χ^2 -test at 95% confidence level. At the 95% confidence level, the correctly classified 143 out of 153 malignant nodules (a 93.5% accuracy), and 163 out of 174 control subjects (a 93.7% accuracy). The overall accuracy using the proposed 3D shape-based CAD system for 95% confidence level is 93.6% in **the first detection of lung nodules**. The classification based on traditional growth rate approach [14] over

one year is 87 out of 153 malignant nodules (a 56.9% accuracy), and 114 out of 174 benign nodules (a 65.7% accuracy) at a 95% confidence level, these results highlight the advantage of the proposed shape-based diagnostic approach.

Another way to measure and test the performance of the proposed diagnostic system is to compute the receiver operating characteristic (ROC). Each point on the graph is generated by using a different cut point (classification threshold). Figure 8 shows the ROC of the two approaches; our proposed shape index based diagnostic approach and the growth rate based diagnostic approach [14]. It is clear from Fig 8 that the area under ROC curve of our present approach is larger ($Az = 0.9782$) than the area under the ROC curve of the growth rate-based diagnostic approach [14] (Az is 0.6757 for one year estimated growth rate). The high sensitivity and specificity of the proposed approach is due to using the estimated number of spherical harmonic to approximate the 3D shape of the detected lung nodule as a new discriminatory feature which is more separable than using the estimated growth rate.

As demonstrated in this paper, the preliminary results justify further elaboration of the proposed alternative method for diagnosing malignant lung nodules. Its novelty lies in using the shape of a segmented 3D nodule instead of the more conventional growth rate as a reliable diagnostic feature. The shape is described in terms of a linear combination of spherical harmonics (SHs).

The proposed nodule shape analysis could lead to more accurate, faster, and more clinically useful diagnostics of detected pulmonary nodules without the need for investigating their temporal development on the successive LDCT images of the same subject collected for a relatively long time. The present C++ implementation on the Intel quad processor (3.2GHz each) with 16 GB memory and 1.5 TB hard drive with the RAID technology takes approximately 7 sec for processing 30 mm lung nodules and less than 3 sec for processing 5 mm lung nodules.

References

1. Chung, M.K., Shen, L., Dalton, K.M., Evans, A.C., Davidson, R.J.: Weighted fourier series representation and its application to quantifying the amount of gray matter. *IEEE Trans. Med. Imag.* 26, 566–581 (2007)
2. Courant, R., Hilbert, D.: *Methods of Mathematical Physics*, vol. II. Interscience, New York (1953)
3. El-Baz, A., Gimel'farb, G.: Em based approximation of empirical distributions with linear combinations of discrete gaussians. In: *Proc. IEEE Int. Conf. Image Process (ICIP 2007)*, San Antonio, Texas, USA, September 16-19, vol. IV, pp. 373–376 (2007)
4. El-Ba, A., Gimel'farb, G.G., Falk, R., Holland, T., Shaffer, T.: A new stochastic framework for accurate lung segmentation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *MICCAI 2008, Part I. LNCS*, vol. 5241, pp. 322–330. Springer, Heidelberg (2008)
5. Fang, Q., Boas, D.: Tetrahedral mesh generation from volumetric binary and gray-scale images. In: *Proc. IEEE Int. Symp. on Biomed. Imag. From Nano to Macro (ISBI 2009)*, Boston, MA, USA, June 28-July 1, pp. 1142–1145 (2009)

6. Farag, A.A., El-Baz, A.S., Gimel'farb, G.G., Falk, R., El-Ghar, M.A., Eldiasty, T., Elshazly, S.: Appearance models for robust segmentation of pulmonary nodules in 3D LDCT chest images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 662–670. Springer, Heidelberg (2006)
7. Furuya, K., Murayama, S., Soeda, H., Murakami, J., Ichinose, Y., Yabuuchi, H., Katsuda, Y., Koga, M., Masuda, K.: New classification of small pulmonary nodules by margin characteristics on high-resolution ct. *Acta Radiologica* 40, 496–504 (1999)
8. Gurney, W., Swensen, S.: Solitary pulmonary nodules: determining the likelihood of malignancy with neural network analysis. *Radiology* 196, 823–829 (1995)
9. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *Int. J. Computer Vision* 1, 321–331 (1987)
10. Kawata, Y., Niki, N., Ohmatsu, H., Kakinuma, R., Eguchi, K., Kaneko, M., Moriyama, N.: Biexponential characterization of prostate tissue water diffusion decay curves over an extended b-factor range. *IEEE Trans. Nucl. Sci.* 45, 2132–2138 (1998)
11. Kawata, Y., Niki, N., Ohmatsu, H., Kusumoto, M., Kakinuma, R., Mori, K., Nishiyama, H., Eguchi, K., Kaneko, M., Moriyama, N.: Computerized analysis of 3-d pulmonary nodule images in surrounding and internal structure feature spaces. In: *Proc. Int. Conf. Image Process.*, pp. 889–892 (2001)
12. Kido, S., Kuriyama, K., Higashiyama, M., Kasugai, T., Kuroda, C.: Fractal analysis of small peripheral pulmonary nodules in thin-section ct: evaluation of the lung-nodule interfaces. *J. Comput. Assist. Tomogr.* 26, 573–578 (2002)
13. Mori, K., Niki, N., Kondo, T., Kamiyama, Y., Kodama, T., Kawada, Y., Moriyama, N.: Development of a novel computer-aided diagnosis system for automatic discrimination of malignant from benign solitary pulmonary nodules on thin-section dynamic computed tomography. *J. Comput. Assist. Tomogr.* 29, 215–222 (2005)
14. Reeves, A., Chan, A., Yankelevitz, D., Henschke, C., Kressler, B., Kostis, W.: On measuring the change in size of pulmonary nodules. *IEEE Trans. Med. Imag.* 25, 435–449 (2006)
15. Shen, L., Chung, M.K.: Large-scale modeling of parametric surfaces using spherical harmonics. In: *Proc. 3rd Int. Symp. 3D Data Process. Visualiz. Transmission*, Chapel Hill, NC, USA, June 14–16, pp. 294–301 (2006)
16. Weir, H.K., Thun, M.J., Hankey, B.F., Ries, L.A., Howe, H.L., Wingo, P.A., Jemal, A., Ward, E., Anderson, R.N., Edwards, B.K.: Annual report to the nation on the status of cancer, 1975–2000. *J. National Cancer Institute* 95, 1276–1299 (2003)

Erratum: Optimal Data-Driven Sparse Parameterization of Diffeomorphisms for Population Analysis

Sandy Durrleman, Marcel Prastawa, Guido Gerig, and Sarang Joshi

SCI Institute, University of Utah, 72 S. Central Campus Dr.,
UT-84112 Salt Lake City

G. Székely and H.K. Hahn (Eds.): IPMI 2011, LNCS 6801, pp. 123–134, 2011.
© Springer-Verlag Berlin Heidelberg 2011

DOI 10.1007/978-3-642-22092-0_64

In the original version, the first name of the first author is incorrect. Instead of "Sandy Durrleman" it should be read as "Stanley Durrleman".

The original online version for this chapter can be found at
http://dx.doi.org/10.1007/978-3-642-22092-0_11

Author Index

- Aach, Til 197
Abo El-Ghar, Mohammed 772
Abugharbieh, Rafeef 612
Alexander, Daniel C. 748
Altinay, Murat 73
Asman, Andrew J. 85
Astley, Sue 510
Avison, Malcolm J. 147
Ayache, Nicholas 463, 735
- Bai, Junjie 61
Barnes, Josephine 748
Bayouth, John 245
Bazin, Pierre-Louis 1
Ben Ayed, Ismail 171, 221
Berks, Michael 510
Blitz, Ari 1
Boucher, Maxime 271
Bouix, Sylvain 320
Brady, Michael 333
Buatti, John 245
Buechert, Martin 346
Buzug, Thorsten M. 308
- Carass, Aaron 371
Cardoso, M. Jorge 159
Caserta, Enrico 398
Chelikani, Sudhakar 525
Chen, Hanbo 97
Chen, Mingqing 61
Chen, Zezhi 510
Cheriet, Farida 660
Clarkson, Matthew J. 159, 748
Cootes, Timothy F. 636
Criminisi, Antonio 25, 184
Cuzzocreo, Jennifer L. 1
- Darkner, Sune 436
de Bruijne, Marleen 49, 475
Deligianni, Fani 296
Deng, Fan 97
Deriche, Rachid 723
Dewan, Maneesh 111
Ding, Zhaohua 147
Dirksen, Asger 49, 475
- Du, Jia 448
Duncan, James S. 37, 525
Durrleman, Sandy 123, E1
- Edwards, A. David 296
El-Baz, Ayman 772
Elnakib, Ahmed 772
Evans, Alan 271
- Falk, Robert 772
Faraco, Carlos 97
Fletcher, P. Thomas 13
Fontejn, Hubert M. 748
Fox, Nick C. 748
- Gao, Jianbin 674
Garvin, Gregory 221
Gee, James C. 674
Gerig, Guido 123, E1
Ghosh, Aurobrata 723
Gimel'farb, Georgy 772
Goela, Aashish 171
Goh, Alvina 257, 448
Golby, Alexandra J. 135
Golland, Polina 135, 735
Gollmer, Sebastian T. 308
Grady, Leo 660
Graff, Christian G. 760
Gramfort, Alexandre 562, 600
Guo, Lei 97, 574
Gur, Yaniv 538
- Hadjidemetriou, Stathis 346
Hamalainen, Matti 600
Han, Dongfeng 245
Han, Junwei 574
Hao, Xiang 13
Haueisen, Jens 600
Hegenbart, Sebastian 498
Hennig, Juergen 346
Ho, Jeffrey 550
Honkela, Antti 735
Hu, Xintao 574
Huang, Xiaolei 411
Hunter III, Allan A. 674

- Iglesias, Juan Eugenio 25, 184
 Islam, Ali 171

 Janoos, Firdaus 398, 588
 Jia, Hongjun 283
 Jiang, Xi 97
 Jiao, Fangxiang 538
 Johnson, Chris R. 538
 Joshi, Sarang 123, E1, 538

 Kadir, Timor 333
 Kaftan, Jens N. 197
 Kanade, Takeo 384
 Ker, Dai Fei Elmer 384
 Khalifa, Fahmi 772
 Kikinis, Ron 320
 Kirschner, Matthias 308
 Klassen, Eric 147
 Klein, Tassilo 359
 Konukoglu, Ender 25, 735
 Kowalski, Matthieu 600
 Kubicki, Marek 320
 Kurtek, Sebastian 147

 Lai, Rongjie 233
 Landman, Bennett A. 85
 Langs, Georg 135
 Lashkari, Danial 135
 Lauze, François 624
 Lehmann, Manja 748
 Leone, Gustavo 398
 Li, Hongsheng 411
 Li, Kaiming 97, 574
 Li, Shuo 171, 221
 Lian, Jun 686
 Liao, Shu 283
 Lin, Ben A. 37
 Liu, Tianming 97, 574
 Lo, Pechin 49, 475
 Lombaert, Herve 660
 Lorenzi, Marco 463
 Lu, Chao 525
 Ludwig, Ute 346
 Lv, Jinglei 574

 Machiraju, Raghu 398, 588
 Maguire, Maureen G. 674
 Menze, Bjoern H. 735
 Metaxas, Dimitri 184
 Michel, Vincent 562

 Miller, L. Stephen 97
 Modat, Marc 159, 748
 MomayyezSiahkal, Parya 711
 Montillo, Albert 25, 184
 Mórocz, Istvan Á. 588
 Myers, Kyle J. 760

 Nambakhsh, Mohammad Saleh 171
 Navab, Nassir 359, 648
 Ng, Bernard 612
 Nielsen, Mads 49, 624
 Nitzken, Matthew 772

 O'Donnell, Thomas 197
 Ourselin, Sebastien 159, 748

 Pearlman, Paul C. 37
 Pécot, Thierry 398
 Pedregosa, Fabian 562
 Penneç, Xavier 463, 624
 Peters, Terry 171
 Petersen, Jens 49, 475
 Pham, Dzong L. 1
 Platsch, Günther 333
 Pluta, John 73
 Polimeni, Jonathan R. 660
 Potesil, Vaclav 333
 Prastawa, Marcel 123, E1
 Prince, Jerry 371
 Punithakumar, Kumaradevan 171, 221

 Qiu, Anqi 257, 448

 Rathi, Yogesh 320
 Rigolo, Laura 135
 Risholm, Petter 699
 Rittscher, Jens 398
 Robinson, Emma 296
 Romano, Walter 221
 Ross, James 699
 Roy, Snehashis 371
 Rueckert, Daniel 296

 Saghir, Zaigham 49
 Schuh, Andreas 197
 Sharp, David J. 296
 Shen, Dinggang 283, 686
 Shen, Tian 411
 Shenton, Martha 320
 Shi, Yonggang 233

- Shiee, Navid 1
Shotton, Jamie 184
Siddiqi, Kaleem 271, 711
Singh, Shantanu 398, 588
Sinusas, Albert J. 37
Soatto, Stefano 424
Sørensen, Lauge 475
Sommer, Stefan 624
Song, Qi 61, 208, 245
Sonka, Milan 61, 208, 245
Soza, Grzegorz 197
Sporring, Jon 436
Srivastava, Anuj 147
Stojkovic, Branislav 208
Strohmeier, Daniel 600
Suh, Jung Wook 73
Sweet, Andrew 135
- Tagare, Hemant D. 37
Taurani, Aakant 245
Taylor, Chris 510
Thirion, Bertrand 296, 562
Tie, Yanmei 135
Tietjen, Christian 197
Toga, Arthur W. 233
Tu, Zhuowen 25
- Uhl, Andreas 498
- Van Leemput, Koen 735
Varoquaux, Gael 296, 562
Vécsei, Andreas 498
Vemuri, Baba C. 550
- Wachinger, Christian 359, 648
Wang, Hongzhi 73, 674
Wang, Qian 686
Washko, George R. 699
Wassermann, Demian 320, 723
Weber, Marc-André 735
Wells, William M. 588, 699
Wesarg, Stefan 308
Westin, Carl-Fredrik 320
Whitaker, Ross T. 13
Winn, John 184
Wu, Guorong 283, 686
Wu, Jue 674
Wu, Xiaodong 61, 208, 245
- Xie, Yuchen 550
Xing, Lei 486
Xiong, Guanglei 486
Xu, Jinhui 208
Xu, Lei 208
- Yang, Xianfeng 257
Yi, Zhao 424
Yigitsoy, Mehmet 648
Yin, Zhaozheng 384
Yuan, Jing 171
Yushkevich, Paul 73
- Zhan, Yiqiang 111
Zhang, Degang 97
Zhang, Pei 636
Zheng, Yuanjie 674
Zhou, Xiang Sean 111
Zhu, Dajiang 97, 574
Zhu, Yongding 208