Segmentation of Thalamus from MR images via Task-Driven Dictionary Learning

Luoluo Liu[a], Jeffrey Glaister[a], Xiaoxia Sun[a], Aaron Carass[a,b], Trac D. Tran[a], and Jerry L. Prince[a]

[a]Dept. of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
[b]Dept. of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA

ABSTRACT

Automatic thalamus segmentation is useful to track changes in thalamic volume over time. In this work, we introduce a task-driven dictionary learning framework to find the optimal dictionary given a set of eleven features obtained from T1-weighted MRI and diffusion tensor imaging. In this dictionary learning framework, a linear classifier is designed concurrently to classify voxels as belonging to the thalamus or non-thalamus class. Morphological post-processing is applied to produce the final thalamus segmentation. Due to the uneven size of the training data samples for the non-thalamus and thalamus classes, a non-uniform sampling scheme is proposed to train the classifier to better discriminate between the two classes around the boundary of the thalamus. Experiments are conducted on data collected from 22 subjects with manually delineated ground truth. The experimental results are promising in terms of improvements in the Dice coefficient of the thalamus segmentation over state-of-the-art atlas-based thalamus segmentation algorithms.

Keywords: Dictionary learning and sparse representation, diffusion tensor imaging, thalamus, segmentation

1. INTRODUCTION

The thalamus is an important subcortical structure in the brain of vertebrates.[1] Its primary function is to relay sensory and motors signals to the cerebral cortex. Thalamic size or volume is useful for tracking the progression of neurodegenerative diseases.[2] There is a need to automatically segment the thalamus from magnetic resonance images (MRI) in order to compute the thalamic volume.

One common method to find the thalamus is to transform the T1-weighted (T1-w) MRI into an atlas domain and use labels from the atlas to identify the subcortical structures in the original image. For example, Bazin and Pham[3] proposed combining statistical and topological atlases to find such subcortical structures. More sophisticated methods incorporate shape or spatial priors, such as shape models, topological correction, or spatial information. A method proposed by Patenaude *et al.*[4] incorporates prior anatomical information using explicit shape models. Fischl *et al.*[5] proposed using spatial information of relative locations of subcortical structures as a spatial prior. However, none of these algorithms incorporate the strong connectivity properties of the thalamus.

Diffusion tensor imaging (DTI) is an MRI sequence used to understand connectivity between structures within the brain.[6] DTI images the diffusion rate of water molecules across multiple directions, which reflects the interactions of water molecules with fibers and membranes and reveals the microscopic details of the tissue structure. Diffusion occurs more strongly along the direction of neural fibers that make up these connections. Since the thalamus has more connections than its immediate surrounding tissues, a rich set of features derived from DTI can help identify the thalamus.

Sparse representation uses a learned dictionary to represent the feature vectors as a linear combination of a small subset of atoms or elements in the dictionary. Sparse Representation Classification (SRC) proposed by Wright *et al.*[7] uses the feature vectors as dictionary atoms and has been applied to segmentation of the prostate from computed tomography images.[8] However, SRC relies on representative training data to form the

Further author information: (Send correspondence to L. Liu)
L. Liu: E-mail: lliu69@jhu.edu

dictionary atoms and hence the performance is affected by noisy training data. Other algorithms proposed for dictionary learning include Label-Consistent K-SVD[9] and Online Dictionary Learning,[10] which has been applied to hippocampus segmentation.[11] However, both methods perform joint optimization by concatenation of the dictionary and the classifier parameters and thus the recovery performance is not guaranteed.[12] Task-Driven Dictionary Learning (TDDL)[13] forms the problem into a bi-level optimization, with the dictionary and linear classifier optimized concurrently using a stochastic gradient descent method. TDDL has been found to produce results comparable to other dictionary learning methods, is robust to noisy training data, and converges quickly.

The proposed algorithm uses a modified TDDL framework to learn a dictionary to classify voxels as belonging to the thalamus class. The dictionary is learned from a set of features extracted from T1-w and DTI MRI. Due to the small size of the thalamus compared to the rest of the brain, a modified sampling scheme is used when learning the dictionary, which preferentially samples voxels belonging to the thalamus and boundary classes, compared to voxels far from the thalamus. When presented with a test data set, the learned dictionary and classifier are used to classify each voxel independently, producing a binary mask consisting of voxels belonging to the thalamus class. A morphological operator and connected-component analysis is applied to the image to obtain the final thalamus segmentation.

2. METHODS

The proposed algorithm consists of three steps. A set of features are extracted at each voxel in the volume. Next, given a set of training data with a manually delineated thalamus, a dictionary is learned to sparsely represent the feature vectors and a linear classifier is trained to distinguish between thalamus and non-thalamus feature vectors. Finally, when provided with a set of test volumes, each voxel is classified by the linear classifier and simple post-processing steps generate the final segmentation.

2.1 Feature extraction

Eleven features are extracted at each voxel from T1-weighted MRI (T1-w) and DTI. Let the vector $\boldsymbol{s}$ be defined as a voxel coordinate in 3D space, then $\boldsymbol{s} \triangleq (x, y, z)^T$. The first set of features depends on the spatial difference between the voxel coordinate and the volume center voxel $\boldsymbol{s}_c$. Let the absolute difference of a voxel to the center voxel be defined as $\Delta\boldsymbol{s} = (|x - x_c|, |y - y_c|, |z - z_c|)^T$. The first three features of $\boldsymbol{d}(\boldsymbol{s})$ are the non-increasing function for each element of $\Delta\boldsymbol{s}$ with σ^2 as a smoothing factor,

$$\boldsymbol{d}(\boldsymbol{s}) = \exp\left\{-\frac{\Delta\boldsymbol{s}}{\sigma^2}\right\}.$$

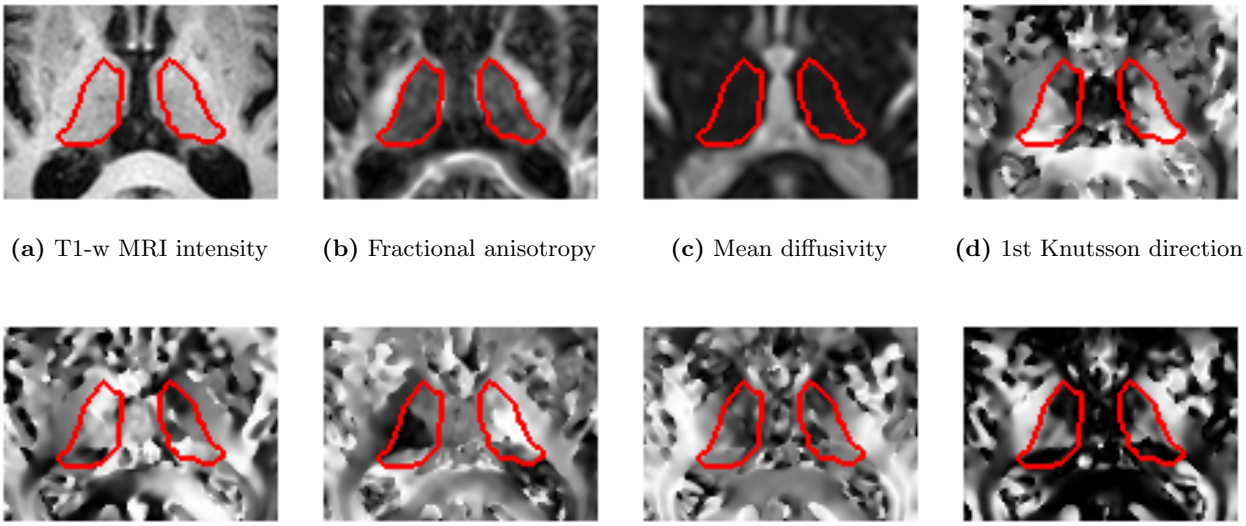
The fourth feature is the T1-w intensity at $\boldsymbol{s}$, denoted as $I(\boldsymbol{s})$. The T1-w volumes are normalized using white-matter peak normalization to account for variations in intensities between different image volumes.

The remaining seven features are derived from the diffusion tensor. The diffusion tensor is a matrix that represents the diffusivity at a voxel and can be expressed in terms of its eigenvalues and eigenvectors. Fractional anisotropy $\mathrm{FA}(\boldsymbol{s})$ and mean diffusivity $\mathrm{MD}(\boldsymbol{s})$ are two features derived from the eigenvalues of the tensor at $\boldsymbol{s}$.

Five features are based on the direction of maximum diffusion, which corresponds to the principal eigenvector $\boldsymbol{u} = (u_1, u_2, u_3)^T$ of the diffusion tensor. However, diffusion occurs in the direction of $\boldsymbol{u}$ and $-\boldsymbol{u}$ with equal probability, so we represent $\boldsymbol{u}$ as an orientation in the Knutsson space,[14]

$$\boldsymbol{K}(\boldsymbol{u}) = (u_1^2 - u_2^2, 2u_1u_2, 2u_1u_3, 2u_2u_3, \frac{1}{\sqrt{3}}(2u_3^2 - u_1^2 - u_2^2))^T.$$

The eleven aforementioned features: $\boldsymbol{d}(\boldsymbol{s}), I(\boldsymbol{s}), \mathrm{FA}(\boldsymbol{s}), \mathrm{MD}(\boldsymbol{s})$, and $\boldsymbol{K}(\boldsymbol{u})$ are concatenated to construct the final feature vector $\boldsymbol{f}(\boldsymbol{s})$. Examples of these features are shown in Fig. 1 and the red contour highlights the ground truth of the thalamus.



(a) T1-w MRI intensity **(b)** Fractional anisotropy **(c)** Mean diffusivity **(d)** 1st Knutsson direction

(e) 2nd Knutsson direction **(f)** 3rd Knutsson direction **(g)** 4th Knutsson direction **(h)** 5th Knutsson direction

Figure 1: Examples of features used in $\boldsymbol{f}(\boldsymbol{s})$. The red contour is the ground truth delineation of the thalamus.

2.2 Algorithm

The proposed algorithm adapts the TDDL framework[13] for use with a large set of training data. Given a dictionary $\mathbf{D}$, the sparse coding procedure aims to find a sparse representation of the feature vector $\boldsymbol{f}(\boldsymbol{s})$ using $\mathbf{D}$, denoted as $\boldsymbol{\alpha}^\star(\boldsymbol{f}(\boldsymbol{s}), \mathbf{D})$. It is found by solving the following minimization problem known as Basis Pursuit[15] or Least Absolute Shrinkage and Selection Operator (LASSO),[16]

$$\boldsymbol{\alpha}^\star(\boldsymbol{f}(\boldsymbol{s}), \mathbf{D}) = \underset{\boldsymbol{\alpha}\in\mathbb{R}^n}{\arg\min} \frac{1}{2}\|\boldsymbol{f}(\boldsymbol{s}) - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1. \tag{1}$$

The sparse vector $\boldsymbol{\alpha}^\star$ and manual delineation $\boldsymbol{y}(\boldsymbol{s})$ are used to jointly learn the dictionary $\mathbf{D}$ and the linear classifier weights $\mathbf{W}$ that minimizes the prediction error of $\boldsymbol{y}(\boldsymbol{s})$. In our implementation, we measure the prediction error as the empirical squared error over a set $\mathcal{S}$ of samples from the training data set. This gives the following objective function,

$$\langle \mathbf{D}^\star, \mathbf{W}^\star \rangle = \underset{\mathbf{D}\in\mathbb{R}^{m\times n}, \mathbf{W}\in\mathbb{R}^{k\times n}}{\arg\min} \frac{1}{2|\mathcal{S}|} \sum_{\boldsymbol{s}\in\mathcal{S}} \|\boldsymbol{y}(\boldsymbol{s}) - \mathbf{W}\boldsymbol{\alpha}^\star(\boldsymbol{f}(\boldsymbol{s}), \mathbf{D})\|_2^2 + \frac{\mu}{2}\|\mathbf{W}\|_F^2, \tag{2}$$

where the first term is the empirical risk and the Frobenius norm penalty prevents over-fitting of the model. Due to the size of the training data, we take $\mathcal{S}$ to be a random subset of voxels from the training data and the stochastic gradient descent algorithm is used to optimize (2).

After learning the dictionary and classifier parameters, each voxel in a new test volume are classified independently as belonging to either the thalamus or non-thalamus class. Let the voxel coordinate of the test voxel be given as $\boldsymbol{s}_{\boldsymbol{\tau}}$. Then, we extract a test feature vector $\boldsymbol{f}(\boldsymbol{s}_{\boldsymbol{\tau}})$ and solve for $\boldsymbol{\alpha}^\star(\boldsymbol{f}(\boldsymbol{s}_{\boldsymbol{\tau}}), \mathbf{D}^\star)$ using (1). Using the classifier weights, the predicted class $\hat{c}(\boldsymbol{s}_{\boldsymbol{\tau}})$ is given by

$$\hat{c}(\boldsymbol{s}_{\boldsymbol{\tau}}) = \underset{r\in\{1,2\}}{\arg\max}\, \mathbf{w}_r^\star \boldsymbol{\alpha}^\star(\boldsymbol{f}(\boldsymbol{s}_{\boldsymbol{\tau}}), \mathbf{D}^\star), \tag{3}$$

where $\mathbf{w}_r^\star$ is the r^{th} row of $\mathbf{W}^\star$.

After voxel-wise classification, post-processing is applied to the resulting binary map to remove small holes and find the largest connected component, which is assumed to be the thalamus. Small holes and small connections between components are removed by applying the morphological opening operator, followed by the morphological closing operator. Both operators use a spherical structuring element with a radius of 2 voxels. Next, the binary map is divided equally into left and right volumes along the midsagittal plane.[17] The largest connected component in either hemisphere is found. These two components are assumed to correspond to the left and right thalamus and voxels not belonging to these components are classified as non-thalamus.

2.3 Implementation Details

The proposed algorithm uses a non-uniform sampling scheme to find the training set $\mathcal{S}$ in stochastic gradient descent because of the discrepancy between the number of non-thalamus voxels and thalamus voxels, This is to encourage the dictionary to learn more information about the thalamus voxels and non-thalamus voxels near the thalamus boundary (henceforth called boundary voxels). To do this, we partition the training data into three subsets: thalamus voxels, boundary voxels, and all other non-thalamus voxels. Non-thalamus voxels within 5 voxels of the thalamus boundary are considered to be boundary voxels. Then training samples are drawn uniformly from each subset with a certain ratio. We set the ratio to be 3/8, 3/8, and 2/8 respectively. We draw 5000 samples per stochastic gradient descent iteration in our experiments.

Other algorithm parameters for the optimization algorithm include: ℓ_1 norm regularization weight $\lambda = 0.1$; Frobenius norm regularization weight $\mu = 0.9$; stochastic gradient descent step size $\rho = 0.0001$; dictionary size $n = 400$; and the total number of iterations of gradient descent for training $T = 8000$. σ^2 of the spatial absolute difference feature $\boldsymbol{d}(\boldsymbol{s})$ is chosen to be 100. All these parameters were determined empirically.

3. EXPERIMENTS

3.1 Dataset

The data consists of 22 patients from a study of cerebellar ataxia with manual delineated ground truth.[18] The subject images were acquired on a 3T MR scanner (Intera, Philips Medical Systems, Netherlands).

3.2 Comparison with state-of-art segmentation algorithms

We compared the proposed algorithm with a set of state-of-art segmentation and classification algorithms. We use a two-fold cross validation scheme for testing. Results are compared with two state-of-art algorithms: TOADS[3] and FreeSurfer.[5] Both are atlas-based subcortical segmentation algorithms that provide labels for the thalamus.

The performance of the segmentation algorithms is quantified using the Dice coefficient (DC):[19]

$$\mathrm{DC} = \frac{2|A \cap G|}{|A| + |G|}.$$

DC measures the overlap between a segmentation algorithm's result A and the ground truth G. Figure 2 shows example results of the proposed algorithm. Box plots showing median and range of Dice coefficients for the proposed and comparison algorithms are in Fig. 3. The proposed algorithm performs better than TOADS and FreeSurfer, with a median DC of 0.8057, compared with 0.6104 for TOADS and 0.6875 for FreeSurfer. A paired Wilcoxen signed rank test comparing our method with TOADS has a p-value < 0.001 and comparing with FreeSurfer has a p-value of 0.0017, indicating significant improvement. However, the proposed algorithm is also prone to producing outliers if the post-processing steps remove too many or too few misclassified voxels. See Fig. 2(d) for an example of an outlier (DC = 0.5740) where a large amount of thalamus voxels were classified as non-thalamus.

4. CONCLUSIONS

In this paper, we explore the a task-driven dictionary learning algorithm to perform voxel-wise classification with features derived from T1-w and DTI MR data. The results show that the algorithm performed better than two state-of-the-art atlas-based methods. However, in cases where the post-processing steps failed to remove misclassified voxels, the algorithm performance can suffer slightly. Future improvements include incorporating more intelligent outlier rejection and segmentation refinement methods.

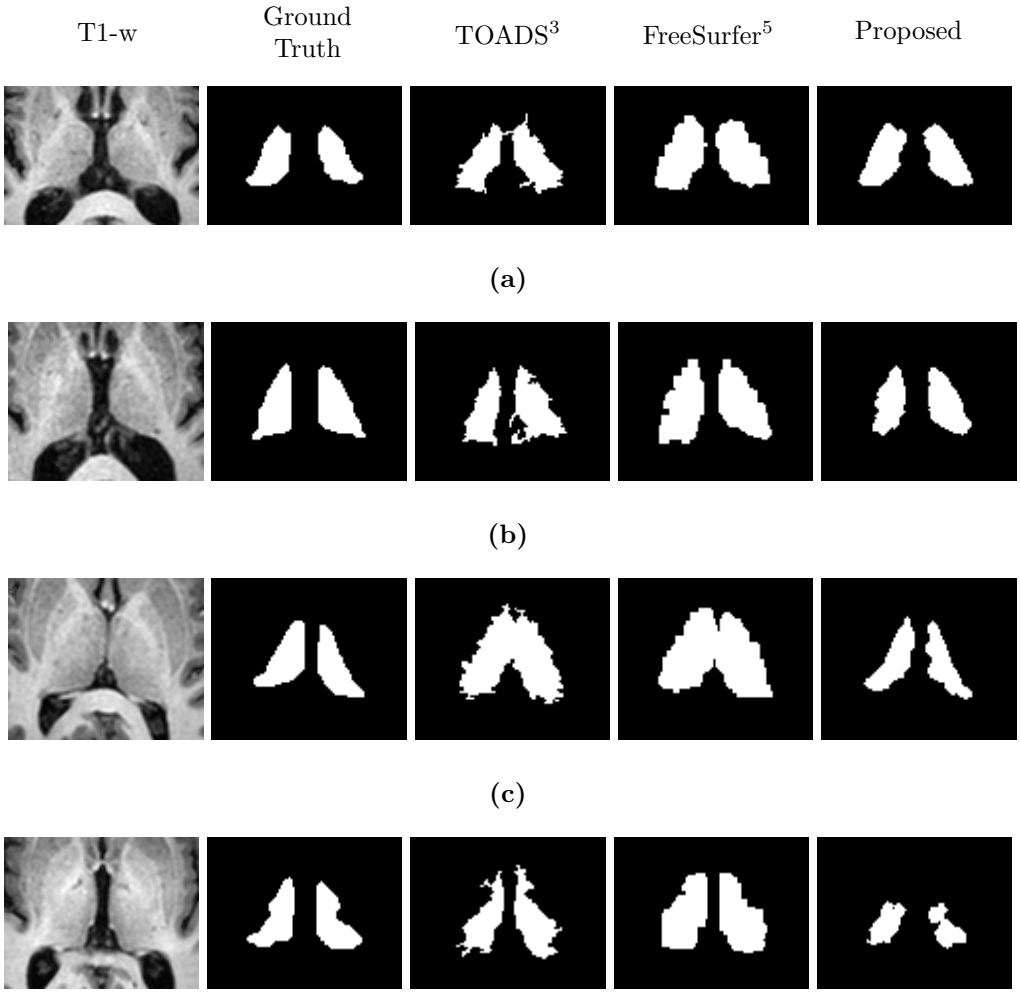


Figure 2: Examples of the resulting segmentation: **(a)** shows our best result with our proposed algorithm; whereas **(b)** and **(c)** show typical results for the proposed algorithm; and **(d)** is an outlier for the proposed method, where post-processing removed too many thalamus voxels.

5. ACKNOWLEDGMENTS

This work has been partially supported by the NIH/NINDS grants R21-NS082891 and R01-NS056307; NSF under Grants CCF-1117545 and CCF-1422995; ARO under Grant 60219-MA; and ONR under Grant N00014-12-1-0765. J. Glaister is supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Sherman, S. M. and Guillery, R., [*Exploring the Thalamus*], Elsevier, San Diego, California (2000).

[2] Cifelli, A., Arridge, M., Jezzard, P., Esiri, M. M., Palace, J., and Matthews, P. M., "Thalamic neurodegeneration in multiple sclerosis," *Annals of Neurology* **52**(5), 650–653 (2002).

[3] Bazin, P.-L. and Pham, D., "Homeomorphic brain image segmentation with topological and statistical atlases," *Med. Image Anal.* **12**, 616–625 (2008).

[4] Patenaude, B., Smith, S. M., Kennedy, D., and Jenkinson, M. A., "A Bayesian model of shape and appearance for subcortical brain segmentation," *NeuroImage* **56**, 907–922 (2011).

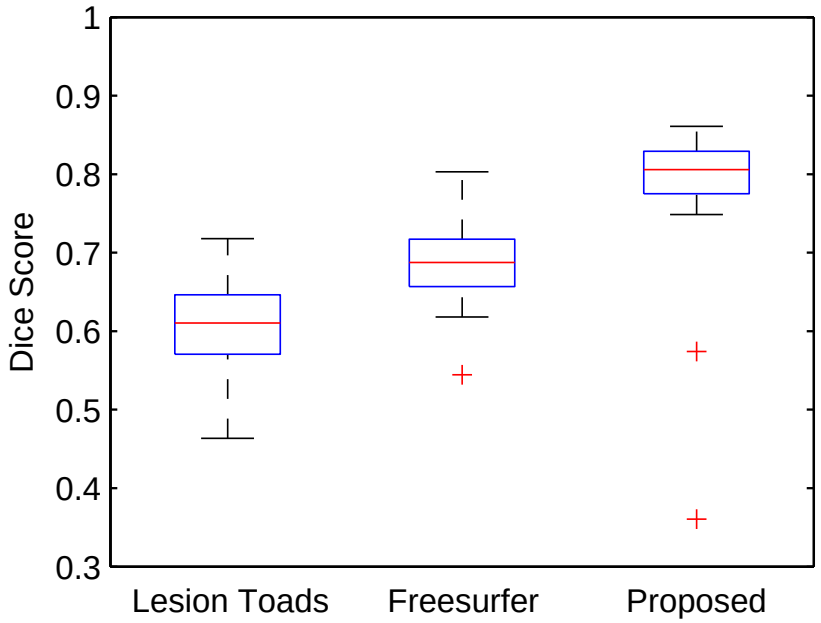


Figure 3: Thalamus segmentation Dice coefficients. The proposed algorithm has a median DC of 0.8057, higher than TOADS (median DC of 0.6104) and FreeSurfer (median DC of 0.6875).

[5] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al., "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron* **33**(3), 341–355 (2002).

[6] Westin, C.-F., Maier, S. E., Mamata, H., Nabavi, A., Jolesz, F. A., and Kikinis, R., "Processing and visualization for diffusion tensor MRI," *Medical image analysis* **6**(2), 93–108 (2002).

[7] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y., "Robust face recognition via sparse representatio," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009).

[8] Gao, Y., Liao, S., and Shen, D., "Prostate segmentation by sparse representation based classification," in [*15th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2012)*], *Lecture Notes in Computer Science* **7512**, 451–458, Springer Berlin Heidelberg (2012).

[9] Jiang, Z., Lin, Z., and Davis, L. S., "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in [*Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on; IEEE*], 1697–1704 (2011).

[10] Mairal, J., Bach, F., Ponce, J., and Sapiro, G., "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research* **11**, 19–60 (2010).

[11] Tong, T., Wolz, R., Coupé, P., Hajnal, J. V., Rueckert, D., Initiative, A. D. N., et al., "Segmentation of mr images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage* **76**, 11–23 (2013).

[12] Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T., "Coupled dictionary training for image super-resolution," *Image Processing, IEEE Transactions on* **21**(8), 3467–3478 (2012).

[13] Mairal, J., Bach, F., and Ponce, J., "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(4), 791–804 (2012).

[14] Knutsson, H., "Producing a continuous and distance preserving 5-D vector representation of 3-D orientation," in [*IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*], 175–182 (1985).

[15] Chen, S. and Donoho, D., "Basis pursuit," in [*Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on; IEEE*], **1**, 41–44 (1994).

[16] Tibshirani, R., "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)* **58**(1), 267–288 (1996).

[17] Anbazhagan, P. et al., "Automatic estimation of midsagittal plane and AC-PC alignment on nonrigid registration," in [*3rd International Symposium on Biomedical Imaging (ISBI 2006)*], 828–831 (2006).

[18] Stough, J. V., Glaister, J., Ye, C., Ying, S. H., Prince, J. L., and Carass, A., "Automatic method for thalamus parcellation using multi-modal feature classification," in [*17th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2014)*], *Lecture Notes in Computer Science* **8675**, 169–176, Springer Berlin Heidelberg (2014).

[19] Dice, L. R., "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).