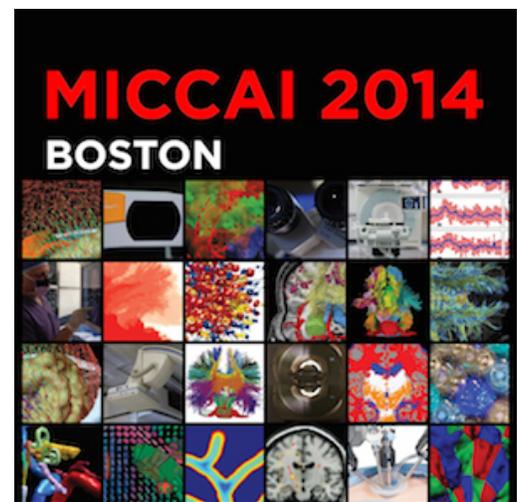




# Sparsity Techniques In Medical Imaging

2<sup>nd</sup> International Workshop  
Held in Conjunction with MICCAI 2014  
Boston, USA



## Preface

The Second International Workshop on Sparsity Techniques in Medical Imaging was held in conjunction with the 17<sup>th</sup> International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) on September 14, 2014, in Boston, USA.

Compressed sensing and sparse methods have played an important role in the medical imaging field, including image reconstruction, image enhancement, image segmentation, anomaly detection, disease classification, and image database retrieval. The inherent property of sparsity in the medical images and the image database introduces essential prior knowledge to facilitate the informatics acquisition, reconstruction and analysis. For example, in MR image reconstruction, sparsity in transformed space such as wavelet has been successfully used to speed up scanning time and improve reconstruction quality. In medical image segmentation, sparse shape prior can preserve the local details that are not significant in the training data and correct the misleading appearance cues. In low-dose dynamic CT, sparsity in the selection of high-dose patches recovers signal from noisy data and improve medical imaging safety by reducing the necessary radiation dose. In neuroscience, structured sparsity has been shown to be useful for predict the relevant features of brain diseases, such as Alzheimer's.

Leveraging the success of the previous workshops, the 2<sup>nd</sup> MICCAI workshop on Sparsity Techniques in Medical Imaging aimed to provide a comprehensive forum for reviewing clinical opportunities in sparsity techniques, and for sharing state-of-art as well as emerging techniques for solving computational challenging image analysis and imaging problems using sparsity, by bringing together leading researchers and clinical scientists from around the world.

In response to a call for papers, a total of 23 papers were initially submitted to the workshop. These papers underwent a rigorous, double-blind peer-review process, with each paper being reviewed by a minimum of 2 reviewers, and in many cases, by 3 expert reviewers from the Program Committee. Based on the results of this review, 14 papers were accepted by the workshop for presentation. All the accepted papers were revised by incorporating the reviewers' comments and re-submitted by the authors to be included in this proceedings volume. 4 papers have been selected for 15-minute podium presentation during the 2 plenary oral sessions. The rest 10 papers have been selected for poster presentation.

The workshop further provided two plenary lectures, one on sparsity techniques in medical imaging, by Dr. Yoram Bresler from University of Illinois, Urbana-Champaign; the other on applications to medical image reconstruction and analysis using sparsity, dictionaries and patches, given by Dr. Daniel Rueckert from Imperial College London. As a result, the workshop successfully provided a forum among participants for the dissemination of state-of-art research and technologies, the exchange of emerging ideas, the initiation of collaborations, and the exploration of new clinical applications for sparsity techniques with dictionary learning in medical imaging and analysis.

We would like to express our sincere appreciation to the authors whose contributions to this workshop proceeding that have required considerable commitment of time and effort. We also thank the members of the Program Committee for their excellent work in reviewing the submitted manuscript on tight schedule, and the members of the program committee for their outstanding job in organizing and compiling the papers in this proceeding volume.

Tsuan Chen  
Ruogu Fang  
Zhi-Pei Liang  
Dimitris Metaxas  
Pina C. Sanelli  
Shaoting Zhang

September 2014

# About

## Scope

Sparsity has been an important modeling tool in compressed sensing, machine learning, image processing, neuroscience and statistics. In the medical imaging field, sparsity methods have been successfully used in image reconstruction, image enhancement, image segmentation, anomaly detection, disease classification, and image database retrieval. The inherent property of sparsity in the medical images and the image database introduces essential prior knowledge to facilitate the informatics acquisition, reconstruction and analysis. For example, in MR image reconstruction, sparsity in transformed space such as wavelet has been successfully used to speed up scanning time and improve reconstruction quality. In medical image segmentation, sparse shape prior can preserve the local details that are not significant in the training data and correct the misleading appearance cues. In low-dose dynamic CT, sparsity in the selection of high-dose patches recovers signal from noisy data and improve medical imaging safety by reducing the necessary radiation dose. In neuroscience, structured sparsity has been shown to be useful for predict the relevant features of brain diseases, such as Alzheimer's. We have also witnessed significant progress in flexible models to incorporate complex feature structure such as structured sparsity and dynamic group sparsity, which substantially enhances the model expressiveness and expands the domain of solvable problems. Sparsity in large-scale data, which deals with millions or billions of possible features or data points, have also been developed to efficiently select the key elements. Developing more powerful sparsity models for a large range of medical imaging and medical image analysis problems as well as efficient optimization and learning algorithm will keep being a main research topic in this field.

## Aims

The goal of this workshop is to publish original, high quality papers on innovation research and development in the analysis of medical image data using sparsity models and methods. This workshop will help advance the scientific research within the field of sparsity methods for medical imaging. It will foster dialogue and debate in this vibrant field covering Compressed Sensing (CS), Sparse Learning (SL), Sparse Representation

(SR) and their applications to medical imaging. It will consist of previously unpublished, contributed and invited papers.

## **Topics**

For this workshop, authors are invited to submit original research papers and high-quality overview and survey articles on topics including, but not limited to:

### **Methodology:**

- Efficient Sparse Learning
- Dictionary Learning
- Shape Prior Modeling
- Convex Optimization on Sparsity Priors
- Group Sparsity
- Structured Sparsity
- Large-scale Sparse Learning
- Multi-Source Sparse Learning
- Statistical Analysis
- Model Selection, etc.

### **Applications:**

- Image / Signal Reconstruction
- Image Segmentation
- Image Enhancement
- Image Registration
- Compressed Sensing Magnetic Resonance Imaging
- Anomaly Detection and Correction

# Workshop Organization

## Organization Committee

Tsuhuan Chen	Cornell University, USA
Ruogu Fang	Florida International University, USA
Zhi-Pei Liang	University of Illinois at Urbana-Champaign, USA
Dimitris Metaxas	Rutgers University, USA
Pina C. Sanelli	Weill Cornell Medical College, USA
Shaoting Zhang	University of North Carolina at Charlotte, USA

## Program Committee

Yoram Bressler	University of Illinois at Urbana-Champaign, USA
Tom Weidong Cai	University of Sydney, Australia
James Gee	University of Pennsylvania, USA
Le Lu	National Institute of Health, USA
Albert Montillo	General Electronic, USA
Richardo Otazo	New York University, School of Medicine, USA
Jens Rittscher	University of Oxford, UK
Daniel Rueckert	Imperial College of London, UK
Dinggang Shen	University of North Carolina at Chapel Hill, USA
Li Shen	Indiana University, USA
Richard Souvenir	University of North Carolina at Charlotte, USA
Pascal Spincemaille	Weill Cornell Medical College, USA
Yi Wang	Cornell University, USA
Cark-Fredrik Westin	Harvard medical School, USA
Yves Wiaux	Heriot-Watt University, UK
Guorong Wu	University of North Carolina at Chapel Hill, USA
Lin Yang	University of Kentucky, USA
Yiqiang Zhan	Siemens Healthcare, USA
Yuanjie Zheng	University of Pennsylvania, USA
Yan Zhou	Elekta, USA

# Workshop Program

## Plenary Talks:

**Dr. Yoram Bresler**, Professor, Department of Electrical and Computer Engineering, Department Bioengineering, University of Illinois, Urbana-Champaign

Talk Title and Abstract: TBD

**Dr. Daniel Rueckert**, Professor, Department of Computing, Imperial College London

Talk Title: **"Sparsity, Dictionaries and Patches: Applications to Medical Image Reconstruction and Analysis"**

**Abstract:** This talk will focus on the convergence medical imaging and machine learning techniques for the discovery and quantification of clinically useful information from medical images. The first part of the talk will describe machine learning techniques based on sparsity that can be used for image reconstruction, e.g. the acceleration of MR imaging. The second part will discuss model-based approaches that employ statistical as well as probabilistic approaches for segmentation. In particular, we will focus on segmentation techniques that combine patch-based approaches such as dictionary learning with sparsity to improve the accuracy and robustness of the segmentation approaches.

## Oral Presentations:

### Medical Image Analysis with Sparsity

Session Chair: Tsuhan Chen, Ruogu Fang

[STMI-O-1] Sparsity Based Spectral Embedding: Application to Multi-Atlas Echocardiography Segmentation

Ozan Oktay, Wenzhe Shi, Jose Caballero, Kevin Keraudren, Daniel Rueckert

[STMI-O-2] Stain Unmixing in Brightfield Multiplex Immunohistochemistry Images

Ting Chen, Chukka Srinivas

### Medical Imaging with Sparsity

Session Chair: Zhi-Pei Liang

[STMI-O-3] Reduced-dose patient to baseline CT rigid registration in 3D Radon space

Guy Medan, Achia Kronman, Leo Joskowicz

[STMI-O-4] Predicting cross-task behavioral variables from fMRI data using the \$k\$-

support norm

Michail Misyrilis, Anna B. Konova, Matthew B. Blaschko, Jean Honorio, Nelly Alia-Klein, Rita Z. Goldstein, Dimitris Samaras

**Poster Presentations:**

[STMI-P-1] Fast MRI for repeated scans

Lior Weizman, Leo Joskowicz, Dafna Ben Bashat

[STMI-P-2] Multisite Disease Classification with Functional Connectomes via Multitask Structured Sparse SVM

Takanori Watanabe, Clayton Scott, Chandra Sripada

[STMI-P-3] Auto-contouring the Prostate for Online Adaptive Radiotherapy

Yan Zhou, Xiao Han

[STMI-P-4] Detection of Multiple Sclerosis Lesions using Sparse Representations and Dictionary Learning

Hrishikesh Deshpande, Pierre Maurel, Christian Barillot

[STMI-P-5] Leveraging Sparsity: A Low-Rank + Sparse Decomposition (LR+SD) Method for Automatic EEG Artifact Removal

Jerome Gilles California, Travis Meyer, P Douglas, PK Douglas

[STMI-P-6] Group Sparse Kernelized Dictionary Learning for the Clustering of White Matter Fibers

Kuldeep Kumar, Christian Desrosiers

[STMI-P-7] 3D Mouse Left Ventricle Reconstruction using Sparse MR Images with Arbitrary Orientations

Yang Yu, Jingjing Liu, Dimitris Metaxas, Leon Axel

[STMI-P-8] Anisotropic Tensor Total Variation Regularization For Low Dose Low CT Perfusion Deconvolution

Ruogu Fang, Tsuhan Chen, Pina Sanelli

[STMI-P-9] Region segmentation for sparse decompositions: better brain parcellations from rest fMRI

Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, Gael Varoquaux

[STMI-P-10] Improved Sparse Shape Composition Model for Multi-shape Prior

Bing Wang, Chonghao Fan, Hongzhi Xie, Lixu Gu

# Sparsity Based Spectral Embedding: Application to Multi-Atlas Echocardiography Segmentation

Ozan Oktay, Wenzhe Shi, Jose Caballero,  
Kevin Keraudren, and Daniel Rueckert

Biomedical Image Analysis Group, Imperial College London, UK  
o.oktay13@imperial.ac.uk

**Abstract.** Echocardiography is one of the primary imaging modalities used in the diagnosis of cardiovascular diseases. It is commonly used to extract cardiac functional indices including the left ventricular (LV) volume, mass, and motion. The relevant echocardiography analysis methods, including motion tracking, anatomical segmentation, and registration, conventionally use the intensity values and/or phase images, which are highly sensitive to noise and do not encode contextual information and tissue properties directly. To achieve more accurate assessment, we propose a novel spectral representation for echo images to capture structural information from tissue boundaries. It is computationally very efficient as it relies on manifold learning of image patches, which is approximated using sparse representations of dictionary atoms. The advantage of the proposed representation over intensity and phase images is demonstrated in a multi-atlas LV segmentation framework, where the proposed spectral representation is directly used in deformable registration. The results suggest that the proposed spectral representation can provide more accurate registration between images. This in turn provides a more accurate LV segmentation. Finally, it is the first time that a multi-atlas approach achieves state-of-the-art results in echo image segmentation.

## 1 Introduction

In the diagnosis of cardiovascular diseases, echocardiography is still the most important and widely used tool due to its high availability and ease of use. It has been used to extract functional and quantitative indices like left ventricular (LV) mass, volume and motion. The accuracy of these measurements depends on the correct delineation of endocardial boundary; thus, automated segmentation tools are more desirable for analysis as manual tracing is subject to inter-observer variability and human error. However, volumetric segmentation is still a challenging task for echo images due to image artefacts and low image quality.

The existing approaches to echocardiography segmentation can be divided into model-based and data-driven. Deformable surface models [2] and active-shape models [5] are two examples of the former category, which require a good model initialization or training to learn shape prior information to subsequently segment the ventricle boundary in target images. Although these approaches

achieve state-of-the-art segmentation results, they are limited by inter-subject anatomical variations due to the extensive training needed to cope with large shape variations. On the other hand, the second category relies on intensity distributions instead of a trained model. Thus, they are less sensitive to inter-subject variability, but are highly susceptible to the noise level and inconsistencies in the intensity distribution. Two common examples are edge-based level sets [12] and multi-atlas segmentation [20]. Particularly, although atlas based approaches have been successful for MRI segmentation [1], large registration errors on echo images prevent them from being effective for echo images.

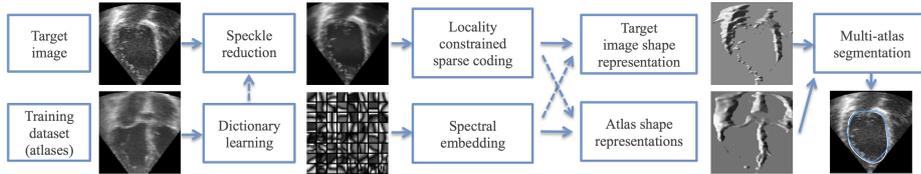
Indeed, intensity and phase images are not representative enough to guide image registration because they do not directly reflect properties of the tissues or their contextual information. In this paper, to address this problem we propose a novel spectral representation for echo images, through which we extend and outperform the multi-atlas segmentation framework proposed in [20]. The new image representation captures structural information and guides the deformable registration to obtain a better tissue alignment. It also reduces the noise sensitivity and removes the need for image compounding, and ultimately achieving higher segmentation accuracies.

Spectral embedding is employed to compute the proposed representation, which has been successfully applied in min-cut segmentation [8], multi-modal image registration [16], and large deformation estimation problems [10]. Nevertheless, spectral embedding is not directly applicable to 3D echo images due to the large number of image patches, resulting in long computation times and intensive memory usage. We therefore propose, as an additional contribution, a more efficient embedding that exploits the redundant nature of echo image patches. The underlying manifold structure is learned only for atoms from a trained dictionary that sparsely represent the image patches. A single over-complete dictionary is assumed to be representative enough for all echo image patches to approximate the low dimensional space and each image patch is mapped to the underlying manifold space as a sparse linear combination of atoms yielding a set of spectral coordinates. To preserve the geodesics and local structure, sparse selection in coding is achieved by enforcing the locality constraint [17], which implies both sparsity and locality as explained in [18].

In the context of this paper, the proposed image patch embedding is referred as spectral representation. The paper is structured as follows: In section 2 of this paper, we introduce the relevant theory of dictionary learning and sparse coding for the manifold approximation. Section 3 presents validation results on the CETUS challenge data [4], which shows significant improvement in segmentation accuracy using the proposed representation over phase and intensity images. In the last section, the paper concludes by a brief discussion of the results.

## 2 Methodology

In the proposed segmentation framework, echo images are first sparsely reconstructed with dictionary atoms for speckle reduction. Secondly, a spectral repre-



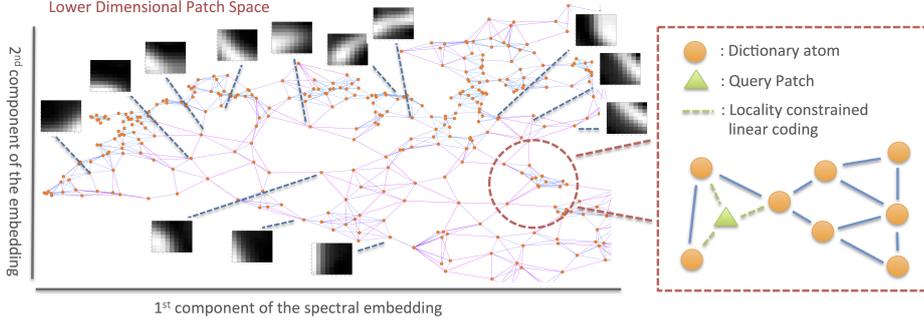
**Fig. 1.** Block diagram of the proposed multi-atlas segmentation framework

segmentation is extracted from the processed images by mapping image patches to the manifold space of the dictionary atoms. Then, atlas labels are propagated to the target image by deformable registration using the spectral representation. The framework is shown in Fig. 1 and detailed below.

**Speckle reduction:** Target echo sequences are preprocessed prior to segmentation to increase signal-to-noise ratio. Instead of relying on standard speckle reduction techniques [6], images are denoised using dictionary learning and sparse coding similar to the image denoising application in [7]. On top of achieving state-of-the-art denoising, dictionary learning provides global patch analysis by building a set of atoms from training data that sparsely represent image patches. For echo images, these atoms have characteristic edge patterns. We use the K-SVD algorithm [7] to approximate image patches  $\mathbf{y}_n \in \mathbb{R}^P$  as sparse combinations  $\mathbf{x}_n \in \mathbb{R}^M$  of atoms from an over-complete dictionary  $\mathbf{C} \in \mathbb{R}^{P \times M}$  with a precision bounded by  $\epsilon$ , namely solving:  $\min_{\mathbf{C}, \mathbf{X}} \sum_{n=1}^N \|\mathbf{x}_n\|_0$  s.t.  $\forall n, \|\mathbf{y}_n - \mathbf{C}\mathbf{x}_n\|^2 \leq \epsilon \mid \epsilon \in \mathbb{R}_+$ . Patches are overlapping and wrap around image boundaries, meaning there are  $N$  patches for an image of  $N$  pixels.

**Spectral representation:** As shown in previous works [10,16], spectral coordinates can be computed using non-linear dimensionality reduction of image patches; this paper particularly focuses on Laplacian Eigenmaps (LE) [3]. The algorithm computes the Laplacian graph  $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$  using the adjacency and degree matrices  $\mathbf{A}, \mathbf{D} \in \mathbb{R}^{N \times N}$  corresponding to all image patches. Then, spectral coordinates are obtained by finding the lowest  $K$  eigenvectors of the matrix  $\mathbf{L}$ . This representation is suited for small datasets such as small stacks of MRI slices, but is prohibitive for 3D echo volumes due to the large amounts of voxels in the image that result in a very large adjacency matrix. Furthermore, finding a fixed low-dimensional space for all images is also challenging and is usually solved by point-matching algorithms.

To overcome these problems, we propose to perform manifold learning on dictionary atoms ( $\mathbf{c}_m \in \mathbb{R}^P$ ) and then the spectral coordinates are approximated by sparse linear combinations of dictionary atoms. For this approximation, two main assumptions are made: (1) echo image patches can be sparsely represented by dictionary atoms and (2) these patches can be expressed in a manifold [11] that groups atoms with similar edge patterns. In that respect, the learned dictionary atoms that are the byproduct from the speckle reduction step are transformed into spectral coordinates with LE. Dictionary atoms of similar shape are



**Fig. 2.** The lowest two spectral coordinates of the dictionary atoms (left), locality constrained linear coding of a query patch to map it to manifold space (right).

grouped together in the spectral coordinates and the variance of the patches is maximized, as shown in Fig. 2. With the embedding learned, each image patch ( $\mathbf{y}_n$ ) is mapped to the lower dimensional space through linear combinations ( $\tilde{\mathbf{x}}_n$ ) of spectral coordinates corresponding to dictionary atoms ( $\mathbf{s}_c \in \mathbb{R}^{M \times K}$ ) solving:  $\mathbf{s}_{y_n} = \tilde{\mathbf{x}}_n^\top \mathbf{s}_c$ . The linear codes are found by minimizing the cost function  $\min_{\mathbf{X}} \sum_n \|\mathbf{y}_n - \mathbf{C}\tilde{\mathbf{x}}_n\|^2 + \lambda \|\mathbf{b}_n \odot \tilde{\mathbf{x}}_n\|^2$  s.t.  $\forall n, \mathbf{1}^\top \tilde{\mathbf{x}}_n = 1$ , where  $\odot$  denotes the element-wise multiplication and  $\lambda \in \mathbb{R}_+$ . This formulation enforces a locality constraint [17] based on pair-wise distances  $\mathbf{b}_n = [b_{(n,1)}, \dots, b_{(n,M)}]$  where  $b_{(n,m)} = \exp(-\|\mathbf{y}_n - \mathbf{c}_m\|^2 / \sigma)$  and  $\sigma$  is the variance term. The penalty term assigns higher weights to dictionary atoms  $\mathbf{c}_m$  that are close to the patch  $\mathbf{y}_n$ .

As explained in [18], the locality constraint also implies sparsity, thus the solution can be considered as a sparse weighting of the dictionary atoms. The sparse codes computed in Euclidean space can be applied in manifold space as long as the locality constraint is applied. A single component of approximated spectral coordinates is displayed in Fig. 3.

**Multi-atlas segmentation:** The proposed spectral representation is used in image registration to perform multi-atlas segmentation on echocardiographic images. Different than the standard multi-atlas approach [1], the image similarity metric in the proposed registration algorithm is based on image descriptors. In that respect, images are aligned to each other by minimizing sum-of-squared differences (SSD) between their spectral coordinates instead of image intensity values.

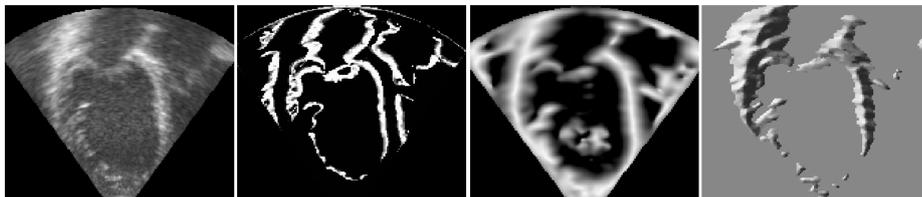
The proposed segmentation framework is described as follows: All atlases collected from the training dataset are linearly aligned to a target image using manually selected three landmarks (left ventricle apex, mid-ventricle, and mitral valve). Similar to the approach in [1], a subset of atlases is selected by computing normalized mutual information (NMI) over a region of interest defined by the atlas labels and target image. The most similar  $L$  atlases are then selected and spectral represented atlases ( $\mathcal{S}_A$ ) are deformable registered to the target spectral image ( $\mathcal{S}_T$ ) with B-spline FFD [14]. The following cost function is minimized:

$\sum_{k=1}^K \|\mathcal{S}_{A_k}(\mathbf{p} + \mathbf{u}) - \mathcal{S}_{T_k}(\mathbf{p})\|^2 + \beta \mathcal{R}(\mathbf{u})$ , where  $\mathbf{p}$ ,  $\mathbf{u}$  and  $k$  denote position, displacement and spectral coordinate dimension. The regularization  $\mathcal{R}$  is defined as bending energy and weighted by  $\beta \in \mathbb{R}_+$ . The algorithm estimates a single common displacement field between volumetric spectral image pairs (in total  $K$ ) while minimizing the cost function. Lastly, the segmentation is decided by majority voting of the propagated atlas labels.

**Local phase images:** The work in [20] on multi-atlas echocardiography segmentation uses local phase images to register atlases to target images. To demonstrate the contribution of the proposed spectral representation, local phase images are evaluated in the same segmentation framework. Images are first converted to a band-pass signal with Laplacian of Poisson filter [19]:  $\mathcal{F}\{LOP\}(\mathbf{w}) = -8\pi^3 |\mathbf{w}|^2 \exp(-2\pi |\mathbf{w}| \rho)$ , where  $\mathbf{w} \in \mathbb{R}^3$  is the position vector in the frequency domain and  $\rho$  controls the central frequency. In our experiments, this filter selection achieved better results compared to Gaussian derivative filter. Afterwards, an analytic signal is obtained by filtering with an isotropic Riesz filter, and the phase image is characterized by the angle between real and imaginary components as explained in [20]. Additionally, local-phase based boundary images [13] are evaluated in the same framework, which are computed in multi-scale ( $\rho \in \mathbb{R}^3$ ) using a monogenic signal. An example of the computed phase and boundary images is displayed in Fig. 3. In the registration step of multi-atlas segmentation, the cost function is defined as  $\mathcal{C} = \omega_1 \cdot \text{NMI}(I_A, I_T) + \omega_2 \cdot \text{NMI}(\phi_A, \phi_T)$ , where  $\phi$  and  $I$  denote phase and intensity images. A similar formulation is used for boundary images by replacing phase images in the cost function.

### 3 Implementation and Results

**Validation dataset:** The proposed segmentation framework is validated on the dataset provided by the MICCAI 2014 CETUS challenge [4], consisting of a set of 3D echo cardiac image sequences acquired from 30 subjects and separated into training (15 subjects) and testing (15 subjects) datasets. As ground truth segmentations for the testing set are not provided, the validation is performed blindly using the evaluation system provided by the CETUS.



**Fig. 3.** Left to right: (1) input image, (2) phase-based boundary detection, (3) local phase image, (4) proposed spectral representation (single component of embedding)

**Table 1.** Multi-atlas segmentation results on training (cross-validation) and testing datasets based on (A) Unprocessed images, (B) Speckle reduced images, (C) Local phase-based boundary images [13], (D) Local phase images [20], and (E) Proposed spectral representation. EF and SV values are reported based on the percentage error measure. Surface distances are given in terms of mean and maximum values.

	Mean (mm)	Max (mm)	Dice Score	Ejection fraction	Stroke volume	
Testing	(A)	3.85±2.06	12.24±5.12	0.80±0.11	0.62±0.21	0.72±0.22
	(B)	2.84±1.07	10.00±3.04	0.85±0.06	0.81±0.14	0.77±0.16
	(C)	2.98±1.20	8.99±3.05	0.84±0.07	0.85±0.11	0.76±0.15
	(D)	2.67±0.92	8.69±2.78	0.85±0.05	0.85±0.10	0.78±0.13
	(E)	2.32±0.78	7.41±1.84	0.87±0.04	0.93±0.05	0.87±0.09
Training	(A)	2.67±0.86	8.81±3.23	0.87±0.06	0.63±0.11	0.58±0.14
	(B)	2.39±0.62	8.55±2.90	0.88±0.05	0.72±0.25	0.74±0.34
	(C)	2.60±0.75	8.48±2.66	0.88±0.05	0.72±0.22	0.72±0.21
	(D)	2.31±0.67	7.71±2.55	0.89±0.04	0.73±0.23	0.74±0.21
	(E)	2.19±0.56	7.63±2.43	0.89±0.04	0.80±0.18	0.86±0.15

**Implementation details:** The patch and dictionary sizes selected for dictionary learning are  $P = 7 \times 7 \times 7$  and  $M = 850$ . The adjacency graph is constructed by linking each dictionary atom to its 8 most similar neighbours in terms of  $\ell_2$  norm distance. In total  $K = 4$  spectral components are selected for spectral representation. In locality constraint coding, parameters are set to  $\lambda = 0.3$  and  $\sigma = 0.2$  for normalized images. The bandpass filter parameter for the computation of phase images is chosen to be  $\rho = 4.5$  and  $\boldsymbol{\rho} = [3.5, 5.0, 7.0]$  for the boundary images. To conclude, the regularization weights in the registration cost functions are defined as  $\omega_1 = \omega_2 = \beta = 1$ , and  $L = 5$  atlases are selected from the training dataset for label propagation.

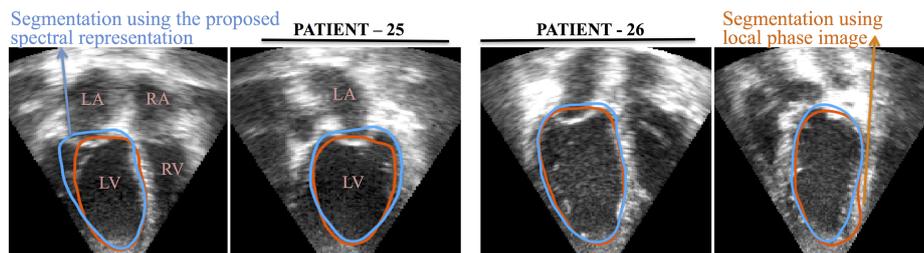
**Validation strategy and results:** In the validation, LV segmentation is done only for the end-diastolic and systolic frames. The accuracy of computed segmentations and clinical indices are used as criteria to evaluate the proposed method and compared against phase and intensity images. In that respect, multi-atlas segmentation is performed on 5 different types of image surrogates, which are provided in Table 1. The evaluation is performed separately for testing and training datasets (cross-validation). As shown, the best result for the testing dataset is obtained using spectral representation based multi-atlas segmentation, which achieves 2.32 mm mean error and 0.87 Dice score. In comparison to intensity and phase images, an improvement of 1.53 and 0.35 mm is observed for the mean surface distance and a similar figure of merit is seen for Dice score results. This suggests that the proposed representations provide more useful information to guide the registration algorithm. Moreover, the results for the cross-validation on the training dataset also demonstrate that spectral representation outperforms segmentation based on speckle reduced intensity images.

**Table 2.** Comparison of the proposed multi-atlas approach (E) against the state-of-the-art echocardiography segmentation: active surfaces [2] and active shape model [5]. Estimated ejection fraction ( $EF$ ) and end-diastolic volume ( $EDV$ ) are compared against their reference values. The correlation accuracy is reported in terms of Pearson’s coefficient ( $R$ ) and Bland-Altman’s limit of agreement ( $BA$ ).

	Mean (mm)	$R_{EF}$	$BA_{EF} (\mu \pm 2\sigma)$	$R_{EDV}$	$BA_{EDV}$	# of Patients
(E)	$2.32 \pm 0.78$	0.923	$-0.74 \pm 6.26$	0.926	$12.88 \pm 35.71$	15
[2]	-	0.907	$-2.4 \pm 23$	0.971	$-24.60 \pm 21.80$	24
[5]	$1.84 \pm 1.86$	-	$0 \pm 19$	-	$3.06 \pm 46.86$	10

To assess the clinical usefulness of the proposed framework, ejection fraction ( $EF_c$ ) and stroke volume ( $SV_c$ ) are evaluated for each patient and compared against their reference values ( $EF_r$ ). The comparison is done by the percentage error measure  $\mathcal{C}_{EF} = 1 - |EF_c - EF_r| / EF_r$ , similarly for the stroke volume  $\mathcal{C}_{SV}$ . The mean value of the percentage errors given in Table 1 demonstrate that the values computed using the spectral representation show a closer agreement with reference values compared to the other representations. Furthermore, a qualitative comparison of the segmentations obtained with the spectral representation and phase image is given in Fig. 4. Finally, the validation results are compared against two state-of-the-art methods in echocardiography segmentation, shown in Table 2. Although the given results are obtained for different datasets, the comparison demonstrates that multi-atlas segmentation can be as successful as the best-performing methods in estimation of important clinical parameter values.

It is observed that the use of a larger number of atlases does not increase the accuracy significantly because the dataset is small and contains large variations, so limiting the number of atlases to 5 in all experiments reduced computation time. Experiments were carried out on a 3.00 GHz quad-core machine, and the approximate computation time per image was 3 min for denoising, 2.5 min for spectral representation, and 30 min for deformable registrations with 5 atlases.



**Fig. 4.** Left ventricle segmentation of two different subjects. Segmentation obtained with the spectral representation (in blue) delineates the endocardium more accurately than the local phased multi-atlas segmentation (in orange).

## 4 Discussion and Conclusion

In this paper, we presented a new spectral representation for echocardiography images based on sparse reconstruction of dictionary atom spectral embeddings. The advantages of this representation are both quantitatively and qualitatively demonstrated in a multi-atlas LV segmentation framework. The results show that it outperforms the local phase and boundary representations in terms of segmentation accuracy. This finding can be related to an improved noise robustness and the explicit use of encoded contextual information, the lack of which in intensity and phase images is an important limitation for guiding the deformable registration in multi-atlas segmentation.

In addition, the proposed representation is computationally efficient and does not require image feature design by hand-crafting as in phase images since distinctive spectral representation is learned from the data itself. Another interesting realization is that spectral representation based multi-atlas segmentation can achieve state-of-the-art results in echocardiography LV segmentation, without requiring any shape prior models. Previous attempts on multi-atlas segmentation required image compounding, and they were not as successful as the proposed framework due to inaccuracies in registration algorithm.

The proposed image descriptors could alternatively be replaced by the sparse coding coefficients without a need for spectral embedding, as proposed in [15] for image segmentation and in [9] for image registration. One particular example would be the use of histogram of sparse codes to represent the image patches. However, the spectral embedding has two main advantages in comparison to the coefficient based representation: Spectral coordinates obtained from different image patches are comparable, and they allow the use of globally smooth distance metrics which are necessary for the registration stage. On the other hand, sparse codes in an overcomplete dictionary, unlike in an orthonormal basis, are non-unique and are therefore unsuitable for image patch comparison. Locality constrained coding yields a locally smooth sparse selection of atoms, meaning that a distance metric could potentially be defined locally at best, but not globally. The other advantage of spectral embedding is the reduced computation time. The spectral coordinates provide a compact and rich representation of patches with a few components, whereas histogram of coefficients would require a large number of computations for comparison and a large vector to represent an image patch.

In conclusion, sparse and parametrizable characteristics of echo images enable us to develop a consistent spectral representation that contains rich structural information. The proposed representation is generic, and can be applied to other ultrasound image applications which require image registration.

## References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage* 46(3), 726–38 (2009)

2. Barbosa, D., et al.: Fast and fully automatic 3-D echocardiographic segmentation using B-spline explicit active surfaces: Feasibility study and validation in a clinical setting. *Ultrasound in medicine & biology* 39(1), 89–101 (2013)
3. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation* 15(6), 1373–96 (2003)
4. Bernard, O., et al.: Challenge on endocardial three-dimensional ultrasound segmentation (CETUS) (2014)
5. Butakoff, C., Sukno, F., Doltra, A., Silva, E., Sitges, M., Frangi, A.F.: Order statistic based cardiac boundary detection in 3D+T echocardiograms. In: *Functional Imaging and Modeling of the Heart*, pp. 359–366. Springer (2011)
6. Coupé, P., Hellier, P., Kervrann, C., Barillot, C.: Nonlocal means-based speckle filtering for ultrasound images. *IEEE T Image Process* pp. 2221–29 (2009)
7. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE T Image Process* 15(12), 3736–45 (2006)
8. Estrada, F.J., Jepson, A.D., Chennubhotla, C.: Spectral embedding and min cut for image segmentation. In: *BMVC*. pp. 1–10 (2004)
9. Guerrero, R., Rueckert, D.: Data-specific feature point descriptor matching using dictionary learning and graphical models. In: *SPIE Medical Imaging*. International Society for Optics and Photonics (2013)
10. Lombaert, H., Grady, L., Pennec, X., Ayache, N.: Spectral log-demons: Diffeomorphic image registration with very large deformations. *IJCV* pp. 1–18 (2013)
11. Peyré, G.: Manifold models for signals and images. *Computer Vision and Image Understanding* 113(2), 249–260 (2009)
12. Rajpoot, K., Grau, V., Alison Noble, J., Becher, H., Szmigielski, C.: The evaluation of single-view and multi-view fusion 3D echocardiography using image-driven segmentation and tracking. *Medical image analysis* 15(4), 514–528 (2011)
13. Rajpoot, K., Grau, V., Noble, J.A.: Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. In: *ISBI*. pp. 783–786. IEEE (2009)
14. Rueckert, D., Sonoda, L., Hayes, C., Hill, D.L., Leach, M., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE TMI* 18(8), 712–21 (1999)
15. Tong, T., Wolz, R., Coupé, P., Hajnal, J.V., Rueckert, D.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *NeuroImage* 76, 11–23 (2013)
16. Wachinger, C., Navab, N.: Entropy and Laplacian images: Structural representations for multi-modal registration. *Medical Image Analysis* 16(1), 1–17 (2012)
17. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: *CVPR*. pp. 3360–67. IEEE (2010)
18. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. In: *NIPS*. vol. 9, p. 1 (2009)
19. Zhang, L., Zhang, D., Guo, Z.: Monogenic-LBP: a new approach for rotation invariant texture classification. In: *ICIP*. pp. 2677–2680. IEEE (2010)
20. Zhuang, X., Yao, C., Ma, Y., Hawkes, D., Penney, G., Ourselin, S.: Registration-based propagation for whole heart segmentation from compounded 3D echocardiography. In: *ISBI*. pp. 1093–1096. IEEE (2010)

# Stain Unmixing in Brightfield Multiplex Immunohistochemistry Images

Ting Chen and Chukka Srinivas

Ventana Medical Systems, Inc. A Member of the Roche Group, USA, 94043

**Abstract.** Multiplex immunohistochemistry (IHC) staining is a newly emerging technique for the detection of multiple biomarkers within a single tissue section and has become more popular due to its significant efficiency and the rich diagnostic information it contains. Therefore, to accurately unmix the IHC image and differentiate all the stains is of tremendous clinical importance since it is the initial key step in multiplex IHC image analysis in digital pathology. Due to the limitation of the CCD color camera, the acquired RGB image only contains three channels, and the unmixing of which into more than three colors is hence a challenging task. To the best of our knowledge, such a problem is barely studied in literature.

This paper presents a novel stain unmixing algorithm for brightfield multiplex IHC images based on a group sparsity model. The proposed framework achieves robust unmixing for more than three chromogenic dyes while preserving the biological constraints of the biomarkers. Typically, a number of biomarkers co-localize in the same cell parts. With this biological information known a priori, the number of stains at one pixel therefore has a fixed up-bound, i.e. equivalent to the number of co-localized biomarkers. By leveraging the group sparsity model, the fractions of stain contributions from the co-localized biomarkers are explicitly modeled into one group to yield least square solution within the group. Sparse solution is obtained among the groups since ideally only one group of biomarkers are present at each pixel. The algorithm is evaluated on both synthetic and clinical data sets and demonstrates better unmixing results than the existing strategies.

## 1 Introduction

A multiplex immunohistochemistry (IHC) slide has the potential advantage of simultaneously identifying multiple biomarkers in one tissue section as opposed to single biomarker labeling in multiple slides. Therefore, it is often used for simultaneous assessment of multiple biomarkers in cancerous tissue. For example, tumors often contain infiltrates of immune cells, which may prevent the development of tumors or favor the outgrowth of tumors [1]. In this scenario, multiple biomarkers are used to target different types of immune cells and the population distribution of each type of them is used to study the clinical outcome of the patients. The biomarkers of the immune cells are stained by different chromogenic

dyes. In order to conduct accurate detection and classification of the cells, the correct unmixing of the IHC digital image to its individual constituent dyes for each biomarker and obtaining the proportion of each dye in the color mixture is a prerequisite step for multiplex IHC image analysis.

Typically, a tissue slide is stained by the multiplex assay. The stained slide is then imaged using a CCD color camera mounted on a microscope or a scanner. The acquired RGB color image is a mixture of the underlying co-localized biomarker expressions. Several techniques have been proposed in the literature to decompose each pixel of the RGB image into a collection of constituent stains and the fractions of the contributions from each of them. Ruifrok *et al.* developed an unmixing method called color deconvolution [2] to unmix the RGB image with *up to three stains* in the converted optical density space. Given the reference color vectors  $x_i \in R^3$  of the pure stains, the method assumes that each pixel of the color mixture  $y \in R^3$  is a linear combination of the pure stain colors and solves a linear system to obtain the combination weights  $b \in R^M$ . The linear system is denoted as  $y = Xb$ , where  $X = [x_1, \dots, x_M]$ ,  $M \leq 3$  is the matrix of reference colors. This technique is currently most widely used in digital pathology domain, however, the maximum number of stains which can be resolved is limited to three, as the linear system is deficient for not having enough equations when there are more than three stains. A multilayer perceptron learning based technique has been proposed in [4] for three color brightfield image unmixing. In [3], Rabinovich *et al.* formulated the color unmixing problem into non-negative matrix factorization and proposed a system capable of performing the color decomposition in a fully automated manner, wherein no reference stain color selection is required. Again, these methods have the same limitation in dealing with large stain numbers due to solving  $y = Xb$ . *To the best of our knowledge, the method of unmixing brightfield IHC image with more than three stains is not available in literature.* In order to compare with the Ruifrok’s method, we divide the color space into several systems with up to three colors in each system based on nearest color matching of each pixel to one of the systems. Ruifrok’s method can therefore be used in solving each individual system. Due to the independent assignment of each pixel into different systems, the spatial continuity is lost in the unmixed images and artifacts such as holes are observed. However, this is the most straightforward modification of Ruifrok’s method to work on more than three color multiplex brightfield image unmixing.

Alternatively, there exists another class of methods for multi-spectral image unmixing that works for a larger number of stain colors [5–9]. In fact, the multi-spectral image differs from the RGB image in terms of image acquisition. Multi-spectral imaging system is used to capture the image using a set of spectral narrow-band filters instead of the CCD color camera. The number of filters  $K$  can be as many as dozens or hundreds, leading to a multi-channel image that provides much richer information than the bright field RGB image. The linear system constructed from it is always an over-determined system with  $X$  being a  $K \times M (K \gg M)$  matrix that leads to a unique solution. However, the scanning process in the multi-spectral imaging system is very time consuming and only a

single field of view, manually selected by the technician, can be scanned instead of the whole slide, the usage of which is thus limited. As an example of the multi-spectral imaging unmixing, the two-stage methods [6, 7] are developed in the remote sensing domain to first learn the reference colors from the image context and then use them to unmix the image. More recently, a sparse model is proposed by Greer in [9] for high dimensional multi-spectral image unmixing. It adopts the  $L_0$  norm to regularize the combination weights  $b$  of the reference colors hence leads to a solution that only a small number of reference colors are contributed to the stain color mixture. This serves as a valuable source of inspiration for selecting regularization terms for the linear system. However, the method proposed in [9] is also designed for multi-spectral image and no prior biological information about the biomarkers are used in that framework which may lead to undesired solution for real data.

In this paper, we propose a novel color unmixing algorithm for multiplex IHC image (scanned using CCD color camera) that can handle more than three stain colors and maintain the biological properties of the biomarkers. Intuitively, the unmixing algorithm for the multiplex IHC image should work as following. (1) Only one group of stains has non-zero contribution in the color mixture for each pixel. (2) Within that group, the fractions of the contributions from each constituent stain should be correctly estimated. These conditions motivate us to model the unmixing problem within the group sparsity [10] framework so as to ensure the sparsity among the group but non-sparsity within the group.

## 2 Methodology

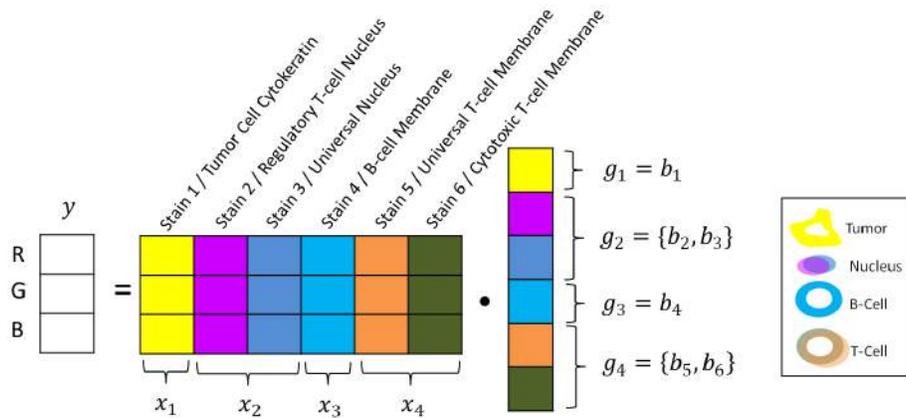


Fig. 1: The group sparsity framework of the unmixing algorithm.

In this section, we present the methodology of our algorithm. We begin with illustrating the basic framework in Fig.1 using the following example. In the analysis of cancerous tissues, different biomarkers are specified to one or more

types of immune cells. For instance, CD3 is a known universal marker for all T-cells and CD8 only stains the membranes of the cytotoxic T-cells. FoxP3 marks the regulatory T-cells in the nuclei and Hematoxylin (HTX) stains all the nuclei. Therefore, the co-localization information of the markers can be inferred from the biological knowledge, i.e. CD3 and CD8 co-locate in the membrane while FoxP3 and HTX may appear in the same nucleus. We can also have tumor marker on the tumor cell’s cytoplasm region and B-cell marker on the B-cell’s membrane. The framework of our proposed algorithm is shown in Fig.1 using the aforementioned immune cell example. Based on this biological co-localization information of the biomarkers, it is straightforward to conclude that only two colors can co-exist at each pixel for this case. The six chromogenic stains are therefore grouped into four different groups where co-localized stains are in the same group, as shown in the right panel of Fig.1.

## 2.1 Optical Density Transform

For the preprocessing, the RGB image  $I$  is converted into the optical density (OD) space using the following formula derived from Beer’s law based on the fact that the optical density is proportional to the stain concentration.

$$O_c = -\log\left(\frac{I_c}{I_{0,c}}\right) \quad (1)$$

where  $c$  is the index of the RGB color channels,  $I_0$  is the RGB value of the white points and  $O$  is the optical density image obtained. As in [2],  $O$  will be image to work with in the rest of the paper.

## 2.2 Group Sparsity Unmixing

We begin with illustrating the notations used in this paper. Let  $\mathbf{y}$  be a pixel of  $O$  and it is a 3-dimensional column vector corresponding to the OD values converted from RGB. Assume there are  $M$  biomarkers available in the multiplex IHC slide. We have  $M$  stain colors. Let  $\mathbf{b}$  be the combination weight vector of the stains and  $b_m, m = 1, \dots, M$  is the  $m_{th}$  element of  $\mathbf{b}$ . The typical unmixing problem thus is formulated as the following:

$$\min_{\mathbf{b}} \|\mathbf{y} - X\mathbf{b}\|_2^2. \quad (2)$$

Each column of  $X$  corresponds to a reference stain color sampled from the control slide of pure stain. As we discussed before, this linear system has solution only when the column of  $X$  is less than or equal to 3 for  $\mathbf{y} \in R^3$ . Therefore, meaningful regularization is needed for the linear system to have a solution.

The biomarker co-localization information provides a partition of  $\mathbf{b}$  into a set of groups  $g_1, g_2, \dots, g_N$ ,  $N$  being the total number of groups. Within each group, the biomarkers are known to have the co-localization possibility. We adopt this biological information to formulate the regularization term of the cost function. Let  $g_i$  be a  $q_i$ -dimensional column vector representing the combination weights of the stains within the  $i_{th}$  group and  $q_i$  be the number of stains within the group

$g_i$ . We thus have  $q_1 + q_2 \dots + q_N = M$ .  $x_i$  denotes the  $i_{th}$  group of reference colors, which is a  $3 \times q_i$  matrix. Fig.1 shows an example of the stain group setting. Six stains are available in this example ( $M = 6$ ). Two of them are co-localized membrane stains and two are co-localized nucleus stains. One is tumor cytokeratin stain and the rest is a membrane stain but only for B-cell. This information allows us to divide the stains into four groups ( $N = 4$ ) as shown in Fig.1. For instance,  $g_2$  contains  $b_2$  and  $b_3$  that are corresponding to the two co-located nucleus stains and  $x_2$  contains the reference color vectors for all the stains within the 2 $_{nd}$  group. However, the 4 $_{th}$  stain of B-cell marker does not co-localize with other biomarkers, so  $g_3$  only has one single member  $b_4$  and  $x_3$  is its reference color vector.

More specifically, the unmixing problem is formulated as the following convex optimization problem with the aforementioned notations:

$$\min_{\mathbf{b}} \|\mathbf{y} - \sum_{i=1}^N x_i g_i\|_2^2 + \lambda \sum_{i=1}^N \sqrt{q_i} \|g_i\|_2 \quad (3)$$

where  $\mathbf{b} = [b_1, b_2, \dots, b_M]^t = [g_1^t, g_2^t, \dots, g_N^t]^t$  and  $\|\cdot\|_2$  is the Euclidean norm with out squared. The first term in Eqn.3 solves for the linear system that is equivalent to [2], which minimize the least square error between the intensity of the raw image and the possible linear combination of the reference colors that approximates the raw image.  $\lambda$  is the regularization parameter that controls the amount of the group sparsity constraint in the second term. This model will act like LASSO at the group level. The entire groups will be dropped out when optimal  $\mathbf{b}$  (or  $\mathbf{g}$ ) is found, that is only a small number of  $g_i$  are non-zero.

Note that when the size of each group  $q_i = 1$ , the model becomes equivalent to lasso. In this case, no biological co-localization information is used in this model however the system remains to be solvable due to the sparsity constraints. The background noise is suppressed in this setting, comparing to the conventional Ruifrok’s method. In the experiment section, we’ll also demonstrate the efficacy of lasso unmixing by limiting the size of the group to 1.

Alternative direction method of multipliers (ADMM) algorithm [11] is used to solve Eqn.3. We implemented the algorithm in C++ to provide fast computation. It costs about 7 seconds to unmix a 750 by 1400 image on an Intel Core i7 1.87GHZ PC.

## 3 Experiments

### 3.1 Synthetic Data Experiment

As ground truth unmixing results are not available for real clinical data, we created a synthetic multiplex image from ground truth unmixed channels to validate our algorithm. We first synthetically generated six unmixed images as shown in the first row of Fig.2 C, following the stain co-localization and grouping rule in the example framework (Fig.1). The vectorized binary masks of the unmixed

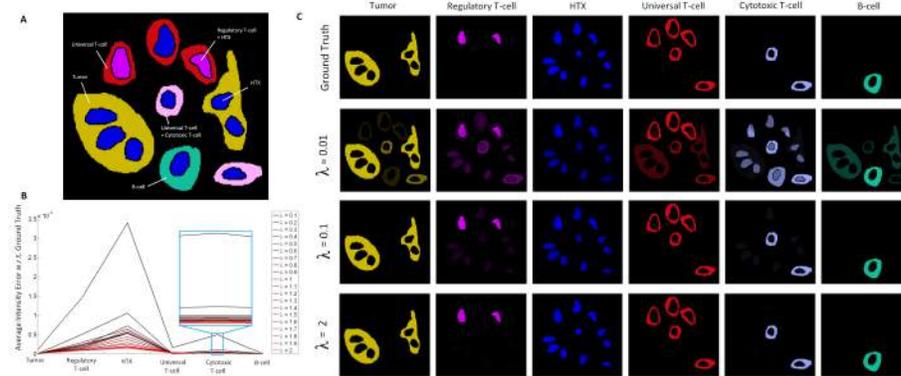


Fig. 2: Toy example. **A:** Image to be unmixed. **B:** The average intensity error of increasing  $\lambda$  for each channel. **C:** Unmixing results with different  $\lambda$ .

channels were multiplied by the reference color matrix to create the multiplex image in Fig.2 A. To demonstrate the algorithm performance w.r.t. the group sparsity regularization parameter  $\lambda$  variation, we plotted the average intensity error between the algorithm outputs and the ground truth unmixed channels in Fig.2 B for  $\lambda$  with in the range 0 to 2. The plot shows that the system has stable solutions when  $\lambda > 0.3$ . In Fig. 2 C, we also show the unmixing results for increasing  $\lambda$ . Note that when  $\lambda = 0.01$ , the system is close to deficient as in Eqn.2, hence unmixing errors are observed as shown in the second row of Fig.2 C.

### 3.2 Clinical Data Experiment

A clinical data set containing several different cancer tissue samples was used to demonstrate the proposed algorithm, including colorectal cancer, non small cell lung cancer and breast cancer that consist of 32 fields of view (FOV). The tissues were stained with the following assay as shown in Fig. 3: yellow chromogen for tumor cell cytoke-  
 Fig.5 shows the unmixing examples of decomposing the multiplexed image into single stain channels using modified Ruifrok’s method based on nearest neighbor color assignment and the proposed group sparsity method. Note that  $\lambda$  is set to be 0.5 through the clinical data experiments. Pixel discontinuities, unmixing errors and artifacts are observed from the modified Ruifrok’s method by solving multiple three color systems using the color similarity for system assignment. The proposed method instead solves one single

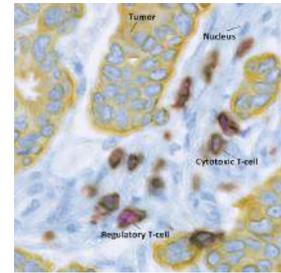


Fig. 3: Multiplexed tissue image real data example.

system for all the pixels hence leads to a smoother unmixed images, meanwhile maintains the biological constraints as wells as reduces the background noises due to the group sparsity regularization.

Since the cytotoxic T-cell is a subset of the universal T-cell, the green cytotoxic T-cell membrane marker always co-localizes with the orange universal T-cell membrane marker, but the orange marker can present alone. Fig. 6 shows an example of the orange only cell and the green and orange co-localized cell. We can see that the aglorithm is able to handle both cases. This demonstrates that the  $L_2$  norm constraint is used within the group to linearly separate the color mixture into different stain channels. Meanwhile, the modified Ruifrok’s method is prone to unmixing errors due to the hard assignment of the unmixing system based on color similarity.

As a special case example, the algorithm can also be used for less than or equal to three color unmixing. When the group size becomes 1, the algorithm is equivalent to Ruifrok’s unmixing plus a sparse constraint on the combination weights. The system can be solved by LASSO. We set the group size to 1 and compared to Ruifrok’s method [2] for two-stain unmixing on a clinical breast cancer data set containing 217 FOVs. The proposed technique consistently shows

better performance than Ruifrok’s method. Example results are shown in Fig.4 and much less background noise is observed using the proposed sparse unmixing method.

## 4 Conclusion

In this paper, we introduce a novel color unmixing strategy for multiplexed bright field histopathology images based on a group sparsity model. The biological co-localization information of the biomarkers is explicitly defined in the regularization term to produce biologically meaningful unmixing results. The

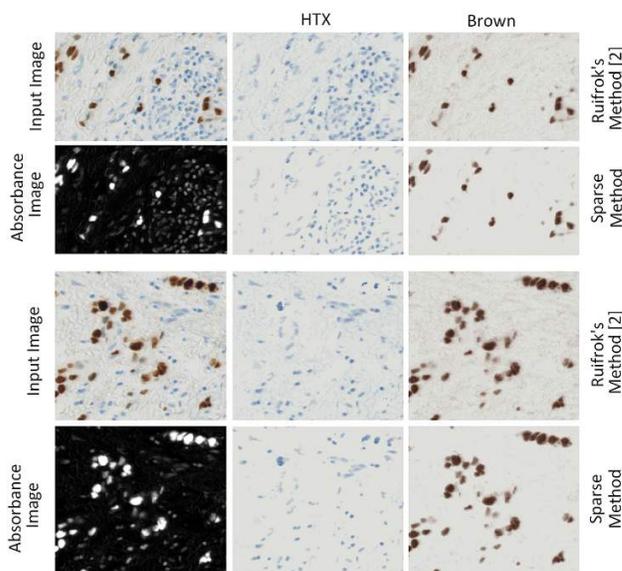


Fig. 4: Two-stain unmixing result comparisons when group size is 1.

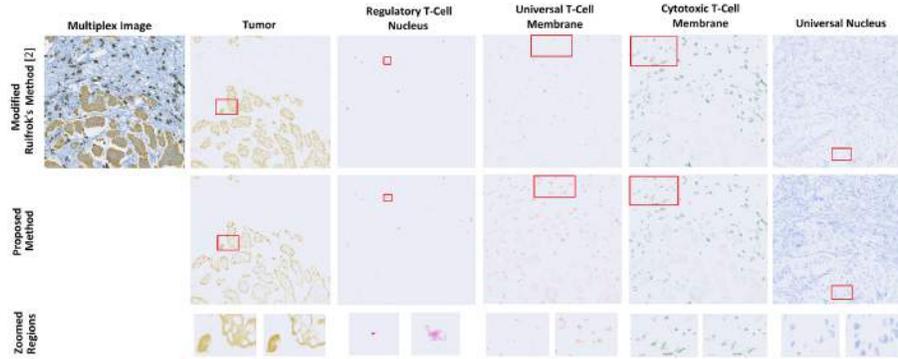


Fig. 5: Comparisons between the proposed group sparsity unmixing method and the modified Ruifrok’s method based on nearest neighbor color assignment. More completed nuclei (purple and blue channels) are observed in group sparsity unmixing results. Incorrect universal T-cell unmixing is observed in modified Ruifrok’s unmixing result due to the lack of co-localization constraint.



Fig. 6: Example unmixing of T-cell membrane co-localization case. **a**: The proposed group sparsity method without co-localization constraint (group size = 1). **b**: The proposed group sparsity method with co-localization constraint (group size = 2 for the two membrane stainings).

experiments of both synthetic and clinical data demonstrate the efficacy of the proposed algorithm in terms of accuracy and stability when compared to the existing techniques.

## References

1. J. Galon, *et al.*, “ Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome”, *Science*, 313(5795):1960-1964, 2006.
2. A. C. Ruifrok, *et al.* “Quantification of Histochemical Staining by Color Deconvolution”, *Anal. Quant. Cytol. Histol.*, 23:291-299, 2001.
3. A . Rabinovich, *et al.*, “Unsupervised Color Decomposition of Histologically Stained Tissue Samples”, *NIPS*,2003.
4. C. Wemmert, *et al.*, “Stain Unmixing in Brightfield Multiplexed Immunohistochemistry”, *ICIP*, 2013.
5. N. Kesheva, “A Survey of Spectral Unmixing Algorithms”. *Lincoln Laboratory Journal*, v.14(1), pp. 55-78, 2003.
6. M. Zortea and A. Plaza, “Spatial Preprocessing for Endmember Extraction”. *IEEE Trans. Geosci. Remote Sens.*, vol.47(8), pp. 2679-2693, 2009.

7. G. Foody and D. Cox, "Sub-pixel Land Cover Composition Estimation Using A Linear Mixture Model and Fuzzy Membership Functions". *Int. J. Remote Sens.*, vol.15(3), pp. 619-631, 1994.
8. Z. Yang et.al, "Blind Spectral Unmixing Based on Sparse Nonnegative Matrix Factorization". *IEEE Trans. Image Proc.*, vol.20(4), pp.1112-1125, 2011.
9. J. B. Greer, "Sparse Demixing of Hyperspectral Images". *IEEE Trans Image Proc.*, vol. 21(1), pp.219-228, 2012.
10. N. Simon, *et al.*, "A Sparse Group Lasso". *Journal of Computational and Graphical Statistics*, vol. 22(2), pp.231-245, 2013.
11. S. Boyd, *et al.*, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers". *Foundations and Trends in Machine Learning*, vol. 3(1), pp.1-122, 2010.

# Reduced-dose patient to baseline CT rigid registration in 3D Radon space

Guy Medan, Achia Kronman, Leo Joskowicz

The Rachel and Selim Benin School of Computer Science and Engineering  
The Hebrew University of Jerusalem, Israel  
gmedan/leo.josko @mail.huji.ac.il

**Abstract.** We present a new method for rigid registration of CT scans in Radon space. The inputs are the two 3D Radon transforms of the CT scans to be registered, one densely sampled and the other sparsely sampled. The output is the rigid transformation that best matches them. The algorithm starts by finding the best matching between each direction vector in the sparse 3D Radon transform and the corresponding direction vector in the dense 3D Radon transform. It then solves the system of linear equations derived from the direction vector pairs. Our method can be used to register two CT scans and to register a baseline scan to the patient with reduced-dose scanning without compromising registration accuracy. Our preliminary simulation results on the Shepp-Logan head phantom dataset and a pair of clinical head CT scans indicates that our 3D Radon space rigid registration method performs significantly better than image-based registration for very few scan angles and comparably for densely-sampled scans.

## 1 Introduction

Rigid registration of CT scans acquired at different times plays a key role in numerous medical applications, including diagnosis, follow-ups, surgery planning and simulations. Rigid registration methods include intensity-based iterative registration methods, fiducial-based registration methods, and frequency-based registration methods. These methods are used routinely in a clinical environment and yield accurate results in most cases.

Rigid registration plays an increasingly important role in image-guided interventional CT procedures. Interventional CT procedures include biopsies, catheter insertion, hematoma evacuation, and many more. Often times, a high-quality CT scan of the patient is available before the procedure. Since the diagnosis and procedure planning is usually performed on this CT scan, it is desirable to use it for guidance during the intervention. In addition, repeated CT scanning is often performed during the intervention to evaluate anatomical changes and determine the location of surgical tools. This results in the exposure of the patient to ionizing radiation, which has been shown to have risks for the patient [1, 2]. It is thus highly desirable to develop methods that reduce the radiation dose required for intraoperative CT registration.

Two main approaches have been developed for rigid registration of CT scans: 1) image-based, and 2) Radon-based. Image-based methods, by far the most popular, perform the registration by comparing the intensity values of both scans. To yield adequate results, they require both CT scans to be of high quality and free of image reconstruction artifacts. Radon-space methods use the CT scans Radon transform representation (sinograms) for the registration. They are not subject to image reconstruction artifacts and have the potential to yield robust and accurate results with reduced-dose scanning.

Previous research addresses rigid registration in Radon space with a variety of methods. Freiman et al. [3] describe a method for 2D/3D registration of X-Ray to CT images. Their method uses invariant features in Fourier space to find the rigid parameters with out-of-plane coarse registration followed by in-plane fine registration. Mao et al. [4] describe a slice-by slice registration method in 2D Radon space and its extension to 3D/3D registration for small angles or with implanted fiducials, and Mooser et al. [5] use an iterative optimization process to find the registration parameters in 3D Radon space. You et al. [6] investigate the mathematical relation between rigid movement in image space and Radon space and its invariants, and in [7] Fourier phase matching technique is applied to this relation to allow recovery of the rigid registration parameters of translation and rotation using approximations for small angles. The parameters are extracted in a stage-by-stage manner that employs the result of the previous stage in the evaluation of the next parameters, by decomposing the 3D problem into a series of 2D in-plane registrations.

In this paper we describe a new method for rigid registration of CT scans in 3D Radon space. The inputs are the two 3D Radon transforms of the CT scans to be registered, one densely sampled and the other sparsely sampled. The output is the rigid transformation that best matches the 3D Radon transforms. The algorithm first finds for each direction vector in the sparse 3D Radon transform the best matching direction vector in the dense 3D Radon transform. It then constructs and solves a system of linear equations from the direction vector pairs. The advantages of our method are: 1) it can be used both to register two CT scans and to register a baseline scan to the patient with reduced-dose scanning without compromising registration accuracy; 2) it supports fast on-line patient to baseline CT scan registration; 3) it is robust to noise, small anatomical differences, and has a wide convergence range because it relies on a closed-form solution of a set of linear equations instead of an iterative process. Our preliminary simulation results on the Shepp-Logan head phantom dataset and a pair of clinical head CT scans indicate that our Radon space method performs significantly better than image-based registration for very few scan angles.

## 2 Method

We first present the mathematical background of the Radon transform and its application to CT scan rigid registration. We then describe our new 3D Radon space method and algorithm details.

**Mathematical background.** We follow the definitions and notations in [6] for parallel-beam scanning. Let  $f : \mathfrak{R}^k \rightarrow \mathfrak{R}$  be an image function that maps  $k$ -dimensional location vectors to intensity values. Let  $H(\mathbf{n}, s)$  be the hyperplane defined by normal direction vector  $\mathbf{n}$  and distance  $s$  from the origin in  $k$ -dimensional space. The Radon transform  $R$  of image function  $f$  is a function  $Rf : S^{k-1} \times \mathfrak{R} \rightarrow \mathfrak{R}$  defined on unit sphere  $S^{k-1}$  of normal direction vector  $\mathbf{n}$  and distance  $s$ :

$$Rf(\mathbf{n}, s) = \int_{H(\mathbf{n}, s)} f(X) d\mu \quad (1)$$

where  $X$  is an  $k$ -dimensional vector and  $d\mu$  is the standard measure on  $H(\mathbf{n}, s)$ . Let  $f, g$  be two image functions such that  $g$  is a similarity transformation of  $f$ :

$$g(X) = f(\rho A_{\mathbf{r}, \theta} X + X_0) \quad (2)$$

where  $\rho > 0$  is the scaling constant,  $X_0 \in \mathfrak{R}^k$  is the constant offset vector, and  $A_{\mathbf{r}, \theta}$  is a unitary  $k \times k$  matrix in which rotations are represented by an axis vector  $\mathbf{r}$  and an angle  $\theta$  of rotation about  $\mathbf{r}$ . A well-known relation between the Radon transforms  $Rf, Rg$  of image functions  $f, g$  is:

$$Rg(\mathbf{n}, s) = \rho^{n-1} Rf(\mathbf{n}', \rho^{-1}(s + \mathbf{n} \cdot X_0)) \quad (3)$$

where  $\mathbf{n}$  and  $\mathbf{n}'$  are normal unit direction vectors satisfying:

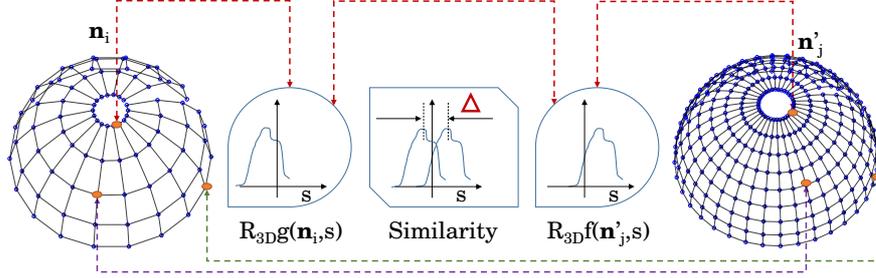
$$\mathbf{n}' = A_{\mathbf{r}, \theta}^{-1} \mathbf{n} \quad (4)$$

This relation can be interpreted as follows. For a given normal unit direction vector  $\mathbf{n}$ , the Radon transforms of  $f$  and  $g$ ,  $Rf(\mathbf{n}, s)$  and  $Rg(\mathbf{n}, s)$  are one-dimensional (1D) intensity signals of the distance  $s$ , which we denote by  $F_{\mathbf{n}}(s) = Rf(\mathbf{n}, s)$  and  $G_{\mathbf{n}}(s) = Rg(\mathbf{n}, s)$ . Without offset and scaling, i.e. when  $X_0 = \mathbf{0}$  and  $\rho = 1$ , Eq. 3 reduces to  $Rg(\mathbf{n}, s) = Rf(\mathbf{n}', s)$ , which means that the 1D signals  $F_{\mathbf{n}'}(s)$  and  $G_{\mathbf{n}}(s)$  are identical for direction vectors  $\mathbf{n}$  and  $\mathbf{n}'$ . That is, the projection in the direction  $\mathbf{n}'$  **before** the image  $f$  is rigidly rotated about the axis  $\mathbf{r}$  is **identical** to the projection in a different direction  $\mathbf{n}$  **after** the rotation, where the direction vectors  $\mathbf{n}, \mathbf{n}'$  are related by the same rotation  $A_{\mathbf{r}, \theta}$ . Furthermore, when the offset is not zero, that is  $X_0 \neq \mathbf{0}$ , we have:

$$G_{\mathbf{n}}(s) = F_{\mathbf{n}'}(s + \mathbf{n} \cdot X_0) \quad (5)$$

which means that  $F_{\mathbf{n}'}(s)$  remains the same and is shifted by  $\Delta = \mathbf{n} \cdot X_0$  for direction vectors  $\mathbf{n}$  and  $\mathbf{n}'$ .

In physical space, the image functions  $f, g$  are volumetric images; their Radon transform,  $R_{3D}f, R_{3D}g$  are 3D, and the direction vectors are points on the unit sphere  $S^2$  (Fig. 1). The spatial rigid transformation that relates  $f$  and  $g$  can be



**Fig. 1.** Illustration of the matching procedure of 3D Radon transforms.  $\mathbf{n}_i, \mathbf{n}_j'$  are direction vectors represented as points on the unit sphere. Each direction vector corresponds to a 1D projection signal  $R_{3D}g(\mathbf{n}_i, s)$  or  $R_{3D}f(\mathbf{n}_j', s)$ .

described by a translational offset  $X_0$ , a rotation axis vector  $\mathbf{r}$ , and a rotation angle about it,  $\theta$ . The goal of the rigid registration is to find the parameters  $(\mathbf{r}, \theta, X_0)$  for which Eq. 2 holds.

The rigid transformation that aligns images  $f$  and  $g$  can be computed by matching their 3D Radon transforms,  $R_{3D}f, R_{3D}g$ , instead of matching the images themselves. This is called rigid registration in 3D Radon space. Furthermore, since Eq. 2 reduces to Eq. 5 without scaling, we can match  $F_{\mathbf{n}_j'}(s)$  and  $G_{\mathbf{n}_i}(s)$  where  $\mathbf{n}_j'$  and  $\mathbf{n}_i$  are the direction vectors of the 3D Radon transforms. When these Radon transforms are equal, that is when  $G_{\mathbf{n}_i}(s) = F_{\mathbf{n}_j'}(s - \Delta_i)$  for offset  $\Delta_i$  and direction vectors  $\mathbf{n}_i, \mathbf{n}_j'$ , then, from Eqs. 4 and 5 we get:

$$\Delta_i = \mathbf{n}_i \cdot X_0 \quad (6)$$

$$\mathbf{n}_j' = A_{\mathbf{r}, \theta}^{-1} \mathbf{n}_i \quad (7)$$

which is a set of linear equations. The desired rigid transformation parameters  $(\mathbf{r}, \theta, X_0)$  can thus be computed by finding the pairs of direction vectors  $\mathbf{n}_i, \mathbf{n}_j'$  that satisfy Eqs. 6 and 7. Three pairs of independent direction vectors suffice to fully determine the resulting linear system of equations.

In general, the similarity between  $F_{\mathbf{n}_j'}$  and  $G_{\mathbf{n}_i}$  does not imply Eqs. 6 and 7. Indeed, two identical 1D signals from two different direction vectors need not correspond to the same region of the images  $f$  and  $g$ ; this similarity may be coincidental. However, such coincidental signal matches are unlikely in CT scans of human anatomy, which is rich in complexity and detail, and is radially asymmetric. For most direction vector pairs, the matchings correspond to the same image region after rigid transformation. In addition, not all direction vector pairs yield rigid transformations within the expected range.

**3D Radon rigid registration method.** Based on these observations, we propose the following method for 3D Radon rigid registration of image  $g$  to

image  $f$ . The inputs are the 3D Radon transforms of  $g$  and  $f$  defined by direction vectors  $\{\mathbf{n}_i\}_{i=1}^K$  and  $\{\mathbf{n}'_j\}_{j=1}^L$ . The goal is to build a set of matching projection pairs with relative displacements (Fig. 1).

For each direction vector  $\mathbf{n}_i$  we find the matching direction vector  $\mathbf{n}'_j$  and relative displacement  $\Delta_i$  for which the corresponding 1D signals  $G_{\mathbf{n}_i}$  and  $F_{\mathbf{n}'_j}$  are most similar. The result is a set of matching pairs of projections, along with their relative displacements  $\left\{ (F_{\mathbf{n}'_j}, G_{\mathbf{n}_i}, \Delta_i) \right\}_{i=1}^K$ .

Substituting each direction vector pair in Eqs. 6 and 7 yields an overdetermined set of linear equations. We compute the desired rigid transformation parameters  $(\mathbf{r}, \theta, X_0)$  by least-squares minimization. Offset  $X_0$  is estimated as  $\hat{X}_0 = (\mathbf{N}^T \mathbf{N})^{-1} \mathbf{N}^T \mathbf{\Delta}$  where  $\mathbf{N} = [\mathbf{n}_1 \dots \mathbf{n}_K]^T$  and  $\mathbf{\Delta} = [\Delta_1 \dots \Delta_K]^T$ . This solution minimizes the term  $\sum_{i=1}^K (\Delta_i - \mathbf{n}_i \cdot X_0)^2$ .

To estimate the rotation matrix  $A_{\mathbf{r}, \theta}$ , we define the  $3 \times 3$  matrix  $M = \sum_{i=1}^K \mathbf{n}'_{j(i)} \mathbf{n}_i^T$  and compute its Singular Value Decomposition  $M = U^T \Sigma V$ . From the values of  $U, V$  we obtain the estimate  $\hat{A}_{\mathbf{r}, \theta} = UV^T$ . This solution minimizes the term  $\sum_{i=1}^K (\mathbf{n}_i - A_{\mathbf{r}, \theta} \mathbf{n}'_{j(i)})^2$ .

A key property of this method is that it does not require a dense set of direction vectors of the 3D Radon transform of image  $g$ . Since the set of linear equations from which the transformation parameters are computed is of dimension 3, the set is overconstrained with more than three direction pairings. Using more direction pairs that are not outliers usually increases robustness and improves accuracy. This is akin to point-based rigid registration, in which more than three point pairs are used. The method is therefore suitable for finding the rigid registration between sparsely and densely sampled set of direction vectors for  $R_{3D}g$  and  $R_{3D}f$ . This is the situation of interventional CT procedures that require registering the patient with his/her earlier CT scan.

**3D Radon rigid registration algorithm.** We now describe a new 3D Radon rigid registration algorithm based on the method described above. The inputs are the two Radon transforms  $R_{3D}f$  and  $R_{3D}g$  of images  $f$  and  $g$ . The output is the rigid transformation  $(\mathbf{r}, \theta, X_0)$ . The algorithm consists of two steps. First, for each direction vector in the sparse  $R_{3D}g$  transform, we find the matching direction vector in the dense  $R_{3D}f$  transform. Then, we construct and solve the set of linear equations obtained by substituting each direction vector pair in Eqs. 6 and 7. We describe each step in detail next.

*1. Direction vectors pairing.* We evaluate the similarity of the two 1D signals from two direction vectors with Normalized Cross Correlation (NCC); the NCC value is the direction vectors pair score. For each direction vector  $\mathbf{n}_i$ , we select the direction vector  $\mathbf{n}'_j$  with the highest NCC score and compute its relative displacement  $\Delta_i$ . We define an index function  $match(i) = argmax_j \{NCC(R_{3D}g(\mathbf{n}_i, s), R_{3D}f(\mathbf{n}'_j, s))\}$  that pairs the direction vectors. To avoid searching all possible direction vectors  $\mathbf{n}'_j$ , we restrict the search to a neighbourhood of  $\mathbf{n}_i$  defined by  $\Phi(\mathbf{n}_i) = \{\mathbf{n}'_j : \cos^{-1}(\mathbf{n}_i \cdot \mathbf{n}'_j) < \varphi\}$  where  $\varphi$  is the largest expected relative orientation offset between the images.

2. *Transformation computation.* We construct and solve the set of linear equations obtained by substituting each direction vector pair in Eqs. 6 and 7 as described above. We use RANSAC to eliminate outliers. Since the resulting set of equations is 3-dimensional, we obtain high-quality results with a large number of RANSAC iterations in a short time. We set the RANSAC inliers threshold  $\psi$  for the relative angle  $\cos^{-1}(\mathbf{n}_i^T \hat{A}_{r,\theta} \mathbf{n}'_j)$  to be half the angular resolution of the densely-sampled set  $R_{3D}f$ .

**Computation of 3D Radon transforms from 2D sinograms.** The 3D Radon transform  $R_{3D}g$  of the baseline image  $f$  can be efficiently computed from the 2D sinograms of the slices as described in [5]. Our algorithm achieves the desired rigid registration with a sparse sampling of  $R_{3D}g$ , which takes place in the CT scanner. The reduced number of direction vectors required thus leads to a significant reduction of the radiation dose without compromising accuracy and without having to reconstruct the image  $g$ .

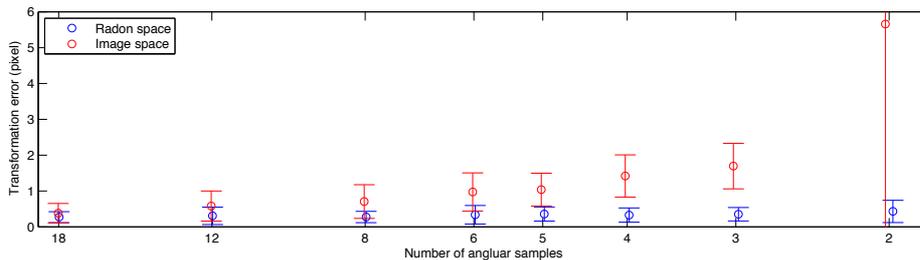
### 3 Experimental results

To evaluate our method, we conducted the following simulation experiments in Matlab. We use the Shepp-Logan head phantom dataset whose size is  $256 \times 256 \times 256$  voxels with intensity values in  $[0, 1]$ . To simulate data acquisition noise, we add  $N(0, 0.05)$  Gaussian noise to the dataset to obtain the baseline image  $f$ . We then apply to  $f$  a series of rigid transformation including both rotations and translations to generate a new set of images  $h$  (Table 1). For each image  $h$ , we generate its sinograms by projection and create a set of new sparsely-sampled images  $\{g_l\}$ . Each image  $g_l$  is created by filtered back projection from 2 to 18 projection directions instead of the usual 180 required for full-resolution reconstruction. The resulting images include significant reconstruction artifacts.

Parameter Setting	Axis vector $\hat{r}$ <i>not normalized</i>	Angle $\theta$ <i>degrees</i>	Translation $X_0$ <i>pixels</i>
1	(1, 2, 100)	1.0	(2, 0, -1)
2	(34, 45, 39)	-7.7	(14, 15.2, -18.5)
3	(23, -12, 1)	13.2	

**Table 1.** Parameters and settings for the ground-truth transformations. A total of 18 rigid transformations (all possible  $3 \times 3 \times 2$  possible combinations).

We then perform two sets of rigid registrations: one in image space using Matlab’s `imregtform` and the other one in Radon space with our method. In image space, we compute the rigid transformation parameters between the original phantom image  $f$  and the reconstructed and transformed phantom images  $g_l$ . In Radon space we applied our registration method on the 3D Radon transforms of  $f$  and  $g_l$ . The 3D Radon transform of  $f$  was computed at an angular resolution of  $1^\circ$  for 180 2D projection directions per slice, for a total of 32,400 direction vectors. The resulting rigid transformations of the image-based and Radon-based



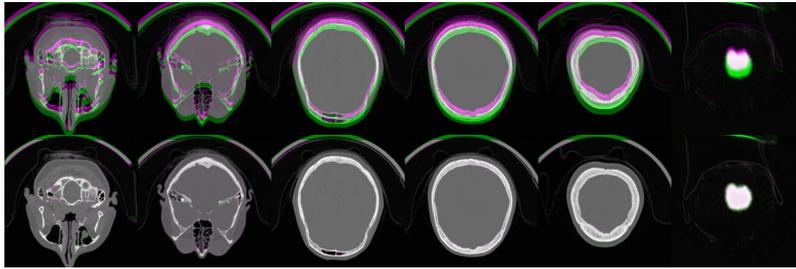
**Fig. 2.** Plot of the image-based (red) and 3D Radon (blue) rigid transformation error with respect to the ground-truth transformation (vertical axis) as a function of the number of scan angles (horizontal axis), 18 to 2.

registration were then applied to the original image  $f$ . The resulting images were compared to the ground-truth rigid transformations of  $f$  by computing the RMS error between 3D voxel coordinates. The experiment was repeated between 3 to 10 times for each rigid transformations and sparse sampling settings. Fig. 2 shows the results. Note that our Radon space method performs significantly better than image-based registration for very few scan angles ( $< 12$ ). Note also that our algorithm handles well rotation offsets  $> 10^\circ$ , which are challenging for other algorithms that rely on small-angle approximations.

In a second experiment, we test our method on a pair of CT scans from a patient head taken at two different times. The voxel sizes of the CT scans are  $0.42 \times 0.42 \times 0.67 \text{ mm}^3$ . Prior to registration, we removed the scanning bed from both images, as the bed is not rigidly attached to the patient and introduces errors in the Radon space signals. In practice, this can be done automatically, since the Radon transform of the bed without the patient is always the same and can be precomputed and subtracted from the patient scan. We then computed the image-based registration of the full-resolution scans and our Radon space registration with the second image from 18 angles using our method and compared the results (Fig. 3). The RMSE between the image space registration and our method is 0.64mm. This indicates that our method yields results comparable to full-resolution image-space registration with about 10% of the radiation dose of the second scan.

## 4 Conclusions

We have presented a new 3D Radon space rigid registration method for CT scans registration. Our method can be used to register two existing CT scans and to register a baseline CT scan to the patient for interventional CT procedures. The key characteristic of our method is that it allows the registration of a full-resolution CT scan to a sparsely-sampled CT scan without compromising the registration accuracy. This results in a significant X-ray dose reduction when registering a diagnostic CT scan to the patient prior to image-guided interventional CT procedures. Another advantage of our method is that it supports fast



**Fig. 3.** Overlay of six representative slices from two head CT scans of the same patient: before registration (top row), after 3D Radon space registration (bottom row).

on-line patient to baseline CT scan registration, as most of the 3D Radon space computation on the baseline image can be performed prior to the intervention. Our preliminary results indicate that a very small number of scan directions are sufficient to obtain voxel size accuracy, that the method has a wide convergence range, and that it is robust to small anatomical differences.

Future work includes extending our formulation of parallel-beam CT acquisition to cone beam and spiral acquisition, conducting more extensive simulation experiments, and conducting studies with actual CT sinograms.

## References

1. Chodick, G., Ronckers, C.M., Shalev, V., Ron, E., et al.: Excess lifetime cancer mortality risk attributable to radiation exposure from computed tomography examinations in children. *IMAJ-RAMAT GAN*- **9**(8) (2007) 584.
2. Mettler Jr, F.A., Wiest, P.W., Locken, J.A., Kelsey, C.A.: CT scanning: patterns of use and dose. *Journal of Radiological Protection* **20**(4) (2000) 353.
3. Freiman, M., Pele, O., Hurvitz, A., Werman, M., Joskowicz, L.: Spectral-based 2d/3d X-ray to CT image rigid registration. In: *SPIE Medical Imaging, International Society for Optics and Photonics* (2011) 79641B–79641B.
4. Mao, W., Li, T., Wink, N., Xing, L.: CT image registration in sinogram space. *Medical physics* **34**(9) (2007) 3596–3602.
5. Mooser, R., Forsberg, F., Hack, E., Székely, G., Sennhauser, U.: Estimation of affine transformations directly from tomographic projections in two and three dimensions. *Machine vision and applications* **24**(2) (2013) 419–434.
6. You, J., Lu, W., Li, J., Gindi, G., Liang, Z.: Image matching for translation, rotation and uniform scaling by the radon transform. In: *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on. Volume 1., IEEE* (1998) 847–851.
7. Lu, W., Fitchard, E.E., Olivera, G.H., You, J., Ruchala, K.J., Aldridge, J.S., Mackie, T.R.: Image/patient registration from (partial) projection data by the fourier phase matching method. *Physics in Medicine and Biology* **44**(8) (1999) 2029.

# Predicting cross-task behavioral variables from fMRI data using the $k$ -support norm

Michail Misyrlis<sup>1,3</sup>, Anna B. Konova<sup>2,3</sup>, Matthew B. Blaschko<sup>4,5</sup>,  
Jean Honorio<sup>6</sup>, Nelly Alia-Klein<sup>3</sup>, Rita Z. Goldstein<sup>3</sup>, and Dimitris Samaras<sup>1</sup>

<sup>1</sup>Department of Computer Science, Stony Brook University, Stony Brook, NY, USA

<sup>2</sup>Department of Psychology, Stony Brook University, Stony Brook, NY, USA

<sup>3</sup>Icahn School of Medicine at Mount Sinai, New York, NY, USA

<sup>4</sup>Center for Visual Computing, École Centrale Paris, France

<sup>5</sup>Équipe Galen, INRIA Saclay, Île-de-France, France

<sup>6</sup>CSAIL, MIT, Cambridge, MA, USA

**Abstract.** Sparsity regularization allows handling the curse of dimensionality, a problem commonly found in fMRI data. In this paper, we compare LASSO ( $\ell_1$  regularization) and the recently introduced  $k$ -support norm on their ability to predict real valued variables from brain fMRI data for cocaine addiction, in a principled model selection setting. Furthermore, in the context of these two regularization methods, we compare two loss functions: squared loss and absolute loss. With the squared loss function,  $k$ -support norm outperforms LASSO in predicting real valued behavioral variables measured on an inhibitory control task given fMRI data from a different task, designed to capture emotionally-salient reward responses. The absolute loss function leads to significantly better predictive performance for both methods in almost all cases and the  $k$ -support norm leads to more interpretable and more stable solutions often by an order of magnitude. Our results support the use of the  $k$ -support norm for fMRI analysis and the generalizability of the I-RISA model of cocaine addiction.

**Keywords:** Functional magnetic resonance Imaging (fMRI), Regularization, Sparse representations

## 1 Introduction

Functional magnetic resonance imaging (fMRI) is a widely used modality within the field of neuroimaging, that measures brain activity by detecting associated changes in blood oxygenation. One of the goals of fMRI data analysis is to detect correlations between brain activation and a task the subject performs during the scan.

The main challenges in statistical fMRI data analysis [1–4] are (i) the curse of dimensionality (ii) a small number of samples, due to the high cost of fMRI acquisition, and (iii) high levels of noise, such as system noise and random neural activity.

Sparsity regularizers are key statistical methods for improving predictive performance in the event that the number of observations is substantially smaller than the dimensionality of the data, as is the case in fMRI analysis. In this paper we compare the most frequently applied sparsity regularizer developed in the statistics literature, LASSO [5], with the  $k$ -support norm [6], a recently introduced method which is less biased towards sparse solutions.

The  $k$ -support norm can be viewed as a generalization of LASSO when  $k = 1$  and ridge regression when  $k = d$ , where  $d$  is the dimensionality of the data. The  $k$ -support norm has previously been used in [6] for classification. It was first used for fMRI data modelling in [7] with a specific choice  $k$  parameter of the norm, that is, the method was not tested in a model selection setting. In both cases the  $k$ -support norm was used with the squared loss function.

We focus on comparing LASSO with the  $k$ -support norm in order to establish the latter regularizer’s superiority in analyzing fMRI data. We use two loss functions, namely the squared error and the absolute error. The advantage of the absolute error loss is that it is more robust, in that it penalizes outliers less than squared loss, while still retaining convexity which guarantees finding the global optimum. We compare the methods not only in their predictive accuracy but also in the interpretability and stability of their results.

Our contribution in this paper is threefold. First, we introduce a novel method, the  $k$ -support norm with absolute error. Second, this is the first attempt to compare LASSO with the  $k$ -support norm in a principled model selection setting. Finally, to the best of our knowledge this is the first application of the  $k$ -support norm to a real valued response variable in a challenging clinical setting where the fMRI signal collected during *one task* is used to predict behavioral responses collected at a *different time* during a *second task*.

The neuroscientific motivation for our experiments is the exploration of human drug addiction. Basic studies have led to a theoretical model of human drug addiction, characterized by Impaired Response Inhibition (RI) and Salience Attribution (SA) (hence, I-RISA) [8]. According to the model, the skew in SA is predictive of impaired RI, together contributing to excessive drug use and relapse, core clinical symptoms of cocaine addiction. We use the fMRI data from a SA task (drug Stroop) in order to predict behavioral data in a RI task (color-word Stroop) collected at a different time, hence providing further evidence to support the I-RISA model.

## 2 Methods

We denote by  $X \in \mathbb{R}^{n \times d}$  the design matrix of  $n$  samples each with  $d$  dimensions; we denote by  $y \in \mathbb{R}^n$  the vector of targets.

A basis of statistical inference is the application of regularized risk, in which a loss function is evaluated over a sample of data and is linearly combined with a regularizer that penalizes some norm of the prediction function as in (Eq. (1)),

where the first term is the loss function and the second is the penalty term:

$$\min_{\beta} f(\beta, X, y) + \lambda J(\beta). \quad (1)$$

The scalar parameter  $\lambda > 0$  controls the degree of regularization and  $J$  is a scalar valued function monotonic in a norm of  $\beta \in \mathbb{R}^n$ . Sparsity regularization is a key family of priors over linear functions that prevents overfitting and aids interpretability of the resulting models [5, 6]. Key to the mathematical understanding of sparsity regularizers is their interpretation as convex relaxations to quantities involving the  $\ell_0$  norm, which simply counts the number of non-zero elements of a vector. One of the most important sparsity regularizers is the LASSO [5], where  $\lambda J(\beta) = \lambda \|\beta\|_1$ . In many learning problems of interest, LASSO has been observed to shrink too many of the  $\beta$  variables to zero. In the presence of a group of highly correlated variables, LASSO may prefer a sparse solution. However including all correlated variables in the model could potentially lead to higher predictive accuracy [6] and the  $k$ -support norm provides a way of calibrating the cardinality of the regression vector  $\beta$  so as to include more variables.

The  $k$ -support norm can be computed as

$$\|\beta\|_k^{sp} = \left( \sum_{i=1}^{k-r-1} (|\beta|_i^\downarrow)^2 + \frac{1}{r+1} \left( \sum_{i=k-r}^d |\beta|_i^\downarrow \right)^2 \right)^{\frac{1}{2}} \quad (2)$$

where  $|\beta|_i^\downarrow$  is the  $i$ th largest element of the vector and  $r$  is the unique integer in  $\{0, \dots, k-1\}$  satisfying

$$|\beta|_{k-r-1}^\downarrow > \frac{1}{r+1} \sum_{i=k-r}^d |\beta|_i^\downarrow \geq |\beta|_{k-r}^\downarrow. \quad (3)$$

In this paper, we consider LASSO and the  $k$ -support norm with two loss functions: the squared error  $f(\beta, X, y) = \|y - X\beta\|_2^2$  and the absolute error  $f(\beta, X, y) = \|y - X\beta\|_1$ .

In practice, we approximate the absolute error with a Huber type smoothing around zero to ensure differentiability.

### 3 Experimental Set-up

In this section we present our experiments and the data sets used in them. Our experiments aim at providing empirical evidence for the support of the I-RISA model.

We use the fMRI drug-word task described in [9, 10]. The neuropsychological experiment for cocaine addiction data set has a block design, which includes eight sessions, with each of them having different conditions. The two varying conditions are the monetary reward (50¢, 25¢, 1¢ and 0¢) and the cue shown (drug words, neutral words). The session consists of an initial screen displaying

the monetary reward and then presenting a sequence of forty words in four different colors (yellow, blue, red or green). The subject was instructed to press one of four buttons matching the color of the word they had just read. The subjects were rewarded for correct performance depending on the monetary condition.

We use the behavioral responses of the same subjects in a color-word task [11], a classic task of inhibitory control. In this task the subjects pressed for ink color of color words printed in either their congruent or incongruent colors. Four colors and words (red, blue, yellow and green) were used in all possible combinations. Both congruent and incongruent stimuli were presented randomly. The subjects performed four consecutive runs of this task. As there were 12 incongruent events in each run of 200 events, each subject’s data contained up to 48 incongruent events.

For 38 control subjects and 74 cocaine abusers, we use the fMRI data from the drug-word task, to predict color-word behavioral variables.

Our experimental setting consists of 500 trials with an 85% / 15% random split between training and test sets. We perform model selection on the training set. That is, for each combination of parameters ( $\lambda \in \{10^i : i = -2, \dots, 8\}$  for LASSO,  $\lambda \in \{10^i : i = -2, \dots, 8\}$ ,  $k \in \{1, 2, 3, 6, 12, 100, 200, 300, 600\}$  for  $k$ -support norm), we do a leave-one-subject-out cross validation on the samples that constitute the training set. We measure the correlation between the predicted and the true response variables on the training set. The parameter setting that leads to the highest correlation is used on the whole training set in order to learn a set of weights for each method, which are then applied on the test set. Finally, we measure the correlation between the predicted and the true response variables on the test set. We report the mean correlation on the holdout test samples and its standard error across the 500 random permutations in Sec. 4. We note that the same sample randomization is used for both LASSO and  $k$ -support norm.

In **experiment 1** we use the fMRI contrast drug > neutral words, averaged over monetary reward condition, to predict the conflict effect in the subjects’ reaction time on the color-word task, defined as the difference in time between correctly performing the task for congruent and incongruent events. We use the Insula, Hippocampus Complex, Amygdala and ACC, part of the brain’s limbic (emotion) circuit, as regions of interest (ROIs) for this experiment. These regions are chosen on the basis of previous studies on independent datasets that showed limbic system modulation by drug-related cues, eg. drug words [12].

In **experiment 2** we use the fMRI contrast 50¢ > 0¢, averaged over word type condition, in order to predict the subjects’ responses on the color-word task, defined as the difference in percent accuracy between performing the task for congruent and incongruent events. We use the Basal Ganglia and Thalamus, part of the brain’s reward circuit, as ROIs for this experiment. We chose these ROIs on the basis of previous studies on independent datasets that showed reward system modulation by primary and secondary reinforcers, including money [13].

## 4 Results

We compare the performance of the two methods in Table 1 for the first experiment and Table 2 for the second experiment.

Mean Correlation,  $D > N$ , Conflict effect on Reaction Time

Control Subjects			
Norm / Loss	Squared	Absolute	$p$
LASSO	0.16 (0.02)	0.27 (0.02)	<0.01
$k$ -support	0.22 (0.02)	0.24 (0.02)	<0.05
$p$	<0.001	0.21	
Cocaine Subjects			
Norm / Loss	Squared	Absolute	$p$
LASSO	0.27 (0.01)	0.37 (0.01)	<0.001
$k$ -support	0.33 (0.01)	0.36 (0.02)	<0.001
$p$	<0.001	0.96	

**Table 1.** Mean (SE) correlation over 500 random permutations of the samples between the predicted and the actual conflict effect on the reaction times for drug > neutral using the limbic ROI, for all combinations of regularizers and loss functions. The  $p$ -values were computed with a Wilcoxon signed rank test between the 500 correlation values for the two combinations of regularizer and loss function in the preceding rows or columns. Based on the  $p$ -values, there is a statistically significant difference between absolute loss predictions and squared loss predictions and between LASSO and  $k$ -support norm with the squared loss function in both cocaine and control subjects.

With the squared loss function, the  $k$ -support norm outperforms LASSO for almost all cases, while when combined with the absolute loss function, the regularizers do not significantly differ in their predictive performance. The absolute loss function, for both regularizers, leads to correlations that are significantly higher than those with the squared loss function in almost all cases.

We report the fraction of non-zero weights that were selected by each method for over 50% of the 500 trials in Tables 3 and 4 for the first and the second experiment respectively.

We average the weights assigned to the voxels over the 500 permutations and then compute the cumulative distribution function (CDF) for those weights. We threshold the CDF at 0.9 and visualize the weights of the voxels up to that threshold<sup>1</sup> in Fig. 1. The overly sparse solutions of the LASSO (Fig. 1(b), 1(d)) lead to models that cannot be interpreted as easily as the solutions of the  $k$ -support norm method (Fig. 1(a), 1(c)).

In the presence of correlated features, the degree of sparsity of the solution can be tuned with the  $k$ -support norm in order to include several highly correlated features. In contrast, LASSO tends to pick one representative feature with

<sup>1</sup> Due to space constraints we include one representative example out of two for each experiment. The omitted results are qualitatively similar.

Mean Correlation,  $50\% > 0\%$ , Conflict effect on Accuracy

Control Subjects			
Norm / Loss	Squared	Absolute	$p$
LASSO	0.25 (0.02)	0.09 (0.02)	<0.001
$k$ -support	0.26 (0.02)	0.09 (0.02)	<0.001
$p$	0.42	0.78	

Cocaine Subjects			
Norm / Loss	Squared	Absolute	$p$
LASSO	0.22 (0.02)	0.42 (0.02)	<0.001
$k$ -support	0.27 (0.01)	0.41 (0.02)	<0.001
$p$	<0.001	0.78	

**Table 2.** Mean (SE) correlation over 500 random permutations of the samples between the predicted and the actual response variables for  $50\% > 0\%$  using the Basal Ganglia, Thalamus ROI, for all combinations of regularizers and loss functions. The  $p$ -values were computed with a Wilcoxon signed rank test between the 500 correlation values for the two combinations of regularizer and loss function in the preceding rows or columns. Based on the  $p$ -values there is a statistically significant difference between absolute loss predictions and squared loss predictions and between  $k$ -support and LASSO with the squared loss in cocaine subjects only.

Voxel Selection Stability,  $D > N$ , Conflict effect on Reaction Time

Control Subjects		
Norm / Loss	Squared	Absolute
LASSO	0.0004	0.0007
$k$ -support	0.0029	0.0018

Cocaine Subjects		
Norm / Loss	Squared	Absolute
LASSO	0	0.0023
$k$ -support	0.0058	0.0734

**Table 3.** Voxel Selection stability over 500 random permutations of the samples for drug  $>$  neutral using the limbic ROI, for all combinations of regularizers and loss functions. The fraction of voxels which are selected for more than 50% of the 500 trials are presented. The higher values reported for  $k$ -support norm indicate that it makes more stable voxel selection than LASSO over different training sets.

no guarantee of consistency in feature selection across different splits of the data samples into training and test sets. In all cases the fraction of non-zero weights selected by the  $k$ -support norm is higher than that of LASSO, indicating that the  $k$ -support norm method leads to more stable solutions as compared to those obtained with LASSO.

---

Voxel Selection Stability, 50% > 0%, Conflict effect on Accuracy

---

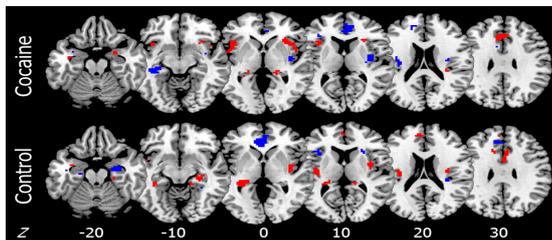
Control Subjects		
Norm / Loss	Squared	Absolute
LASSO	0.0004	0.0050
$k$ -support	0.0037	0.0083
Cocaine Subjects		
Norm / Loss	Squared	Absolute
LASSO	0.0008	0.0013
$k$ -support	0.0223	0.0122

**Table 4.** Voxel Selection stability over 500 random permutations of the samples for 50% > 0% using the Basal Ganglia, Thalamus ROI, for all combinations of regularizers and loss functions. The fraction of voxels which are selected for more than 50% of the 500 trials are presented. The higher values reported for  $k$ -support norm indicate that it makes more stable voxel selection than LASSO over different training sets.

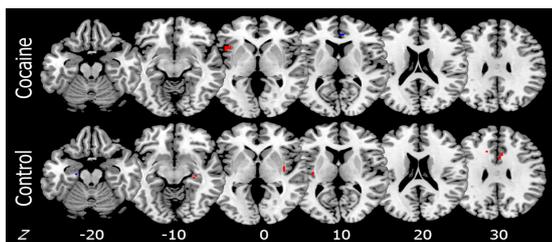
## 5 Discussion

In our experiments, in almost all cases, the  $k$ -support norm outperforms LASSO in predicting the behavioral measures given fMRI data when combined with squared loss, while when combined with the absolute loss, the predictive accuracy of the two regularizers does not differ significantly. The absolute loss led to higher predictions than squared loss for both regularizers for almost all cases. The LASSO leads to sparse solutions, since it tends to pick one feature per group of correlated features. On the other hand, the  $k$ -support norm allows calibrating the cardinality of the solutions and thus can select more interpretable groupings of correlated features and also leads to more stable results across different training sets. Thus, our results support the further exploration of the  $k$ -support norm for fMRI analysis.

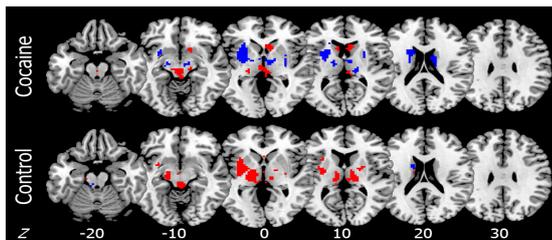
We also provide further evidence to support the I-RISA model of drug addiction, whereby the skew in SA in cocaine abusers, as indexed by fMRI response to drug words and monetary rewards, two motivationally salient stimuli, is predictive of RI, as indexed by response slowing and accuracy on a task requiring inhibitory control (the color-word Stroop). Specifically, we show that in cocaine users, response to drug words in voxels located in limbic brain regions, such as the anterior insula and ACC implicated in emotion processing and emotion regulation, was predictive of slower responses on the RI task (Exp. 1), while response to money in voxels located in reward-related brain regions, such as the putamen implicated in habits, was predictive of lower accuracy on the RI task (Exp. 2).



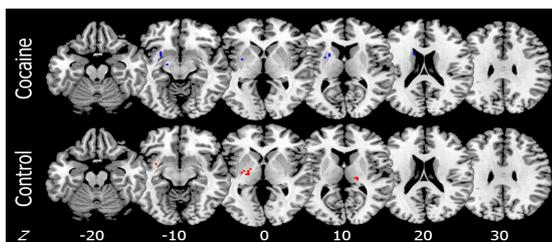
(a) Most predictive voxels in Exp. 1 using the  $k$ -support norm with the Absolute Loss



(b) Most predictive voxels in Exp. 1 using the LASSO with the Absolute Loss



(c) Most predictive voxels in Exp. 2 using the  $k$ -support norm with the Squared Loss



(d) Most predictive voxels in Exp. 2 using the LASSO with the Squared Loss

**Fig. 1.** Visualization of the most predictive voxels in Exp. 1 and Exp. 2 over the 500 permutations, where red indicates positive, blue indicates negative in the relationship between the tasks. The degree of sparsity of the solution can be tuned with the  $k$ -support norm, thus leading to models ((a), (c)) that are easier to interpret than those of LASSO ((b), (d)). (Best viewed in color)

## References

1. J Honorio, D Tomasi, R Goldstein, H Leung, and D Samaras, “Can a single brain region predict a disorder?,” *IEEE Transactions on Medical Imaging*, 2012.
2. S Song, Z Zhan, Z Long, J Zhang, and L Yao, “Comparative study of svm methods combined with voxel selection for object category classification on fmri data,” *PloS one*, vol. 6, no. 2, pp. e17191, 2011.
3. Andreas Bartels and Semir Zeki, “Brain dynamics during natural viewing conditions a new guide for mapping connectivity in vivo,” *Neuroimage*, vol. 24, no. 2, pp. 339–349, 2005.
4. D R Hardoon, J Mourao-Miranda, M Brammer, and J Shawe-Taylor, “Unsupervised analysis of fmri data using kernel canonical correlation,” *NeuroImage*, vol. 37, no. 4, pp. 1250–1259, 2007.
5. R Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
6. A Argyriou, R Foygel, and N Srebro, “Sparse prediction with the  $k$ -support norm,” *Neural Information Processing Systems*, 2012.
7. K Gkirtzou, J Honorio, D Samaras, RZ Goldstein, M Blaschko, et al., “fmri analysis of cocaine addiction using k-support sparsity,” in *International Symposium on Biomedical Imaging*, 2013.
8. RZ Goldstein and ND Volkow, “Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex,” *The American journal of psychiatry*, vol. 159, no. 10, 2002.
9. RZ Goldstein, D Tomasi, S Rajaram, LA Cottone, L Zhang, T Maloney, F Telang, N Alia-Klein, and ND Volkow, “Role of the anterior cingulate and medial orbitofrontal cortex in processing drug cues in cocaine addiction,” *Neuroscience*, vol. 144, no. 4, pp. 1153–1159, 2007.
10. AB Konova, SJ Moeller, D Tomasi, MA Parvaz, N Alia-Klein, ND Volkow, and RZ Goldstein, “Structural and behavioral correlates of abnormal encoding of money value in the sensorimotor striatum in cocaine addiction,” *European Journal of Neuroscience*, vol. 36, no. 7, 2012.
11. SJ Moeller, D Tomasi, J Honorio, ND Volkow, and RZ Goldstein, “Dopaminergic involvement during mental fatigue in health and cocaine addiction,” *Translational psychiatry*, vol. 2, no. 10, pp. e176, 2012.
12. HW Chase, SB Eickhoff, AR Laird, and L Hogarth, “The neural basis of drug stimulus processing and craving: an activation likelihood estimation meta-analysis,” *Biological psychiatry*, vol. 70, no. 8, pp. 785–793, 2011.
13. X Liu, J Hairston, M Schrier, and J Fan, “Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies,” *Neuroscience & Biobehavioral Reviews*, vol. 35, no. 5, pp. 1219–1236, 2011.

# Fast MRI for repeated scans

Lior Weizman<sup>1</sup>, Leo Joskowicz<sup>1</sup> and Dafna Ben-Bashat<sup>2</sup>

<sup>1</sup> School of Eng. and Computer Science, Hebrew University of Jerusalem, Israel

<sup>2</sup> Tel-Aviv Sourasky Medical Center, Tel-Aviv, Israel  
lweizm45@cs.huji.ac.il

**Abstract.** Radiological assessment of brain disease progression and response to therapy is often performed with repeated MRI scans acquired every few weeks/months. In these longitudinal studies, each scan is acquired anew without taking into account the information present in previous scans. However, many image regions remain unchanged from one time point to the next, making the difference image between the time points sparse. We present a new algorithm that speeds up the MR acquisition process of the repeated scan by using the data acquired in the baseline scan of the same patient. Our method uses the baseline scan to identify regions of interest in the repeated scan. These regions are partially acquired, followed by reconstruction process that speeds up the entire scanning procedure. Our experimental study on 16 pairs of baseline/follow-up MR scans shows that the image quality of the MR scans produced by our method with a speedup factor of up to 3.5 are within the imaging variability of the scanner.

## 1 Introduction

Radiological assessment of brain disease progression and response to therapy is often performed with repeated MRI scans acquired every few weeks/months [1, 2]. The MR scanning protocol consists of several pulse sequences, resulting in various imaging contrasts such as T1 and T2 weighted images. At each time point, an entire, multi-sequence scan is acquired anew without taking into account the information present in the previous scans. However, many image regions remain unchanged from one time point to the next. This results in unnecessarily long scanning times. Our hypothesis is that the difference between the current scan and the previous scan data can be sparse, and the previous scan can be advantageously used to speed up the scanning time of a repeated scan with minimal compromise of the image quality.

During the scan acquisition, the MR signals are stored in a spatial frequency domain called *k-space* [3]. The speed at which the *k-space* values can be acquired is inherently limited by the required image contrast, resolution and coverage, the properties of human tissues, hardware limitations, and safety issues [4]. The demand for multi-sequence MRI under these fundamental speed limits has given rise to plethora of methods for MRI speed-up.

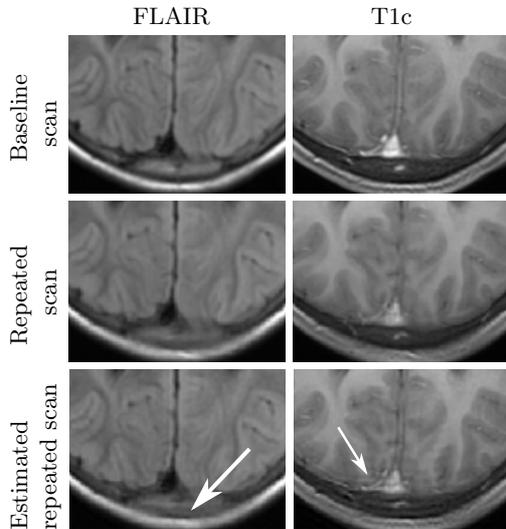


Fig. 1: FLAIR (left) and contrast-enhanced T1 (right) full-time scanning baseline (top) and repeat (middle) slices are presented vs. fast acquired one (bottom), with a speed-up factor of 3.5. Note that the images are very similar with the exception of very small artifacts, indicated by the arrows. The reconstruction quality of contrast-enhanced T1 (T1c) images falls below that of FLAIR images, due to changes in contrast agent injection rate between baseline and repeated scans.

Methods for speeding-up MR acquisition with existing hardware are mostly based on partial k-space acquisition. This approach consists of selectively sampling the k-space followed by the estimation of the missing k-space samples using a priori knowledge on the image. Keyhole methods update data in the center of the k-space more frequently than in other parts, thus providing high temporal resolution but lower spatial resolution [3]. Methods for fast dynamic MRI use the previous image frames in the time-series to complete the missing k-space values [5, 6]. However, these methods compromise the spatial coverage and/or spatial image resolution.

In Compressed sensing (CS) MRI [7, 8], the basic premise is that MRI can be sparsely represented in a transform domain, thereby requiring only a subset of the k-space for reconstruction. The sparsity of MRI in different transform domains has been used by others for various applications. For example, Bilgic et al. [9] exploit the fact that certain characteristics of the scanned object do not change across pulse sequences. They propose a reconstruction algorithm that relies on Bayesian compressed sensing to jointly reconstruct a set of images from under-sampled k-space data.

One concept that has not been previously researched is the use of the patient's baseline scan to speed up the acquisition of his/her repeated scan. In many clinical diagnostic applications, patients are longitudinally scanned to determine pathology changes between time points and to evaluate treatment efficacy. In most cases, there is substantial similarity between the baseline and the repeated scans. The changes usually occur in a confined region around the tumor or pathology, while the rest of the image remains the same. Consequently, the data from the baseline scan can be advantageously used to speed up the scanning time of a repeated scan without compromising image quality. Fig. 1 shows representative results of our method.

In this paper we present a new method to effectively exploit the data from the baseline scan to reduce the acquisition time of the repeated scan. The main contributions of this paper are: 1) the use of baseline scan data for repeated scan acquisition speed-up; 2) no compromise on image quality in clinically important regions; and 3) experimental results obtained from 16 MR clinical brain show reliable reconstruction results with speedup factor of 3.5 or less. To the best of our knowledge, this is the first attempt to speed-up an MR scan with the same patient’s baseline scan.

## 2 Method

The inputs to our method are the baseline brain scan of the patient, consisting of two or more MRI pulse sequences with several contrasts, such as T1-weighted and T2-weighted, and a single pulse sequence from a repeated scan. The outputs are the remaining imaging contrasts of the repeated scan, acquired in fast acquisition mode. In the following, we assume for simplicity that: 1) the same MR pulse sequences are acquired in the baseline and repeated scans; 2) all the pulse sequences have same number of slices, denoted by  $N_s$ ; 3) the k-space is sampled with Cartesian sampling trajectories; 4) the differences between scans of the same patient are mainly due to pathological changes and; 5) the acquisition of the repeated scan is spatially matched to the baseline scan. We discuss the validity of these assumptions later, in Section 4.

The method consists of two step. First, we use the first sequence in the repeated scan to detect the changes from the baseline scan so that slices with significant changes will be acquired in full-time mode. Second, we acquire slices that are similar to their corresponding ones in the baseline scan in fast acquisition mode, thus speeding up the entire scanning process.

### 2.1 Detection of slices for fast acquisition

This step automatically detects slices with significant changes based on the baseline scan and the first imaging contrast of the repeated scan. We compare the slices with the following measure. The difference between two corresponding  $s_i$ -th slices of these pulse sequences is defined as:

$$I_{diff}(s_i) = I_f(s_i) - I_b(s_i) \quad (1)$$

where  $I_b$  and  $I_f$  are the corresponding baseline and repeated matching imaging contrasts.

We focus on the outliers of  $I_{diff}(s_i)$  to define a measure of difference between the scans. In the literature we find many approaches for outliers detection, such as Chauvenet’s criterion and Grubbs’ test. For simplicity, we use the interquartile range method [10] to identify and reject outliers from data. The outliers in  $I_{diff}(s_i)$  are the voxels  $O = \{o_1, \dots, o_N\}$ .

We then perform a connected components analysis on  $O$  to obtain  $L = \{l_1, \dots, l_K\}$  regions of outliers. Regions with less than  $N_v$  voxels are automatically rejected and considered to be spatially isolated outliers. For the remaining

regions,  $I_b(s_i)$  and  $I_f(s_i)$ , we model the grey-level distribution of pathologies with a Gamma distribution with parameters  $k$  and  $\theta$  [11]. Since there is no closed form for the parameter's Maximum Likelihood Estimation, we use the approximation in [12]:

$$\hat{k} = \frac{a_0 + a_1Q + a_2Q^2}{Q(b_0 + b_1Q + Q^2)} \quad (2)$$

$$\hat{\theta} = \frac{1}{k \cdot N_{l_j}} \sum_{i=1}^{N_{l_j}} x_i; \quad (3)$$

where  $X = \{x_i\}_{i=1}^{N_{l_j}}$  are the outliers of group  $l_j$ ,  $Q$  is :

$$Q = \ln\left(\frac{1}{N_{l_j}} \sum_{i=1}^{N_{l_j}} x_i\right) - \frac{1}{N_{l_j}} \sum_{i=1}^{N_{l_j}} \ln(x_i) \quad (4)$$

and  $a_i, b_i$  are as defined in [12]. We estimate these values for both the baseline and the repeated scans imaging contrasts, since we model change as either progression or regression of the pathology.

As was observed by Prastawa et al. [11] the closer the distribution of the outliers group is to the Gamma distribution, the higher the probability that its slice contains changes in pathology. Therefore, to identify outliers that represent pathology changes, we measure the distance of every group of voxels from Gamma distribution defined by the estimated parameters of the group with the KL-distance [13]. The difference measure for slice is the sum of the measurements of different regions in the slice.

This results in a measure of difference between a previously acquired slice and its corresponding repeated scan slice. Slices are then sorted by this measure, and the  $N_{full}$  most different slices will be fully scanned in the remaining pulse sequences to avoid compromising image contrast in them. The remaining slices,  $G_s^{fast} = s_1, \dots, s_{N_{fast}}$ , will be scanned in a fast scanning procedure. The user-defined parameter  $N_{full}$  defines the trade-off between the fast acquisition and the number of slices acquired slices in full image scan mode.

## 2.2 Fast acquisition of selected slices

The input to this step is a list of slices to be acquired in the fast acquisition mode. For these slices, the k-space lines are randomly sampled with variable density, so that the sampling density is higher near the k-space origin. The missing k-space lines are taken directly from the baseline scan.

Specifically, let  $S_i^p(k_m)$  and  $S_i^r(k_m)$  be the k-spaces of  $i$ -th slice of the baseline and the repeated scans, and let  $m$  the index of the phase encode line number  $k_m$ . Let  $C$  be the set of random sampled k-space lines and let  $N_k$  be the number of items in  $C$ . The estimation of the k-space for the slices rapidly acquired is:

$$\hat{S}_i^r(k_m) = \begin{cases} S_i^r(k_m) & k_m \in C \\ S_i^p(k_m) & \text{otherwise} \end{cases} \quad (5)$$

for  $i \in G_s^{fast}$ . The inverse Fourier transform of  $\hat{S}_i^r$  is the estimated  $i$ -th image slice. The estimated repeated scan,  $\hat{I}_f$ , consists of  $N_{full}$  full-time acquired slices and  $N_{fast}$  estimated slices.

### 2.3 Speed-up factor computation

We compute the speedup of our method at the pulse sequence level, where we assume that the acquisition of a single line in the k-space takes the same time for all pulse sequence types. Let  $N_l$  be the number of k-space lines required for Nyquist rate acquisition. We define the speedup factor by the time required to acquire a pulse sequence at Nyquist rate, which is the number of slices,  $N_s$ , times  $N_l$ , divided by the acquisition time of the new method. This value consists of the number of full-time acquires slice,  $N_{full}$  times  $N_l$ , plus the number of fast-acquired slices,  $N_{fast}$  times the number of k-space lines acquired with the proposed method,  $N_k$ :

$$F = \frac{N_s \cdot N_l}{N_{full} \cdot N_l + N_{fast} \cdot N_k} \quad (6)$$

For example, with typical values of  $N_s = 40$ ,  $N_l = 320$ ,  $N_{full} = 7$ ,  $N_k = 43$ , our method can acquire a scan 3.5 times faster than sampling at Nyquist rate.

## 3 Experimental Results

We conducted a retrospective quantitative evaluation of our method with clinical MRI datasets. Experiments involved six patients, three of them with Optic Pathway Gliomas (OPG) and three with Glioblastoma Multiforme (GBM). Each patient was scanned with a 1.5T General Electric MRI system, with a multi-sequence protocol at intervals of several months at the Tel-Aviv Medical Center, Israel. In total, 16 pairs of scans were acquired. Each scan consisted of T2-weighted, contrast-enhanced T1 (T1c), and FLAIR images. Each dataset has  $512 \times 512 \times 38$  voxels with voxel size of  $0.5mm \times 0.5mm \times 5.0mm$ .

Studies have shown that T2-weighted images are most sensitive for detecting brain pathology [14]. Therefore, we set this image contrast to be fully acquired with no speed-up in the repeated scan. The acquisitions of the remaining imaging contrasts, T1c and FLAIR were accelerated with our method.

The k-space samples of the scans were generated synthetically from images obtained at the Nyquist rate by applying an inverse Fourier transform. We set the minimum number of outlier voxels in a group to  $N_v = 100$  and the number of slices to be fully scanned to  $N_{full} = 10$ . The parameters  $N_s = 38$  and  $N_l = 512$  are explicitly derived from the dimensions of the data. In our experiments, data intensity values were normalized to the range of  $[0, 1]$  to compensate for grey-level variations between time-points. Experiments were performed with the original data, where no noise was added or filtered.

We performed two experiments. First, we set  $N_k = 15$  to obtain a speed-up factor of method to 3.5 and visually examined the results. Fig. 1 illustrates

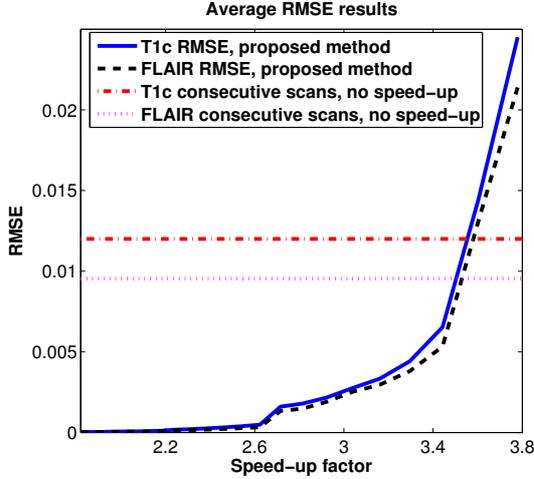


Fig. 2: Average RMSE of 16 baseline-follow-up pairs of scans for various speed-up factors. The horizontal lines are the RMSE values between two scans with no radiological changes.

the resulting images. Note that the images are very similar to each other, with the exception of very small artifacts which arise from misregistration errors and grey-level differences between the baseline and repeated scans.

In the second experiment, we examined values of  $N_k$  in the range of 2 and 200, (corresponding to speed-up factors in the range of 1.8 and 3.7), and quantitatively evaluated the performance of the method as a function of the speed-up factor in terms of root mean square error (RMSE) vs. full-time scanning at Nyquist rate. The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_j (I_f(j) - \hat{I}_f(j))^2}{\sum_j (I_f(j))^2}} \quad (7)$$

where  $j$  is the spatial slice index.

To provide a RMSE reference value, we additionally computed the RMSE between a different set of 16 pairs of registered scans acquired at Nyquist rate of patients who exhibited no radiological changes between scans. The average RMSE values measure the variability between two scans of the same patient in which there are no actual changes between scans. The resulting values, shown as the red and pink horizontal lines in Fig. 2, are  $RMSE_v^{T1c} = 1.2 \times 10^{-2}$  for the contrast-enhanced T1, and  $RMSE_v^{FLAIR} = 9.5 \times 10^{-3}$  for the FLAIR.

Fig. 2 shows the tradeoff between the speed-up factor and the RMSE. The horizontal lines show the reference RSME values described above. We observe that for a speed-up factor of up to 3.5 the RMSE values are within the variability of the scanner. Our method's performance is higher for the FLAIR images than for the T1c images because the grey-level values of the T1c images are highly depend on the contrast agent injection rate during acquisition, which may vary between scans. As a result, the k-space values of the T1c baseline scan used to estimate part of the repeated scan's k-space of this image produce some imaging

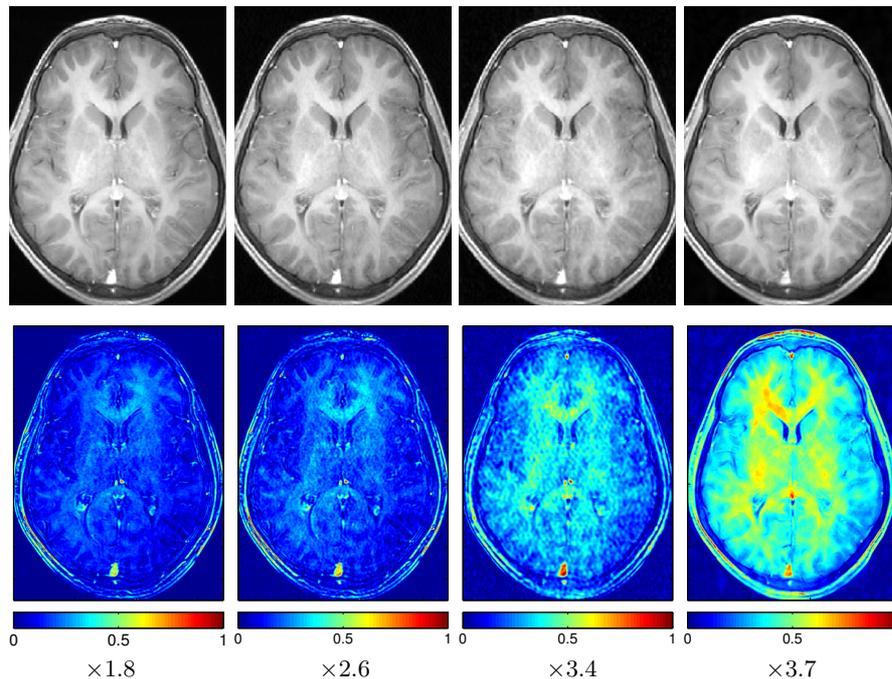


Fig. 3: Reconstruction results (top) and absolute difference images vs. full-time scanned image (bottom) of T1c representative image, for speed-up factors of (left to right): 1.8, 2.6, 3.4 and 3.77. The color bars at the bottom represent the grey level percentage estimation error divided by 100.

artifacts, despite the normalization performed in our experiments. Fig. 3 shows reconstruction results of T1c for representative speed-up factors.

#### 4 Discussion and Conclusions

We have described a new method for MR acquisition speed-up of a repeated brain scan. Our method finds the most similar slices between the baseline scan and the repeated scan and speeds-up their acquisition in the repeated scan. Our results show that a speedup of up to 3.5 is achievable within the imaging scanner variability. To the best of our knowledge, this is the first attempt to speed-up an MR scan with baseline patient scans.

We now address two practical issues regarding the implementation of the method. First, the method assumes that the baseline and repeated scans are spatially matched. This spatial matching can be obtained by reproducing the past scan’s slice positions for the scan being acquired. This feature is currently offered by some MRI vendors [15].

Second, we assume that changes between baseline and repeated scans are caused due to pathology changes. However, changes may be the result of dif-

ferences in field inhomogeneity, coil properties, different scanners, different sequences, etc. In our method we normalize the grey level intensity values of the scans to match the same scale, in order to minimize the effect of external resources on the changes between the scans.

We note that in the special case of longitudinal studies, scans are in many cases acquired in the same scanning site with the same scanning protocol to minimize the effect of external parameters on the resulted clinical follow-up. This assumption, together with refined image normalization, is sufficient to avoid reconstruction artifacts which may arise due to mixing k-space samples of two scans acquired with a few months gap.

While our method may compromise on image quality to speed up the acquisition process, this compromise is limited to regions that may have lower clinical relevance, as slices with high clinical importance are fully scanned. This is in contrast to existing methods that make the compromise across the entire image.

In addition, the proposed method is independent with and complimentary to CS methods for rapid MRI and can work in conjunction with them to speed-up the acquisition. Future work includes speeding-up additional pulse sequences and implementing our method on a real MR scanner.

## References

1. Weizman, L., Ben Sira, L., Joskowicz, L., Constantini, S., Precel, R., Shofty, B., Ben Bashat, D.: Automatic segmentation, internal classification, and follow-up of optic pathway gliomas in mri. *Medical image analysis* **16**(1) (2012) 177–188
2. Weizman, L., Sira, L.B., Joskowicz, L., Rubin, D.L., Yeom, K.W., Constantini, S., Shofty, B., Bashat, D.B.: Semiautomatic segmentation and follow-up of multicomponent low-grade tumors in longitudinal brain mri studies. *Medical physics* **41**(5) (2014) 052303
3. Bernstein, M.A., King, K.F., Zhou, X.J.: *Handbook of MRI pulse sequences*. Access Online via Elsevier (2004)
4. Chronik, B.A., Rutt, B.K.: Simple linear formulation for magnetostimulation specific to MRI gradient coils. *Magnetic resonance in medicine* **45**(5) (2001) 916–919
5. Liang, D., Ying, L.: Compressed-sensing dynamic MR imaging with partially known support. In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, IEEE* (2010) 2829–2832
6. Liang, D., et al.: k-t ISD: Dynamic cardiac MR imaging using compressed sensing with iterative support detection. *Magnetic resonance in medicine* **68**(1) (2012) 41–53
7. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid mr imaging. *Magnetic resonance in medicine* **58**(6) (2007) 1182–1195
8. Huang, J., Zhang, S., Metaxas, D.: Efficient mr image reconstruction for compressed mr imaging. *Medical Image Analysis* **15**(5) (2011) 670–679
9. Bilgic, B., Goyal, V.K., Adalsteinsson, E.: Multi-contrast reconstruction with bayesian compressed sensing. *Magnetic Resonance in Medicine* **66**(6) (2011) 1601–1615
10. Upton, G., Cook, I.T.: *Understanding statistics*. Oxford University Press (1996)

11. Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., Gerig, G.: Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology* **10**(12) (2003) 1341–1348
12. Greenwood, J.A., Durand, D.: Aids for fitting the gamma distribution by maximum likelihood. *Technometrics* **2**(1) (1960) 55–65
13. Kullback, S., Leibler, R.A.: On information and sufficiency. *The Annals of Mathematical Statistics* **22**(1) (1951) 79–86
14. Edelman, R.R., Hesselink, J.R., Zlatkin, M.B., Crues, J.V.: *Clinical magnetic resonance imaging. Volume 2.* Saunders Elsevier Philadelphia (2006)
15. McMillan, K., Uike, M., Tao, X., Kosugi, S., Okuda, H.: MR efficiency becomes critical as healthcare costs, scanner time demand increases. *Signa Pulse of MRI Autumn 2010* (2010) S10–S13

# Multisite Disease Classification with Functional Connectomes via Multitask Structured Sparse SVM

Takanori Watanabe<sup>1</sup>, Daniel Kessler<sup>2</sup>, Clayton Scott<sup>1</sup>, Chandra Sripada<sup>2</sup>

<sup>1</sup>Department of EECS, <sup>2</sup>Department of Psychiatry  
University of Michigan, Ann Arbor, MI 48109, USA  
{takanori,kesslerd,clayscot,csripada}@umich.edu

**Abstract.** There is great interest in developing imaging-based methods for diagnosing neuropsychiatric conditions. To this end, multiple data-sharing initiatives have been launched in the neuroimaging field, where datasets are collected across multiple imaging sites. While this enables researchers to study the disorders of interest with substantial sample size, it also creates new challenges since the data aggregation process introduces various sources of site-specific heterogeneities. To address this issue, we introduce a multitask structured sparse support vector machine (SVM) that uses resting state functional connectomes (FCs) as the features for predicting diagnostic labels. Specifically, we employ a penalty that accounts for the following two-way structure that exists in a multisite FC dataset: (1) the 6-D *spatial structure* in the FCs captured via either the GraphNet, fused Lasso, or the isotropic total variation penalty, and (2) the *inter-site* structure captured via the multitask  $\ell_1/\ell_2$ -penalty. To solve the resulting high dimensional optimization problem, we introduce an extension to a recently proposed algorithm based on the alternating direction method. The potential utility of the proposed method is demonstrated on the multisite ADHD-200 dataset.

**Keywords:** Multitask learning, structured sparsity, support vector machine, resting-state fMRI, alternating direction method

## 1 Introduction

In this work, we are interested in a supervised classification problem, where the goal is to predict the diagnostic status of an individual using functional connectomes (FCs) derived from resting-state fMRI (rs-fMRI) [4]. Fortunately, with various data sharing projects emerging in the neuroimaging community [12, 15], we have access to training data of unprecedented sample size. However, such community-wide collaborative efforts typically involve aggregating data from multiple imaging sites, which introduces several sources of systematic confounds, such as variability in the scanner quality, image acquisition protocol, subject demographics, etc. In order to effectively make use of these multisite datasets, it is important to train the classifiers in a way that accounts for these site-specific heterogeneities. To this end, we propose a classification framework that adopts a multitask learning (MTL) approach [5, 8, 10, 13].

The idea behind MTL is to *jointly* train multiple tasks in order to improve classification performance, under the assumption that the tasks are related to each other in some sense. Recently, MTL methods have been successfully applied in brain decoding [8, 13], where the *participants* from a multi-subject fMRI study are treated as the tasks. The underlying assumption here is that the brain regions that are activated from a stimulus will share similar patterns across different tasks/subjects. In contrast to these works, the method we propose in this work treats the *sites* from which the rs-fMRI scans are collected as the tasks.

## 2 Material and Methods

To generate the FCs, we used the grid-based parcellation scheme adopted by Watanabe *et al.* in [16], which involves 347 nodes defined on the standard MNI template; Fig. 1 provides a schematic representation of this parcellation scheme. Each node represents a 15mm diameter sphere with 33 voxels, and is placed throughout the entire brain with a spacing of 18×18×18mm (voxel resolution is 3×3×3mm). A regional time-series is assigned on each node by spatially averaging the BOLD signals, and FCs of size  $p = \binom{347}{2} = 60,031$  are obtained by computing all pairwise Pearson correlations between the time-series of the nodes.

### 2.1 Supervised Learning and the Multitask Framework

Suppose we are given  $K$  supervised learning tasks, where for each task  $k = 1, \dots, K$ , we are given  $n_k$  input/output pairs  $\{(\mathbf{x}_i^k, y_i^k)\}_{i=1}^{n_k} \in (\mathbb{R}^p \times \{\pm 1\})^{n_k}$ . In the context of our work,  $\mathbf{x}_i^k$  and  $y_i^k$  represent the FC and the diagnostic label of the  $i$ -th subject from the  $k$ -th site, respectively. The goal is to jointly learn  $K$  linear classifiers of the form  $f_k(\mathbf{x}) = \text{sign}(\langle \mathbf{w}^k, \mathbf{x} \rangle)$ , where  $\mathbf{w}^1, \dots, \mathbf{w}^K \in \mathbb{R}^p$  are task-specific weight vectors obtained by solving the following optimization problem:

$$\arg \min_{\mathbf{w}^1, \dots, \mathbf{w}^K \in \mathbb{R}^p} \sum_{k=1}^K \frac{1}{n_k} \sum_{i=1}^{n_k} \ell(y_i^k \langle \mathbf{w}^k, \mathbf{x}_i^k \rangle) + \mathcal{R}(\mathbf{w}^1, \dots, \mathbf{w}^K).$$

The first term here is the *pooled empirical risk* of a convex margin-based loss  $\ell: \mathbb{R} \rightarrow \mathbb{R}_+$  and the second term  $\mathcal{R}: \mathbb{R}^{pK} \rightarrow \mathbb{R}_+$  is a penalty function that enforces certain kind of structure on the weight vectors. In this work, we employ the *hinge-loss*  $\ell(t) = \max(1 - t, 0)$  from the well known support vector machine (SVM) classifier, although other convex margin-based losses can be used as well.

For brevity, we define a functional  $\mathcal{L}(\mathbf{Y}^k \mathbf{X}^k \mathbf{w}^k) := \sum_{i=1}^{n_k} \ell(y_i^k \langle \mathbf{w}^k, \mathbf{x}_i^k \rangle)$  which aggregates the empirical loss from the  $k$ -th task, where  $\mathbf{X}^k \in \mathbb{R}^{n_k \times p}$  denotes the design matrix for the  $k$ -th task and  $\mathbf{Y}^k \in \{\pm 1\}^{n_k \times n_k}$  is defined as  $\mathbf{Y}^k := \text{diag}(y_1^k, \dots, y_{n_k}^k)$ . Also for conciseness, let  $\underline{\mathbf{w}} \in \mathbb{R}^{pK}$  denote the vector obtained by stacking the weight vectors  $\{\mathbf{w}^k\}_{k=1}^K$  together. In this work, we focus on convex penalty functions of the form:  $\mathcal{R}(\underline{\mathbf{w}}) = \gamma \sum_{k=1}^K \mathcal{R}_1(\mathbf{w}^k) + \lambda \mathcal{R}_2(\underline{\mathbf{w}})$ , where  $\gamma, \lambda \geq 0$  are hyperparameters. Thus the objective function can be written as:

$$\arg \min_{\underline{\mathbf{w}} \in \mathbb{R}^{pK}} \sum_{k=1}^K \frac{1}{n_k} \mathcal{L}(\mathbf{Y}^k \mathbf{X}^k \mathbf{w}^k) + \gamma \sum_{k=1}^K \mathcal{R}_1(\mathbf{w}^k) + \lambda \mathcal{R}_2(\underline{\mathbf{w}}). \quad (1)$$

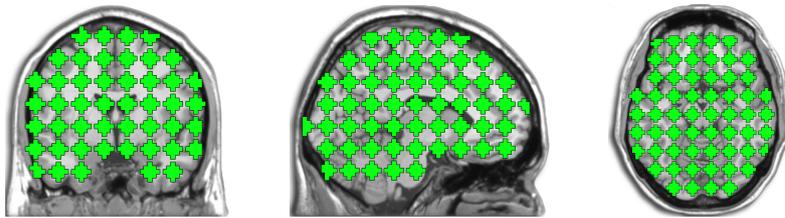


Fig. 1: The brain parcellation scheme adopted in this work. The green regions represent (pseudo)-spherical nodes each encompassing 33 voxels.

The first penalty  $\mathcal{R}_1$  allows us to encode prior knowledge about the *intra-task* structure of the data. While various penalties such as GraphNet (GN), fused Lasso (FL), and isotropic total variation (TV) have been applied successfully in the fMRI literature [1, 6, 9, 16], these penalties by themselves do not account for the *inter-task* structure of the dataset (FL is also known as anisotropic total variation). Thus a second penalty  $\mathcal{R}_2$  is included in (1), which allows us incorporate a notion of “task-relatedness” by enforcing some form of structure on  $\mathbf{w}$ .

For the intra-task penalty  $\mathcal{R}_1$ , following the recent work of [16], we account for the 6-D spatial structure of FCs (defined by pairs of points in 3-D) by employing either the GN or FL penalty, which can be expressed in the following form:

$$\mathcal{R}_1(\mathbf{w}^k) = \frac{1}{q} \|\mathbf{C}\mathbf{w}^k\|_q^q = \begin{cases} \text{GraphNet} & \text{if } q = 2 \\ \text{Fused Lasso} & \text{if } q = 1, \end{cases}$$

where  $\mathbf{C}$  denotes a 6-D finite differencing matrix. The idea behind GN and FL is to promote spatial contiguity by penalizing the differences among neighboring coordinates of the FC. Similarly, the TV penalty, which is a rotationally invariant counterpart of the FL penalty, can also be used to encourage spatial contiguity; see [9] for its closed form expression.

## 2.2 Structured Sparsity with Group Variable Selection

We propose to integrate the *structured sparsity* framework introduced in [16] with the popular multitask  $\ell_1/\ell_2$ -penalty [5, 10]. Specifically, for the inter-task penalty  $\mathcal{R}_2$ , we use  $\mathcal{R}_2(\mathbf{w}) = \sum_{j=1}^p \|\mathbf{w}_j\|_2$ , which is the so-called  $\ell_1/\ell_2$ -penalty. Here  $\mathbf{w}_j \in \mathbb{R}^K$  is a vector formed by stacking the  $j$ -th weight vector coefficients across the  $K$  tasks. This penalty has the appealing *group variable selection* property [5, 10], which promotes learning features that are relevant across all sites, thereby simplifying interpretation of the selected features. At the same time, the actual weights associated with a given correlation can vary across site, in contrast to training a single classifier over a pooled dataset.

## 2.3 Optimization Algorithm

To solve the proposed large scale optimization problem, we apply the *alternating direction method of multipliers* (ADMM) algorithm [2] introduced in [16], but with a minor modification. The complete algorithm is outlined in Alg. 1. We note that this section focuses on GN and FL, but the ADMM algorithm for TV differs only in line 5 of Alg. 1, but the details are omitted for lack of space.

**Alg. 1** ADMM for Multitask Structured Sparse SVM

---

```

1: Initialize variables, assign hyperparameters  $\lambda, \gamma \geq 0$ 
2: repeat
3:   for  $k = 1, \dots, K$  do
4:      $\mathbf{w}^k \leftarrow ((\mathbf{X}^k)^T \mathbf{X}^k + 2\mathbf{I}_p)^{-1} \{ (\mathbf{Y}^k \mathbf{X}^k)^T (\mathbf{v}_1^k - \mathbf{u}_1^k) (\mathbf{v}_2^k - \mathbf{u}_2^k) + \mathbf{A}^T (\mathbf{v}_4^k - \mathbf{u}_4^k) \}$ 
        $\triangleright$  solve using matrix inversion Lemma
5:      $\mathbf{v}_3^k \leftarrow \begin{cases} \text{apply Equation (3)} & \text{if } q=1 \text{ (FL)} \\ \rho(\gamma \mathbf{B} + \rho \mathbf{I})^{-1} \tilde{\mathbf{C}} (\mathbf{v}_4^k - \mathbf{u}_3^k) & \text{if } q=2 \text{ (GN)} \end{cases}$ 
6:      $\mathbf{v}_1^k \leftarrow \text{Prox}_{\ell/(\rho n_k)}(\mathbf{Y}^k \mathbf{X}^k \mathbf{w}^k + \mathbf{u}_1^k) \triangleright \text{Prox}_{\tau \ell}(t) := \begin{cases} t & \text{if } t > 1 \\ 1 & \text{if } 1 - \tau \leq t \leq 1 \\ t + \tau & \text{if } t < 1 - \tau \end{cases}$ 
7:      $\mathbf{v}_4^k \leftarrow (\tilde{\mathbf{C}}^T \tilde{\mathbf{C}} + \mathbf{I}_{\tilde{p}})^{-1} (\tilde{\mathbf{C}}^T [\mathbf{v}_3^k + \mathbf{u}_3^k] + \mathbf{A} \mathbf{w}^k + \mathbf{u}_4^k) \triangleright$  solve using FFT
8:   end for
9:   for  $j = 1, \dots, p$  do
10:     $\mathbf{v}_{2,j} \leftarrow \text{vsoft}_{\lambda/\rho}(\mathbf{w}_j + \mathbf{u}_{2,j}) \triangleright \text{vsoft}_{\tau}(t) := \max(1 - \frac{\tau}{\|t\|_2}, 0) t, \quad t \in \mathbb{R}^K$ 
11:   end for
12:   for  $k = 1, \dots, K$  do  $\triangleright$  dual variable update
13:     $\mathbf{u}_1^k \leftarrow \mathbf{u}_1^k + \mathbf{Y}^k \mathbf{X}^k \mathbf{w}^k - \mathbf{v}_1^k$ 
14:     $\mathbf{u}_2^k \leftarrow \mathbf{u}_2^k + \mathbf{w}^k - \mathbf{v}_2^k$ 
15:     $\mathbf{u}_3^k \leftarrow \mathbf{u}_3^k + \mathbf{v}_3^k - \tilde{\mathbf{C}} \mathbf{v}_4^k$ 
16:     $\mathbf{u}_4^k \leftarrow \mathbf{u}_4^k + \mathbf{A} \mathbf{w}^k - \mathbf{v}_4^k$ 
17:   end for
18: until stopping criterion is met

```

---

To apply Alg. 1, we employ the *data augmentation+masking* strategy that was proposed in [16]. In brief, the idea behind this method is that as it stands, the ADMM algorithm for solving the objective function (1) with the GN, FL, or TV penalty will require the inversion of the Laplacian matrix  $\mathbf{C}^T \mathbf{C}$ , which is prohibitively large. Thus we rewrite the GN/FL penalty as  $\mathcal{R}_1(\mathbf{w}^k) = \|\mathbf{B} \tilde{\mathbf{C}} \mathbf{A} \mathbf{w}^k\|_q^q$ , where  $\mathbf{A}$  is an *augmentation matrix*,  $\tilde{\mathbf{C}}$  is the finite differencing matrix for the augmented  $\mathbf{w}^k$ , and  $\mathbf{B}$  is a diagonal masking matrix that ensures the penalty remains unaffected, *i.e.*,  $\|\mathbf{B} \tilde{\mathbf{C}} \mathbf{A} \mathbf{w}^k\|_q^q = \|\mathbf{C} \mathbf{w}^k\|_q^q$ . This results in a new Laplacian matrix  $\tilde{\mathbf{C}}^T \tilde{\mathbf{C}}$ , which possesses a special structure known as *block-circulant with circulant-blocks*, whose matrix inverse can be evaluated efficiently via the fast Fourier Transform (FFT) (line 7, Alg. 1; see [16] for more details).

Using this augmentation+masking strategy, we can rewrite the objective as:

$$\min_{\mathbf{w}} \sum_{k=1}^K \frac{1}{n_k} \mathcal{L}(\mathbf{Y}^k \mathbf{X}^k \mathbf{w}^k) + \frac{\gamma}{q} \sum_{k=1}^K \|\mathbf{B} \tilde{\mathbf{C}} \mathbf{A} \mathbf{w}^k\|_q^q + \lambda \sum_{j=1}^p \|\mathbf{w}_j\|_2,$$

which can be converted into the following canonical ADMM form [2]:

$$\begin{aligned} & \min_{\{\mathbf{w}^k, \mathbf{v}_1^k, \mathbf{v}_2^k, \mathbf{v}_3^k, \mathbf{v}_4^k\}} \sum_{k=1}^K \frac{1}{n_k} \mathcal{L}(\mathbf{v}_1^k) + \frac{\gamma}{q} \sum_{k=1}^K \|\mathbf{B} \mathbf{v}_3^k\|_q^q + \lambda \sum_{j=1}^p \|\mathbf{v}_{2,j}\|_2 \\ & \text{s.t. } \mathbf{Y}^k \mathbf{X}^k \mathbf{w}^k = \mathbf{v}_1^k, \mathbf{w}^k = \mathbf{v}_2^k, \tilde{\mathbf{C}} \mathbf{v}_4^k = \mathbf{v}_3^k, \mathbf{A} \mathbf{w}^k = \mathbf{v}_4^k \quad \forall k = 1, \dots, K. \end{aligned} \quad (2)$$

It is straightforward to show that the above two problems are equivalent, and Alg. 1 follows from applying the standard ADMM iteration on (2). We emphasize that all the updates in Alg. 1 can be carried out efficiently in analytical form. For example, line 5 in Alg. 1 is a simple diagonal matrix inversion in the case of GN, and for the FL case we have the following closed form update:

$$[\mathbf{v}_3^k]_s \leftarrow \begin{cases} \text{soft}_{\gamma/\rho}([\tilde{\mathbf{C}}(\mathbf{v}_4^k - \mathbf{u}_3^k)]_s) & \text{if } \mathbf{B}_{s,s} = 1 \\ [\tilde{\mathbf{C}}(\mathbf{v}_4^k - \mathbf{u}_3^k)]_s & \text{if } \mathbf{B}_{s,s} = 0, \end{cases} \quad (3)$$

where  $\text{soft}_\tau(t) := \max(1 - \frac{\tau}{|t|}, 0) \cdot t$  denotes the *soft-threshold operator* and  $[\cdot]_s$  indexes the  $s$ -th element of a vector. Finally, we note  $\text{Prox}_{\tau\ell}(t)$  in line 6 is an elementwise update corresponding to the proximal operator of the hinge-loss.

### 3 Experiments

*The ADHD-200 Dataset.* We used the publicly available ADHD-200 competition dataset [15], which contains rs-fMRI scans of subjects diagnosed as either typically developing (TD) or with ADHD. The dataset is collected across seven sites and consists of two parts: a training set and a validation test set (Brown site excluded from our study as the subject labels are not released). Analyses were limited to participants with: (1) MPRAGE anatomical images with consistent near-full brain coverage with successful registration; (2) complete phenotypic information for main phenotypic variables (diagnosis, age, handedness); (3) mean framewise displacement (FD) within two standard deviation (SD) of the sample mean; (4) full IQ within two SDs of the ADHD-200 sample mean. After applying these sample selection criteria, we analyzed resting state scans from 628 individuals (TD=416, ADHD=212) in the training set and 106 subjects (TD=65, ADHD=41) in the test set. Functional images were reconstructed, slice-time corrected, motion corrected, and co-registered to the MNI space using SPM8.

*Experimental Results.* To assess the validity of the proposed method, we compared the performance of various SVM-based classifiers using the ADHD-200 dataset, where resting-state FCs were produced using the parcellation scheme described in Sec. 2. For the intra-task penalty  $\mathcal{R}_1$ , we compared four different regularization schemes: Elastic-net (EN) [5] with  $\mathcal{R}_1(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|_2^2$ , GN, FL, and TV. For the inter-task penalty  $\mathcal{R}_2$ , we compared three different approaches:

1. **Pooled**  $\ell_1$ : a single classifier is trained on the entire ADHD-200 dataset ( $\mathcal{R}_2(\mathbf{w}) = \|\mathbf{w}\|_1$  with  $\mathbf{w} \in \mathbb{R}^p$  as  $K = 1$ ).
2. **Single-task**  $\ell_1/\ell_1$ : equivalent to training separately across sites due to the separability of the penalty across sites ( $\mathcal{R}_2(\mathbf{w}) = \sum_{j=1}^p \|\mathbf{w}_j\|_1$ ).
3. **Multitask**  $\ell_1/\ell_2$ : *jointly* train the classifiers by solving (1).

The regularization parameters  $\{\lambda, \gamma\}$  are tuned by conducting a 5-fold cross-validation (CV) on the training set over the following two-dimensional grid:  $\lambda, \gamma \in \{2^{-13}, 2^{-12}, \dots, 2^{-3}\}$ . The final weight vector estimate is obtained by re-training the classifiers on the entire training set using the  $\{\lambda, \gamma\}$  values that maximized the CV classification accuracy; for validation, we predicted

the labels of the test set subjects using this weight vector. All methods were solved using ADMM with the algorithm terminated when the condition  $\|\mathbf{w}^{\text{new}} - \mathbf{w}^{\text{old}}\|_2 \leq 5 \cdot 10^{-3} \times \|\mathbf{w}^{\text{old}}\|_2$  was met or the iteration count reached 400.

To evaluate the quality of the classifiers, we analyzed the following set of performance measures for both the 5-fold CV and the validation test set results:

- Classification accuracy (ACC)
- Area under the ROC curve (AUC)
- Balanced score rate (BSR) = (sensitivity+specificity)/2
- Stability score (Stab.) = a measure of feature selection stability (see [1, 14])
- P-value (PVAL) computed from binomial test.
- Sparsity level (SP%) =  $100 \cdot \frac{|\# \text{ non-zero features}|}{pK}$

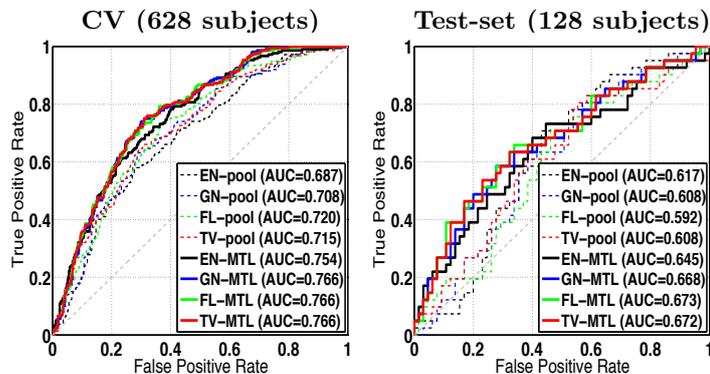
The *AUC* and *BSR* are analyzed since *ACC* by itself can be misleading when the dataset labels are imbalanced (*ACC*, *AUC*, and *BSR* are averaged across the tasks); the ROC curves are constructed by varying the threshold of the classifiers. *Stability score* is a measure introduced in [14] which quantifies the stability of the features selected across the CV folds (see [1, 14] for its precise definition). Classifier performance on the test set was compared to random guessing via a binomial test based on a binomial distribution  $B(p, n)$  with  $p=0.5$  and  $n=109$  samples, with *PVAL* evaluated via an one-sided binomial test [7]; the alternative approach of permutation test was not pursued due to its severe computational cost. Finally, *sparsity level* is the fraction of features selected in the final model.

Table 1 presents the classification results from the 5-fold CV and validation on the test-set, and Fig. 1 displays the corresponding ROC curves. These results demonstrate that training a single classifier via the “pooling” approach yields the worst performance in terms of accuracy, AUC, and BSR, suggesting that blindly aggregating the datasets across different sites can be problematic for accurate disease classification. Comparison between the single-task and the multitask approaches shows that the  $\ell_1/\ell_2$ -penalized approach yields superior performance in terms of AUC, although no striking difference can be observed in terms of accuracy and BSR.

In addition to the performance gain with the  $\ell_1/\ell_2$ -penalty, the set of weight vector estimates  $\{\hat{\mathbf{w}}^k\}_{k=1}^K$  all share a common support of length  $p$  with this multitask approach. This is invaluable for interpretation, as the selected features can be viewed as edges that are informative across all sites. For visualization, we grouped the indices of this support according to the network parcellation scheme proposed by Yeo *et al.* in [17], and reshaped them into a  $347 \times 347$  symmetric matrix with zeroes on the diagonal. The resulting support matrices for the  $\text{EN} + \ell_1/\ell_2$  and the  $\text{FL} + \ell_1/\ell_2$ -penalized SVM are presented in Fig. 3 (results for  $\text{GN} + \ell_1/\ell_2$  and  $\text{TV} + \ell_1/\ell_2$  were very similar to  $\text{FL} + \ell_1/\ell_2$ ). An interesting observation here is that the support structure from the  $\text{FL} + \ell_1/\ell_2$ -penalized SVM shows concentrated connectivity patterns in the intra-frontoparietal (6-6) and the intra-default network (7-7) regions; Fig. 3 provides a brain space representation of these connections (figures generated using BrainNet Viewer, [www.nitrc.org/projects/bnv/](http://www.nitrc.org/projects/bnv/)). These network regions are frequently reported to exhibit disrupted connectivity patterns in resting state studies of ADHD [3],

Table 1: The classification results from the 5-fold CV and the validation test-set.

	CV (628 subjects)				Test-set (106 subjects)				
	ACC	AUC	BSR	Stab.	ACC	AUC	BSR	PVAL	SP%
EN ( $\ell_1$ )	.689	.687	.630	.277	.557	.617	.476	.143	2.54%
GN ( $\ell_1$ )	.704	.708	.631	.253	.594	.608	.494	.032	28.88%
FL ( $\ell_1$ )	.688	.720	.586	.059	.632	.592	.530	.004	64.85%
TV ( $\ell_1$ )	.701	.715	.620	.005	.623	.608	.521	.007	90.32%
EN ( $\ell_1/\ell_1$ )	.709	.752	.649	.276	.623	.609	.530	.007	0.28%
GN ( $\ell_1/\ell_1$ )	.713	.750	.652	.165	.642	.613	.573	.002	67.14%
FL ( $\ell_1/\ell_1$ )	.715	.750	.659	.329	.632	.634	.547	.004	1.30%
TV ( $\ell_1/\ell_1$ )	.718	.753	.661	.345	.642	.654	.550	.002	1.61%
EN ( $\ell_1/\ell_2$ )	.720	.754	.657	.217	.651	.645	.556	.001	0.25%
GN ( $\ell_1/\ell_2$ )	.720	.766	.657	.320	.642	.668	.546	.002	1.03%
FL ( $\ell_1/\ell_2$ )	.718	.766	.653	.315	.642	.673	.546	.002	0.79%
TV ( $\ell_1/\ell_2$ )	.720	.766	.658	.316	.642	.672	.546	.002	0.80%

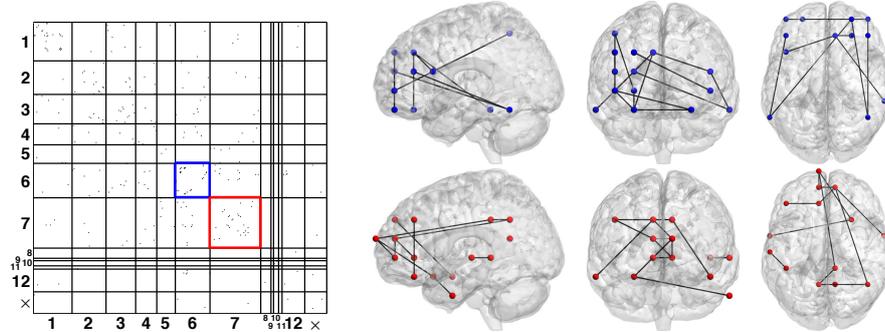
Fig. 2: Table 1 classifiers' ROC ( $\ell_1/\ell_1$ -curves omitted to improve curve visibility).

although the accuracies obtained from our classifiers are not at the level where the selected features can be interpreted as reliable ADHD biosignatures.

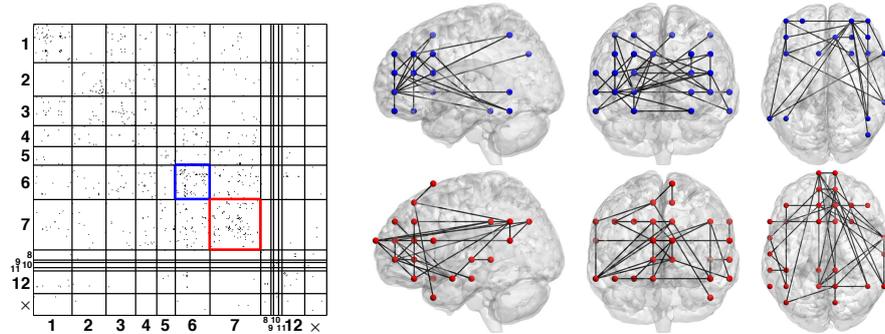
Finally, we note that most of the accuracies reported on the validation test-set in Table 1 exceeded the highest result from the actual ADHD-200 competition (which was 61.54% [15]). However, there are two major caveats: (1) the results in this work cannot be directly compared with the official competition results due to the subject screening procedure we applied on the test set (the criteria such as the FD-based one is important for avoiding confounds from excessive head motion), and (2) the participants in the actual competition were required to predict the labels of 26 subjects from the Brown site, despite the fact that no training data were provided from this site, making it harder to predict the labels for these subjects. The second caveat also implies that most MTL methods, including the  $\ell_1/\ell_2$ -penalty employed in this work, cannot be applied since there are no means to train a weight vector for a task whose data are not provided. An alternative approach such as *transfer learning* [11] may be considered for this.

Table 2: Network parcellation scheme of the brain proposed by Yeo *et al.* in [17].

Network membership Table ( $\times$ is “unlabeled”)			
1. Visual	2. Somatomotor	3. Dorsal Attention	4. Ventral Attention
5. Limbic	6. Frontoparietal	7. Default	8. Striatum
9. Amygdala	10. Hippocampus	11. Thalamus	12. Cerebellum



(a) Multitask Elastic-net SVM result



(b) Multitask Fused Lasso SVM result

Fig. 3: Weight vectors estimated from the  $\text{EN} + \ell_1/\ell_2$  and  $\text{FL} + \ell_1/\ell_2$ -penalized SVM. **Left:** support matrices of the selected features (rows/cols grouped by network membership). **Right:** brain space representation of the selected edges in the intra-frontoparietal (6-6: blue) and the intra-default network (7-7: red).

## 4 Conclusion

We presented a multitask structured sparse SVM, a multitask extension to the connectome-based disease classification method introduced in [16], where the imaging sites are treated as *tasks*. Experimental results on the multisite ADHD-200 dataset suggest that the multitask approach using the  $\ell_1/\ell_2$ -penalty can provide improvement in classification performance over the naive *pooling approach*, where a single classifier is trained on the entire multisite dataset. In addition, the  $\ell_1/\ell_2$ -penalty achieved higher AUC scores than the single-task  $\ell_1/\ell_1$ -penalty, and the *group variable selection* property of the multitask approach gives a more interpretable model by selecting the same set of features across sites, which can be visualized compactly in brain space.

## References

1. Baldassarre, L., Mourao-Miranda, J., Pontil, M.: Structured sparsity models for brain decoding from fMRI data. *Proc. PRNI* pp. 5–8 (2012)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* 3, 1–122 (2011)
3. Castellanos, F., Proal, E.: Large-scale brain systems in ADHD: beyond the prefrontalstriatal model. *Trends Cogn. Sci.* 16, 17 (2012)
4. Castellanos, F.X., Martino, A.D., Craddock, R.C., Mehta, A.D., Milham, M.P.: Clinical applications of the functional connectome. *NeuroImage* 80(0), 527 – 540 (2013)
5. Chen, X., He, J., Lawrence, Carbonell, J.G.: Adaptive multi-task sparse learning with an application to fMRI study. In: *SDM* (2012)
6. Grosenick, L., Klingenberg, B., Katovich, K., Knutson, B., Taylor, J.E.: Interpretable whole-brain prediction analysis with GraphNet. *NeuroImage* 72(0), 304 – 321 (2013)
7. Heinzle, J., Wenzel, M.A., Haynes, J.D.: Visuomotor functional network topology predicts upcoming tasks. *J. Neurosci.* 32(29), 9960–9968 (2012)
8. Marquand, A., Brammer, M., Williams, S., Doyle, O.: Bayesian multi-task learning for decoding multi-subject neuroimaging data. *NeuroImage* 92(0), 298 – 311 (2014)
9. Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Thirion, B.: Total variation regularization for fMRI-based prediction of behavior. *IEEE Trans. Med. Imag.* 30(7), 1328–1340 (2011)
10. Obozinski, G., Taskar, B., Jordan, M.I.: Joint covariate selection and joint subspace selection for multiple classification problems. *Stat. Comput.* 20(2), 231–252 (2010)
11. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22(10), 1345–1359 (2010)
12. Poline, J.B., Breeze, J.L., Ghosh, S.S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., Helmer, K.G., Marcus, D.S., Poldrack, R.A., Schwartz, Y., Ashburner, J., Kennedy, D.N.: Data sharing in neuroimaging research. *Front. Neuroinformatics* 6(9) (2012)
13. Rao, N.S., Cox, C.R., Nowak, R.D., Rogers, T.T.: Sparse overlapping sets lasso for multitask learning and its application to fMRI analysis. *NIPS* pp. 2202–2210 (2013)
14. Rasmussen, P.M., Hansen, L.K., Madsen, K.H., Churchill, N.W., Strother, S.C.: Model sparsity and brain pattern interpretation of classification models in neuroimaging. *Pattern Recognition* 45(6), 2085 – 2100 (2012)
15. The ADHD-200 Consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Front. Syst. Neurosci.* 6(62) (2012)
16. Watanabe, T., Kessler, D., Scott, C., Angstadt, M., Sripada, C.: Disease prediction based on functional connectomes using a scalable and spatially-informed support vector machine. *NeuroImage* 96(0), 183 – 202 (2014)
17. Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., Hollinshead, M., Roffman, J., Smoller, J., Zöllei, L., Polimeni, J., Fischl, B., Liu, H., Buckner, R.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106(3), 1125–65 (2011)

# Auto-contouring the Prostate for Online Adaptive Radiotherapy

Yan Zhou<sup>1</sup> and Xiao Han<sup>1</sup>

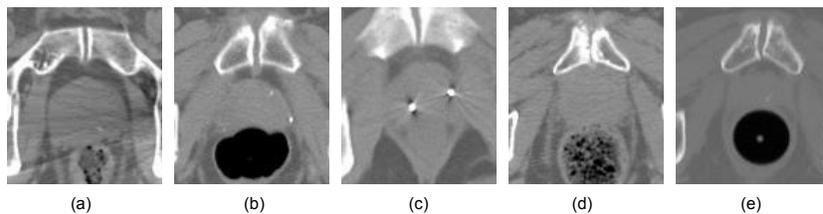
Elekta Inc., Maryland Heights, MO, USA  
yan.zhou@elekta.com, xiao.han@elekta.com,

**Abstract.** Among all the organs under cancer treatment, prostate is a very important one in male pelvic region but very difficult to segment, due to the poor contrast of 3D CT images, invisible boundaries between the prostate and its neighboring organs (e.g. bladder, rectum), and artifacts produced by prostate seeds etc. Furthermore, the same patient’s organ conditions (e.g. size, shape and location) can significantly change throughout the whole treatment course. In this paper, we propose a learning-based approach to deal with both inter-patient and intra-patient variation for auto-contouring the prostate in adaptive radiotherapy. In general, the method starts with learning population-based characteristics, and adaptively incorporate patient-specific knowledge as the same patient’s subsequent treatment images become available. Specifically, we learn a population-based boundary classifier and a sparse shape dictionary based on a set of already contoured patients. For intra-patient treatment tasks, previously segmented prostates of the same patient are utilized to adaptively update the boundary classifier and sparse shape dictionary. The updating process is fully automatic and completely offline, which won’t affect run-time efficiency. The proposed method has been extensively evaluated on 44 3D CT images of 11 patients, each with more than 3 daily treatment images. Our method produces superior performance over two other state-of-art auto-contouring methods, which is promising for online adaptive treatment planning.

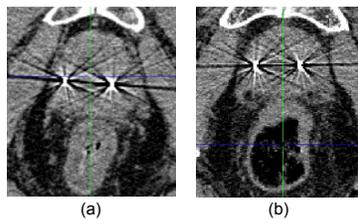
## 1 Introduction

In radiotherapy procedures, a patient needs to take a series of 3D CT images for treatment planning, re-planning, online dose delivery throughout the whole treatment course. To create a new treatment plan, a physician often needs to contour the image from scratch. It is a time consuming task which often induces large inter/intra institutional variation. Ideally an automatic contouring method not only reduces contouring time for the physicians, but also improves the accuracy and consistency for treatment. Among all the organs under cancer treatment, prostate is a very important one in male pelvic region but very difficult to segment. Major challenges for inter-patient cases are the following: (1) low contrast of 3D CT images, which makes large portion of the prostate boundary invisible (Figure 1), (2) image artifacts produced by prostate seeds (Figure

1 (a) (c)), and (3) large area of gas/feces/coil filling in the rectum (Figure 1 (b)(d)(e)). For intra-patient cases, the same patient's organ condition may vary a lot at different treatment times (Figure 2) due to (1) size of volume changes in response to treatment, (2) relative position change between neighboring organs, (3) shape deformation due to filling state of neighboring organs (e.g. bladder or rectum). Therefore, the desired auto-contouring method should be capable of handling both inter-patient and intra-patient variations. Additionally, the online processing should be computationally light weighted to adapt to the fast pace of online treatment procedures.



**Fig. 1.** Some prostate CT images from clinics: (a) an image with streaking artifacts, (b) a patient with large rectum gas, (c) a patient with seeds causing bright spots and streaking artifacts, (d) a patient with feces filling, (e) a patient with a rectum coil.



**Fig. 2.** An example of intra-patient variation of two images taken at different treatment times. (a) a patient's image at time  $t_1$ , (b) the same patient's image at time  $t_2$ .

Intuitively, the contouring process can be made easier by considering the same patient's previous contours as prior knowledge/reference. However, currently when designing a new plan or during treatment, physicians usually do not utilize the same patient's previous plans. Even when previous plans are used, they are incorporated by registration to map the previous contours to the current image. One common method is rigid registration. This method only provides a few degrees of freedom. Thus, the registered contours may not be precise. Deformable registration [1][2][3][4] may be employed to improve the accuracy by calculating non-linear organ deformations. In general, the accuracy of

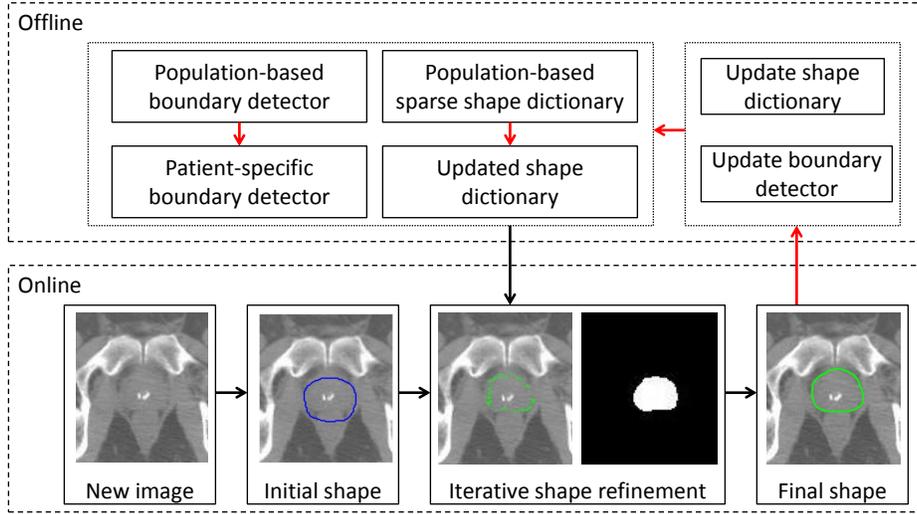
the contours may depend on the number of reference images (atlases) used [4]. There is, however, an increased computational cost proportional to the number of reference images (atlases), which makes it difficult to use for online adaptive planning. Alternatively, a few machine learning based methods [5][6][7] tried to use context information [5] or several layers of feature abstraction [7] to make the prostate region more discriminative for segmentation. They either focused on inter-patient segmentation or intra-patient segmentation. In [7], the authors evaluated those methods for prostate segmentation on MRI images and produced promising results.

In this paper, we propose a unified learning-based framework to accommodate both inter-patient and intra-patient variations for adaptive radiotherapy. In particular, a population-based boundary classifier and a population-based sparse shape dictionary can be trained. The trained boundary classifier and the sparse shape dictionary can then be used to perform auto-contouring in a patient’s planning image. As more treatment images are collected, the system may automatically update the boundary classifier and the sparse shape dictionary to incorporate patient-specific information. Once a new treatment image is received, the system may perform auto-contouring of the interested organ on the fly. One advantage of this approach is its high accuracy. In the online auto-contouring stage, the method is able to achieve a mean Dice value of 0.93 for the prostate. Another advantage of this approach is its ability to handle extremely low quality 3D CT images (e.g., Figure 12), since consistent artifacts/low quality can be learned as part of the patient-specific knowledge. Additionally, this method is computationally very fast when applying to a new scan. Because the learning process of previous scans is completely off-line, which can be done any time when the machine is vacant. Online auto-contouring takes the same amount of time no matter how many previous images were used for training. This is an advantage over deformable registration methods with multiple atlases, in which case the amount of auto-contouring time increases with the number of atlases used. The proposed learning-based prostate segmentation method has been extensively evaluated on 44 images of 11 patients, each with more than 3 daily treatment 3D CT images. It produces superior performance over two other state-of-the-art segmentation methods. The learning framework is very accurate and fast with the flexibility of working on any quality images, which is well designed for online adaptive radiotherapy in clinics.

In section 2, we will discuss the methodology in detail. In section 3, enormous experimental results are given, which is followed by section 4 conclusions.

## 2 Methodology

Figure 3 shows the flowchart of this learning-based system. The top part shows the components for off-line training/updating, while the bottom part shows the online auto-contouring components. Algorithm flow involved in off-line training/updating is marked with red arrows. Algorithm flow involved in online



**Fig. 3.** Algorithm flowchart. Algorithm flow involved in off-line training/updating is marked with red arrows. Algorithm flow involved in online auto-contouring process is marked with black arrows.

auto-contouring process is marked with black arrows. Details regarding each component are discussed in the following:

## 2.1 Population-based Boundary Classifier

In order to detect the boundary surface of the prostate for a new image, we first train a population-based boundary classifier. For each of the voxels being detected, the job of the classifier is to make a decision as to whether the current voxel “is” or “is not” on the boundary. From a set of patients’ images, we collect positive and negative samples according to the manual contours provided by experts. Boundary voxels on the contours are selected as positive samples, while voxels far away from the contour are selected as negative samples. For each training sample, rotation invariant 3D steerable features [8] are extracted and stored as a feature vector. We use the random forest algorithm [9] to train the boundary classifier on the collected samples. During online contouring stage, we use the Demons [2] method with a single atlas to get the initial contour. For each voxel on the surface of the initial contour, we apply boundary classification in a neighborhood along the normal direction. The boundary classifier will return a probability value for each voxel being searched. Then we select the one with the highest probability as the new boundary location.

## 2.2 Population-based Sparse Shape Refinement

The detected 3D boundary from the boundary classifier is very noisy. Thus a shape model is needed here to constrain the solution space. Among all the recently promoted shape models, the sparse shape model [10][11][12][13] is known to be able to handle complex shape variations, model non-Gaussian errors and preserve local detailed information of the input image, which well fits our needs. So we adopt this model to refine the detected boundary. Specifically, we first construct a sparse shape dictionary from a set of patients. Once we have a new input boundary shape, we use the dictionary as the shape prior to refine the shape. In particular, it selects a sparse set of 3D shapes in the shape dictionary and composes them together to represent the input shape. This model leverages two sparsity observations of the input shape instance: (1) the input shape can be approximately represented by a sparse linear combination of shapes in the shape dictionary; (2) parts of the input shape may contain gross errors but such errors are sparse. For each refinement iteration, the algorithm minimizes the following optimization function:

$$\arg \min_{x,e,\beta} \|T(v_S, \beta) - SDx - Se\|_2^2 + \gamma_1 \|x\|_1 + \gamma_2 \|e\|_1 \quad (1)$$

Where  $v_S$  is a subset of points on the input shape,  $D$  is the shape dictionary that represents all training shapes,  $T(v_S, \beta)$  is a global transformation operator with parameter  $\beta$ , which aligns the input shape to the same space of  $D$ .  $x$  denotes the weighting coefficients of the linear combination, and  $e$  is a vector that models the large residual errors.  $S$  is a binary diagonal matrix which indicates if a certain point is in the subset  $v_S$ . Here, in our implementation, each input boundary shape is represented by a 3D mesh with 4096 surface points. Each surface point is represented by its three dimensional coordinates. The solved shape is then sent back to the boundary detectors for another round of shape refinement (Figure 3). The iterative process stops once 1) it reaches a certain number of iterations or 2) it reaches a certain minimal residual error.

## 2.3 Adapting Patient-specific Information

Once a new treatment image is collected, it is necessary to update the boundary classifier and the sparse shape dictionary accordingly to incorporate patient-specific information. To update the boundary classifier, we collect training samples from a certain number of the same patient’s previously treated images ( $n = 3$  in our implementation). If a patient doesn’t have as many treatment images available yet, we then compare the structure similarities between the current image and the pool of images from all the patients, and select the most similar ones. We use the method in [14] to handle this image retrieval task. Only the selected images are used for training the new boundary classifier. The updating process is off-line and doesn’t require human intervention.

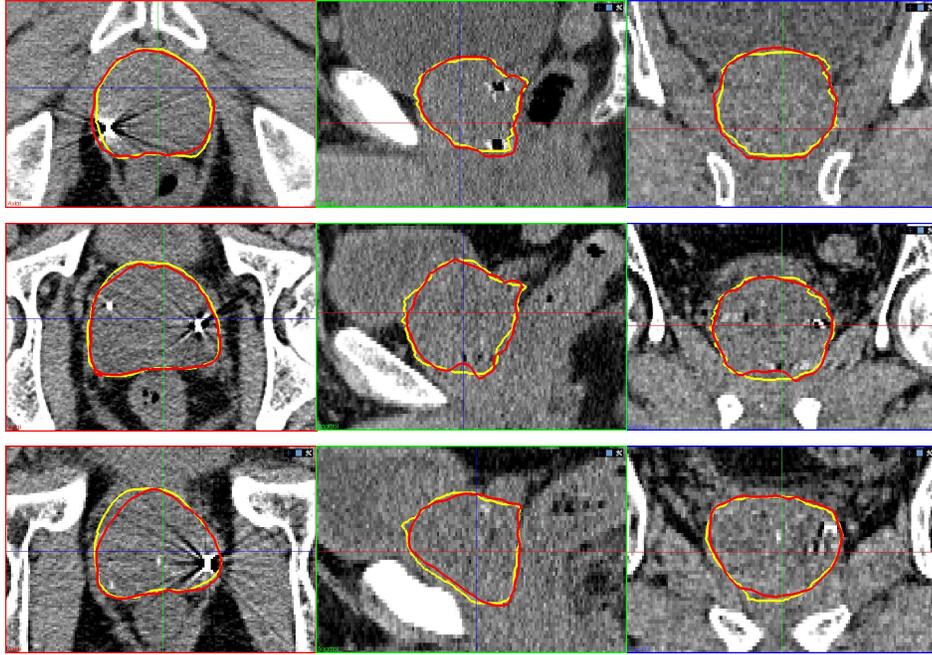
While updating the boundary classifier is quite efficient since not many images are required for training, updating the sparse shape dictionary is a completely different story. To handle large shape variation even from the same pa-

tient, we want a general shape dictionary that can comprehensively capture shape variations in the shape space. Thus we don't want to limit the number of training shapes. However, we want to include the patient's most recent images to the dictionary to gain patient-specific knowledge. Training the shape dictionary from scratch is very time consuming. To improve the computational efficiency, dictionary learning techniques have also been employed to train a compact dictionary instead of using all training shapes. We use an online learning method [12] to adaptively and efficiently incorporate new shapes. When new training shapes come, instead of re-constructing the dictionary from scratch, we update the existing one using a block-coordinates descent approach. Using the dynamically updated dictionary, the sparse shape dictionary can be gracefully scaled up to model shape prior from a large number of training shapes without sacrificing run-time efficiency.

### 3 Experiments

The population-based boundary classifier and sparse shape dictionary are trained from 21 3D CT images of 21 patients across 5 hospitals. Each contour is represented by a 3D mesh with 4096 points. We evaluated the trained boundary classifier and sparse shape model on 44 3D CT images from 11 patients, and each patient has at least 3 treatment images. Figure 4 shows auto-contouring results for 3 patients from top to bottom. Each patient has three snapshots of its axial, sagittal and coronal planes. We compared the auto-contouring results (in red) with the experts' manual contours (in yellow). Despite the two bright seeds inside the prostate producing significant artifacts on the images, the auto-contouring results are very close to the ground-truth contours.

To get a quantitative overview of the performance compared with other methods on the whole testing dataset, Figure 5 shows the Dice values of three different methods. We started by using the Demons method [2] with only one atlas. For each patient's first 3D CT image, we randomly pick one atlas from the population and use the Demons deformable registration method for auto-contouring. Then for the same patient's subsequent images, we use their previous segmented image as the atlas. As shown in Figure 5 (a), the Dice values have large variation mainly due to the bad performance of the first segmented image. With randomly picked single atlas, it can hardly fit the target patient's anatomical structures well. Thus we increased the number of atlases and applied the STAPLE [3] strategy to fuse the deformable registration results from all the atlases. Similarly, for the patient's first planning image, we randomly picked 5 atlases from the population. For the same patient's subsequent images, we included all his/her latest segmented images as atlases. As shown in Figure 5 (b), the box-plot has less outliers and the average performance got improved. But without a patient-specific boundary detector, the algorithm can hardly precisely drive the detected contours to the desired borders. Additionally, the computational cost of using 5 atlases is tremendous. Our method auto-segments the patient's first planning image by the population-based classifier and the population-based sparse

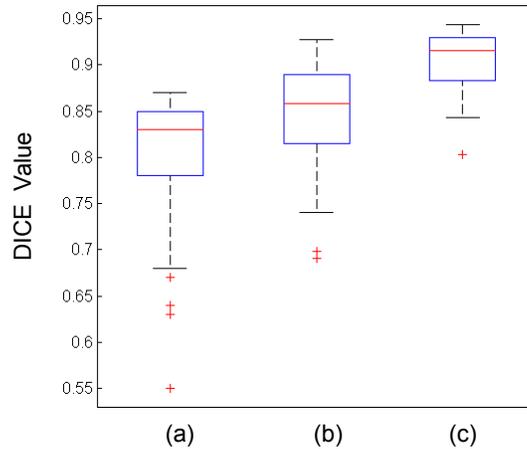


**Fig. 4.** Auto-segmentation results of three patients (from top to bottom) shown as snapshots on the axial, sagittal and coronal planes (from left to right). The results of the proposed method are shown in red, and the experts' manual contours (ground-truth) are shown in yellow.

shape dictionary, and adaptively learn the patient-specific knowledge during subsequent treatments without sacrificing run-time efficiency. The performance gets improved in terms of accuracy and computational cost (Figure 5 (c)).

## 4 Conclusions

We proposed a learning-based auto-contouring method for online adaptive radiotherapy. Following the nature of the tasks, the learning system is not only able to generate population-based information, but also capable of adaptively gaining patient-specific knowledge throughout the whole treatment course. The training and online updating steps are fully automatic and completely off-line, so that it doesn't reduce runtime efficiency. It achieves higher accuracy for the prostate segmentation when compared with two other state-of-the-art methods, which makes it very promising for online adaptive treatment planning.



**Fig. 5.** Dice values of three methods under comparison: (a) the Demons method with only one atlas; (b) the Demons method using 5 atlases with the STAPLE label fusion; (c) Our proposed method.

## References

1. Zitov, B., Flusser, J.: Image registration methods: a survey. *Image and Vision Computing* **21**(11) (2003) 977 – 1000
2. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration (2008)
3. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. Volume 23. (July 2004) 903–921
4. Han, X., Hibbard, L.S., Oconnell, N.P., Willcut, V.: Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion (2010)
5. Gao, Y., Liao, S., Shen, D.: Prostate segmentation by sparse representation based classification. *Medical Physics* **39**(10) (2012) 6372–6387
6. Li, W., Liao, S., Feng, Q., Chen, W., Shen, D.: Learning image context for segmentation of prostate in CT-guided radiotherapy. In: MICCAI (3). Volume 6893. (2011)
7. Liao, S., Gao, Y., Oto, A., Shen, D.: Representation learning: A unified deep learning framework for automatic prostate MR segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2013*. Volume 8150. (2013) 254–261
8. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-d cardiac CT volumes using marginal space learning and steerable features. *Medical Imaging, IEEE Transactions on* **27**(11) (Nov 2008) 1668–1681
9. Breiman, L., Schapire, E.: Random forests. In: *Machine Learning*. (2001) 5–32
10. Shaoting Zhang, Yiqiang Zhan, M.D.J.H.D.N.M.X.S.Z.: Towards robust and effective shape modeling: Sparse shape composition. In: *Medical Image Analysis*. Volume 16. (2012) 265–277

11. Shaoting Zhang, Yiqiang Zhan, D.N.M.: Deformable segmentation via sparse representation and dictionary learning. In: *Medical Image Analysis*. Volume 16. (2012) 1385–1396
12. Zhang, S., Zhan, Y., Zhou, Y., Uzunbas, M., Metaxas, D.: Shape prior modeling using sparse representation and online dictionary learning. In Ayache, N., Delingette, H., Golland, P., Mori, K., eds.: *Medical Image Computing and Computer-Assisted Intervention*. Volume 7512 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg (2012) 435–442
13. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09, ACM (2009) 689–696
14. Zhou, Y., Peng, Z., Zhou, X.S.: Automatic view classification for cardiac MRI. In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. (2012) 1771–1774

# Detection of Multiple Sclerosis Lesions using Sparse Representations and Dictionary Learning

Hrishikesh Deshpande<sup>1,2,3,4</sup>, Pierre Maurel<sup>1,2,3,4</sup>, Christian Barillot<sup>1,2,3,4</sup>

1. University of Rennes 1, Faculty of Medecine, F-35043 Rennes, France
2. INSERM, U746, F-35042 Rennes, France
3. CNRS, IRISA, UMR 6074, F-35042 Rennes, France
4. Inria, VISAGES Project-team, F-35042 Rennes, France

**Abstract.** The manual delineation of Multiple Sclerosis (MS) lesions is a challenging task pertaining to the requirement of neurological experts and high intra- and inter-observer variability. It is also time consuming because large number of Magnetic Resonance (MR) image slices are needed to obtain 3-D information. Over the last years, various models combined with supervised and unsupervised classification methods have been proposed for segmentation of MS lesions using MR images. Recently, signal modeling using sparse representations (SR) has gained tremendous attention and is an area of active research. SR allows coding data as sparse linear combinations of the elements of over-complete dictionary and has led to interesting image recognition results. The dictionary used for sparse coding plays a key role in the classification process. In this work, we have proposed to learn class specific dictionaries and developed a new classification scheme, to automatically detect MS lesions in 3-D multi-channel MR images.

## 1 Introduction

Multiple sclerosis is a chronic, autoimmune disease of the central nervous system, characterized by structural damages of axons and their myelin sheathes. During progression of the disease, certain areas of brain develop MS lesions. The evolution of MS lesions is highly variable and is not fully known. MS is more common in North America and Europe and is more prevalent in young adult population, causing non-traumatic disabilities.

Magnetic Resonance Imaging (MRI) holds the capability of detecting abnormalities in 95% of the patients with MS and is the best paraclinical method for imaging MS [1]. These images are analyzed to find the number and spatial patterns of the lesions, appearance of new lesions and the total lesion load, which are key parameters in the current MS diagnostic setup. However, manual segmentation of MS lesions is a laborious and time consuming process and is prone to high intra- and inter-expert variability. Therefore, there is a need for fully automated MS lesion detection methods that can handle large variety of MR data and which can provide results that correlate well with expert analysis [2].

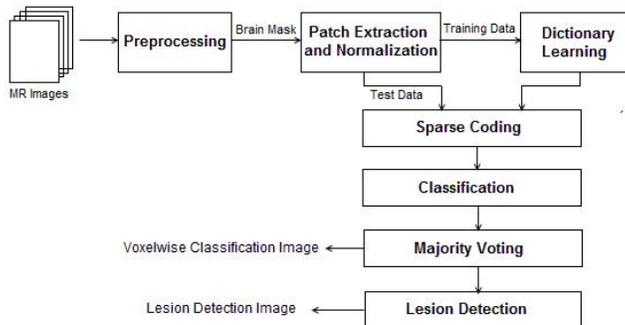
Over the past years, various approaches for semi-automatic and automatic segmentation of MS lesions have been proposed. In these methods, different image features, classification methods and models have been tried, but they usually suffer from high sensitivity to the imaging protocols and so usually require tedious parameter tuning or specific normalized protocols [3]. More recently, sparse representation has evolved as a model to represent an important variety of natural signals using few elements of an overcomplete dictionary. Many publications have demonstrated that sparse modeling can achieve state-of-the-art results in image processing applications such as denoising, texture segmentation and face recognition [4, 5]. In [5], given multiple images of individual subjects under varying expressions and illuminations, the images themselves were used as dictionary elements, for classification. Such a method uses dictionary learning to analyze image as a whole. Mairal et al [6] proposed to learn discriminative dictionaries better suited for local image discrimination tasks. In medical imaging, local image analysis is of prime importance and it could be interesting to see the performance of sparse representation and dictionary learning based classification methods in the context of disease detection. Some researchers have reported works on segmentation of endocardium and MS lesions using dictionary learning [7, 8]. Weiss et al. proposed an unsupervised approach for MS lesion segmentation, in which a dictionary learned using healthy brain tissue and lesion patches is used as basis for classification [7].

Our approach differs from this method in several ways. In [7], authors use only FLAIR MR images for analysis of clinical data. However, MS lesions appear in different intensity patterns in various MR sequences, which include T1 (T1-w MPRAGE) and T2-weighted, (T2-w) and Proton Density (PD). The complementary information in these MR images can further assist in classifying MS lesions. We build our analysis using above mentioned MR sequences. Our major contribution is however that we learn class specific dictionaries for healthy brain tissues and lesions that promote the sparse representation of healthy and lesion patches. The lesion patches are well adapted to its own class dictionary, as opposed to the other. Thus, we can use the reconstruction error derived from sparse decomposition of test patch on to these dictionaries for classification. In the dataset, the healthy class patches outnumber lesion patches and exhibit more variability. Thus we use different dictionary lengths for modelling individual class patches. In this manner, we take into consideration the data variability and class imbalance in healthy and lesion classes. Finally, supervised approach for detection results in omission of tuning of one parameter as mentioned in [7], making this method fully automatic. In the following sections, we describe our new approach and provide its evaluation using clinical images.

## 2 Methodology

As shown in Figure 1, we first preprocess MR images for noise removal and then extract the image patches of predefined size using brain mask. These patches are normalized and are divided into the training and test sets for healthy brain

tissue and lesion classes, with the help of manual segmentation images. Using training signals, we derive different classification approaches by either learning single dictionary or two separate dictionaries for both the classes. Finally, for a given test patch, the reconstruction error based classification method is developed, followed by voxel-wise classification and lesion detection. The following subsections briefly describe these steps.



**Fig. 1.** Flowchart of MS Lesion Detection using Dictionary Learning

## 2.1 Patch Extraction and Training Set

We divide the intracranial MR volume into several 3-D patches and flatten them into one dimensional concatenated vectors representing intensities of T1-w MPRAGE, T2-w, PD and FLAIR images. Keeping the computational complexity of further analysis in mind, we extract a patch every  $M$  voxels in each direction. As described earlier, we develop supervised approach by labelling these patches as belonging to either healthy or lesion class. If, in a patch, the number of voxels manually labelled as lesions exceeds a threshold  $T_L = 6mm^3$ , it is included in a lesion set, or in healthy set otherwise. For every subject, we obtain around  $1.5 \times 10^6$  patches for healthy and  $10^3$  to  $10^5$  patches for lesion class, depending on the lesion load for each patient. These patches are finally normalized to limit their individual norms below or equal to unity, as per constraint imposed by dictionary learning.

## 2.2 Sparse Representations and Dictionary Learning

Sparse representation of the data allows the decomposition of signal into linear combination of few basis elements in an overcomplete dictionary. Consider a signal  $\mathbf{x} \in \mathbb{R}^N$  and an overcomplete dictionary  $\mathbf{D} \in \mathbb{R}^{N \times K}$ . The sparse coding problem can be stated as  $\min_{\mathbf{a}} \|\mathbf{a}\|_0$  s.t.  $\mathbf{x} = \mathbf{D}\mathbf{a}$  or  $\|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 \leq \varepsilon$ , where  $\|\mathbf{a}\|_0$  is  $l_0$  norm of the sparse coefficient vector  $\mathbf{a} \in \mathbb{R}^K$  and  $\varepsilon$  is error in representation.

Basis pursuit algorithm solves the convex approximation of the problem above by replacing  $l_0$  norm with  $l_1$  norm that also results in sparse solution [9]. Thus, the sparse coding problem can be given by

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1, \quad (1)$$

where  $\lambda$  controls the trade-off between representation error and sparsity.

The fixed dictionaries like wavelets can be efficient if a background analytical model can be inferred. On the other hand, the dictionary learning from underlying data has produced exciting results with greater data adaptability and has replaced the use of generic models. For a set of signals  $\{\mathbf{x}_i\}_{i=1,\dots,m}$ , the dictionary learning problem is to find  $\mathbf{D}$  such that each signal can be represented by sparse linear combination of its atoms. This can be stated as the following optimization problem

$$\min_{\mathbf{D}, \{\mathbf{a}_i\}_{i=1,\dots,m}} \sum_{i=1}^m \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda \|\mathbf{a}_i\|_1. \quad (2)$$

The optimization is carried out as two-step process involving the sparse coding step with fixed  $\mathbf{D}$  and the dictionary update step with fixed  $\mathbf{a}$ .

### 2.3 Classification

**(a) Using Single Dictionary** : In the context of MS lesion classification, the simplest idea, similar to [7], could be to use a single dictionary learned from healthy and lesion class patches. As the lesions are outliers with respect to the healthy brain intensities, the decomposition of lesion patch using this dictionary would result in higher representation error than that for the healthy tissue patch. For a given test patch, we calculate the sparse coefficients and reconstruction error, and assign it to the lesion class if this error is greater than chosen threshold. The threshold is selected by observing the histogram of the error map.

**(b) Using Class Specific Dictionaries (Same Length)** : Here, we learn class specific dictionaries  $D_1$  and  $D_2$  for healthy and lesion classes, respectively. Given a test patch  $\mathbf{x} \in \mathbb{R}^N$ , classification is performed in two steps: In the first step, sparse coefficients  $\mathbf{a}_i$  are obtained using Eq (1) for each class  $i=1$  (Healthy) and 2 (Lesion). The test patch is then assigned to class  $c$  such that

$$c = \operatorname{argmin}_i \|\mathbf{x} - \mathbf{D}_i \mathbf{a}_i\|_2^2. \quad (3)$$

**(c) Using Class Specific Dictionaries (Different Lengths)** : The dictionaries learned using above mentioned approach does not take into account the data variability between two-classes. The size of the dictionary plays a major role in the data representation. For healthy class data with more variability and number of training samples than that for the lesion class, we allow larger dictionary length for healthy class data and study its effect on MS lesion classification.

## 2.4 Voxel-wise Classification and Lesion Detection

As already stated, there is some overlap between patches. However, to obtain voxel-wise classification, each voxel needs to be assigned to either of the classes. This is achieved using majority voting, in which, the voxel under consideration is classified as healthy or lesion, using majority votes of all patches which contain that voxel.

The voxelwise classification image is further processed to obtain the lesion based detection image. A lesion is said to be detected if  $\frac{R_D \cap R_{GT}}{R_{GT}} \geq T_O$ , where  $R_D$  and  $R_{GT}$  are respectively the candidate regions in the classification image and the ground truth, whereas  $T_O$  is the threshold indicating overlap between them as a fraction of ground truth lesion.

## 3 Dataset and Preprocessing

The proposed approach was validated on MRI volumes of 14 MS patients acquired by Verio 3T Siemens scanner. T1-w MPRAGE, T2-w, PD and FLAIR MR modalities were chosen for the experiment. The volume size for T1-w MPRAGE and FLAIR is  $160 \times 256 \times 256$  and voxel size is  $1 \times 1 \times 1 \text{ mm}^3$ , whereas for T2-w and PD scans, the volume size is  $192 \times 256 \times 44$  and voxel size is  $1 \times 1 \times 3 \text{ mm}^3$ . The manual segmentation images obtained from neurological experts are referred to as ground truth lesion masks.

For MR images of each patient, the imaging artifacts are corrected by denoising using non-local means and Intensity Inhomogeneity Correction (IIH). The images so obtained are then registered with respect to T1-w MPRAGE volume and are processed further to extract the intra-cranial region.

## 4 Results and Discussions

We implemented our method using MATLAB and Python. The packages AN-IMA and N3 ITK were used for denoising, registration and IIH correction, respectively [10–12]. We used the neuroimaging software Brain Extraction Tool (BET) for brain extraction [13]. For dictionary learning and sparse coding, we used SPArse Modeling Software (SPAMS) package [14].

We performed the experiments on 14 subjects using Leave-One-Subject-Out-Cross-Validation. Different parameters have been tested for the methods. It was found that image patch of size  $5 \times 5 \times 5$ , with a patch every 2 voxels in each direction, was optimal with respect to the classification efficiency. The dictionary length of 5000 and sparsity parameter  $\lambda = 0.95$  were optimal selections for dictionary learning method. For voxel-wise classification method, we then recorded the number of voxels that belong to True Positives (TP), False Negatives (FN), False Positives (FP) or True Negatives (TN) and the classification methods were finally validated by calculating sensitivity =  $\frac{TP}{TP+FN}$  and Positive Predictive Value (PPV) =  $\frac{TP}{TP+FP}$ .

In the first method, we studied the classification by learning single dictionary with the help of both healthy brain tissue and lesion patches. We chose sparse penalty factor  $\lambda = 0.85$  in the sparse coding step and performed the classification for various threshold values on the histogram of error map, as explained previously. We then selected the threshold for which the best voxelwise classification results were obtained in terms of both sensitivity and PPV. It was observed that the method suffered with a very large number of false positive detections.

Next, we learned class specific dictionaries for healthy and lesion classes, each. We used dictionary lengths of 5000 for signal representation of each class. The mean sensitivity and PPV obtained using this approach were 91.5% and 7.5%. This method performs better than the previous method but still contains many false positives. The primary reason behind this can be the difference in the data variability of each class signals. The healthy class patches have more variability in terms of representation of white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF), as compared to the variations in the representation of lesions. Hence, we adopted different dictionary lengths for representation of these classes. We used dictionary lengths of 5000 and 1000 respectively, for healthy and lesion classes. Table 1 summarizes the results of the voxelwise classification for the three methods described above.

**Table 1.** Voxel-wise classification results using: (a) Single Dictionary (SD), with 5000 atoms learned using healthy and lesion class data, (b) Class Specific Dictionaries with Same Lengths (CSD SL): 5000 atoms each and (c) Class Specific Dictionaries with Different Lengths (CSD DL): 5000 atoms for healthy class and 1000 atoms for lesion class. Sensitivity and Positive Predictive Value (PPV) (%) are given for each method.

Patient	(a) SD		(b) CSD SL		(c) CSD DL	
	Sens.	PPV	Sens.	PPV	Sens.	PPV
1	42	1	97	3	53	31
2	74	1	98	2	66	41
3	73	1	91	2	63	27
4	91	2	98	17	57	68
5	61	1	95	10	54	65
6	91	7	89	29	38	55
7	78	1	85	3	20	32
8	72	1	98	3	69	21
9	80	1	36	2	4	9
10	66	1	97	9	61	52
11	89	2	98	12	66	41
12	75	1	99	8	52	36
13	78	1	100	3	77	31
14	59	1	100	2	78	17
Average	<b>73.5</b>	<b>1.57</b>	<b>91.5</b>	<b>7.5</b>	<b>54.14</b>	<b>37.57</b>

It can be seen that using class specific dictionaries with the same dictionary length improves both sensitivity and PPV, as compared to the first method. But PPV in the second method is still low, indicating that there are still large number of false positives, which can explain higher sensitivity. Using different dictionary lengths, as implemented in third method, drastically reduces the number of false positives, which can be seen by the significant increment in PPV, while keeping the sensitivity in the acceptable limit.

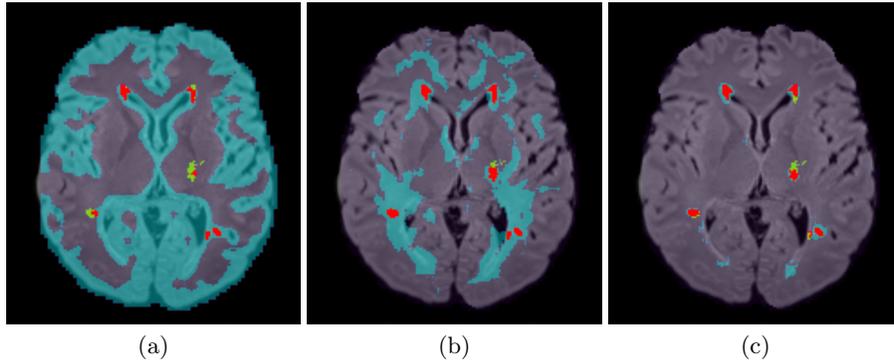
The mean PPV and sensitivity for lesion detection with class specific dictionaries of different lengths are shown in Table 2 for various overlap thresholds  $T_O$ . To be consistent with the threshold  $T_L$  incorporated in learning stage (Refer Section 2.1), we ignore very small lesions with volumes less than  $T_L$ . It can be seen that we detect 61% of the lesions with the overlap threshold of 1% . Moreover, in 49% of the lesions detected, at least 40% of the voxels are correctly classified by the method.

**Table 2.** Performance analysis for lesion detection using Class Specific Dictionaries with Different Lengths (CSD DL) for each class, with 5000 atoms for healthy class dictionary and 1000 atoms for lesion class dictionary.

	$T_O = 0.01$	$T_O = 0.1$	$T_O = 0.2$	$T_O = 0.3$	$T_O = 0.4$
PPV (%)	61.67	58.41	56.53	54.31	49.40
Sensitivity (%)	60.97	57.58	56.67	54.56	49.94

In Figure 2, we show the results for patient 8, for all the methods discussed above. The detection image is superimposed on FLAIR MR image. It can be observed that methods (a) and (b) have large number of false positives. We get the best classification results using class specific dictionaries with different dictionary lengths. But, in terms of voxelwise classification, there are still few false positives and true negatives around the actual lesion. This does not pose a major problem for lesion detection as long as significant portion of the actual lesion is being classified correctly. There are, however, some false positive lesion detections.

We are aware that we do not have a very large population for training. Hence we investigated the incorporation of longitudinal database into our analysis by considering MR sequences at 3 time points (M0, M3 and M6) for all the patients. As the lesions evolve over the course of time, it is fair to consider that each new dataset will enrich our learning model. Thus, we modified the training data, for each patient, in two ways: (1) Data at time-points M0 and M3, with 26 datasets and (2) Data at time-points M0, M3 and M6, with 39 datasets. However, the lesion detection experiments for the same test subjects, as in previous experiments, using class specific dictionaries with the lengths of 5000 and 1000 for healthy and lesion class respectively, did not show any significant improvement in the sensitivity and PPV. This suggests that the population for training the dictionaries earlier was sufficient and the dictionaries should be adapted to learn



**Fig. 2.** Classification results for Patient 8. For illustration purpose, one slice has been arbitrarily selected. True Positives are in red, False Positives are in cyan, False Negatives are in green. Methods (a), (b) and (c) are the same as in Table 1.

more specific structures viz. WM, GM and CSF versus lesions to help improve the detection.

## 5 Conclusion

In this paper, we have proposed a new supervised approach to automatically detect multiple sclerosis lesions using dictionary learning. We investigated the performance of three methods which either use one dictionary, treating lesions as outliers, or use class specific dictionaries for healthy and lesion classes, wherein the underlying data for each class is represented by the dictionary and sparse coefficients. We further studied the effect of using different dictionary lengths, allowing larger dictionaries to represent the complex data and concluded that such method minimizes the false positive detections in the classification.

Although the method using class specific dictionaries follows supervised approach, contrary to the single dictionary based classification method, which does not necessarily require training data, it is worth mentioning that the former method eliminates one parameter: threshold on error map. This crucial parameter is not easy to tune and could lead to worse classification results for small errors in the brain extraction procedure.

To further improve the results, it would be interesting not to learn only one dictionary for healthy brain tissues, but derive dictionaries more specific to WM, GM and CSF, in addition to the lesions. One could also study the role of sparse coefficients in addition to the reconstruction error, in the classification step.

## References

1. Grossman, R., McGowan, J.: Perspectives on Multiple Sclerosis. *AJNR* (1998) 19(7):1251–1265

2. Polman, C., et al.: Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria
3. García-Lorenzo, D., Francis, S., et al.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *MedIA* **17**(1) (2013) 1 – 18
4. Li, S., Fang, L., Yin, H.: An Efficient Dictionary Learning Algorithm and Its Application to 3-D Medical Image Denoising. *Biomedical Engineering, IEEE Transactions on* **59**(2) (Feb 2012) 417–427
5. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust Face Recognition via Sparse Representation. *PAMI IEEE Transactions* **31**(2) (Feb 2009) 210–227
6. Mairal, J., Bach, F., et al.: Discriminative learned dictionaries for local image analysis. In: *IEEE CVPR*. (June 2008) 1–8
7. Weiss, N., Rueckert, D., Rao, A.: Multiple Sclerosis Lesion Segmentation Using Dictionary Learning and Sparse Coding. In: *MICCAI 2013*. Volume 8149.
8. Rosas-Romero, R., Tagare, H.D.: Segmentation of endocardium in ultrasound images based on sparse representation over learned redundant dictionaries. *Engineering Applications of Artificial Intelligence* **29**(0) (2014) 201 – 210
9. Chen, S.S., Donoho, D.L., Michael, Saunders, A.: Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* **20** (1998) 33–61
10. Coupé, P., Yger, P., et al.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE TMI* **27**(4) (April 2008) 425–41
11. William, W., Paul, V., et al.: Multi-modal volume registration by maximization of mutual information. *MedIA* **1**(1) (1996) 35 – 51
12. Tustison, N., Avants, B., , et al.: N4ITK: Improved N3 Bias Correction. *IEEE TMI* **29**(6) (June 2010) 1310–1320
13. Smith, S.: Fast robust automated brain extraction. In: *HBM 17*(3). (2002) 143–155
14. Mairal, J., Bach, F., et al.: Online Dictionary Learning for Sparse Coding. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML '09, New York, NY, USA, ACM (2009) 689–696

# A Low-Rank + Sparse Decomposition (LR+SD) Method for Automatic EEG Artifact Removal

J. Gilles<sup>1</sup>, T.Meyer<sup>1</sup>, and P.K. Douglas<sup>2,3</sup>

<sup>1</sup> Department of Mathematics, University of California, Los Angeles, CA 90095, USA

<sup>2</sup> Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, California, USA 90024

<sup>3</sup> Wellcome Trust Centre for Neuroimaging at UCL, London WC1N 3BG

**Abstract.** In theory, multimodal EEG-fMRI recordings represent an excellent tool for studying bioelectric-hemodynamic coupling in the human brain without incurring added complexity due to nonstationarity. However, ballistocardiogram (BCG) artifacts as opposed to magnetic gradient noise have made analysis of EEG data collected in the MRI environment very challenging. Conventionally, BCG artifacts have been removed only partially after meticulous user-guided identification of independent components associated with noise. In this paper, we present a novel method for automatically removing BCG artifact from event related EEG data by leveraging sparsity in the time domain. Our method, low rank + sparse decomposition (LR+SD) extends robust PCA and requires tuning of only a single regularization parameter. We apply this method first to simulated data, and then to real simultaneous EEG-fMRI data, collected while subjects viewed photic stimuli. We found that LR+SD improved the signal-to-noise ratio by 34 and 36 percent, as compared to either manual or automatic IC methods respectively. This method appears quantitatively superior to IC methods, and may improve the feasibility of analyzing event related EEG-fMRI data collected concurrently.

**Keywords:** Concurrent EEG-fMRI, ballistocardiogram, artifact removal, robust, PCA, low rank, sparsity, sparse decomposition

## 1 Introduction

Independently, electroencephalogram (EEG) and functional MRI (fMRI) offer either rich temporal (EEG) or spatial (fMRI) information related to neuronal dynamics in the brain. Ideally, these imaging techniques could be combined in a complementary fashion to harness their respective strengths [4], [16], and potentially improve our ability to localize epileptiform generators [15]. However, analysis of EEG data collected in the MRI environment has proven quite challenging, given a number of artifacts introduced during concurrent recordings.

EEG recordings putatively reflect the superposition of electric dipoles associated with synchronous activity from neural populations measured at the

scalp [3]. When collected inside the MR scanning environment, these signals are corrupted by noise due to the switching of magnetic fields, which creates a prominent gradient artifact. This gradient signal initially appeared problematic, however, a number of template-subtraction methods have been developed that can effectively remove this large signal [8].

More troublesome to this analysis is the quasi-periodic signal known as the ballistocardiogram (BCG) artifact, which cannot be easily removed with template based methods. The BCG is generated as EEG electrodes move due to pulsatile motion during the cardiac cycle, and presents broadly in the spectral frequencies often analyzed in EEG (0.5-25 Hz) [5]. The presence of these artifacts can dramatically change the spectral properties of the signal, and obscure ability to perform trial-by-trial analyses.

Thus far, a variety of techniques have been tested to remove BCG artifact from these data including template based average artifact subtraction based on cardiac r-wave timing [1], filtering [12], independent component analysis (ICA) [10], optimal basis sets (OBS) [14], clustering [17], and combined methods [5], many of which can be applied using manual or automated algorithms. Although each of these methods have achieved some degree of efficacy, most of these either require collection of additional data (e.g. ECG data) for template characterization purposes, require tedious manual artifact component identification, or require cardiac signal identification within the EEG itself, which is often only intermittently identifiable throughout a recording session.

Here, we introduce a new algorithm for removing artifact from EEG signals that uses low rank + sparse decomposition (LR+SD). To do so, we propose a mathematical model based on a reasonable experimental assumption that artifact components will be mathematically expressed differently than the data themselves. Specifically, we focus on event related or stimulus-locked EEG events, and assume that these will be represented sparsely in the time domain. Importantly, this method obviates the need for any reference or template artifact signal. As such, the combined effects of many types of artifacts can be removed in a single decomposition without the need for manual identification of artifact components in the data. We then assess the utility of this new algorithm on simulated and real data.

## 2 Methods

### 2.1 Low Rank + Sparse Decomposition Method

We denote  $\{\tilde{f}_i(k)\}_{i=1}^N$  the set of recorded EEG signals. The index  $i$  corresponds to the channel index, assuming we have a total of  $N$  electrodes distributed over the scalp. Moreover, we assume that each signal is recorded over  $K$  samples, i.e.  $k \in \{1, \dots, K\}$ . We consider  $J$  (unknown) artifacts, and will denote them  $\{f_j^A(k)\}_{j=1}^J$  where the index  $j$  identifies different artifacts. The goal of the artifact removal procedure is then to retrieve cleaned EEG signals,  $\{f_i(k)\}_{i=1}^N$ .

In the proposed method we assume the following model: each recorded EEG

channel is a linear combination of its cleaned version and the different artifacts. This model is equivalent to write:

$$\tilde{f}_i(k) = f_i(k) + \sum_{j=1}^J a_{ij} f_j^A(k), \quad (1)$$

where the mixing coefficients  $a_{ij}$  are unknown. In the following, we use a matrix formalism to model the global processes. To do so, we cast each EEG and artifact channels as columns of  $K \times N$  matrices:

$$\underbrace{\begin{pmatrix} | & | & & | \\ \tilde{f}_1 & \tilde{f}_2 & \dots & \tilde{f}_N \\ | & | & & | \end{pmatrix}}_{\tilde{F}} = \underbrace{\begin{pmatrix} | & | & & | \\ f_1 & f_2 & \dots & f_N \\ | & | & & | \end{pmatrix}}_F + \underbrace{\begin{pmatrix} | & | & & | \\ \tilde{f}_1^A & \tilde{f}_2^A & \dots & \tilde{f}_N^A \\ | & | & & | \end{pmatrix}}_{\tilde{F}^A}, \quad (2)$$

where  $\tilde{f}_i^A(k) = \sum_{j=1}^J a_{ij} f_j^A(k)$ . The matrix  $\tilde{F}$  contains all recorded EEG channels,  $F$  the wanted cleaned EEG signals and  $\tilde{F}^A$  contains the mixing of all artifacts. The latter can be written as  $\tilde{F}^A = \sum_{j=1}^J F_j^A$  where

$$F_j^A = \begin{pmatrix} | & | & & | \\ a_{1j} \tilde{f}_j^A & a_{2j} \tilde{f}_j^A & \dots & a_{Nj} \tilde{f}_j^A \\ | & | & & | \end{pmatrix}. \quad (3)$$

The key to our method is to notice that each matrix  $F_j^A$  has its columns proportional to the same vector  $\tilde{f}_j^A$  implying that  $\text{rank}(F_j^A) = 1$  and consequently  $\text{rank}(\tilde{F}^A) \leq J$ . Otherwise, the matrix  $F$  should contain events resulting from true EEG data. In our case, we focus on event related or stimulus-locked EEG events, which occur at specific times and affect a limited number of electrodes. Therefore, it is reasonable to assume that  $F$  is a sparse matrix. Thus the artifact removal problem is equivalent to performing a low-rank + sparse decomposition of  $\tilde{F}$ . The resulting sparse component therefore corresponds to the cleaned EEG signals. Such decomposition can be done by solving the following minimization problem:

$$\begin{aligned} (F, F^A) &= \arg \min \|F^A\|_* + \lambda \|F\|_1 \\ &\text{such that } \tilde{F} = F + F^A, \end{aligned} \quad (4)$$

where  $\|\cdot\|_*$  denotes the nuclear norm of a matrix (i.e. the sum of its singular values),  $\|\cdot\|_1$  denotes the sum of the absolute value of the matrix entries and  $\lambda$  is a positive parameter allowing us to control the balance of the sparsity of the  $F$

matrix and rank of  $F^A$ . This framework is very similar to Robust PCA, which has been actively studied in the mathematics community in recent years. Here, we extend the Lin et al. (2009) algorithm for artifact removal [11], and select the regularization parameter that maximizes signal to noise ration by running a sweep across all possible ranks. Upon publication, all code developed for this project will be available on the NITRC repository.

## 2.2 Simulated Dataset

In general, there is no ground truth EEG signal when data are empirical, making it difficult to assess the utility of artifact-removal algorithms. We therefore created a simulated dataset using the free BESA (Brain Electrical Source Analysis) to generate simulated EEG signals generated by three distributed dipole sources corrupted by known artifacts using a spherical head forward model. In order to add realistic noise to the data, we used ECG, EMG, and right and left EOG reference artifact recordings extracted from the free sample of the SHHS Polysomnography Database. These reference artifacts were normalized and added to the pure simulated EEGs using randomized mixing coefficients accordingly to a uniform distribution.

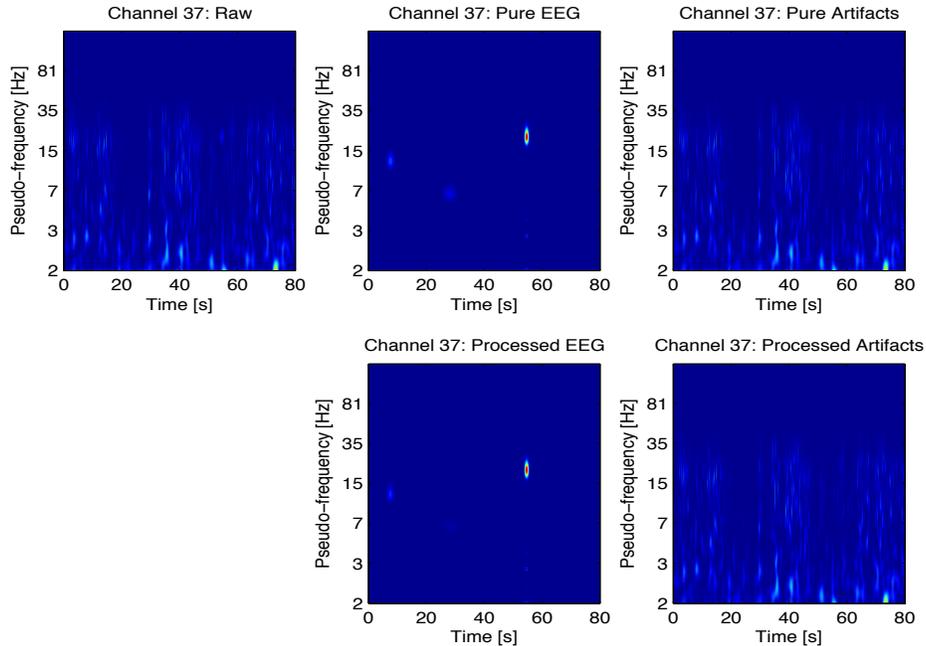
## 2.3 Empirical Data: Concurrent EEG-fMRI

Twenty healthy individuals (ages 23-30, 12 male) provided written informed consent to participate in this study, approved by the UCLA IRB. Concurrent recordings took place while subjects passively viewed 140 Gabor flashes, presented via an MR projector screen with a varied inter stimulus interval of 13.85 +/- -2.8 sec, a task known to generate reproducible occipital ERSPs in the alpha (8-12 Hz) spectral band [9]. EEG were recorded using a 256-channel GES 300 Geodesic Sensor Net (Electrical Geodesics, Inc.) at 500 Hz. MRI clock signals were synced with EEG data collection for subsequent MR artifact removal. Functional scans were acquired using 3-T Siemens Trio MRI Scanner using echo planar imaging gradient-echo sequence with echo time (TE) of 25msec, repetition time of 1s, 6mm slices, 2mm gap, flip angle 90 degrees, with 3mm in-plane resolution, ascending acquisition. EEG data then underwent MR gradient artifact removal by subtracting an exponentially weighted moving average template, according to methods described in [8].

We compare LR+SD to the established InfoMax ICA cleaning method, as implemented in Brain Analyzer v.2.0.2 software (Brain Products) using manual identification of cardiac signal within the EEG followed by the automated solutions procedure for identifying IC components correlated with cardiac signal. For comparison purposes, we also collected single modality EEG data outside the MR environment using the same stimuli and parameters in a copper shielded room (referred to as "Outside Scanner" in figures).

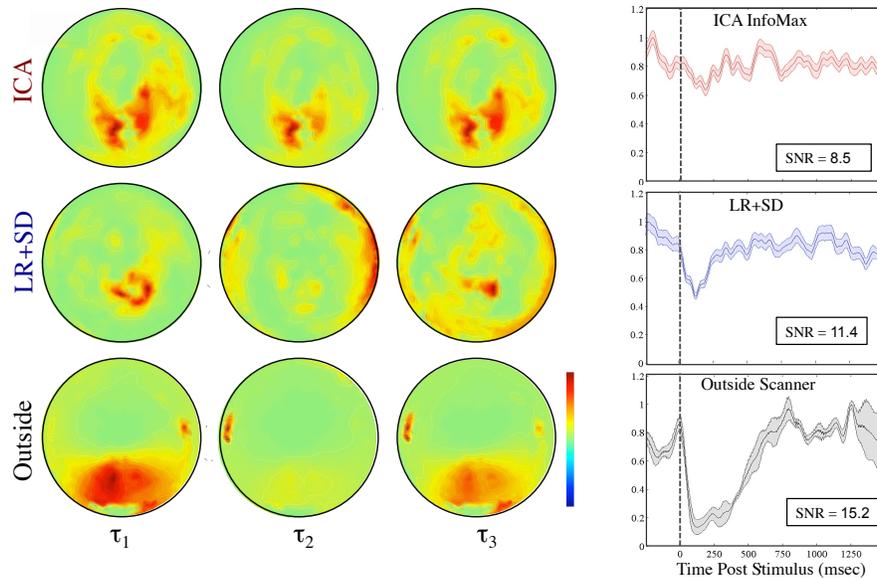
### 3 Results

**3.1 Simulated Data Results.** In the case of simulated data, we know the “true” solution. We adopt the time-frequency representation (TFR) to visualize results computed via a continuous wavelet transform (CWT) using the Morlet wavelet, to assess the efficiency of the proposed method. Figure 1 shows the TFRs for simulated data arising from three distributed dipole sources. The TFRs corresponding to the pure and artifact signals are depicted in the two upper right plots while the TFR obtained from the raw EEGs (pure EEGs mixed with pure artifacts) is given on the upper left plot. Notice that the time-frequency energy corresponding to pure EEGs is nearly undetectable due to the artifact energy. The signatures of each event are not visible in the raw EEGs’ TFRs yet they are clearly visible in the sparse component. In both the single and multiple source experiments the regularization parameter  $\lambda$  had value  $5 \cdot 10^{-3}$  which resulted in ranks of 4 and 5 for the low-rank artifact components, respectively. The proposed method shows excellent results in separating the artifact parts from the EEG signals of interest.

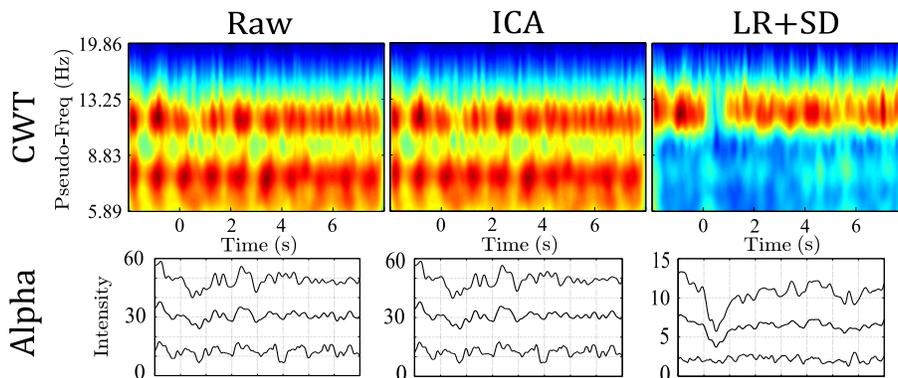


**Fig. 1.** Simulated Data Results. Three sources measured with an electrode located close to the primary motor cortex (source 3). (Top Panel) CWT of original simulated data consisting of three dipole sources corrupted by ECG, EMG and right and left EOG template artifacts (left), the true EEG data alone (middle), and artifact alone (right). (Bottom Panel) LR+SD result TFRs separating the cleaned EEG (left) from artifact (right). Values are normalized to maximum for each display.

**3.2 EEG-fMRI Empirical Data Results.** Group level ERSF results for LR+SD artifact removal of our experimental data are summarized and compared to ICA artifact removal as well as out of scanner data in Figure 2(a-d). Signal-to-noise ratio (SNR) was computed by calculating the ratio of the maximum absolute signal diminution in alpha power from 0 to 500 msec following stimulus presentation to the standard deviation of alpha power from the following 1000msec post stimulus. SNR was 8.5, 11.4, and 15.2 for ICA, LR+SD, and out of scanner data respectively. Figure 2(d) shows group level alpha spectral EEG data projected topographically for pre-stimulus (-250 msec), ERSF (50msec), and post stimulus (500 msec), with timings with respect to the stimulus occurring at time equal to zero. Figure 3 shows single-patient alpha power averaged over all stimuli for a window of 2sec pre stimulus to 8sec post. The raw data is shown in comparison with the sparse component from LR+SD and ICA using an average of the time-frequency intensity over alpha band frequencies. The strength at the specific frequency of 10Hz with bounds of one standard deviation is also shown for each dataset.



**Fig. 2.** Group Level Results comparing independent component analysis (ICA) and LR+SD based cleaning to EEG data collected outside of the MR scanner environment. (TOP PANEL) Normalized Alpha Power time courses derived from the ocular EEG channel averaged across all subjects plotted with mean  $\pm$  SEM for ICA, LR+SD and Outside Scanner Results. Signal-to-Noise ratios are shown in the lower left corner for each result with the stimulus occurring at time equal to zero. (LOWER PANEL) Group level alpha power results projected topographically for 500msec prior to stimulus onset (PRE), 50 msec following stimulus onset (ERSF), and 500msec following stimulus onset (POST).



**Fig. 3.** Single-subject alpha band TFR averaged over stimuli. Columns denote (left) the raw MRI signal, (center) the ICA-cleaned signal (right) the sparse component from LR+SD. For each dataset is shown (top row) TFR over a frequency range about the alpha band and (bottom row) the signal strength at 10Hz with single standard deviation bounds.

## 4 Discussion

In this paper, we introduce a novel method for removing artifacts from EEG signals, which decomposes data into the sum of two matrices: a sparse matrix representing the cleaned EEG data, and a low-rank matrix, which corresponds to the artifact portion of the data. We applied this algorithm to remove artifact from both simulated and empirical data. Overall, LR+SD was quantitatively more effective than ICA from an SNR perspective, and qualitatively more robust at recovering the diminution in alpha power topographically. Overall, the LR+SD algorithm is quite similar to the robust PCA algorithm which has been applied for background subtraction purposes [2], operating on the assumption that the EEG data events themselves are sparsely represented across channels in the time domain. Artifacts are conversely assumed to be broadly distributed at the channel level across the scalp. Here, we used only the Infomax ICA algorithm for comparison, since previous studies have shown that this algorithm is most effective at BCG removal. However, it should be noted that results using ICA for artifact removal of concurrent EEG-fMRI have varied [7].

After optimization, we found that a rank of 25 was required for the low rank matrix to describe the BCG artifact. Lower ranks were ineffective at isolating the BCG noise, due to the complexity of the BCG signal itself. In the clinical setting, the ECG cardiac signal is often low-pass filtered and gross changes of its signature (e.g. ST segment elevation) are examined. However, lower amplitude changes in higher frequencies of the ECG exist and have been shown to correlate with cardiac ischemia even up to 250Hz [6]. Given this broad spectral signature, it is unsurprising that more than a few sparse components were required to capture

the artifact signals in real data. In summary, LR+SD provides an automatic method for parsing data and artifact into separate groups, with the need to tune only one regularization parameter. Further work may use spatial information from electrode topographies as further constraints.

**Acknowledgements** The authors want to thank the Keck Foundation for their support.

## 5 Bibliography

- P.J. Allen et al. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *NeuroImage*. 8, 229-239 (1998)
- Cui,X. et al. Background Subtraction Using Low Rank and Group Sparsity Constraints *ECCV*. 1, 612-625 (2012)
- S. Dahne et al. Integration of Multivariate Data Streams With Bandpower Signals. *IEEE Transactions on Multimedia*. 15(5), 1001–1013 (2013)
- J. Daunizeau and H. Laufs and K.J. Friston. *EEGfMRI Information Fusion: Biophysics and Data Analysis*: Springer Press: EEG-fMRI. 511–526 (2009)
- S. Debener et al. Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI. *NeuroImage*. 34(2), 587–97 (2007)
- P.K. Douglas et al. Temporal and postural variation of 12-lead high-frequency QRS. *J Electrocardiol*. 39(3), 259–265 (2006)
- T. Eichele et al. Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *PNAS*. 102(49), 17798-803 (2005)
- R.I. Goldman et al. Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport*. 13(18), 2487–2492 (2002)
- S.A. Huettel et al. Linking hemodynamic and electrophysiological measures of brain activity. *Cerebral Cortex*. 14(2), 165–173 (2004)
- Tzzy-Ping Jung et al. Removing Electroencephalographic Artifacts by Blind Source Separation. *Psychophysiology*. 37, 163–178 (2000)
- Z. Lin et al. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *UILU-ENG-09-2215*. arXiv:1009.5055v2 (2009)
- A.J. Masterton et al.: Measurement and reduction of motion and BCG from simultaneous EEG and fMRI recordings. *NeuroImage*. 37, 202–211 (2007)
- K.J. Mullinger et al. Identifying the sources of the pulse artifact in EEG recordings made inside the MRI Scanner. *NeuroImage*. 71(1), 75–83 (2013)
- R.K. Niazy et al. Removal of fMRI environment artifacts from EEG data using optimal basis sets. *NeuroImage*. 28(3), 720–737 (2010)
- R. Thornton et al. Epileptic networks in focal cortical dysplasia revealed using electroencephalography-functional magnetic resonance imaging. *NeuroImage*. 70(5),822–837 (2011)
- P.A. Valdes-Sosa et al. Model driven EEG/fMRI fusion of brain oscillations. *Human Brain Mapping*. 30(9),2701–2721 (2009)
- Y. Zou et al. Automatic EEG artifact removal based on ICA and hierarchical clustering. *IEEE ICASSP*. 649–652 (2012)

# Group Sparse Kernelized Dictionary Learning for the Clustering of White Matter Fibers

Kuldeep Kumar and Christian Desrosiers

Ecole de technologie supérieure (ETS)  
1100 Notre-Dame W., Montreal, Canada H3C1K3  
kuldeep.kumar@etsmtl.ca, christian.desrosiers@etsmtl.ca

**Abstract.** This paper presents a novel method that combines kernelized dictionary learning and group sparsity to efficiently cluster white matter fiber tracts obtained from diffusion Magnetic Resonance Imaging (dMRI). Instead of having an explicit feature representation for the fibers, this method uses a non-linear kernel and specialized distance measures that can better learn complex bundles. Through the use of a global sparsity prior, the method also provides a soft assignment of fibers to bundles, making it more robust to overlapping fiber bundles and outliers. Furthermore, by using a group sparsity prior, it can automatically discard small and uninteresting bundles. We evaluate our method both qualitatively and quantitatively using expert labeled data, and compare it with state of the art approaches for this task.

## 1 Introduction

Due to its ability to infer the orientation of white matter fibers in-vivo and non-invasively, diffusion tensor imaging (DTI) has become an essential tool to study the microstructure of white matter in the brain. While extracting the individual fiber tracts from DTI data, a process known as tractography, is important to visualize the connection pathways in the brain, this process typically produces a large number of tracts which makes their analysis complex. To facilitate this analysis, it is often necessary to group the individual tracts into larger clusters, called bundles.

Methods proposed for the fiber clustering problem can be categorized in terms of the features and distance measures used to group the fibers into bundles. Features proposed to represent fibers include the distribution parameters (mean and covariance) of points along the fiber [2] and B-splines [11]. Approaches using such explicit features typically suffer from two problems: they are sensitive to the length and endpoint positions of the fibers and/or are unable to capture their full shape. Instead of using explicit features, fibers can also be compared using specialized distance measures. Popular distance measures for this task include the Hausdorff distance, the Minimum Direct Flip (MDF) distance and the Mean Closest Points (MCP) distance [3, 12]. Fiber clustering approaches can also be divided with respect to the clustering methods used, which include manifold embedding based approaches like spectral clustering and normalized cuts [2], agglomerative approaches like hierarchical clustering [3], k-means, and k-nearest

neighbors [12]. Several studies have also focused on incorporating anatomical features into the clustering [14] and on clustering large multi-subject datasets [9].

Recently, several researchers have studied the connection between clustering and factorization problems like dictionary learning [15] and non-negative matrix factorization [10]. For instance, dictionary learning has been shown to be a generalization of the traditional clustering problem, in which objects can be assigned to more than one cluster. In fiber clustering, such soft assignments are desirable since fiber bundles often overlap each other. Using a soft clustering, instead of hard one, can also make the method more robust to outliers (e.g., false fibers generated during tractography) that do not belong to any real cluster. Moreover, researchers have also recognized the advantages of applying kernels to existing clustering methods, like the k-means algorithm [4], as well as to dictionary learning approaches [13]. Such “kernelized” methods better capture the non-linear relations in the data.

The major contribution of this paper is a novel fiber clustering approach based on kernelized dictionary learning. By modeling the fiber clustering task as a dictionary learning problem, this approach provides a soft assignment of fibers to bundles, which makes it more robust to overlapping bundles and outliers. Furthermore, through the use of a non-linear kernel, it avoids the need to specify explicit features for the fibers, and can facilitate the separation of clusters in a manifold space. Also, by having both a global and group sparsity prior, our approach can control the minimum membership of fibers to bundles as well as the size of these bundles. This makes it more robust to the selection of the number of clusters in the output, a parameter which can be hard to tune, and allows it to automatically discard insignificant clusters. To our knowledge, this work is the first to combine group sparsity and kernelized dictionary learning. Our results on the fiber clustering problem show the potential of this approach for other medical imaging applications.

## 2 The proposed approach

### 2.1 The clustering problem

Before presenting our proposed approach, we first define the clustering problem and underline its link to dictionary learning. Let  $\mathbf{X} \in \mathbb{R}^{d \times n}$  be the data matrix of  $n$  fibers, where each column contains the feature vector  $\mathbf{x}_i \in \mathbb{R}^d$  of a fiber tract  $i$ . The traditional (hard) clustering problem can be defined as assigning each fiber to a bundle from a set of  $k$  bundles, such that fibers are as close as possible to their assigned bundle’s prototype (i.e., cluster center). Let  $\Psi^{k \times n}$  be the set of all  $k \times n$  cluster assignment matrices (i.e., matrices in which each row has a single non-zero value equal to one), this problem can be expressed as finding the matrix  $\mathbf{D}$  of  $k$  bundle prototypes and the fiber-to-bundle assignment matrix  $\mathbf{W}$  that minimize the following cost function:

$$\min_{\substack{\mathbf{D} \in \mathbb{R}^{d \times k} \\ \mathbf{W} \in \Psi^{k \times n}}} \frac{1}{2} \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F^2. \quad (1)$$

This formulation of the clustering problem can be seen as a special case of dictionary learning, where  $\mathbf{D}$  is the dictionary and  $\mathbf{W}$  is constrained to be a cluster assignment matrix, instead of constraining its sparsity.

While solving this clustering problem is NP-hard, optimizing  $\mathbf{W}$  or  $\mathbf{D}$  individually is easy. Thus, for a given dictionary  $\mathbf{D}$ , the optimal  $\mathbf{W}$  assigns each fiber  $i$  to the prototype  $k$  closest to its feature vector:

$$w_{ki} = \begin{cases} 1 & \text{if } k = \arg \min_{k'} \|\mathbf{x}_i - \mathbf{d}_{k'}\|^2, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Likewise, for a fixed  $\mathbf{W}$ , the optimal dictionary is found by solving a simple linear regression problem:

$$\mathbf{D} = \mathbf{X}\mathbf{W}^\top (\mathbf{W}\mathbf{W}^\top)^{-1}. \quad (3)$$

This suggests the following heuristic: starting with a dictionary containing a random subset of the columns of  $\mathbf{X}$ , optimize  $\mathbf{D}$  and  $\mathbf{W}$  alternatively, until convergence.

This clustering problem and simple heuristic correspond to the well-known k-means algorithm. With respect to dictionary learning, the dictionary update step described above is known as the Method of Optimal Directions (MOD) [1]. Although k-SVD [1] could also be used for this task, this technique focuses on learning large dictionaries efficiently and sacrifices the optimality of the dictionary update step to do so. In our case, the dictionary size corresponds to the number  $k$  of bundles (i.e., *clusters*), which is quite small in comparison to the number of tracts. Thus, updating the dictionary using MOD is quite fast.

## 2.2 Group sparse kernel dictionary learning

The k-means approach described in the previous section suffers from four important problems. First, it requires to encode fibers as a set of features, which is problematic due to the variation in their length and endpoints. Second, it assumes linear relations between the fibers and bundle prototypes, while these relations could be better defined in a non-linear subspace (i.e., the manifold). Third, it performs a hard clustering of the fibers, which can lead to poor results in the presence of overlapping bundles and outliers. Finally, it may find insignificant bundles (e.g., bundles containing only a few fibers) when the parameter controlling the number of clusters is not properly set.

To overcome these problems, we present a new clustering method based on group sparse kernelized dictionary learning. Let  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^q$  be a fiber mapping function such that  $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$  corresponds to a similarity kernel. Moreover, denote by  $\Phi$  the matrix of mapped fiber tracts, i.e.,  $\Phi = \phi(\mathbf{X})$ , and let  $\mathbf{K} = \Phi^\top \Phi$  be the kernel matrix. We reformulate the clustering problem as finding the dictionary  $\mathbf{D}$  and non-negative weight matrix  $\mathbf{W}$  minimizing the following problem:

$$\min_{\substack{\mathbf{D} \in \mathbb{R}^{q \times k} \\ \mathbf{W} \in \mathbb{R}_+^{k \times n}}} f(\mathbf{D}, \mathbf{W}) = \frac{1}{2} \|\Phi - \mathbf{D}\mathbf{W}\|_F^2 + \lambda_1 \|\mathbf{W}\|_1 + \lambda_2 \|\mathbf{W}\|_{2,1} + \frac{\lambda_3}{2} \|\mathbf{D}\|_F^2. \quad (4)$$

In this formulation,  $\|\mathbf{W}\|_1 = \sum_{i=1}^K \sum_{j=1}^N |w_{ij}|$  is an  $L_1$  norm prior which enforces global sparsity of  $\mathbf{W}$ , and  $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^K \|\mathbf{w}_i\|_2$  is a mixed  $L_{2,1}$  norm prior imposing the vector of row norms to be sparse. Concretely, the  $L_1$  norm prior limits the “membership” of fibers to a small number of bundles, while the  $L_{2,1}$  prior penalizes the clusters containing only a few fibers. The Frobenius norm prior on  $\mathbf{D}$  is used to avoid numerical problems when  $\mathbf{W}$  is singular (i.e., when one or more clusters are empty). Parameters  $\lambda_1, \lambda_2, \lambda_3 \geq 0$  control the trade-off between these three properties and the reconstruction error (i.e., the first term of the cost function).

Using an optimization approach similar to k-means, we alternate between updating the dictionary  $\mathbf{D}$  and the weight matrix  $\mathbf{W}$ . Since the dictionary prototypes are defined in the kernel space,  $\mathbf{D}$  cannot be computed explicitly. To overcome this problem, we follow the strategy proposed in [13] and define the dictionary as  $\mathbf{D} = \Phi\mathbf{A}$ , where  $\mathbf{A} \in \mathbb{R}^{n \times k}$ . Using this formulation,  $\mathbf{A}$  can be computed as follows:

$$\mathbf{A} = \mathbf{W}^\top (\mathbf{W}\mathbf{W}^\top + \lambda_3 \mathbf{I})^{-1}. \quad (5)$$

Matrix  $\mathbf{A}$  is initialized as a random selection matrix (i.e., random subset of columns in the identity matrix), which is equivalent to using a random subset of the transformed fibers (i.e., subset of columns in  $\Phi$ ) as the initial dictionary.

To update  $\mathbf{W}$ , we use an Alternating Direction Method of Multipliers (ADMM) method. First, we separate the problem in two sub-problems, one considering only the reconstruction error and the second considering only the (group) sparsity terms and non-negativity constraints, by introducing ancillary matrix  $\mathbf{Z}$ . The problem can then be reformulated as follows:

$$\min_{\substack{\mathbf{W} \in \mathbb{R}^{k \times n} \\ \mathbf{Z} \in \mathbb{R}_+^{k \times n}}} \frac{1}{2} \|\Phi - \Phi\mathbf{A}\mathbf{W}\|_F^2 + \lambda_1 \|\mathbf{Z}\|_1 + \lambda_2 \|\mathbf{Z}\|_{2,1}, \quad \text{s.t. } \mathbf{W} = \mathbf{Z}. \quad (6)$$

We then convert this constrained problem using an Augmented Lagrangian formulation with multipliers  $\mathbf{U}$ :

$$\min_{\substack{\mathbf{W}, \mathbf{U} \in \mathbb{R}^{k \times n} \\ \mathbf{Z} \in \mathbb{R}_+^{k \times n}}} \frac{1}{2} \|\Phi - \Phi\mathbf{A}\mathbf{W}\|_F^2 + \lambda_1 \|\mathbf{Z}\|_1 + \lambda_2 \|\mathbf{Z}\|_{2,1} + \frac{\mu}{2} \|\mathbf{W} - \mathbf{Z} + \mathbf{U}\|_F^2. \quad (7)$$

In an inner loop, we update  $\mathbf{W}$ ,  $\mathbf{Z}$  and  $\mathbf{U}$  alternatively, until convergence (i.e.,  $\|\mathbf{W} - \mathbf{Z}\|_F^2$  is below some threshold). To update  $\mathbf{W}$ , we derive the objective function with respect to this matrix and set the result to 0, yielding:

$$\mathbf{W} = (\mathbf{A}^\top \mathbf{K} \mathbf{A} + \mu \mathbf{I})^{-1} (\mathbf{A}^\top \mathbf{K} + \mu(\mathbf{Z} - \mathbf{U})). \quad (8)$$

Optimizing  $\mathbf{Z}$  corresponds to solving a group sparse proximal problem (see [7]). This can be done in two steps. First, we do a  $L_1$ -norm shrinkage by applying the non-negative soft-thresholding operator to each element of  $\mathbf{W} + \mathbf{U}$ :

$$\hat{z}_{ij} = S_{\frac{\lambda_1}{\mu}}^+(w_{ij} + u_{ij}) = \max \left\{ w_{ij} + u_{ij} - \frac{\lambda_1}{\mu}, 0 \right\}, \quad i \leq K, j \leq N. \quad (9)$$

Then,  $\mathbf{Z}$  is obtained by applying a group shrinkage on each row of  $\hat{\mathbf{Z}}$ :

$$\mathbf{z}_{i\cdot} = \max \left\{ \|\hat{\mathbf{z}}_{i\cdot}\|_2 - \frac{\lambda_2}{\mu}, 0 \right\} \cdot \frac{\hat{\mathbf{z}}_{i\cdot}}{\|\hat{\mathbf{z}}_{i\cdot}\|_2}, \quad i \leq K. \quad (10)$$

Finally, as in standard ADMM methods, the Lagrangian multipliers are updated as follows:

$$\mathbf{U}' = \mathbf{U} + (\mathbf{W} - \mathbf{Z}). \quad (11)$$

### 2.3 Algorithm summary and complexity

The clustering process of our proposed method is summarized in Algorithm 1. In this algorithm, the user provides a matrix  $\mathbf{Q}$  of pairwise fiber distances (see Section 3 for more details), the maximum number of clusters  $k$ , as well as the trade-off parameters  $\lambda_1, \lambda_2, \lambda_3$ , and obtains as output the dictionary matrix  $\mathbf{A}$  and the cluster assignment weights  $\mathbf{W}$ . At each iteration,  $\mathbf{W}$ ,  $\mathbf{Z}$  and  $\mathbf{U}$  are updated by running at most  $T_{\text{in}}$  ADMM loops, and are then used to update  $\mathbf{A}$ . This process is repeated until  $T_{\text{out}}$  iterations have been completed or the cost function  $f(\mathbf{D}, \mathbf{W})$  converged. The soft assignment of  $\mathbf{W}$  can be converted to a hard clustering by assigning each fiber  $i$  to the bundle  $k$  for which  $w_{ik}$  is maximum.

The complexity of this algorithm is mainly determined by the initial kernel computation, which takes  $O(n^2)$  operations, and updating the assignment weights in each ADMM loop, which has a total complexity in  $O(T_{\text{out}} \cdot T_{\text{in}} \cdot k^2 \cdot n)$ . Since  $T_{\text{out}}$ ,  $T_{\text{in}}$  and  $k$  are typically much smaller than  $n$ , the main bottleneck of the method lies in computing the pairwise distances  $\mathbf{Q}$  used as input. For datasets having a large number of fibers (e.g., more than  $n = 100,000$  fibers), this matrix could be computed using an approximation strategy such as the the Nyström method [6].

## 3 Experiments

We evaluated the performance of our proposed method on a dataset of expert labeled bundles, provided by the Sherbrooke Connectivity Imaging Laboratory (SCIL). The source dMRI data was acquired from a 25 year old healthy right-handed volunteer and is described in [5]. We used 10 of the largest bundles, consisting of 4449 fibers identified from the cingulum, corticospinal tract, superior cerebellar peduncle and other prominent regions. Figure 2(b) shows the coronal and sagittal plane view of the ground truth set.

Although our method has several parameters, only two of them require data specific tuning:  $\lambda_1$  and  $\lambda_2$ . The RBF kernel parameter  $\gamma$  depends on the distance measure used, not the dataset. For these experiments, we used the Mean Closest Points (MCP) distance [3] to compute the pairwise fiber distances  $\mathbf{Q}$ , and set  $\gamma$  to 0.01. Also,  $\lambda_3$  and  $\mu$  correspond to regularization parameters and should be set to a small positive value. In our experiments, we have used  $\lambda_3 = 10^{-6}$  and  $\mu = 0.01$  for these parameters. According to Eq. 9,  $\lambda_1/\mu$  corresponds to a minimum threshold for the assignment weights. As shown in Figure 1(a), this

---

**Algorithm 1:** ADMM method for group sparse kernelized clustering

---

**Input:** Pairwise fiber distance matrix  $\mathbf{Q} \in \mathbb{R}^{n \times n}$ ;  
**Input:** The maximum number of fiber bundles  $k$ ;  
**Input:** The RBF kernel parameter  $\gamma$ ;  
**Input:** The cost trade-off parameters  $\lambda_1, \lambda_2, \lambda_3$  and Lagrangian parameter  $\mu$ ;  
**Input:** The maximum number of inner and outer loop iterations  $T_{\text{in}}, T_{\text{out}}$ ;  
**Output:** The dictionary  $\mathbf{A} \in \mathbb{R}^{n \times k}$  and assignment weights  $\mathbf{W} \in \mathbb{R}_+^{n \times k}$ ;

Initialize the kernel matrix:  $k_{ij} = \exp(-\gamma \cdot q_{ij}^2)$ ;  
Initialize  $\mathbf{A}$  as a random selection matrix and  $t_{\text{out}}$  to 0;

**while**  $f(\mathbf{D}, \mathbf{W})$  not converged and  $t_{\text{out}} \leq T_{\text{out}}$  **do**

Initialize  $\mathbf{U}$  and  $\mathbf{Z}$  to all zeros and  $t_{\text{in}}$  to zero;

**while**  $\|\mathbf{W} - \mathbf{Z}\|_F^2$  not converged and  $t_{\text{in}} \leq T_{\text{in}}$  **do**

Update  $\mathbf{W}$ ,  $\mathbf{Z}$  and  $\mathbf{U}$ :

$\mathbf{W} \leftarrow (\mathbf{A}^\top \mathbf{K} \mathbf{A} + \mu \mathbf{I})^{-1} (\mathbf{A}^\top \mathbf{K} + \mu (\mathbf{Z} - \mathbf{U}))$ ;

$\hat{z}_{ij} \leftarrow \max \left\{ w_{ij} + u_{ij} - \frac{\lambda_1}{\mu}, 0 \right\}, \quad i \leq K, j \leq N$ ;

$z_{i\cdot} \leftarrow \max \left\{ \|\hat{z}_{i\cdot}\|_2 - \frac{\lambda_2}{\mu}, 0 \right\} \cdot \frac{\hat{z}_{i\cdot}}{\|\hat{z}_{i\cdot}\|_2}, \quad i \leq K$ ;

$\mathbf{U} \leftarrow \mathbf{U} + (\mathbf{W} - \mathbf{Z})$ ;

$t_{\text{in}} \leftarrow t_{\text{in}} + 1$ ;

Update dictionary:  $\mathbf{A} \leftarrow \mathbf{W}^\top (\mathbf{W} \mathbf{W}^\top + \lambda_3 \mathbf{I})^{-1}$  ;  
 $t_{\text{out}} \leftarrow t_{\text{out}} + 1$ ;

**return**  $\{\mathbf{A}, \mathbf{W}\}$  ;

---

value can be used to control the mean number of non-zero weights per fiber (i.e., how soft or hard is the clustering). Likewise,  $\lambda_2/\mu$  is a minimum threshold on the total membership to a bundle and, as shown in Figure 1(b), controls the size of bundles in the output. Finally, following the convergence rate shown in Figure 1(c), we have used  $T_{\text{out}} = 20$  for the maximum number of iterations. The same value was used for the number of inner loop iterations (i.e.,  $T_{\text{in}} = 20$ ).

Figure 2(a) shows the mean Adjusted Rand Index (**ARI**) [12] obtained by our method, denoted by **MCP+L1+L21**, over 5 runs with different random initializations. We compared this method with two well-known fiber clustering approaches: QuickBundles (**QB**) [8] and Normalized cuts (**Ncuts**) [2]. QuickBundles recursively groups fibers between which the Minimum Direct Flip (**MDF**) distance is below a specified threshold. Ncuts performs a spectral embedding of the fibers encoded as the mean and covariance parameters of the points distribution, and then clusters the embedded fibers using a recursive partitioning strategy or k-means. Based on earlier results, we used 25 eigenvectors for the embedding and k-means for clustering. We also tested our method without the group sparsity prior (i.e., using  $\lambda_2 = 0$ ) and called **MCP+L1** this simplified model.

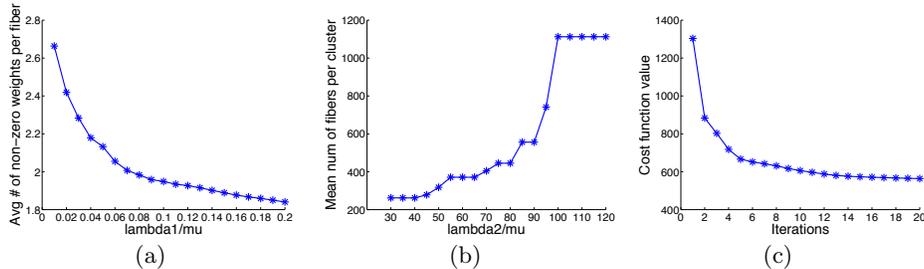


Fig. 1: **(a)** Mean number of non-zero assignment weights per fiber, for  $\lambda_2/\mu = 80$  and increasing  $\lambda_1/\mu$ . **(b)** Mean number of fibers per bundle, for  $\lambda_1/\mu = 0.1$  and increasing  $\lambda_2/\mu$ . **(c)** Cost function value at each iteration of a sample run.

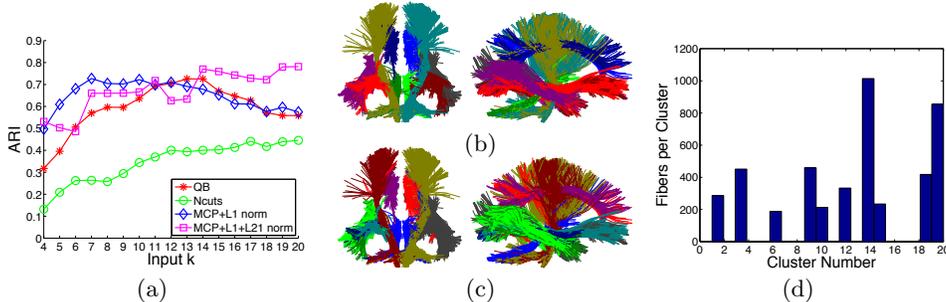


Fig. 2: **(a)** Mean ARI of QB, Ncuts, MCP+L1 ( $\lambda_1/\mu = 0.1$ ,  $\lambda_2/\mu = 0$ ) and MCP+L1+L21 ( $\lambda_1/\mu = 0.1$ ,  $\lambda_2/\mu = 80$ ), for increasing  $k$ . **(b)**-**(c)** Ground truth bundles and clustering output of MCP+L1+L21 for  $k = 20$ . **(d)** Distribution of bundle sizes corresponding to this output.

From these results, we see that Ncuts performs worse than all other methods. This is possibly due to the fact that the features used to encode the fibers do not fully capture their shape. Moreover, we observe that the peak ARI of QuickBundles is similar to that of MCP+L1, but the latter peaks closer to the true number of bundles (i.e., 10). Finally, we see that the MCP+L1+L21 method, which also considers group sparsity, obtains the highest ARI and is less sensitive to the value of  $k$  given as input. The bundles obtained by this method for  $k = 20$  are presented in Figure 2(c). As shown in Figure 2(d), this clustering contains the same number of clusters as the ground truth, even though the maximum number of clusters was set to  $k = 20$ .

## 4 Conclusion

We have presented a new fiber clustering approach based on dictionary learning. This approach uses a non-linear kernel which avoids having to define features for the fibers and can represent complex bundles. Furthermore, by using an  $L_1$  norm prior, instead of hard clustering constraints, it is more robust to overlap-

ping bundles and outliers. Finally, since it also includes a group sparsity prior, our approach can find more interesting bundles than other methods for this task. Experiments conducted on expert labeled data show our methods to outperform state of the art fiber clustering approaches such as QuickBundles and Normalized Cuts. In future work, we will extend the proposed model to incorporate anatomical information in the form of atlases.

## References

1. Aharon, M., Elad, M., Bruckstein, A.: k-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing* 54(11), 4311–4322 (2006)
2. Brun, A., Knutsson, H., et al.: Clustering fiber traces using normalized cuts. In: *MICCAI 2004*, pp. 368–375. Springer (2004)
3. Corouge, I., Gouttard, S., Gerig, G.: Towards a shape model of white matter fiber bundles using diffusion tensor mri. In: *IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2004*. pp. 344–347. IEEE (2004)
4. Dhillon, I., Guan, Y., Kulis, B.: Kernel k-means: spectral clustering and normalized cuts. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 551–556. ACM (2004)
5. Fortin, D., Aubin-Lemay, C., et al.: Tractography in the study of the human brain: a neurosurgical perspective. *The Canadian Journal of Neurological Sciences* 39(6), 747–756 (2012)
6. Fowlkes, C., Belongie, S., Chung, F., Malik, J.: Spectral grouping using the nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(2), 214–225 (2004)
7. Friedman, J., Hastie, T., Tibshirani, R.: A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736* (2010)
8. Garyfallidis, E., Brett, M., et al.: Quickbundles, a method for tractography simplification. *Frontiers in neuroscience* 6 (2012)
9. Guevara, P., Duclap, D., et al.: Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *Neuroimage* 61(4), 1083–1099 (2012)
10. Kim, H., Park, H.: Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23(12), 1495–1502 (2007)
11. Maddah, M., Crimson, W., Warfield, S.: Statistical modeling and em clustering of white matter fiber tracts. In: *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*. pp. 53–56. IEEE (2006)
12. Moberts, B., Vilanova, A., van Wijk, J.: Evaluation of fiber clustering methods for diffusion tensor imaging. In: *Visualization, 2005*. pp. 65–72. IEEE (2005)
13. Nguyen, H., Patel, V., Nasrabadi, N., Chellappa, R.: Kernel dictionary learning. In: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. pp. 2021–2024. IEEE (2012)
14. O’Donnell, L., Westin, C.F.: Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Transactions on Medical Imaging* 26(11), 1562–1575 (2007)
15. Sprechmann, P., Sapiro, G.: Dictionary learning and sparse coding for unsupervised clustering. In: *ICASSP 2010*. pp. 2042–2045. IEEE (2010)

# 3D Mouse Left Ventricle Reconstruction using Sparse MR Images with Arbitrary Orientations

Yang Yu<sup>1</sup>, Jingjing Liu<sup>1</sup>, Dimitris Metaxas<sup>1</sup>, Leon Axel<sup>2</sup>

<sup>1</sup>Department of Computer Science, Rutgers University, Piscataway, NJ, USA

<sup>2</sup>Radiology Department, New York University, NY, USA

**Abstract.** This paper presents a framework for 3D left ventricle reconstruction using sparse magnetic resonance (MR) images with different orientations. Due to the inadequate inter-slice resolution, both short- and long-axis cardiac MR images are commonly acquired to reveal the left ventricle shape and motion. The contours in these images show different profiles of left ventricle and contain its essential shape information. In this paper, we propose a new deformable model to segment left ventricle on 2D slices with different orientations, and reconstruct its 3D model that matches all the contours in the images. An alternating optimization algorithm is proposed to efficiently solve the problem. The framework is applied on mouse cardiac MR data and shows promising results.

## 1 INTRODUCTION

In recent years, magnetic resonance imaging (MRI) is frequently used for the analysis of cardiac function. It enables the generation of 3D deformable models of the heart, from which accurate diagnostic information can be derived. However, it is hard to acquire high-resolution 3D cardiac MR images in animals due to the fast beating heart and the breathing, especially from the experimental small animals, like a mouse. The mouse heart is about 1000th the size of a human heart and beats much faster, at 400-600 beats per minute (bpm), than a human heart, with 60-80 bpm. Although currently available MRI instruments for mouse imaging operate at a higher magnetic field strength (4.7T or above) than clinical MRI scanners, they are still unable to provide adequate spatial resolution in 3D. In practice, only sparse good quality images on a few short- and long-axis slices are acquired to visualize the cardiac motion. They provide enough information for experts to visually analyze the cardiac motion, while it is still very challenging to reconstruct a 3D heart model based on these sparse slices [15].

Most of the previous work focuses on left ventricle reconstruction based on short-axis [9–11]. Since the short-axis images are parallel to each other, they are usually combined to a 3D volume. However, due to the limited number of slices, the inter-slice resolution is usually much lower than intra-slice directions. The distance between slices is about 10 times the pixel distance inside each slice. During the MR image acquisition, the long-axis images usually are first generated to localize the heart position, while they are rarely used for the segmentation for the left ventricle [5, 12]. These images have higher resolution in the long axis, which can help overcome the low inter-slice resolution on short-axis images. Therefore, we utilize both short- and long-axis MR

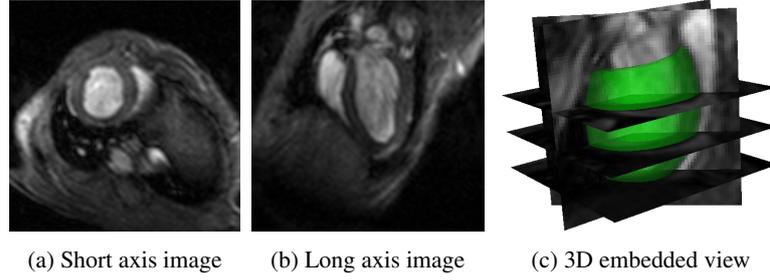


Fig. 1: The mouse left ventricle has a ring shape on the short-axis images (a) and a U shape on the long-axis images (b). By mapping them to the 3D anatomic space, they intersect with the reconstructed left ventricle model exactly on the left ventricle boundaries (c).

images simultaneously in this work for 3D left ventricle reconstruction. Each image at different position provides different contours of the model. Meanwhile, the consistency among them improves the robustness of the reconstruction.

The short- and long-axis images are instances of the same volume of different orientations. Fradkin et al. [6] utilized their consistency to infer the short-axis image position based on the long-axis segmentation result. However, the spatial relationship is only used for initialization. The short-axis contours are then deformed independently. The contours after deformation may be inconsistent with the long-axis ones. Koikkalainen et al. [7] reconstructed a 3D heart model based on parallel MR images from the short and long axes. Different from usual long-axis slices, which are radially placed, they acquired parallel images in the long axis. The slices with different orientations are considered as volume data of the same region with different resolutions. A reference model is registered with them simultaneously to overcome the insufficient sampling for each single volume data. Since most long-axis images are not parallel in MR acquisition, their method will require an additional protocol for heart reconstruction. van Assen et al. [1] proposed a left ventricle reconstruction algorithm based on multiple shape priors. Based on active shape models (ASMs), they first build a point distribution model from training shapes, and then fit this model to all the 2D images to refine the segmentation. The images generate forces on the intersection of the 3D model with the corresponding 2D plans. Similarly, sparse shape composition [16] is used to represent shape models based on sparse reconstruction. The methods, like ASM, represent the shapes based on a large number of training samples, but the training shapes are not always available in clinical applications.

To address the limitations in previous efforts to incorporate 2D slices with arbitrary orientations for 3D left ventricle reconstruction, we introduce a new reconstruction framework. The main contributions of the work are as follows. First, all the slices are segmented simultaneously with a 3D left ventricle model. The 2D contours are just the projection of the model on the corresponding images, so we handle the inconsistency among all the contours, i.e., the short- and long-axis contours are not exactly intersected. Second, there are no restrictions on the position and orientation of each s-

lice. Any additional slice will help improve the segmentation accuracy and robustness. Third, only an elastic shape prior [4] is required in our framework. The reference shape can be generated with one sample data or built manually by expert without any sample. Different from the methods based on multiple shape priors, which ensure that the shape follows a point distribution model, we constrain the non-rigid deformation of the reference shape. The 3D shape regularization term is integrated into all the 2D image segmentations to form a unified problem, which is efficiently solved by our proposed alternating optimization algorithm.

## 2 METHODOLOGY

Given a group of 2D cardiac MR images  $I_i$ , which have known transformations  $T_i$  to the 3D anatomical coordinate system, we expect to reconstruct a 3D left ventricle shape model  $T(S_{ref})$ , where  $S_{ref}$  is a reference left ventricle model and  $T$  is a non-rigid transformation. The projection of the reconstructed model  $T(S_{ref})$  onto image  $I_i$  is defined as  $P_i(T(S_{ref}))$ . It should match with the left ventricle area in the image. The fitness of the model to each image  $I_i$  is measured by the energy function  $E_{img}$ . Since the slices are sparse in the 3D volume, the reconstruction problem is under constrained with only the image information. Therefore, we further assume the model is deformed from the reference model  $S_{ref}$  with a smooth non-rigid deformation  $T$ . The model reconstruction is formulated as the following optimization problem:

$$\min_T \left\{ \sum_i E_{img}(P_i(T(S_{ref})), I_i) + \gamma R(T) \right\} \quad (1)$$

where  $E_{img}$  is the energy term for the fitness to each image  $I_i$ ,  $R(T)$  is the regularization term for the deformation  $T$  and  $\gamma$  is a trade-off parameter.

The image energy term  $E_{img}$  is defined based on both the shape and appearance information. The conventional active contour models focus only on the boundaries of the models. They deform the contours to fit locations that have high probabilities to be boundaries. In our model, we also consider the appearance of the interior region. The appearance statistics are adaptively learned during the deformation. The model is updated based not only on the edge information, but also the region statistics to ensure the appearance consistency of the new territory. The region-based deformable model is defined based on free form deformation. Instead of deformable contours, the whole interior region is deformed to optimize both the edge and the region energy function:

$$E_{img} = E_{edg} + \mu E_{reg} \quad (2)$$

where  $E_{edg}$  is the edge energy term,  $E_{reg}$  is the region energy term and  $\mu$  is a constant that balances the contributions from the two terms. In our formulation, we are able to omit the model smoothness term in 2D images since the whole model smoothness is regularized by the smooth non-rigid transformation of the 3D model.

The model is attracted to edge feature with high image gradient via the edge energy term  $E_{edg}$ . A distance map to the edge feature is built based on gradient vector field [14]. The edge force moves the contour to the minimum of the distance map. Therefore, the edge energy term  $E_{edg}$  is defined as:

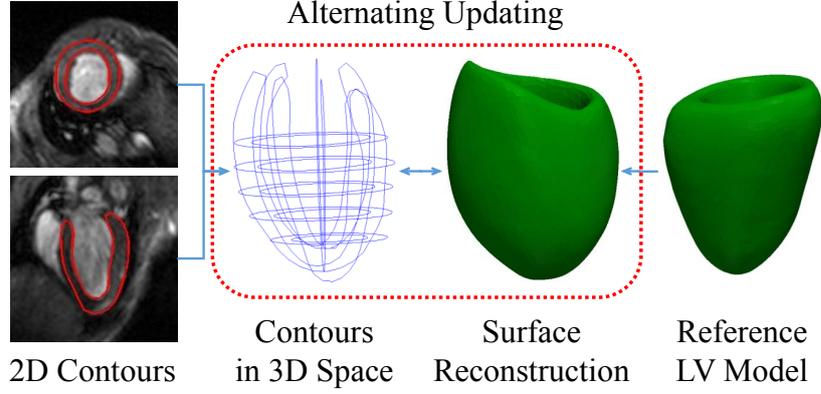


Fig. 2: The pipeline of our 3D left ventricle system. The 3D surface model is deformed from a reference model to fit 2D contours, while the 2D contours is constrained by both image cues and 3D model prior. They are updated alternately to reconstruct the left ventricle model.

$$E_{edg} = \int_C \Phi(\mathbf{x}) d\mathbf{x} \quad (3)$$

where  $C$  is the contour in a 2D image and  $\Phi$  is the distance map function.

The probability of each pixel belonging to the model is defined based on the interior intensity distribution from last iteration. The region energy term deform the model toward areas with high probability. It is defined as:

$$E_{reg} = \int_R \log P(\mathbf{x}) d\mathbf{x} \quad (4)$$

where  $R$  is the interior region of the contour and  $P$  is the probability of each pixel as the interior region of the model.

The 2D contours are projections of one 3D left ventricle model to the corresponding images. Therefore, different from the 2D deformation regularization term in previous segmentation algorithms, we employ a 3D shape prior to constrain all the 2D segmentations simultaneously. The 3D model is defined based on the deformation of a reference left ventricle model  $T(S_{ref})$ . We regularize the non-rigid deformation  $T$  to ensure that the new model is still similar to the reference one. The smoothness of transformation  $T$  is defined as:

$$R(T) = \int_{\mathbb{R}^3} \frac{\tilde{T}(\tilde{\mathbf{x}})}{\tilde{G}(\tilde{\mathbf{x}})} d\tilde{\mathbf{x}} \quad (5)$$

where  $G$  is Gaussian kernel function and  $\tilde{G}$  is its Fourier transform. Function  $\tilde{T}$  indicates the Fourier transform of the deformation function  $T$  and  $\tilde{\mathbf{x}}$  is a frequency domain variable. Gaussian kernel is used as a low-pass filter to regularize the high frequency part of the deformation and enforce the smoothness.

**Algorithm 1** 3D left ventricle reconstruction

---

**Input:** The sparse images  $I_i$  with arbitrary orientation, and the reference left ventricle model  $S_{ref}$

**Output:** The data-specific 3D left ventricle model

Initialize the 2D contours  $C_i$  with graph cuts

**repeat**

Transform the contours  $C_i$  to 3D anatomic space

Deform the 3D reference model  $S_{ref}$  based on (7)

Find the model-plane intersections  $P_i(T(S_{ref}))$

Deform the contours  $C_i$  based on (8)

**until**  $C_i$  and  $T$  converge.

---

**2.1 Deformable Model Implementations**

The image forces are only defined on the intersection of the model in each plane. They are not applied directly to the vertices of the model. Therefore, we introduce the contours of the left ventricle on the images  $C_i$  and reformulate the energy function as:

$$\min_{C_i, T} \left\{ \sum_i [E_{img}(C_i, I_i) + \lambda D(C_i, P_i(T(S_{ref})))] + \gamma R(T) \right\} \quad (6)$$

where  $D$  is the distance between the contour  $C_i$  and the projection of the left ventricle model  $P_i(T(S_{ref}))$ . In this formulation, instead of deforming the reference model directly, the image forces only deform the 2D contours. Therefore, the whole energy function is separated into two parts. The 2D contours and the 3D model can be optimized alternately with Algorithm 1.

We initialize the 2D segmentation via graph cuts [2, 3]. It is very effective to generate a coarse segmentation, while it requires lots of interaction to refine the result. In our work, we use a two-stage segmentation for short-axis images based on its donut shape [13]. We use only a few strokes to indicate the blood pool. Then the left ventricle is automatically segmented with no further interaction. Furthermore, the long-axis images are also segmented via graph cuts, which initialized based on its relative position with short-axis images. The regional segmentation results on all the images are then translated into boundary ones and refined by Metamorphs [13].

The initial contours are first transformed to the 3D anatomic space. Then assuming the contours  $C_i$  are fixed, the reference left ventricle model is deformed to the contours. The energy function is reduced to:

$$\min_T \left\{ \lambda \sum_i D(C_i, P_i(T(S_{ref}))) + \gamma R(T) \right\} \quad (7)$$

We use coherent point drift [8] to optimize (7). The result model maintains the shape of the reference model, and balances the differences among the contours in different slices.

The deformed model  $T(S_{ref})$  is then projected to the 2D spaces. We use them as shape priors and optimize the contours  $C_i$ . In this step, the energy function is independent for each slice:

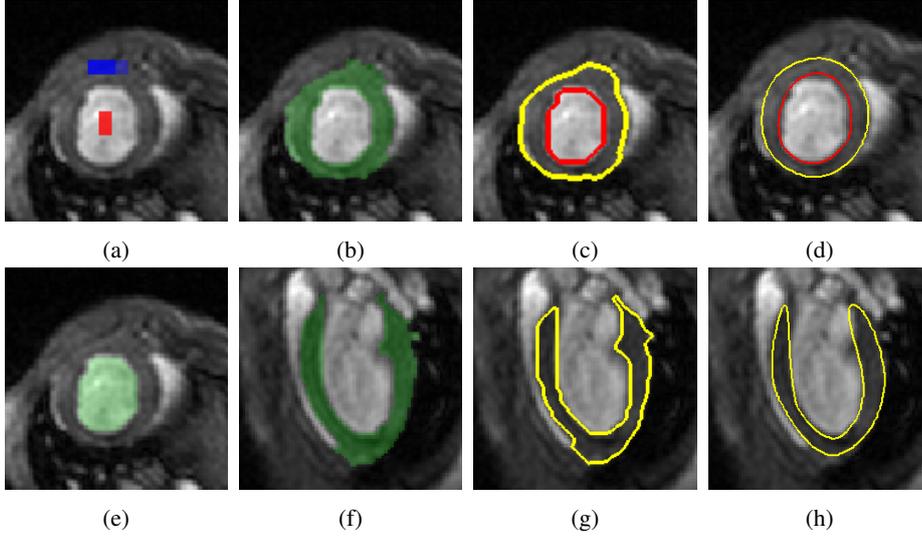


Fig. 3: (a) The initial label for graph cuts, (e) the blood pool segmentation result, (b, f) the left ventricle region (green) from graph cuts on both short- and long-axis images, (c, g) the boundaries based on graph cuts and (d, h) the final result based on our framework.

$$\min_{C_i} \{E_{img}(C_i, I_i) + \lambda D(C_i, P_i(T(S_{ref})))\} \quad (8)$$

where the distance function  $D$  are defined by the distance maps of the model projection on the slices. This will make the contours more consistent with the left ventricle model.

During the alternating optimization, the parameter  $\lambda$  will increase to further enforce the consistency between the left ventricle model and all the contours. When  $\lambda \rightarrow \infty$ , the alternating algorithm (6) will converge to (1).

### 3 EXPERIMENTS

We test our reconstruction algorithm on mouse cardiac MR images. Sparse short- and long-axis images are acquired from the C57BL/6 mice. For each data, there are four to six short-axis slices that are parallel to each other with equal intervals, and four long-axis slices that are radially spaced every  $45^\circ$ . Their positions in the anatomic space are recorded during the acquisition.

We use a few strokes inside and outside the blood pool area, as shown in Fig. 3a, to initialize the segmentation, and get the blood pool area in Fig. 3e. This step is relatively stable due to the high intensity difference between the blood and heart wall. Then the region just outside the blood pool is set as the left ventricle. Graph cuts is used to produce a rough segmentation of the left ventricle on both short- and long-axis

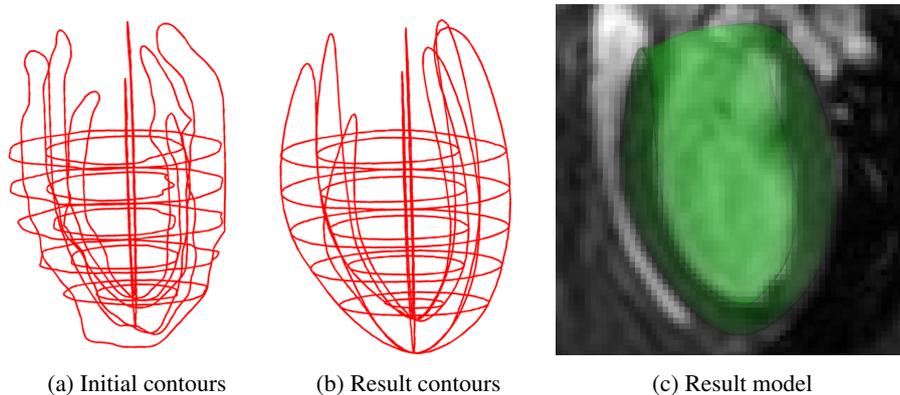


Fig. 4: (a) The initial 2D contours mapped onto 3D anatomic space, (b) the result contours after deformation based on our framework and (c) the resulted 3D model embedded onto a long-axis image.

images. The result of this step often leaks out to other tissues due to the similar intensity among them (Fig. 3b, f). It consequently affects the corresponding boundaries refined by Metamorphs, which cannot correct the region with heavy leak (Fig. 3c, g). Our proposed deformable model overcome these problems with 3D shape constraints. It achieves better segmentation result in 2D images (Fig. 3d, h).

We apply our alternating reconstruction algorithm to generate 3D left ventricle model based on the initial 2D contours. It is noticeable in Fig. 4a that the contours from the short- and long-axis images do not intersect with each other based on only 2D information. Our model introduces a 3D shape model to regularize all the contours. It improves the 2D segmentation results on different slices (Fig. 3d, h). Meanwhile, different from the initial contours projected into anatomic space, the our results balance their differences and make them consistent with each other (Fig. 4b). The 3D left ventricle model is also constructed based on our model. It is embedded into a long axis image in Fig. 4c. The model is smooth and match left ventricle wall in the image.

## 4 CONCLUSIONS

We have presented a new framework for 3D left ventricle reconstruction using sparse short- and long-axis images based on only one shape prior. Less MR images are required to acquire by using our method. This is not only very important for mouse cardiac imaging, but also desired for human data acquisition, since it will reduce the potential risk of strong magnetic field and improve the patient’s comfort. In the future, we will test our framework on human cardiac MR data. Meanwhile, we will introduce a left ventricle detection module to substitute graph cuts-based initialization and build a fully automatic system.

## References

1. van Assen, H.C., Danilouchkine, M.G., Frangi, A.F., Ordás, S., Westenberg, J.J.M., Reiber, J.H.C., Lelieveldt, B.P.F.: SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. *Medical Image Analysis* 10(2), 286–303 (2006)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(11), 1222–1239 (2001)
3. Boykov, Y., Jolly, M.P.: Interactive organ segmentation using graph cuts. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) *Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, vol. 1935, pp. 276–286. Springer Berlin Heidelberg (2000)
4. Chen, S., Cremers, D., Radke, R.J.: Image segmentation with one shape prior – A template-based formulation. *Image and Vision Computing* 30(12), 1032–1042 (2012)
5. Ciofolo, C., Fradkin, M.: Segmentation of pathologic hearts in long-axis late-enhancement MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) *Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, vol. 5241, pp. 186–193. Springer Berlin Heidelberg (2008)
6. Fradkin, M., Ciofolo-Veit, C., Mory, B., Hautvast, G., Breeuwer, M.: Fully automatic segmentation of short and long axis cine cardiac MR. In: 12th Annual SCMR Scientific Sessions, *Journal of Cardiovascular Magnetic Resonance*, vol. 11 (2009)
7. Koikkalainen, J., Pollari, M., Lötjönen, J., Kivistö, S., Lauerma, K.: Segmentation of cardiac structures simultaneously from short- and long-axis MR images. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science*, vol. 3216, pp. 427–434. Springer Berlin Heidelberg (2004)
8. Myronenko, A., Song, X.: Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(12), 2262–2275 (2010)
9. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis* 15(2), 169–184 (2011)
10. Radau, P., Lu, Y., Connelly, K., Paul, G., Dick, A., Wright, G.: Evaluation framework for algorithms segmenting short axis cardiac MRI. *MICCAI Workshop: Cardiac MR Left Ventricle Segmentation Challenge* (2009)
11. Suri, J.S.: Computer vision, pattern recognition and image processing in left ventricle segmentation: The last 50 years. *Pattern Analysis & Applications* 3(3), 209–242 (2000)
12. Üzümcü, M., van der Geest, R.J., Sonka, M., Lamb, H.J., Reiber, J.H.C., Lelieveldt, B.P.F.: Multiview active appearance models for simultaneous segmentation of cardiac 2- and 4-chamber long-axis magnetic resonance images. *Investigative Radiology* 40(4), 195–203 (2005)
13. Uzunbas, M.G., Zhang, S., Pohl, K.M., Metaxas, D., Axel, L.: Segmentation of myocardium using deformable regions and graph cuts. In: *International Symposium on Biomedical Imaging: From Nano to Macro*. pp. 254–257 (2012)
14. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. *IEEE Transactions on Signal Processing* 7(3), 359–369 (1998)
15. Yu, Y., Zhang, S., Li, K., Metaxas, D., Axel, L.: Deformable models with sparsity constraints for cardiac motion analysis. *Medical Image Analysis* 18(6), 927–937 (2014)
16. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis* 16(1), 265–277 (2012)

# Anisotropic Tensor Total Variation Regularization For Low Dose Low CT Perfusion Deconvolution

Ruogu Fang<sup>1</sup>, Tsuhan Chen<sup>1</sup>, Pina C. Sanelli<sup>2,3</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Cornell University, Ithaca, NY

<sup>2</sup> Department of Radiology, Weill Cornell Medical College, New York, NY

<sup>3</sup> Department of Public Health, Weill Cornell Medical College, New York, NY

**Abstract.** Tensor total variation (TTV) regularized deconvolution has been proposed for robust low radiation dose CT perfusion. In this paper, we extended TTV algorithm with anisotropic regularization weighting for the temporal and spatial dimension. We evaluated TTV algorithm on synthetic dataset for bolus delay, uniform region variability and contrast preservation, and on clinical dataset for reduced sampling rate with visual and quantitative comparison. The extensive experiments demonstrated promising results of TTV compared to baseline and state-of-art algorithms in low-dose and low sampling rate CTP deconvolution with insensitivity to bolus delay. This work further demonstrates the effectiveness and potential of TTV algorithm’s clinical usage for cerebrovascular diseases with significantly reduced radiation exposure and improved patient safety.

## 1 Introduction

Cerebrovascular disease, or stroke, is the second leading cause of death worldwide after cancer. CT perfusion (CTP) is one of the most widely used imaging modality for disease diagnosis and assessment of treatment response in cerebrovascular diseases. However the radiation exposure in CTP is has caused significant concerns in the community [1].

Numerous efforts have been proposed to reduce the necessary radiation dose to meet the “as low as reasonably achievable” (ALARA) principle, including lowering the radiation dose level, reducing the exposure time, using effective shielding for the patients and increasing the distance between the body and the radiation source.

Recently, tensor total variation (TTV) regularization algorithm has been proposed for robust low-dose CTP deconvolution [2] by reducing the tube current-exposure time product measured in mAs, which varies linearly with the radiation dosage. TTV has shown promising results in correcting the over-estimation of cerebral blood flow (CBF) and under-estimation of mean transit time (MTT) at significantly reduced radiation level (8%) compared to commercially available methods, such as standard singular value decomposition (sSVD) [3], block-circulant singular value decomposition (bSVD) [4], Tikhonov regularization (Tikh) [5], as well as the state-of-art learning-based method sparse perfusion deconvolution (SPD) [6]. However, the robustness of TTV to bolus delay, contrast preservation and low sampling rate has not been fully explored.

In this paper, we use both synthetic simulation and in-vivo clinical data to extensively evaluate TTV algorithm in low-dose, low-sampling rate CTP data with bolus delay. Block-circulant TTV algorithm shows encouraging performance in various tasks,

including contrast preservation, estimation at reduced sampling rate and robustness to tracer arrival time in arterial input function. This work further demonstrates the effectiveness and potential of TTV algorithm’s clinical usage in cerebrovascular diseases with significantly reduced radiation exposure and improved patient safety.

## 2 Anisotropic Tensor Total Variation Regularized Deconvolution

TTV algorithm regularizes the convolution model with tensor total variation term to reduce the oscillation and error in the recovered residue functions.

Let’s denote  $A \in \mathbb{R}^{L \times L}$  as the block-circulant Teoplitz matrix of the zero-padded arterial input function  $c_{art}(t) \in \mathbb{R}^{T \times 1}$ ,  $C \in \mathbb{R}^{L \times N}$  as the zero-padded contrast concentration curves  $c_{voi} \in \mathbb{R}^{T \times 1}$  in the region of interest (ROI) with  $N$  voxels, and  $K \in \mathbb{R}^{L \times N}$  as the flow-scaled residue impulse functions, where  $L \geq 2T$ , and  $T$  is the time length of the measured signal.  $\|\cdot\|_{TV}$  is the total variation norm. To estimate  $K$ , TTV algorithm optimizes

$$K_{ttv} = \arg \min_{K \in \mathbb{R}^{T \times N}} \left( \frac{1}{2} \|AK - C\|_2^2 + \|K\|_{TV} \right) \quad (1)$$

Here we use anisotropic weighting for the tensor total variation term, instead of isotropic weighting in [2]. It is based on the assumption that the piecewise smooth residue functions in CTP should have small total variation, and the smoothness in the temporal and spatial dimensions should be different. The tensor total variation term is defined as

$$\begin{aligned} \|K\|_{TV} = \sum_{t,i,j,k} & \left[ \gamma_t |\tilde{K}_{t+1,i,j,k} - \tilde{K}_{t,i,j,k}| + \gamma_x |\tilde{K}_{t,i+1,j,k} - \tilde{K}_{t,i,j,k}| \right. \\ & \left. + \gamma_y |\tilde{K}_{t,i,j+1,k} - \tilde{K}_{t,i,j,k}| + \gamma_z |\tilde{K}_{t,i,j,k+1} - \tilde{K}_{t,i,j,k}| \right] \end{aligned} \quad (2)$$

where  $\tilde{K} \in \mathbb{R}^{T \times N_1 \times N_2 \times N_3}$  is the 4-D volume obtained by reshaping matrix  $K$  based on the spatial and temporal dimension. Here  $N = N_1 \times N_2 \times N_3$ . The tensor total variation term here uses the forward finite difference operator using  $L_1$  norm. The  $L_1$  norm here imposes sparsity in TV regularization term, which imposes smoothness while preserving the edges. The regularization parameter  $\gamma_i, i = t, x, y, z$  controls the regularization strength for the temporal and spatial dimension, and the larger the  $\gamma_i$ , the more smoothing the TV term imposes on the residue function in  $i^{th}$  dimension.

We propose an algorithm to efficiently solve the problem in Eq. 1 inspired by the framework of [7], as described in Algorithm 1.

## 3 Experiments

### 3.1 Synthetic Evaluation

Because the clinical CTP does not have ground truth perfusion parameter values for comparison, we first use synthetic data to evaluate the proposed algorithm, using the synthetic experiment setup in [4].

---

**Algorithm 1** The framework of TTV algorithm.

---

**Input:** Regularization parameters  $\gamma_i, i = t, x, y, z$

**Output:** Flow-scaled residue functions  $K \in \mathbb{R}^{T \times N_1 \times N_2 \times N_3}$ .

$K^0 = 0$

$t^1 = r^1 = K^0$

**for**  $n = 1, 2, \dots, N$  **do**

(1) Steepest gradient descent

$$K_g = r^n + s^{n+1}(A^T(C - Ar^n))$$

where  $s^{n+1} = \frac{Q^T Q}{(AQ^T)(AQ)}$ ,  $Q \equiv A^T(Ar^n - C)$

(2) Proximal map:

$$K^n = \text{prox}_\gamma(2\|K\|_{TV})(K_g)$$

where  $\text{prox}_\rho(g)(x) := \arg \min_u \left\{ g(u) + \frac{1}{2\rho} \|u - x\|^2 \right\}$

(3) Update  $t, r$

$$t^{n+1} = (1 + \sqrt{1 + 4(t^n)^2})/2$$

$$r^{n+1} = K^n + ((t^n - 1)/t^{n+1})(K^n - K^{n-1})$$

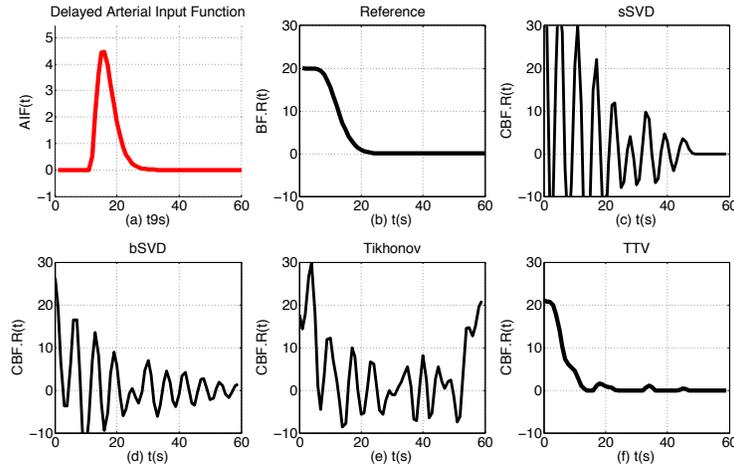
**end for**

---

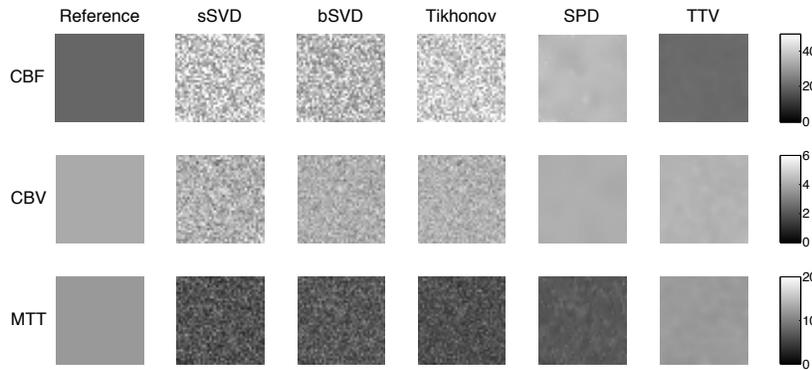
**Bolus Delay In Arterial Input Function:** Circular deconvolution has been used to correct the delay effect using circular representation of  $c_{art}$  and  $c_{voi}$ , but with limited improvement, as shown in Fig. 1, where the arterial input function is delayed by 5 s. bSVD and TTV use the block-circulant version of  $c_{art}$  and  $c_{voi}$ , while sSVD and Tikhonov use standard deconvolution. Though bSVD shows relatively improved performance compared to sSVD and Tikh, the estimated CBF (the maximum value of residue function) is still over-estimated to be around 30 mL/100g/min. On the other hand, TTV is able to correct the bolus delay and estimate accurate residue function.

**Uniform Region Variability:** From the recovered residue function, perfusion parameters CBF and MTT can be estimated. We generate a small region containing  $40 \times 40$  voxels with the same perfusion characteristics, and compute the mean and standard deviation of the perfusion parameters over this region.

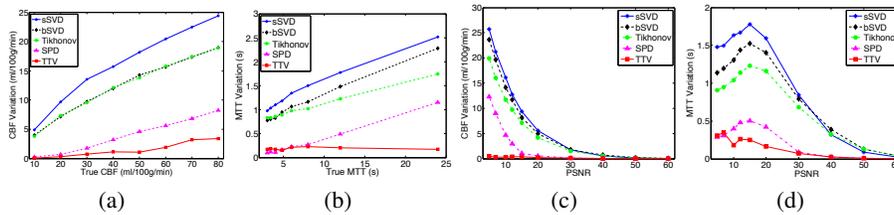
1) The ideal variability of the uniform region should be zero. Fig. 2 shows the estimated perfusion maps of the reference and four methods on the uniform region. While the baseline methods behave poorly in recovering the smooth region, TTV results in uniform perfusion maps for all three parameters at PSNR = 15. SPD reduces the noise level in estimating the three perfusion parameter maps compared to other baseline methods, but the over-estimation in CBF and under-estimation in MTT could not be corrected using SPD. In comparison, TTV not only decreases the noise standard deviation in the



**Fig. 1.** (a) The delayed arterial input function with 5 s delay compared to tracer arrival at the tissue. (b)-(f) The recovered residue functions by baseline methods and TTV. The parameters used for residue function recovery are the simulation is  $CBV = 4 \text{ mL}/100 \text{ g}$ ,  $CBF = 20 \text{ mL}/100 \text{ g}/\text{min}$ ,  $PSNR=25$ .



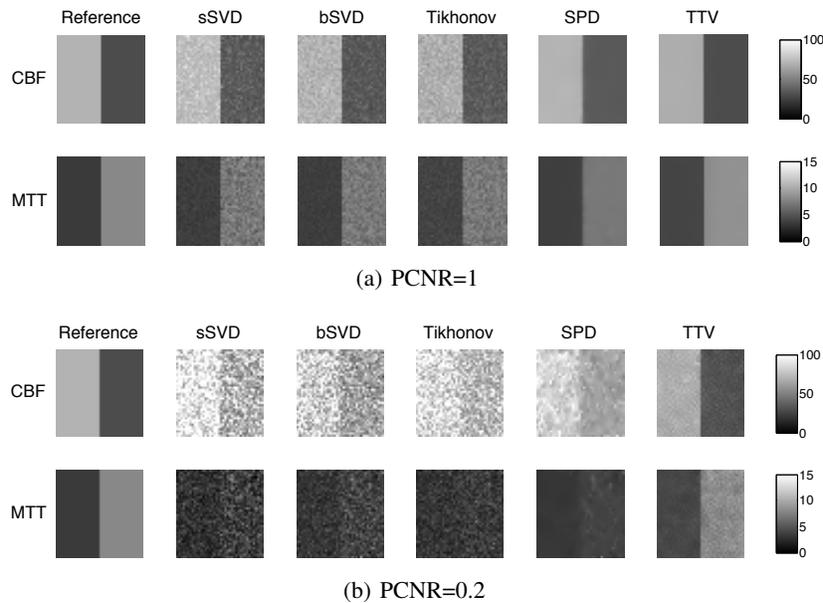
**Fig. 2.** Visual comparison in a uniform regions of perfusion parameter estimation using baseline methods and TTV. The ideal variation is 0. The reference is the ground truth at  $CBV = 4 \text{ mL}/100 \text{ g}$ ,  $CBF = 20 \text{ mL}/100 \text{ g}/\text{min}$ ,  $MTT = 12 \text{ s}$ ,  $PSNR = 15$ .



**Fig. 3.** Comparisons of reducing variations over homogeneous region of (a) CBF at different CBF values with  $PSNR = 15$ . (b) MTT at different true MTT values with  $PSNR = 15$ . (c) CBF at different PSNR values with true  $CBF = 20 \text{ mL}/100 \text{ g}/\text{min}$ . (d) MTT at different PSNR values with true  $MTT = 12 \text{ s}$ .

estimated perfusion maps, but also estimates the accurate quantitative parameters for CBF and MTT.

2) Quantitative comparison is shown in Fig. 3(a)-(b) (where CBF or MTT varies) and Fig. 3 (c)-(d) (where PSNR varies). All figures show that TTV produces lower CBF and MTT variations than the sSVD, bSVD and Tikhonov methods. SPD achieves lower variation than TTV in MTT estimation at different true MTT values in Fig. 3(c), but the mean estimated value of MTT in Fig. 2(b) shows under-estimation of MTT, compared to the ground truth.



**Fig. 4.** Comparisons of CBF and MTT estimated by the different deconvolution algorithms in preserving edges between two adjacent regions at PCNR=1 and 0.2. CBV is not shown because it is uniform in the region. True CBF is 70 and 30 mL/100 g/min on the left and right halves of the region. CBV is uniform in the region at 4 mL/100g. True MTT is 3.43 and 8 s on the left and right halves. Temporal resolution is 1 sec and total duration of 60 sec.

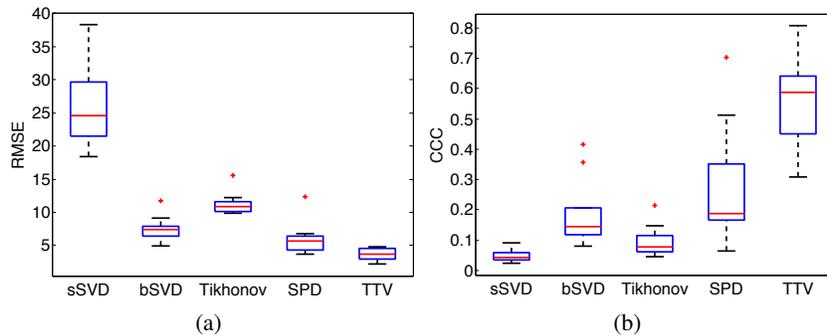
**Contrast preserving:** Contrast is an important indicator of how well two neighboring different regions can be distinguished. The contrast of perfusion parameters between the normal and abnormal tissue computed using the deconvolution algorithm from the noisy data should be comparable to that of the noise-free CTP data. To compare the performance of the baseline methods and TTV in preserving contrast, we generate synthetic CTP data spatially containing two  $40 \times 20$  uniform regions with different perfusion characteristic. Peak contrast-to-noise ratio (PCNR) is defined as  $PCNR = \max |I_1 - I_2| / \sigma$ , where  $I_1$  and  $I_2$  are the perfusion parameter values of then two images to be compared for contrast.

Fig. 4 shows the estimated CBF and MTT by the different algorithms when PCNR=1 and 0.2. The corresponding  $\sigma=40$  and 200. While baseline methods sSVD, bSVD and Tikhonov perform poorly at both PCNR levels, SPD and TTV yield improved CBF and MTT maps with regard to the reference. When the PCNR = 1 and the noise level is moderate, both SPD and TTV are capable of removing the noise and preserving the contrast. However the spatial resolution at the boundary of two regions is smoothed by SPD, compared to the clear-cut boundary using TTV. When the PCNR is as low as 0.2, the contrast to noise ratio is extremely low. sSVD, bSVD and Tikhonov generate severely biased perfusion parameters. SPD reduces the noise level to certain extent, but is unable to correct the estimation bias in CBF and MTT. TTV performs favorably compared to all baseline methods in preserving the edges between two adjacent regions in CBF and MTT, as well as accurate estimation of perfusion parameters.

### 3.2 Clinical Evaluations

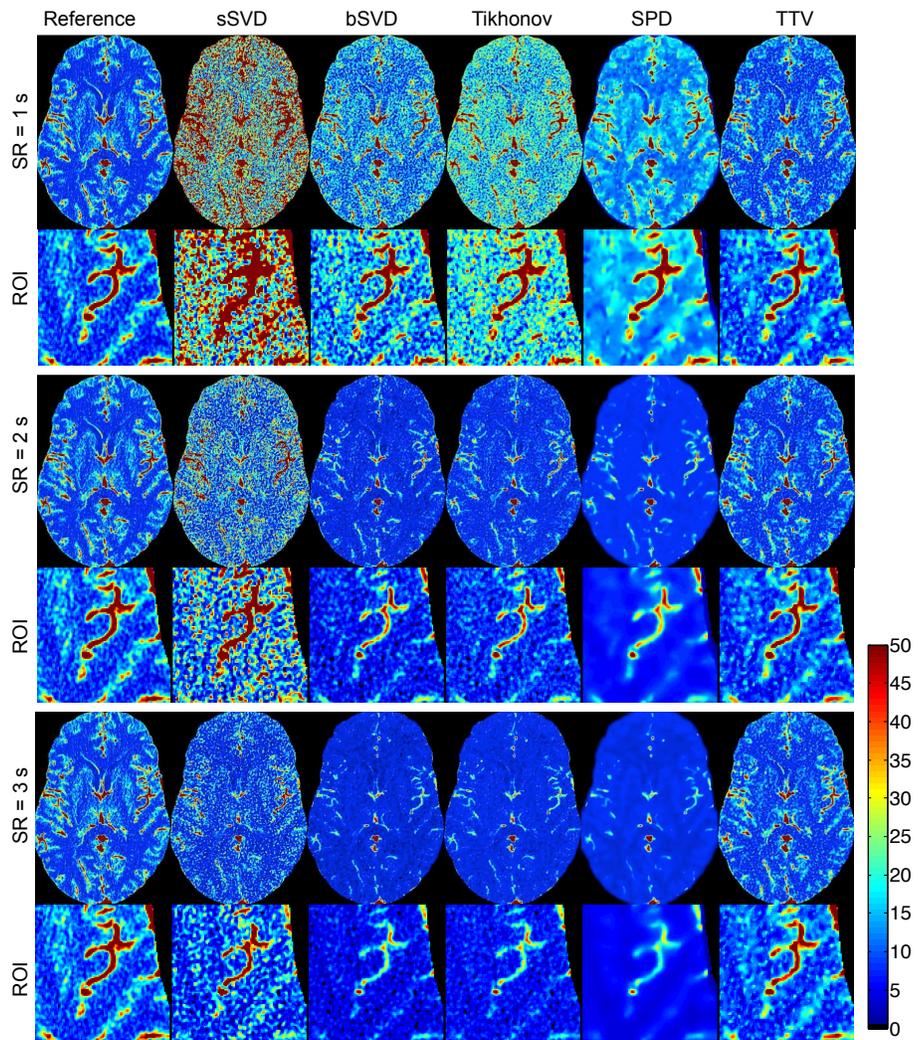
Retrospective review of consecutive CTP exams performed on aneurysmal subarachnoid hemorrhage patients enrolled in an IRB-approved and HIPAA-compliant clinical trial from August 2007-Dec 2013 was used. Ten consecutive patients (9 women, 1 men) admitted to the Weill Cornell Medical College, with mean age (range) of 54 (35-83) years were included. 5 patients had brain deficits shown in the CTP images and the other 5 patients had normal brain images.

Because repetitive scanning of the same patient under different radiation levels is unethical, low-dose Perfusion maps are simulated from the high-dose 190 mAs by adding correlated statistical noise [8] with standard deviation of  $\sigma_a = 25.54$ , which yields PSNR=40. The maps calculated using bSVD from the 190 mAs high-dose CTP data is regarded as the “gold standard” or reference images in clinical experiments.



**Fig. 5.** Comparisons of RMSE and Lin's CCC among the four methods. TTV results in significant ( $P < 0.001$ ) lower RMSE and higher Lins CCC compared with all the baseline methods.

**Visual Comparison:** At normal sampling rate of 1 s and reduced temporal sampling rate of 2 s and 3 s, the errors of CBF estimation in the four baseline algorithms increase, while TTV maintains accurate estimation for CBF value at all sampling rates.

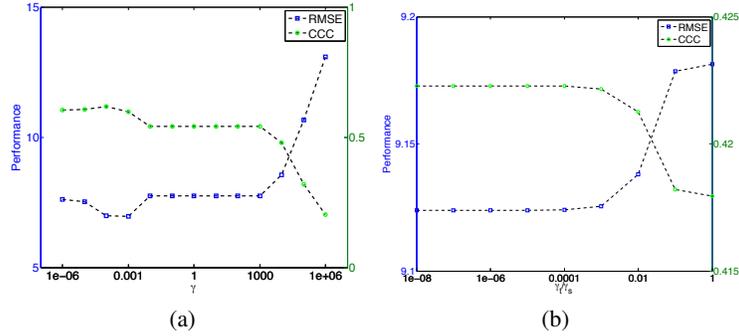


**Fig. 6.** The CBF maps with roomed ROI regions of a patient computed using different deconvolution methods at sampling rate (SR) of 1 s, 2 s and 3 s with 15 mAs tube current. At normal sampling rate 1 s, baseline methods over-estimate CBF values. At reduced sampling frequency 2 s, sSVD still over-estimate while bSVD, Tikhonov and SPD under-estimate CBF values. At reduced sampling rate of 3 s, baseline algorithms under-estimate CBF values. At all sampling rates, TTV accurately estimate the CBF values. (Color image)

**Quantitative Comparison:** Fig. 5 shows significant improvement in image fidelity between the low-dose CBF maps and the high-dose CBF maps by using the TTV algorithm compared to the baseline methods. On average, the root-mean-square-error (RMSE) decreases by 40%, Lin’s CCC increases by 89% from the best performance by using the baseline methods. The quantitative values are computed with the vascular pixel elimination to exclude the influence of high blood flow values in the blood vessels.

### 3.3 Parameters

In the TTV algorithm, there is only a single type of tunable parameter: the TV regularization weight. If the spatial and temporal regularization are treated equally, only one weighting parameter  $\gamma$  needs to be determined. Fig. 7(a) show the RMSE at different  $\gamma$  values. When  $\gamma < 10^3$ , RMSE does not change much. The optimal  $\gamma$  is between  $10^{-4}$  to  $10^{-3}$ .



**Fig. 7.** Performance in terms of root-mean-square-error (RMSE) for different parameters (a)  $\gamma$  and (b) ratio  $\gamma_t/\gamma_s$ .

Since the temporal and the spatial dimensions of the residue impulse functions have different scaling, regularization parameters for  $t$  and  $x, y, z$  should be different too. We set the spatial  $\gamma_s = \gamma_{x,y,z} = 10^{-4}$  since the spatial dimensions have similar scaling, and tune the ratio between the temporal weight  $\gamma_t$  and spatial weight  $\gamma_s$ . Fig. 7(b) shows that when the ratio  $\gamma_t/\gamma_s < 10^{-4}$ , the performance is stable. Compared to isotropic TTV, anisotropic TTV with the ratio of spatial and temporal regularization weight set to  $10^{-4}$  output improved result from Fig. 7(b). Thus we set  $\gamma_t = 10^{-8}$  and  $\gamma_s = 10^{-4}$  for all experiments.

## 4 Conclusion

In this paper, we extended the tensor total variation regularized (TTV) deconvolution algorithm with anisotropic regularization weighting for the temporal and spatial dimensions. We evaluated TTV algorithm for bolus delay, uniform region variability and

contrast preservation on synthetic dataset, as well as for reduced sampling rate with visual and quantitative comparison on clinical dataset. The extensive experiments demonstrated the superiority of TTV compared to baseline and state-of-art algorithms in low-dose and low-sampling-rate CTP deconvolution with insensitivity to tracer arrival time. Future research include evaluation of TTV algorithm on larger-scale clinical datasets with acute stroke and other cerebrovascular diseases.

## References

- [1] Wintermark, M., Lev, M.: FDA investigates the safety of brain perfusion ct. *American Journal of Neuroradiology* **31**(1) (2010) 2–3
- [2] Fang, R., Sanelli, P.C., Zhang, S., Chen, T.a.: Tensor total-variation regularized deconvolution for efficient low-dose ct perfusion. In Barillot, Golland, H.H., ed.: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2014. Lecture Notes in Computer Science*. Springer Berlin Heidelberg (2014)
- [3] Østergaard, L., Weisskoff, R.M., Chesler, D.A., Gyldensted, C., Rosen, B.R.: High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. *Magnetic Resonance in Medicine* **36**(5) (1996) 715–725
- [4] Wu, O., Østergaard, L., Weisskoff, R.M., Benner, T., Rosen, B.R., Sorensen, A.G.: Tracer arrival timing-insensitive technique for estimating flow in MR perfusion-weighted imaging using singular value decomposition with a block-circulant deconvolution matrix. *Magnetic Resonance in Medicine* **50**(1) (2003) 164–174
- [5] Fieselmann, A., Kowarschik, M., Ganguly, A., Hornegger, J., Fahrig, R.: Deconvolution-based CT and MR brain perfusion measurement: theoretical model revisited and practical implementation details. *Journal of Biomedical Imaging* **2011** (2011) 14
- [6] Fang, R., Chen, T., Sanelli, P.C.: Towards robust deconvolution of low-dose perfusion ct: Sparse perfusion deconvolution using online dictionary learning. *Medical image analysis* (2013)
- [7] Huang, J., Zhang, S., Metaxas, D.: Efficient mr image reconstruction for compressed mr imaging. *Medical Image Analysis* **15**(5) (2011) 670–679
- [8] Britten, A., Crotty, M., Kiremidjian, H., Grundy, A., Adam, E.: The addition of computer simulated noise to investigate radiation dose and image quality in images with spatial correlation of statistical noise: an example application to x-ray ct of the brain. *British journal of radiology* **77**(916) (2004) 323–328

# Region segmentation for sparse decompositions: better brain parcellations from rest fMRI

Alexandre ABRAHAM<sup>1,2</sup>, Elvis DOHMATOB<sup>1,2</sup>, Bertrand THIRION<sup>1,2</sup>, Dimitris SAMARAS<sup>3,4</sup>, and Gael VAROQUAUX<sup>1,2</sup>

<sup>1</sup> Parietal Team, INRIA Saclay-Île-de-France, Saclay, France  
`alexandre.abraham@inria.fr`

<sup>2</sup> CEA, DSV, I<sup>2</sup>BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France

<sup>3</sup> Stony Brook University, NY 11794, USA

<sup>4</sup> Ecole Centrale, 92290 Châtenay Malabry, France

**Abstract** Functional Magnetic Resonance Images acquired during resting-state provide information about the functional organization of the brain through measuring correlations between brain areas. Independent components analysis is the reference approach to estimate spatial components from weakly structured data such as brain signal time courses; each of these components may be referred to as a *brain network* and the whole set of components can be conceptualized as a *brain functional atlas*. Recently, new methods using a sparsity prior have emerged to deal with low signal-to-noise ratio data. However, even when using sophisticated priors, the results may not be very sparse and most often do not separate the spatial components into brain regions. This work presents post-processing techniques that automatically sparsify brain maps and separate regions properly using geometric operations, and compares these techniques according to faithfulness to data and stability metrics. In particular, among threshold-based approaches, hysteresis thresholding and random walker segmentation, the latter improves significantly the stability of both dense and sparse models.

**Keywords:** region extraction, brain networks, clustering, resting state fMRI

## 1 Introduction

Functional connectivity between brain networks observed during resting state functional Magnetic Resonance Imaging (R-fMRI) is a promising source of diagnostic biomarkers, as it can be measured on impaired subjects such as stroke patients [9]. However, its quantification highly depends on the choice of the brain atlas. A brain atlas should be *i)* consistent with neuroscientific knowledge *ii)* as faithful as possible to the original data and *iii)* robust to inter-subject variability.

Publicly available atlases (such as structural [8] or functional [13] atlases) went through a quality assessment process and are reliable. To extract a data driven atlas from R-fMRI, Independent Component Analysis (ICA) remains the

reference method. In particular, it yields some additional flexibility to adapt the number of regions to the amount of information available. Networks extracted by ICA are full-brain and require a post-processing step to extract the salient features, i.e., brain regions, which is often done manually [5] (see figure 3). To avoid post-processing and directly extract regions, more sophisticated approaches rely on sparse, spatially-structured priors [1]. Indeed, maps of functional networks or regions display a small number of non-zero voxels, and thus are well characterized through a sparsity criterion, even in the case of ICA [11,3]. However, sophisticated priors such as structured sparsity come with computational cost and still fail to split some networks into separate regions. Altogether, region extraction is unavoidable to go from brain image decompositions to Regions-of-Interest-based analysis [6].

A simple approach to obtain sharper maps is to use hard thresholding, which is a good sparse, albeit non convex, recovery method [2]. We improve upon it by introducing richer post-processing strategies with spatial models, to avoid small spurious regions and isolate each salient feature in a dedicated region. Based on purely geometric properties, these take advantage of the spatially-structured and sparsity-inducing penalties of recent dictionary learning methods to isolate regions. These can also be used in the framework of computationally cheaper ICA algorithms. In addition to these automatic methods that extract brain atlases, we propose two metrics to quantitatively compare them and determine the best one. The paper is organized as follows. In section 2, we introduce the region extraction methods. Section 3 presents the experiments run to compare them. Finally, results are presented in section 4.

## 2 Region extraction methods

Extracting regions to outline objects is a well-known problem in computer vision. For the particular problem of extracting regions of interest (ROIs) out of brain maps, we want a method that *i*) handles 3D images *ii*) processes one image while taking into account the remainder of the atlas (e.g., region extraction for a given image may be different depending on the number of other regions) and *iii*) isolates each salient feature from a smooth image in an individual ROI without strong edges or structure (see figure 1). Here, we assume that a given set of *brain maps* has been obtained by a multivariate decomposition technique.

Most of the following methods allow overlapping components after region extraction. In fact, multivariate decomposition techniques most often decompose the signal of one voxel as a linear mixture of several signal components. In practice, these overlapping regions are small and located in areas of low confidence. Voxels that belong to no component are left unlabeled.

### 2.1 Foreground extraction

Let  $\mathcal{I} = \{I_1, \dots, I_k\}$  be a set of brain maps (3D images), or atlas.  $I(p)$  designates the value for image  $I$  at point  $p$ . We define by  $\mathcal{F}(I)$  the set of foreground points

of image  $I$ , i.e., the points that are eligible for region extraction. We propose two strategies to extract the foreground.

*Hard assignment.* Hard assignment transforms a set of maps into a brain segmentation with no overlap between regions. That means that each voxel will be represented by a unique brain map from the atlas. This map is the one that has the highest value for this voxel. The result is a segmentation from which we can extract connected components.

$$\mathcal{F}_{hard}(I_i) = \{p \in I_i \mid \operatorname{argmax}_{j \in [1, k]} I_j(p) = i\}$$

*Automatic thresholding.* Thresholding is the common approach used to extract ROIs from ICA. However, the threshold is usually set manually and is different for each map. In order to propose an automatic threshold choice, we consider that on average, an atlas assigns each voxel to one region. For this purpose, we set the threshold  $t^k(\mathcal{I})$  so that the number of nonzero voxels corresponds to the number of voxels in the brain:

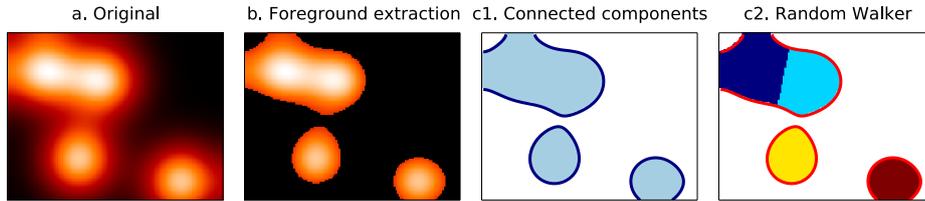
$$\mathcal{F}_{automatic}(I_a) = \{p \in I_a, I(p) > t^k(\mathcal{I})\}$$

## 2.2 Component extraction

*Connected components.* Let  $\mathcal{N}(p)$  be the set of neighbors of point  $p$ . Two points  $p_1$  and  $p_n$  are  $\mathcal{N}$ -connected if  $p_n$  can be reached from  $p_1$  by following a path of consecutive neighboring points:

$$(p_1, p_n) \text{ } \mathcal{N}\text{-connected} \equiv \exists(p_2, \dots, p_{n-1}) : p_{i+1} \in \mathcal{N}(p_i), \forall i \in [1, n-1]$$

We define a connected component as a maximum set of foreground points that are  $\mathcal{N}$ -connected. The set of all  $\mathcal{N}$ -connected components for a given image  $I$  (see figure 1.c1) is written  $ccs(\mathcal{N}, I)$ . Extraction of connected components can be done after hard assignment or automatic thresholding to obtain ROIs (figures 2 and 3). In the following methods, we consider the points extracted with automatic thresholding as foreground ( $\mathcal{F} = \mathcal{F}_{automatic}$ ) and use more sophisticated priors to extract ROI.



**Figure 1.** Example of region extraction. Foreground pixels (b) are extracted from the original image (a). Regions are extracted using connected component extraction (c1) or random walker (c2).

*Hysteresis thresholding.* Hysteresis thresholding is a two-threshold method where all voxels with value higher than a given threshold  $t_{high}$  are used as seeds for the regions. Neighboring voxels with values between the high threshold  $t_{high}$  and the low threshold  $t_{low}$  are added to these seed regions. In our setting, the high threshold can be seen as a minimal activation value over the regions in order to sort out regions of marginal importance. Each brain map has its own optimal value but, in practice, cross validation has shown that keeping the 10% highest foreground voxels as seeds gives the best results. The automatic thresholding strategy described above is used to set the low threshold  $t_{low}$ .

Conserving connected components that have their maximum value over  $t_{high}$  is done at component extraction:

$$ccs_{hysteresis}(\mathcal{N}, I) = \{c \in ccs(\mathcal{N}, I) \mid \max(c) \geq t_{high}\}$$

*Random Walker.* Random Walker is a seed-based segmentation algorithm similar to watershed. It calculates, for each point  $p$ , the probabilities to end up in each of the seeds by doing a random walk across the image starting from  $p$ . The original version of the algorithm [4] was of probabilistic nature, whereby the probability to jump to a neighboring point is driven by the gradient magnitude between them. After convergence the point is attached to the seed with the highest probability.

Random Walker can also be seen as a diffusion process. It is equivalent to hysteresis thresholding where regions that have grown enough to share a boundary are not allowed to be merged. The probabilities to reach each of the seeds can be computed using the laplacian matrix of the graph associated with the map. Due to space limitations we refer the reader to [4] for the complete description of the algorithm. We suppose  $seed(p)$  returns the seed associated with point  $p$ . We refine our neighborhood relationship by considering two points as neighbours only if they are associated to the same seed:

$$\mathcal{N}_{rw}(p) = \{q \in \mathcal{N}, seed(p) = seed(q)\}; ccs_{rw}(I) = ccs(\mathcal{N}_{rw}, I)$$

Note that, in our setting, a high value in the map means a high confidence. So, instead of using the finite difference gradient, we consider the max of the image minus the lowest voxel. Therefore, diffusion is facilitated in areas of high confidence and more difficult elsewhere. We take the local maxima of the smoothed image as seeds for the algorithm.

### 3 Experiments

Experiments are made on a subset of the publicly available Autism Brain Imaging Database Exchange<sup>5</sup> dataset. Preprocessing is done with SPM and includes slice timing, realignment, coregistration to the MNI template and normalization. We select 101 subjects suffering from autism spectrum disorders and 93 typical

<sup>5</sup> [http://fcon\\_1000.projects.nitrc.org/indi/abide/](http://fcon_1000.projects.nitrc.org/indi/abide/)

controls from 4 sites and compute brain atlases on 10 cross-validation iterations by taking a random half of the dataset as the train set. We extract regions from these atlases and quantify their performance on the other half of the dataset with two metrics.

We investigate two decomposition methods to extract brain maps from resting-state fMRI: **ICA** –independent component analysis– that yields full brain continuous maps, and **MSDL** –multi-subject dictionary learning–, [1], that directly imposes sparsity and structure on the maps thanks to the joint effect of  $\ell_1$  norm and total variation minimization. Our goal is to compare the effects of region extraction on sparse and non-sparse sets of maps.

To quantify the usefulness of a set of regions extracted automatically, we consider metrics that characterize two different aspects of the segmentation: the ability to explain newly observed data and the reproducibility of the information extracted, as in the NPAIRS framework [7]. We use Explained Variance (EV) to measure how faithful the extracted regions are to unseen data. Stability with regards to inter-subject variability is measured using Normalized Mutual Information (NMI) over models learned on disjoint subsets of subjects.

Following [10], we extract  $k = 42$  maps. For the metrics to be comparable, we need to apply them on models of similar complexity, i.e. with the same number of regions. For this purpose, we assume that there must be on average 2 symmetric regions per map (some of them may have more, and some of them may have only one inter-hemispheric region). We therefore aim at extracting  $2k$  regions, and take the largest connected components after region extraction. In the end, some maps may not contribute to the final atlas.

### 3.1 Data faithfulness – Explained variance

The explained variance measures how much a model accounts for the variance of the original data. The more variance is explained, the better the model explains the original data. Linear decomposition models original data  $y_{orig}$  by decomposing them into two matrices. In our case, these matrices are brain networks  $\mathcal{I}$  and their associated time series  $y_{model}$ . Time series of regions are measured using least square fitting instead of simple averaging to handle mixed features in region overlaps. Explained variance of these series is then computed over the original ones.

$$y_{orig} = \mathcal{I} \times y_{model} + y_{\varepsilon} ; \text{EV}(y_{model}) = 1 - \frac{\text{Var}(y_{\varepsilon})}{\text{Var}(y_{orig})} = \frac{\text{Var}(y_{orig}) - \text{Var}(y_{model})}{\text{Var}(y_{orig})}$$

### 3.2 Stability – Normalized Mutual Information

To assess model stability, we rely on Normalized Mutual Information, a standard clustering similarity score, applied on hard assignments [12]: given two hard assignments  $U$  and  $V$  with marginal entropy  $H(U)$  and  $H(V)$  respectively,

$$\text{NMI}(U, V) = \frac{H(U) + H(V) - H(U, V)}{\sqrt{H(U) * H(V)}} ; H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i)$$

## 4 Results

Figure 2 presents region extraction results using each method on the same map. In all figures, the threshold applied during region extraction is shown in a given slice to help understanding. Results for each metric are displayed on the right. We vary parameters for each model (smoothing for ICA, 3 parameters of MSDL) and, for each region extraction method, display the best 10% results across parametrization. Figure 3 shows 2 networks out of 42 extracted.

*Region shape* The regions extracted by hard assignment (figure 2.a) present salient angles and their limits do not follow a contour line of the original map. The straight lines are the results of two maps in competition with each other. The 1D cut shows that the threshold applied when using hard thresholding is not uniform on the whole image. The other methods look smoother and follow actual contour lines of the original map. On this particular example, automatic thresholding (figure 2.b) extracts 2 regions: a large one on the left and a very small one on the right. This is one of the drawbacks of thresholding: small regions can appear when their highest value is right above the threshold. Thanks to its high threshold, hysteresis thresholding (figure 2.c) gets rid of the spurious regions but still fails to separate the large region on the left. Random Walker (figure 2.d) manages to split the large region into two subregions.

Similarly, in figure 3 we can see that Random Walker manages to split the default mode network into 3 components, where other methods extract two.

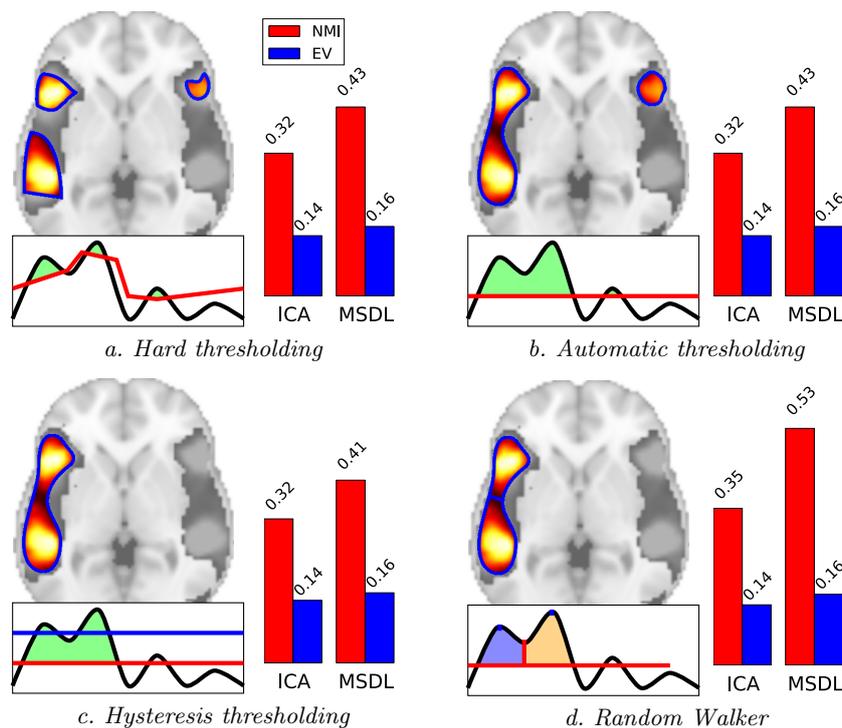
*Stability.* Random Walker dominates the stability metric. It uses local maxima to get regions seeds, and will thus split regions even if they are *connected* after thresholding. Its performance is statistically significant for both dense and sparse atlases and any parametrization. The stability improvement is larger for sparse than for dense maps. This could be due to the inability of random walker to compensate for the original instabilities of the models.

*Data fidelity.* The explained variance scores on best performing models, shown in figure 2, are similar for all methods. In poorly performing models, we observe that automatic and hysteresis thresholdings are slightly above random walker (about 2%), exhibiting the same trade-off as in [1].

## 5 Discussion and conclusion

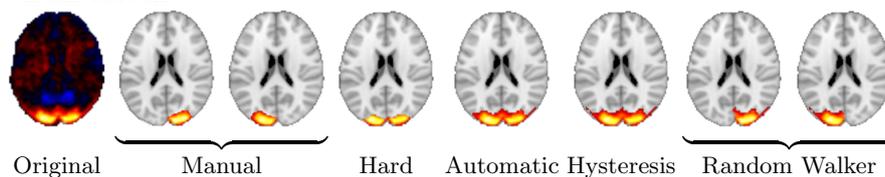
Functional atlases extracted using ICA or sparse decomposition methods are composed of continuous maps and sometimes fail to separate symmetric functional regions.

Starting from hard thresholding [2], we introduce richer strategies integrating spatial models, to avoid small spurious regions and isolate each salient feature in a dedicated region. Indeed, the notion of regions is hard to express with convex penalties. Relaxations such as total-variation used in [1] only captures it

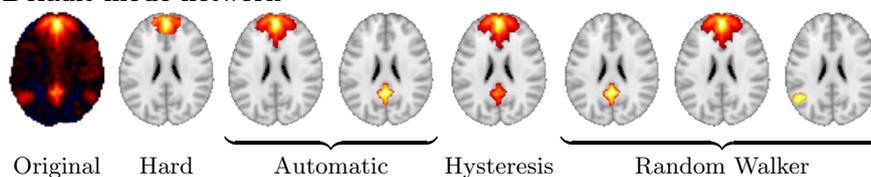


**Figure 2.** Comparison of region extraction methods (after selection of  $2k$  regions). Brain maps obtained with MSDL are located on the left. The activated regions are symbolically represented below in a height map. The bars on the right of each image represent the Normalized Mutual Information and Explained variance obtained on dense maps (ICA) and sparse maps (MSDL). Random walker is the most stable method.

#### Visual cortex



#### Default mode network



**Figure 3.** Region extraction from ICA maps using different approaches. For each network, Random Walker is better at extracting ROIs

partially, while a non-convex segmentation step easily enforces regions. We find that a Random-Walker based strategy brings substantial increase in stability of the regions extracted, while keeping very good explanatory power on unseen data. Finer results and interpretation may arise by using more adapted metrics, for example a version of DICE that can deal with overlapping fuzzy regions. This point is under investigation.

**Acknowledgments** We acknowledge funding from the NiConnect project and NIDA R21 DA034954, SUBSample project from the DIGITEO Institute, France.

## References

1. Abraham, A., Dohmatob, E., Thirion, B., Samaras, D., Varoquaux, G.: Extracting brain regions from rest fMRI with total-variation constrained dictionary learning. In: MICCAI, p. 607 (2013)
2. Blumensath, T., Davies, M.E.: Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis* 27, 265 (2009)
3. 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)
4. Grady, L.: Random walks for image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28(11), 1768–1783 (2006)
5. Kiviniemi, V., Starck, T., Remes, J., et al.: Functional segmentation of the brain cortex using high model order group PICA. *Hum Brain Map* 30, 3865 (2009)
6. Nieto-Castanon, A., Ghosh, S.S., Tourville, J.A., Guenther, F.H.: Region of interest based analysis of functional imaging data. *Neuroimage* 19, 1303 (2003)
7. Strother, S., Anderson, J., Hansen, L., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S., Rottenberg, D.: The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage* 15, 747 (2002)
8. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273 (2002)
9. 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: MICCAI, pp. 200–208 (2010)
10. Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B.: Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In: *Inf Proc Med Imag.* pp. 562–573 (2011)
11. Varoquaux, G., Keller, M., Poline, J., Ciuciu, P., Thirion, B.: ICA-based sparse features recovery from fMRI datasets. In: ISBI. p. 1177 (2010)
12. Vinh, N., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research* 11, 2837–2854 (2010)
13. Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysio* 106, 1125 (2011)

# Improved Sparse Shape Composition Model for Multi-shape Prior

Bing Wang<sup>1</sup>, Chonghao Fan<sup>2</sup>, Hongzhi Xie<sup>3</sup>, Lixu Gu<sup>2,4</sup>

<sup>1</sup> College of Mathematics and Computer Science, Hebei University, China

<sup>2</sup> Multi-disciplinary Research Center, Hebei University, China

<sup>3</sup> Department of Cardiovascular, Peking Union Medical College Hospital, China

<sup>4</sup> School of Biomedical Engineering, Shanghai Jiao Tong University, China

**Abstract.** Shape prior modeling is a challenging and crucial component in various image segmentation applications. Most existing methods aim at dealing with single object's shape variation, which are not directly applicable for multi-shape prior modeling. In this paper, we present an extension of recently proposed Sparse Shape Composition model (SSC) for multi-shape prior modeling. In this extension, multiple shapes of one patient are regarded as a group. A sparse linear composition of training groups is computed iteratively to infer/refine the input group. Thus, not only the a-priori information of each shape but also the a-priori codependency information among different shapes is implicitly incorporated on-the-fly. To validate the efficacy of our method, a 2D left ventricular endocardium and epicardium localization experiment was conducted. The localization result demonstrates that the utilization of our method can achieve more accurate and stable localization compared with SSC.

## 1 Introduction

In various image segmentation applications, especially in the field of medical image segmentation, methods solely relying on image appearance cues usually tend to achieve unsatisfactory result. The fact that objects usually contain strong shape priors, gives rise to various shape model based segmentation methods. Leventon et al. [1] defined a probability distribution over the variances of training shapes, and utilized it to restrict the flow of the geodesic active contour. In [2], Cremers et al. incorporated statistical shape knowledge in the evolution process of a Mumford-Shah based segmentation [3]. Ali et al. [4] integrated prior shape constraints into a graph cuts framework for kidney segmentation. Such segmentation approaches have been proven to be one of the most successful methods in practice, and outperform the conventional methods in both robustness and accuracy owing to the integration of a-priori information.

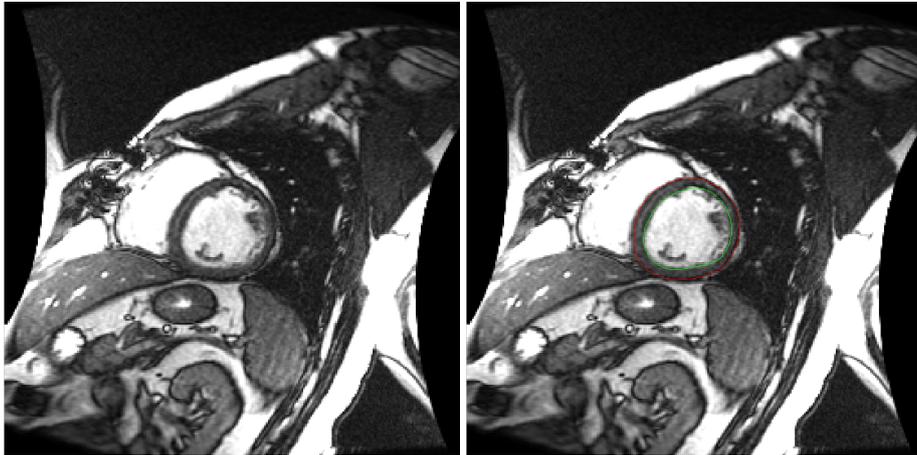
Shape prior modeling plays a significant role in these methods, and is very crucial for the final accuracy and robustness. A straightforward approach is to learn from a number of training shapes by statistical means, leading to statistical shape models (SSMs) [5]. Active Shape models [6] and Active Appearance models [7] proposed by

Cootes et al. in 1995 and 2001, probably are two of the most popular methods in this area. Another widely-used method is level set shape prior model [8]. Subsequently, plenty of adaptations of these algorithms were proposed.

Recently, sparsity theory was introduced into shape prior modeling by [9, 10]. In their model, a sparse composition of training shapes is computed adaptively to infer/refine an input shape. Thus, it was named as Sparse Shape Composition model. With such a setting, it is capable of modeling complex shape variations, and preserving local details very well. Furthermore, when modeling a sparse error vector, it becomes quite robust to sparse non-Gaussian errors.

However, most of existing models are focusing on single shape prior modeling, and not directly applicable to deal with multiple shapes due to the lack of prior co-dependency information utilization. Such co-dependency among different shapes is of great value for various medical image analysis tasks. For instance, the implicit spatial relationship between endocardium and epicardium of left ventricle can be utilized as a supplementary to their shape priors for accurate localization or segmentation purpose. Endocardium and epicardium of left ventricle are very important for quantitative analysis of global and regional cardiac function, such as ejection fraction (EF), left ventricle myocardium mass (MM), and stroke volume (SV) [11]. In Fig.1 a 2D instance from cardiac cine-MR short axis images is shown.

Inspired by [9], we present an extension of SSC which aims at dealing with multi-shape prior modeling. In our method, multiple interested shapes are regarded as a group, and modeled together by a sparse linear combination of training groups. With such a mechanism, the a-priori spatial constraint among different shapes is also implicitly applied. It has the same advantages of SSC, due to the utilization of the same basic idea and optimization framework proposed in SSC.



**Fig. 1.** Left: 2D cardiac cine-MR short axis image. Right: Manually delineated endocardial contour (drawn as green) and epicardial contour (drawn as red) of left ventricle.

## 2 Improved Sparse Shape Composition model

Our model aims at modeling any number of complicated shapes simultaneously with a pre-defined training repository. Shapes may refer to shapes of different objects (e.g., shapes of femur bone, femur cartilage, tibia bone and tibia cartilage of one patient) or shapes of a time-varying object (e.g., shapes of heart), and regarded as a group in our method. The details of how we advanced SSC to our method will be presented in this section.

### 2.1 From shape representation to group representation

Following SSC, explicit parametric shape representation<sup>1</sup> is employed in this method. Specifically, shape instance is represented by a column vector concatenated by coordinates of all its vertices. For instance, column vector *shape* of a 3D mesh which contains 100 vertices is concatenated as Eq. (1).

Let  $m$  represents the number of shapes required to be modeled. Vertex number and column vector of the  $i$ th shape are notated by  $k_i$  and  $shape^i \in \mathbb{R}^{k_i \times d}$  respectively, where  $d$  stands for the dimension of shapes. Then, a column vector  $g$  which represents the group consisted of these  $m$  shapes can be constructed by concatenating  $shape^i$  for  $i = 1, 2, \dots, m$ , as shown in Eq. (2).

$$shape \triangleq [x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2 \ \dots \ x_{100} \ y_{100} \ z_{100}]^T \quad (1)$$

$$g \triangleq [shape^{1T} \ shape^{2T} \ \dots \ shape^{iT}]^T \in \mathbb{R}^{\sum_i k_i \times d} \quad (2)$$

### 2.2 Matrix of training repository

Assume there are  $n$  group samples with manual delineation in the training repository, which are quite sufficient to model variations of shapes and the spatial relationship among them. A matrix which represents the training repository can be constructed based on these samples. An illustration can be found in Fig. 2. It should be noticed that there should be a consistency among columns through this matrix. Specifically, vertex numbers of  $m$  shapes should keep consistent, and shape vertices should be one-to-one corresponding through these groups. Two methods are introduced to acquire consistency of shapes in [9]. Both of them can be extended to groups quite intuitively. We assume this consistency is already achieved here.

After the conversion of all these groups into column vectors, apply pre-alignment to eliminate the position and orientation difference and transform them into a standard coordinate system. Pre-alignment is a two-step procedure based on the generalized Procrustes analysis [12]: first, select a group vector as reference and align others to it, take the transformed vectors and the reference vector as initial aligned groups; second, in order to remove the bias caused by the selection of reference, compute the mean vector of these initial aligned groups, take it as the new reference and align

---

<sup>1</sup> 2D and 3D shapes are represented by curves or meshes composed of a number of vertices.

others to it to get the final aligned groups. These final aligned groups vectors are notated by  $\tilde{g}_j$  for  $j = 1, 2, \dots, n$ . Shape vectors of  $\tilde{g}_j$  are notated as  $\tilde{s}_j^i$  for  $i = 1, 2, \dots, m$ .

Finally, aligned group vectors are assembled together parallelly to concrete the matrix of training repository which is denoted as  $D = [\tilde{g}_1 \tilde{g}_2 \dots \tilde{g}_n]$ .

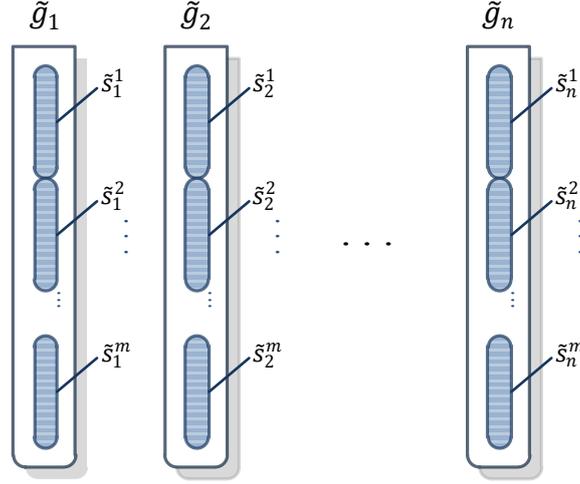


Fig. 2. Diagram illustrating the matrix of training repository.

### 2.3 Problem formulation and optimization framework

The basic idea of SSC can be intuitively extended to groups: for any input group  $y_g = [y_1^T y_2^T \dots y_m^T]^T$ , an optimal sparse linear combination of existing training groups can be found to approximately represent it. The weights or coefficients are denoted as  $x \in R^n$ , and the optimal value of it is denoted as  $x_{opt} \in R^n$ .  $x_{opt}$  is found by solving Eq. 3 utilizing the optimization method proposed in [9].

$$\arg \min_{x,e,\beta} \|T(y_g, \beta) - Dx - e\|_2^2, \text{ s. t. } \|x\|_0 \leq k_1, \|e\|_0 \leq k_2 \quad (3)$$

### 2.4 Difference from SSC

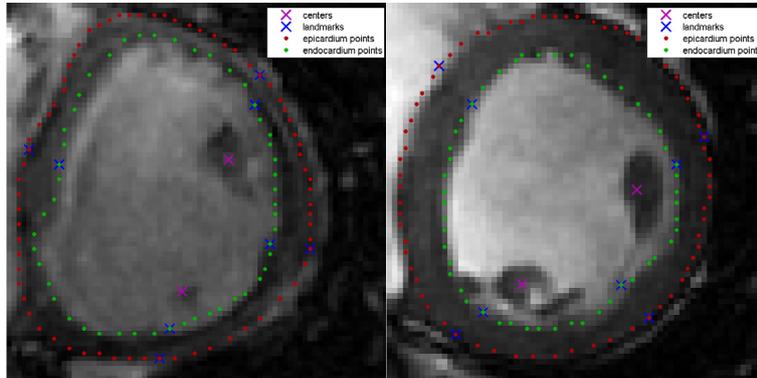
When modeling  $m$  shapes from a group, SSC required to be conducted for  $m$  times. Each time an optimal sparse weight vector is computed for a single input shape. Compared with the unique optimal weight vector  $x_{opt}$  calculated based on our method, these  $m$  vectors tend to differ from each other in practice. Thus, group constraint or co-dependency among different shapes is not utilized in the modeling process.

### 3 Endocardium and epicardium localization

Following the organ localization framework proposed in [9], we conducted an experiment to verify our method, i.e., left ventricular endocardium and epicardium localization from 2D cardiac cine-MR images [12], and compared it with the original SSC. In this experiment, endocardium and epicardium are regarded as a group and localized simultaneously; training repository is consisted of 15 groups from different patients; 91 images from 28 patients are tested.

#### 3.1 Details of localization experiment

This experiment is focusing on the 2D images with papillary muscles, for the convenience of achieving consistency of group vectors. These images are capable of defining 8 landmarks through centers of the two largest papillary muscles. Two intersection points of the line which passes through these centers and the manual endocardial contour, are considered as endocardium landmarks; then, the midperpendicular of these landmarks can be found, which intersect with endocardium at the last two endocardium landmarks. The epicardium landmarks are defined in the same method. After the determination of landmarks, a certain number of vertices are interpolated between two neighboring landmarks along both manual contours to form the group vectors in the training matrix. In our experiment, the vertex number of endocardial contour in a group vector is 50, and 70 for epicardial contour. Two training samples from the repository are shown in Fig. 3. After the determination of landmarks and group vertices, the matrices of training landmarks and groups can be formed based on the method described in section 2.1 and 2.2, and notated as  $D_L$  and  $D_G$ .



**Fig. 3.** Two images with papillary muscle centers, landmarks, endocardium and epicardium contour vertices.

Given a testing image, the localization procedure is as follows: first, manually labeling its landmarks, notated as  $y_L$ ; then,  $x_{opt}$  and  $\beta_{opt}$  are computed by optimizing

Eq. (3) with  $D_L$  and  $y_L$ ; at last, transforming  $D_G x_{opt}$  back to the coordinate system of testing image as the group location.

In order to compare our method with the original SSC, SSC was also employed in the same localization framework with same training data and input landmarks to separately localize the endocardium and epicardium of left ventricle.

### 3.2 Evaluation and comparison

#### Evaluation

In addition to visual evaluation of localization accuracy, three quantitative measures are employed: average perpendicular distance (APD) [12], standard deviation of perpendicular distances (SPD), and dice metric (DM) [13]. As the name implies, APD and SPD measures the perpendicular distances from points on the localization result to manual contour, and calculate the average and standard deviation of them. Higher APD or SPD implies that localization result doesn't match closely to the manual contour. DM measures the overlap rate of the areas surrounded by localization and manual contour. It ranges between 0 and 1, with higher DM indicating better match.

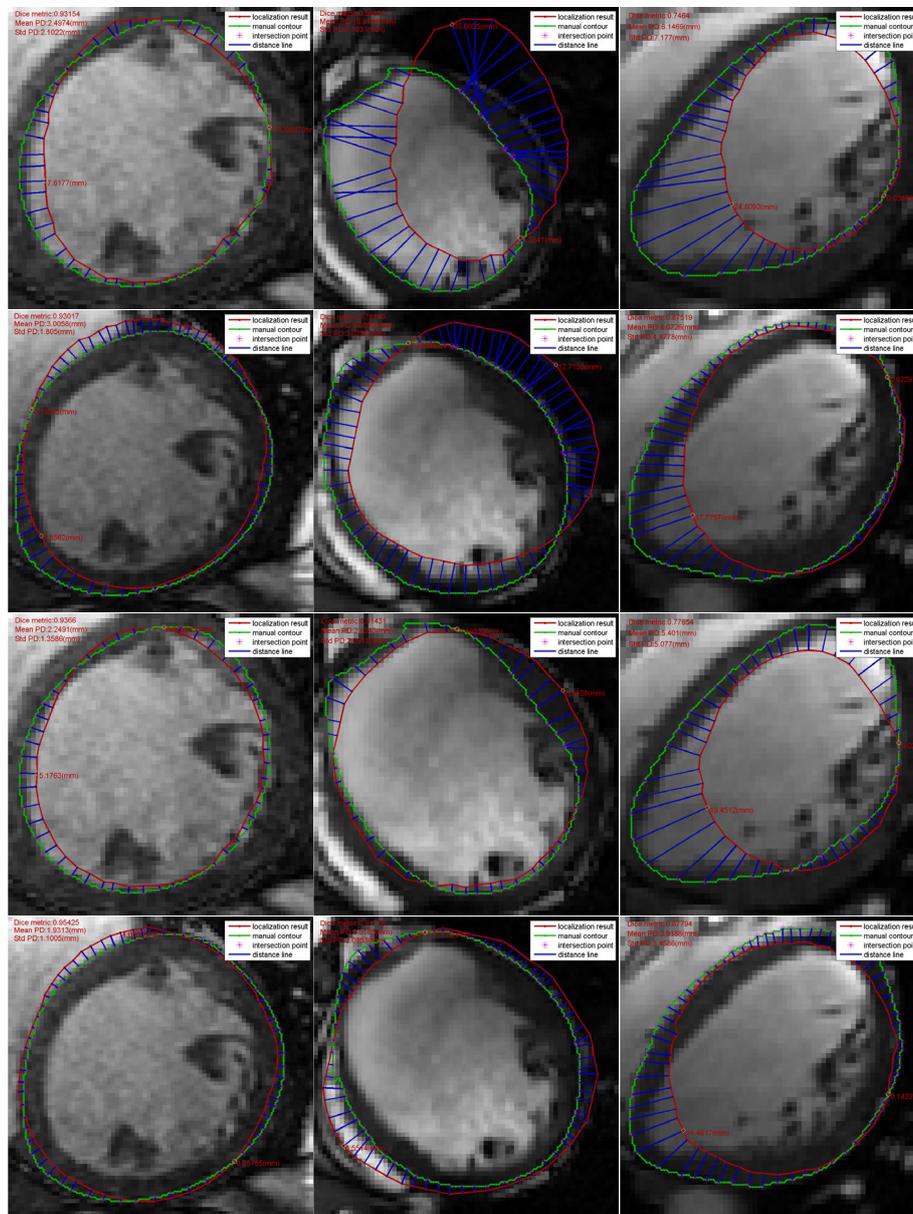
#### Comparison

Accuracy and robustness of endocardium and epicardium localization based on our method and the original SSC are compared in this section. Three localization cases from different patients are shown in Fig. 4. The left column shows a case that both SSC and our method achieve acceptable result. The middle column shows a case that SSC fails in the localization, whereas our method still performs well. The right column shows an extreme case that neither method locates the endocardium and epicardium accurately due to the insufficiency of training samples, which is quite rare in our experiment. Despite the failure in the third case, our method outperforms the original SSC in all three cases from both visual and quantitative point of view. Furthermore, in order to compare them from the big picture, global APD, SPD and DM of our method and SSC are calculated statistically on 91 testing images, as shown in Table 1.

**Table 1.** Global APD, SPD and DM

Measures	SSC		Improved SSC	
	Endocardium	Epicardium	Endocardium	Epicardium
Global APD	3.3554	2.7553	2.3493	2.2032
Global SPD	2.4982	1.9412	1.6836	1.5741
Global DM	0.8766	0.9224	0.9124	0.9363

From the evidence provided in Fig. 4 and Table 1, we can reasonably arrive at a conclusion that our method can achieve more accurate and stable localization in this study compared with SSC.



**Fig. 4.** Three localization cases. First and second row: endocardium and epicardium localization utilizing SSC. Third and fourth row: endocardium and epicardium localization utilizing our method.

## 4 Conclusion

In this paper, we proposed an extension of Sparse Shape Composition model for multi-shape prior modeling. In our method, multiple interested shapes from one patient are regarded as a group and modeled together to incorporate the co-dependency among different shapes. It is validated on a 2D endocardium and epicardium localization task, and exhibits more accurate and stable performance compared with original SSC. The success of our method is mainly relying on the incorporation of co-dependency among shapes.

In the future, we intend to apply this extension to various multi-shape segmentation tasks in clinical practices, especially to 3D shape sequences of time-varying organs.

## References

1. Leventon, M. E., Grimson, W. E. L., & Faugeras, O. (2000). Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on* (Vol. 1, pp. 316-323). IEEE.
2. Cremers, D., Tischhäuser, F., Weickert, J., & Schnörr, C. (2002). Diffusion snakes: Introducing statistical shape knowledge into the Mumford-Shah functional. *International journal of computer vision*, 50(3), 295-313.
3. Mumford, D., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5), 577-685.
4. Ali, A. M., Farag, A. A., & El-Baz, A. S. (2007). Graph cuts framework for kidney segmentation with prior shape constraints. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007* (pp. 384-392). Springer Berlin Heidelberg.
5. Heimann, T., & Meinzer, H. P. (2009). Statistical shape models for 3D medical image segmentation: A review. *Medical image analysis*, 13(4), 543-563.
6. Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models—their training and application. *Computer vision and image understanding*, 61(1), 38-59.
7. Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6), 681-685.
8. Rousson, M., & Paragios, N. (2002). Shape priors for level set representations. In *Computer Vision—ECCV 2002* (pp. 78-92). Springer Berlin Heidelberg.
9. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D. N., & Zhou, X. S. (2012). Towards robust and effective shape modeling: Sparse shape composition. *Medical image analysis*, 16(1), 265-277.
10. Zhang, S., Zhan, Y., & Metaxas, D. N. (2012). Deformable segmentation via sparse representation and dictionary learning. *Medical Image Analysis*, 16(7), 1385-1396.
11. Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright GA. "Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI." The MIDAS Journal -Cardiac MR Left Ventricle Segmentation Challenge, <http://hdl.handle.net/10380/3070>
12. Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 285-339.
13. Lynch, M., Ghita, O., & Whelan, P. F. (2008). Segmentation of the left ventricle of the heart in 3-D+t MRI data using an optimized nonrigid temporal model. *Medical Imaging, IEEE Transactions on*, 27(2), 195-203.