Brain Tumor Image Segmentation Using Kernel Dictionary Learning

Jeon Lee, *Member, IEEE*, Seung-Jun Kim, *Senior Member, IEEE*, Rong Chen, and Edward H. Herskovits

***Abstract*— Automated brain tumor image segmentation with high accuracy and reproducibility holds a big potential to enhance the current clinical practice. Dictionary learning (DL) techniques have been applied successfully to various image processing tasks recently. In this work, kernel extensions of the DL approach are adopted. Both reconstructive and discriminative versions of the kernel DL technique are considered, which can efficiently incorporate multi-modal nonlinear feature mappings based on the kernel trick. Our novel discriminative kernel DL formulation allows joint learning of a task-driven kernel-based dictionary and a linear classifier using a *K*-SVD-type algorithm. The proposed approaches were tested using real brain magnetic resonance (MR) images of patients with high-grade glioma. The obtained preliminary performances are competitive with the state of the art. The discriminative kernel DL approach is seen to reduce computational burden without much sacrifice in performance.**

I. INTRODUCTION

Brain tumor segmentation in magnetic resonance (MR) images plays an important role in diagnosis, surgery or radiotherapy planning, and tumor growth monitoring. Because there is no automated segmentation method acceptable to clinicians, the task is usually done manually by the radiologists, which can take anywhere between 30 minutes to 3.5 hours per patient [1]. The manual segmentation results in the intra-rater volume variability of $(20 \pm 15)\%$ and the inter-rater volume variability of $(28 \pm 12)\%$ [2]. Thus, (semi-)automatic segmentation with high accuracy and reproducibility holds a big potential in MR-based brain tumor diagnosis and treatment. However, automated brain tumor segmentation is very challenging as the pixel intensities and textures corresponding to normal and tumorous regions are often similar [3]. Furthermore, the size, shape and locations of tumors vary considerably from case to case.

Current methods for brain tumor segmentation roughly fall into either generative or discriminative methods. Generative methods utilize prior knowledge on the morphologies and spatial structures of tumorous and normal tissues, and thus require significant effort to encode expert's semantic interpretation into appropriate probabilistic models [4], [5]. Discriminative methods directly learn the characteristic differences between tumor and normal tissues from annotated training images, but may be sensitive to image acquisition protocols and accuracy of the registration process [6], [7].

Recently, competing algorithms for brain tumor detection and segmentation were evaluated through the Multi-modal Brain Tumor Segmentation (BRATS) Challenges [8]. Discriminative probabilistic approaches based on the random forest classifier were shown to yield a good performance [6], [9]. Discriminative approaches including dictionary learning (DL)-based ones have also received attention [8].

DL attempts to construct a set of overcomplete bases such that each datum can be explained well using only a small number of basis vectors chosen from the set. Compared to using predefined bases such as the Fourier or wavelet bases, employing dictionaries learned from data was seen to improve the performances of various signal processing and machine learning tasks significantly [10], [11].

There have been a number of extensions to the plain vanilla DL idea in the literature. First, it was observed that dictionaries can be further tailored for particular tasks. For example, in the context of classification, jointly designing the dictionary and the classifier often improves the classification performance, since the dictionary can then better capture *discriminative* features in the data [12], [13]. Another interesting extension is to kernelize the DL technique. By mapping the data to a high-dimensional feature space and seeking salient subspaces in this transformed domain, kernel DL can cope with highly nonlinear features effectively. Resorting to the kernel trick, existing DL algorithms can be efficiently kernelized without actually computing the transformed features [14], [15]. Discriminative linear DL and a special case of (reconstructive) kernel DL were recently tested for hippocampus and brain tumor segmentations in MR images, respectively [16], [17].

In this study, the kernel extensions of the DL approach are adopted to tackle the challenges associated with the tumor segmentation task in brain MR images. Both reconstructive and discriminative versions of the kernel DL techniques are considered. Our novel formulation for discriminative kernel DL allows joint learning of a kernel-based task-driven dictionary and a linear classifier to detect normal/tumorous regions, using adapted orthogonal matching pursuit (OMP) [18] and *K*-SVD algorithms [19].

The rest of the paper is organized as follows. The general DL and its (reconstructive) kernel extension are introduced in Sec. II. The discriminative kernel DL formulation and the *K*-SVD-based solution method are presented in Sec. III. Preliminary test results using the USPS hand-written digit images and the real MR images from the BRATS dataset are reported in Secs. IV and V, respectively. Conclusions and

Jeon Lee and Seung-Jun Kim (corresponding author) are with Dept. of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250, USA. Phone: +1-410-455-3372. Fax: +1-410-455-3969. E-mail: {jeonlee, sjkim}@umbc.edu.

Rong Chen is with Dept. of Diagnostic Radiology and Nuclear Medicine, University of Maryland, 22 South Greene Street, Baltimore, MD 21201, USA. E-mail: rong.chen.mail@gmail.com

Edward Herskovits is with Dept. of Radiology, University of Maryland, 22 South Greene Street, Baltimore, MD 21201, USA. E-mail: eherskovits@umm.edu.

future work are discussed in Sec. VI.

II. DL and Its Kernel Extension

Given N data (feature) vectors denoted by $\{\mathbf{x}_n \in \mathbb{R}^P\}$, DL aims at obtaining a (possibly overcomplete) basis set $\mathbf{D} \in \mathbb{R}^{N\times K}$ to approximate the data via a linear combination of few basis vectors (called *atoms*) $\{\mathbf{d}_k\}_{k=1}^K$ in $\mathbf{D}$. This can be accomplished by the constrained optimization [19]

$$\min_{\mathbf{D}\in\mathscr{D}\subset\mathbb{R}^{P\times K},\{\alpha_n\in\mathbb{R}^K\}_{n=1}^N} \sum_{n=1}^N \|\mathbf{x}_n - \mathbf{D}\alpha_n\|_2^2 \quad (1a)$$

$$\text{subject to } \|\alpha_n\|_0 \le K_0,\ n=1,2,\ldots,N \quad (1b)$$

where $\mathscr{D} := \{\mathbf{D} := [\mathbf{d}_1,\mathbf{d}_2,\ldots,\mathbf{d}_K] : \|\mathbf{d}_k\|_2 = 1\ \forall k\}$, α_n is the sparse coefficient vector to represent $\mathbf{x}_n$, and K_0 is the maximum number of nonzero coefficients in α_n for any n. Vector α_n is sparse in the sense that $K_0 \ll K$. This problem can be solved effectively via the K-SVD algorithm [19], which alternates sparse coding and dictionary update steps. The algorithm has been successfully applied for various signal processing and machine learning tasks.

Recently, kernelized DL algorithms have also been developed [14], [15]. They mitigate the limitation of linear DL that the features belonging to similar subspaces are difficult to discriminate, by transforming the features through a nonlinear mapping. The feature space may even be infinite-dimensional, but the *kernel trick* comes handy to avoid any direct computation of the transformed features. Upon denoting the mapping as Φ, a relevant formulation is

$$\min_{\Omega\in\mathscr{O}\subset\mathbb{R}^{N\times K},\{\alpha_n\in\mathbb{R}^K\}_{n=1}^N} \sum_{n=1}^N \|\Phi(\mathbf{x}_n) - \Phi(\mathbf{X})\Omega\alpha_n\|_2^2 \quad (2a)$$

$$\text{subject to } \|\alpha_n\|_0 \le K_0,\ n=1,2,\ldots,N \quad (2b)$$

where $\Phi(\mathbf{X}) := [\Phi(\mathbf{x}_1),\Phi(\mathbf{x}_2),\ldots,\Phi(\mathbf{x}_N)]$ and $\mathscr{O} := \{\Omega := [\omega_1,\omega_2,\ldots,\omega_K] : \|\Phi(\mathbf{X})\omega_k\|_2 = 1\ \forall k\}$. Note that since the transformed features are embedded in a high-dimensional space, it is prudent to confine the search for the dictionary in the subspace spanned by the features, which is reflected in the cost function in (2a) [15].

III. Discriminative Kernel DL

Formulations (1) and (2) train dictionaries based on the criterion of faithful reconstruction of (transformed) features. The resulting dictionaries have been shown to be powerful for reconstructive tasks such as image denoising and inpainting [10]. Although it was observed that such a dictionary is also useful for discriminative tasks such as classification [11], tailoring the dictionary for specific supervised learning tasks may potentially lead to better performance [12], [21].

A. Problem Formulation

Online learning algorithms for discriminative DL was derived for the linear and kernelized cases in [12] and [20], respectively, based on iterative stochastic gradient descent methods. Here, the goal is to derive a K-SVD variant for discriminative kernel DL. Inspired by [13], which developed a K-SVD-type algorithm for discriminative linear DL, our kernelized formulation is described next. Let C be the number of classes, and $\mathbf{y}_n \in \{1,0\}^C$ the class label vector for datum $n=1,2,\ldots,N$, whose c-th entry is equal to 1 if datum n belongs to class c, and 0 otherwise, for $c=1,2,\ldots,C$. Let $\mathbf{W}\in\mathbb{R}^{C\times K}$ denote a linear classifier. Then, the proposed formulation is

$$\min_{\Omega,\{\alpha_n\},\mathbf{W}} \sum_{n=1}^N \left[\|\Phi(\mathbf{x}_n) - \Phi(\mathbf{X})\Omega\alpha_n\|_2^2 + \gamma\|\mathbf{y}_n - \mathbf{W}\alpha_n\|_2^2\right] + \mu\|\mathbf{W}\|_F^2 \quad (3a)$$

$$\text{subject to } \|\alpha_n\|_0 \le K_0,\ n=1,2,\ldots,N \quad (3b)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix and γ and μ are scalars controlling the contribution of their corresponding terms. Note that μ is eliminated in the final formulation (4), and does not need to be tuned. Problem (3) learns jointly the dictionary $\Phi(\mathbf{X})\Omega$ and the classifier $\mathbf{W}$, which encourages the dictionary to be tailored to the discriminative task at hand.

In order to apply the kernel trick and find the solution to (3) using a K-SVD-type procedure, (3) is reformulated. Upon defining $\mathbf{X} := [\mathbf{x}_1,\ldots,\mathbf{x}_N]$, $\mathbf{Y} := [\mathbf{y}_1,\ldots,\mathbf{y}_N]$, and $\mathbf{A} := [\alpha_1,\ldots,\alpha_N]$, consider

$$\min_{\Omega,\mathbf{A},\mathbf{W}} \left\| \begin{bmatrix}\Phi(\mathbf{X})\\ \sqrt{\gamma}\mathbf{Y}\end{bmatrix} - \begin{bmatrix}\Phi(\mathbf{X})\Omega\\ \sqrt{\gamma}\mathbf{W}\end{bmatrix}\mathbf{A}\right\|_F^2 \quad (4a)$$

$$\text{subject to } \|\alpha_n\|_0 \le K_0,\ n=1,2,\ldots,N \quad (4b)$$

where it is understood that the ℓ_2-norms of the columns of $[(\Phi(\mathbf{X})\Omega)^T\ \sqrt{\gamma}\mathbf{W}^T]^T$ are normalized to unity. Thanks to this normalization, the explicit regularization term $\mu\|\mathbf{W}\|_F^2$ in (3a) is omitted in (4). Furthermore, it is noted that (4) is in the same form as (1), leading to a K-SVD-based solution.

B. K-SVD-Based Algorithm

Let $k(\cdot,\cdot)$ be a positive semidefinite kernel satisfying $k(\mathbf{x},\mathbf{x}') = \langle\Phi(\mathbf{x}),\Phi(\mathbf{x}')\rangle$. Define also the Gram matrix $\mathbf{K}(\mathbf{X},\mathbf{X}) \in \mathbb{R}^{N\times N}$ whose (n,m)-entry is $k(\mathbf{x}_n,\mathbf{x}_m)$, and row vector $\mathbf{k}(\mathbf{x},\mathbf{X}) := [k(\mathbf{x},\mathbf{x}_1),k(\mathbf{x},\mathbf{x}_2),\ldots,k(\mathbf{x},\mathbf{x}_N)]$. The desired discriminative kernel K-SVD algorithm iterates between the sparse coding and the dictionary/classifier learning steps. The sparse coding step computes K_0-sparse coefficients $\{\alpha_n\}$ given the most recent iterates of dictionary $\Phi(\mathbf{X})\Omega^{(\ell-1)}$ and classifier $\mathbf{W}^{(\ell-1)}$ using the OMP algorithm [18], where ℓ is the iteration index. The dictionary/classifier update step computes $\Omega^{(\ell)}$ and $\mathbf{W}^{(\ell)}$ jointly based on the eigendecomposition with the sparse codes $\mathbf{A}$ fixed.

The detailed steps are listed in Tables I and II. Table I is the OMP-based sparse coding algorithm. Table II is the overall discriminative kernel DL algorithm. In the Tables, $\mathbf{v}_S$ for vector $\mathbf{v}$ and index set S denotes a vector constructed from the entries in $\mathbf{v}$, whose indices are in S. Similarly, $\mathbf{M}|_S$ collects the columns in matrix $\mathbf{M}$, whose indices belong to S. Also, $\mathbf{a}_k^T$ denotes the k-th row of $\mathbf{A}$. It can be seen from the tabulated algorithm that computing the transformed features $\Phi(\mathbf{X})$ is not necessary; only their inner products appear thanks to the kernel trick.

TABLE I

SPARSE CODING ALGORITHM

`discr_kernel_OMP`$(\mathbf{x},\mathbf{y},\Omega,\mathbf{W})$

1: Initialize $I_0=\varnothing$, $\alpha^{(0)}=\mathbf{0}$
2: For $s=1,2,\ldots,K_0$
3: Compute for $k\in\{1,2,\ldots,K\}\backslash I_{s-1}$

$$\tau_k=\left(\mathbf{k}(\mathbf{x},\mathbf{X})-(\alpha^{(s-1)})^T\Omega^T\mathbf{K}(\mathbf{X},\mathbf{X})\right)\omega_k+\gamma(\mathbf{y}^T-(\alpha^{(s-1)})^T\mathbf{W}^T)\mathbf{w}_k$$

4: $k^*=\arg\max_{k\notin I_{s-1}}|\tau_k|$
5: $I_s=I_{s-1}\cup\{k^*\}$
6: $\alpha^{(s)}|_{I_s}=\left[\Omega|_{I_s}^T\mathbf{K}(\mathbf{X},\mathbf{X})\Omega|_{I_s}+\gamma\mathbf{W}|_{I_s}^T\mathbf{W}|_{I_s}\right]^{-1}\cdot\left[\Omega|_{I_s}^T\mathbf{k}(\mathbf{x},\mathbf{X})^T+\gamma\mathbf{W}|_{I_s}^T\mathbf{y}\right]$ and $\alpha^{(s)}|_{I_s^c}=\mathbf{0}$
7: Next s
8: Return $\alpha^{(K_0)}$

TABLE II

DISCRIMINATIVE KERNEL DL ALGORITHM

Input: $\mathbf{X},\mathbf{Y},K_0,\gamma$
Output: $\Omega,\mathbf{W}$

1: Initialize $\Omega^{(0)}$ and $\mathbf{W}^{(0)}$
2: For $\ell=1,2,\ldots$
% Sparse coding via discriminative kernel OMP
3: For $n=1,2,\ldots,N$
4: $\alpha_n=$ `discr_kernel_OMP`$(\mathbf{x}_n,\mathbf{y}_n,\Omega^{(\ell-1)},\mathbf{W}^{(\ell-1)})$
5: Next n
% Dictionary/classifier update via discriminative kernel K-SVD
% Note that $\mathbf{A}:=[\alpha_1,\ldots,\alpha_N]=[\mathbf{a}_1^T,\ldots,\mathbf{a}_K^T]^T=[\alpha_{kn}]$
6: For $k=1,2,\ldots,K$
7: $S_k=\{n\in\{1,2,\ldots,N\}:\alpha_{kn}\neq 0\}$
8: $\mathscr{E}_k^R=(\mathbf{I}-\sum_{j\neq k}\omega_j\mathbf{a}_j^T)|_{S_k}$
9: $\mathbf{E}_k^R=\sqrt{\gamma}(\mathbf{Y}-\sum_{j\neq k}\mathbf{w}_j\mathbf{a}_j^T)|_{S_k}$
10: Compute the largest eigenvalue σ_1^2 and the corresponding eigenvector $\mathbf{v}_1$ of $(\mathscr{E}_k^R)^T\mathbf{K}(\mathbf{X},\mathbf{X})\mathscr{E}_k^R+(\mathbf{E}_k^R)^T\mathbf{E}_k^R$
11: Set $\mathbf{a}_k|_{S_k}=|\sigma_1|\mathbf{v}_1$ and $\mathbf{a}_k|_{S_k^c}=\mathbf{0}$
12: $\omega_k=|\sigma_1|^{-1}\mathscr{E}_k^R\mathbf{v}_1$
13: $\mathbf{w}_k=\sqrt{\gamma}^{-1}|\sigma_1|^{-1}\mathbf{E}_k^R\mathbf{v}_1$
14: Next k
15: $\Omega^{(\ell)}=[\omega_1,\omega_2,\ldots,\omega_K]$, $\mathbf{W}^{(\ell)}:=[\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_K]$
16: Next ℓ
17: Let Λ be a diagonal matrix with diagonal entries equal to those in $\Omega^{(\infty)T}\mathbf{K}(\mathbf{X},\mathbf{X})\Omega^{(\infty)}$
18: Return $\Omega=\Omega^{\infty}\Lambda^{-\frac{1}{2}}$ and $\mathbf{W}=\mathbf{W}^{(\infty)}\Lambda^{-\frac{1}{2}}$

IV. TEST WITH HANDWRITTEN DIGIT IMAGES

The kernel DL algorithms were first tested using the USPS handwritten digit dataset [22]. The dataset contains $9,298$ images of handwritten digits $0,1,\ldots,9$ in 16×16 resolution. A subset of 100 images for each digit (totaling $N=1,000$ images) was randomly chosen to train dictionaries. We constructed ten dictionaries ($K=60$) and one dictionary ($K=600$) for the kernel DL and the discriminative kernel DL, respectively. The sparsity level K_0 was set to 5 and $\gamma=40$ was used. A polynomial kernel of degree 4 was employed. When tested over $4,000$ test samples, the discriminative kernel DL yielded slightly higher accuracy (94.3%) than the kernel DL (93.4%). Besides, to classify the digit images, the former needed to perform the kernel OMP once per sample while the latter did ten times.

V. TEST WITH REAL BRAIN TUMOR MR IMAGES

To evaluate the proposed methods for the tumor image segmentation task, the MR images of 20 patients with high-grade glioma retrieved from the BRATS2013 dataset were used. The dataset consists of multi-modal MR images including the FLAIR, T1-weighted (T1), T1-contrast (T1c), and T2-weighted (T2) MR images along with the ground truth images [8]. To assess the test performance, a 4-fold cross-validation was performed, that is, 20 MR image sets were split into 15 training image sets and 5 test image sets in four different combinations. The performance indices were averaged across the four experiments.

A feature vector was generated as the normalized intensity values corresponding to a $5\times5\times5$ cubic patch around a voxel. A Gaussian kernel $\exp(-\|\mathbf{x}-\mathbf{x}'\|_2^2/2\sigma^2)$ was adopted with $\sigma=1.5$. To fully exploit the multi-modal imagery, the kernel matrix was composed from the ensemble kernel as $\mathbf{K}=\mathbf{K_F}\odot\mathbf{K_{T1}}\odot\mathbf{K_{T1c}}\odot\mathbf{K_{T2}}$, where $\mathbf{K_F}$, $\mathbf{K_{T1}}$, $\mathbf{K_{T1c}}$, and $\mathbf{K_{T2}}$ represent the kernel matrices for the FLAIR, T1, T1c, and T2 images, respectively, and $\odot$ denotes the Hadamard product. For pre-processing, the image intensities were scaled so that the peak levels of the Flair and T2 histograms equal to 0.5 and 0.35, respectively, and the maximum downslope of the T1 and T1c histograms equal to 0.75 and 0.55, respectively. Any normalized pixel intensity exceeding 1.0 was re-normalized to 1.0.

To represent the training set compactly, 2,000 centroids of the feature vectors calculated by the k-means method for each class were used as inputs to the both kernel DL algorithms. For the reconstructive kernel DL, $N=2,000$, $K=1,000$, and $K_0=5$ were used to learn each of dictionaries $\mathbf{D}_{Tumor}$ and $\mathbf{D}_{Normal}$. For discriminative kernel DL, the labels for the tumor and normal tissue voxels were defined as $\mathbf{y}_{Tumor}=[1\ \ 0]^T$ and $\mathbf{y}_{Normal}=[0\ \ 1]^T$, and $N=4,000$, $K=1,000$, $K_0=5$, and $\gamma=1.0$ were used to learn a dictionary $\mathbf{D}_{Discr}$ and a classifier $\mathbf{W}$. Note that only one dictionary is involved in the discriminative approach.

To test the trained algorithms, 20,000 tumor and 20,000 normal voxels from the test set were randomly chosen. For each feature vector corresponding to a test voxel, the sparse code was computed using the kernel OMP algorithm [15]. For the reconstructive kernel DL, the voxel was classified to the class that yielded the smallest reconstruction error. For the discriminative kernel DL, the estimated label vector $\hat{\mathbf{y}}=[\hat{y}_1\ \hat{y}_2]^T$ was calculated by multiplying $\mathbf{W}$ to the sparse code. When $\hat{y}_1>\hat{y}_2$, the test sample was classified into the tumor class and, otherwise, into the normal class. Example segmentation results using both methods are depicted in Fig. 1. The performance indices are summarized in Table III. The formulas for computing the indices can be found in [23]. Compared to the latest studies [17], [24], [25], both approaches show comparable or better performances. Although the discriminative version did not outperform the reconstructive counterpart, it provided a comparable performance with lower computational burden in the testing phase. That is, it needed to perform the kernel OMP routine–the most time-consuming step–once per a test sample while the reconstructive kernel DL did it twice. On a separate note, significant performance degradation was observed when 5 specific MR image sets (HG0022, HG0024-HG0027) having some major bias field and blurring in T1 modality were assigned together as a test set. For the discriminative kernel DL, the false negative (FN) error caused by them correspond to 58.7% of the total FN error. This underlines the importance of preprocessing, which will be one of our future work items. Also, as can be deduced from Figs. 1(f) and 1(g), some of the misclassified voxels would be easily corrected by a post-processing step incorporating spatial smoothness.

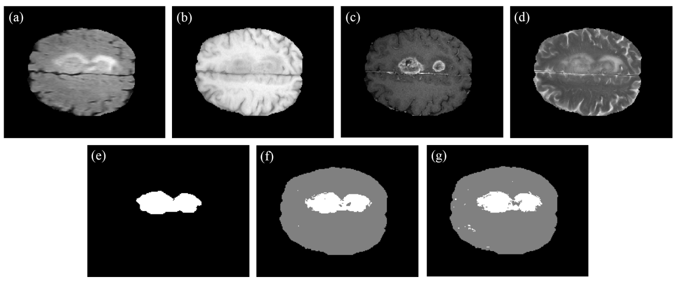

Fig. 1. An example of brain tumor segmentation (HG0007). (a) Flair (b) T1-weighted (c) T1-contrast (d) T2-weighted (e) ground truth (f) segmentation using reconstructive kernel DL (Jaccard = 81.5%) (g) segmentation using discriminative kernel DL (Jaccard = 83.0%)

TABLE III

SUMMARIZED PERFORMANCE (UNIT: %)

	Sensitivity	Specificity	Dice	Jaccard
Kernel DL	86.3	94.0	89.3	81.4
Discriminative kerel DL	86.2	92.5	88.6	80.2

VI. CONCLUSION

Novel automated brain tumor segmentation methods based on kernel DL have been proposed. Both reconstructive and discriminative kernel DL approaches were considered. Novel K-SVD-type algorithm for discriminative kernel DL was derived. To exploit the multi-modality of the MR imagery, an ensemble kernel based on Gaussian kernel functions was constructed. Thanks to the diversity manifested in multi-modal images, and nonlinear mapping of features through kernel techniques, the proposed methods showed very promising segmentation performances, which are competitive with or even beat the state-of-the-art alternatives. The discriminative kernel DL method was found to yield a performance almost the same as the reconstructive counterpart with reduced computational burden. The planned future work includes fine tuning of parameters including the use different kernel parameters for different modalities, as well as extending the proposed approach to multi-class segmentation of tumor regions.

REFERENCES

[1] M. Vaidyanathan, L. P. Clarke, L. O. Hall, C. Heidtman, R. Velthuizen, K. Gosche, S. Phuphanich, H. Wagner, H. Greenberg, and M. L. Silbiger, "Monitoring brain tumor response to therapy using MRI segmentation," *Magn. Reson. Imaging*, vol. 15, no. 3, pp. 323–334, 1997.

[2] G. P. Mazzara, R. P. Velthuizen, J. L. Perlman, H. M. Greenberg, and H. Wagner, "Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 59, no. 1, pp. 300–312, Apr. 2004.

[3] A. Kassner and R. E. Thornhill, "Texture analysis: A review of neurologic MR imaging applications," *Am. J. Neuroradiol.*, vol. 31, pp. 809–16, May 2010.

[4] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based bias field correction of MR images of the brain," *IEEE Trans. Med. Imaging*, vol. 18, pp. 885–896, Oct. 1999.

[5] M. Prastawa, E. Bullitt, S. Ho, and G. Gerig, "A brain tumor segmentation framework based on outlier detection," *Med. Image Anal.*, vol. 8, pp. 275–283, Sep. 2004.

[6] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *Proc. MICCAI*, Nice, France, 2012, pp. 369–376.

[7] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *Proc. MICCAI*, Toronto, Canada, 2011, pp. 354–361.

[8] B. Menze, J. Andras, B. Stefan, K. Jayashree, F. Keyvan, K. Justin, B. Yuliya, et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Trans. Med. Imaging*, 2014, *to appear*.

[9] D. Zikic, B. Glocker, E. Konukoglu, J. Shotton, A. Criminisi, D. Ye, C. Demiralp, O. M. Thomas, T. Das, R. Jena, and S. J. Price, "Context-sensitive classification forests for segmentation of brain tumor tissues," in *Proc. MICCAI-BRATS*, Nice, France, 2012.

[10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representation over learned dictionaries," *IEEE Trans. Image Proc.*, vol. 15, no. 12, pp. 3736–3645, Dec. 2006.

[11] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[12] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.

[13] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. of the 23rd IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)*, San Franscisco, CA, Jun. 2010, pp. 2691–2698.

[14] S. Gao, I. W-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Proc. European Conf. Comput. Vis. (ECCV)*, 2010.

[15] H. V. Nguyen, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Design of nonlinear kernel dictionaries for object recognition," *IEEE Trans. Image Proc.*, vol. 22, no. 12, pp. 5123–5135, Dec. 2013.

[16] T. Tong, R. Wolz, P. Coupé, J. V. Hajnal, D. Rueckert, and The Alzheimer's Disease Neuroimaging Initiative, "Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, Aug. 2013.

[17] J. J. Thiagarajan, K. N. Ramamurthy, D. Rajan and A. Spanias, "Kernel sparse models for automated tumor segmentation," *Intl. J. Artif. Intell. Tools*, vol. 23, no. 3, 18 pages, Jun. 2014.

[18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311-4322, Nov. 2006.

[20] S.-J. Kim, "Online kernel dictionary learning," *submitted to the 3rd IEEE GlobalSIP Conf.*, Orlando, FL, Dec. 2015.

[21] M. J. Gangeh, A. Ghodsi, and M. S. Kamel, "Kernelized supervised dictionary learning," *IEEE Trans. Sig. Proc.*, vol. 61, no. 19, Oct. 2013.

[22] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.

[23] V. Labatut and H. Cherifi, "Evaluation of Performance Measures for Classifiers Comparison," *Ubiquit. Comput. and Commun. J.*, vol. 6, pp. 21–34, Dec. 2011.

[24] S. D. S. Al-Shaikhli, M. Y. Yang, and B. Rosenhahn, "Coupled dictionary learning for automatic multi-label brain tumor segmentation in Flair MRI images," in *Advances in Visual Computing*, pp. 489–500. Springer, 2014.

[25] M. Huang, W. Yang, Y. Wu, J. Jiang, W. Chen, and Q. Feng, "Brain tumor segmentation based on local independent projection-based classification," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2633–2645, Oct. 2014.