

SHAPE-INCLUDED LABEL-CONSISTENT DISCRIMINATIVE DICTIONARY LEARNING: AN APPROACH TO DETECT AND SEGMENT MULTI-CLASS OBJECTS IN IMAGES

Mahdi Marsousi, Xingyu Li, Konstantinos N. Plataniotis

Multimedia Lab, The Edward S. Rogers Department of Electrical and Computer Engineering
University of Toronto, 10 King's College Road, Toronto, Canada

ABSTRACT

This paper introduces a segmentation approach, where a discriminative dictionary with objects' shape information is learned, followed by a sparse representation based segmentation process. In contrast with state-of-the-art sparse representation classification methods using discriminative dictionary learning, the proposed method learns a discriminative dictionary containing both intensity and shape information of object classes, in which shape information is collected and represented in the form of binarized masks. Object segmentation is achieved through an iterative process, including sparse representation, shape estimation, and shape refinement. The introduced method is evaluated and compared to state-of-the-art sparse representation based segmentation methods, and demonstrated better segmentation performance.

Index Terms— Object detection and Segmentation, Sparse Representation, Discriminative Dictionary Learning

1. INTRODUCTION

Recent years have witnessed the promising benefits of sparse representation in image processing and computer vision research. Based on an over-complete dictionary $D \in \mathbb{R}^{n \times K}$ which contains K atoms (columns), a signal $y \in \mathbb{R}^n$ is represented as a sparse linear combination of dictionary atoms. That is, $y \approx Dx$ under representation error constraint as, $\|y - Dx\|_2 \leq \epsilon$, or under sparsity constraint as, $\|x\|_0 \leq T$, where x is a sparse vector. Note, D in sparse representation can be constructed analytically from parametric functions such as wavelet basis, or it can be learned adaptively from data [1]. The later method to generate an over-complete dictionary, so-called dictionary learning, is preferred in many image processing applications, because it provides better sparsity level and lower representation error for specific data [1].

To build a data-dependent dictionary for sparse representation, early works used manually-selected signals as atoms, however they suffer from sub-optimality [2]. Later, methods for learning a compact dictionary from a set of training signals were proposed. In brief, given a set of n -dimensional N_p signals $P = \{p_1, \dots, p_{N_p}\}$, learning a reconstructive dictionary D with K basis can be achieved by solving

$$\langle D, X \rangle = \arg \min_{D, X} \|P - DX\|_F^2 \text{ s.t. } \forall i \|x_i\|_0 \leq T, \quad (1)$$

using the K-SVD method [3], where X is a sparse matrix for training data P . Since (1) focuses on signal representation only, there is no guarantee that the resulting D is discriminative for subsequent signal analysis. To address this problem, signal category information H and label-consistent constraint Q are introduced in the learning process, leading to a new dictionary learning algorithm which is formulated as [4, 5],

$$\langle D, X, A, W \rangle = \arg \min_{D, X, A, W} \|P - DX\|_F^2 + \alpha \|Q - AX\|_F^2 + \beta \|H - WX\|_F^2, \text{ s.t. } \forall i \|x_i\|_0 \leq T, \quad (2)$$

where the first term targets the representation error, and the second and third terms are added to minimize the classification error. This optimization problem, so-called the label consistent discriminative dictionary learning (LC-KSVD) method [4, 5], is solved using the K-SVD method.

With the progress of sparse representation and dictionary learning in image classification [2, 6], some efforts have been also made to combine these two techniques for object detection/segmentation. In the deformable segmentation study, an abstract shape dictionary, which is trained following (1) based on shape prior knowledge of the target objects, is used to refine a coarse segmentation in medical images [7]. To detect/segment human motion in videos, video background is assumed relatively stable and can be sparsely reconstructed by a background dictionary, while the representation residue is considered corresponding to human motion [8]. By casting the motion model into (1) for background dictionary construction, human motion is detected based on coding residue. To take advantage of LC-KSVD, in the study of MRI hippocampus labeling [9], image patches from training data are first used to generate atlas dictionary following (2). Then the center pixel of each query patch is assigned a label based on obtained sparse coefficients. In [10] that focuses on prostate segmentation in 3D MRI images, a shape dictionary and an appearance dictionary are learned individually via (2). By combining sparse representations obtained by the two dictionaries through ensemble learning, prostate pixels are detected. Recently, in the work of multi-organ segmentation in abdominal CT images, an abstract atlas is generated using LC-KSVD. After classifying a query image based on sparse coding coefficients, graph cut is applied to the resulting indexed image for segmentation refinement [11]. In the kidney

detection work which cascades a dictionary learning and the neural network (NN), sparse coefficients are used as features for NN for object detection from background[12].

It should be noted that though the discriminative dictionary learning improves sparse representation-based segmentation, it has a major limitation. As image pixels are classified individually based on patch representations, correlation between image pixels are lost, and hence, the resulting segmentation suffers from high discontinuity and low smoothness.

1.1. Contributions

This paper focuses on image multi-object segmentation scenario, and introduces a new method to detect and segment multiclass objects based on shape included label consistent discriminative dictionary learning (SI-LC-KSVD). Particularly, since shape prior information is a significant clue for segmentation, we include object's shape knowledge into discriminative dictionary learning, provide a complete dictionary update algorithm, and introduce a novel class-specific segmentation process, where information of classification and coarse shape masks obtained from sparse coding is combined/refined for final segmentation.

The rest of this paper is organized as follows. Section 2 presents the proposed SI-LC-KSVD based segmentation method. Experimental results and discussions are given in Section 3, followed by conclusions in Section 4.

2. METHODOLOGY

2.1. Segmentation Strategy

In this paper, a sparse representation based object detection/segmentation method is developed, where dynamically-learned texture and shape information of objects in different classes are used to segment objects in a query image. Briefly, the proposed method consists of two phases: *training* and *segmentation*. In the training phase, a dictionary learning process is performed to collect discriminative texture and shape information in dictionaries. Then the subsequent segmentation process uses the learned dictionaries to reconstruct a query image, and meanwhile to estimate and to refine objects shapes in a recursive manner. Details on the proposed method is provided in the rest of this section.

2.2. Image and Patch Domains

The dictionary learning process and sparse representation process in the proposed method work on image patch domain. For each image $I \in \mathbb{R}^{s_x \times s_y}$, let $I_M \in \mathbb{R}^{s_x \times s_y}$ be its binarized shape mask, where pixels of objects' boundaries are labeled by ones, and background pixels by zeros. Further, we define $P = [p_1, p_2, \dots, p_{N_p}] \in \mathbb{R}^{n_p \times N_p}$ and $M = [m_1, m_2, \dots, m_{N_p}] \in \mathbb{R}^{n_p \times N_p}$ as patches extracted from I and I_M respectively, where $n_p = (2r + 1)^2$ is the number of pixels in a square patch with radius r , and N_p is the number of patches in an image. Since patches are extracted every other κ pixels horizontally and vertically from

I and I_M , the number of patches along x - and y - axes in an image are $n_x = \text{floor}(\frac{s_x - 2r}{\kappa})$ and $n_y = \text{floor}(\frac{s_y - 2r}{\kappa})$, where $\text{floor}(\cdot)$ rounds a fractional number to its lower integer. Thus, $N_p = n_x \times n_y$. In this study, the operations of patch extraction and image reconstruction are defined as $\text{ext}(\cdot, \kappa, r)$ and $\text{rec}(\cdot, \kappa, r, s_x, s_y)$, respectively.

2.3. Shape Included Label-Consistent Discriminative Dictionary Learning (SI-LC-KSVD)

Recently, LC-KSVD has been widely used to classify image pixels for objects' detection and segmentation. As no prior information on objects' shapes is used in LC-KSVD, the obtained segmentation results suffer from high discontinuity and low smoothness, as discussed in Section 1. To address this short-coming, we introduce shape information of objects in forms of binarized shape masks in the proposed SI-LC-KSVD learning process. Let $D^p \in \mathbb{R}^{n_p \times K} = [d_1^p, \dots, d_K^p]$ and $D^m \in \mathbb{R}^{n_p \times K} = [d_1^m, \dots, d_K^m]$ represent dictionaries with K atoms learned from training image patches and their shape masks, respectively. Then the optimization problem of the proposed SI-LC-KSVD is defined as,

$$\begin{aligned} \langle D^p, D^m, A, W \rangle = \arg \min_{D^p, D^m, A, W} & \|P - D^p X\|_F^2 + \alpha \|Q - AX\|_F^2 \\ & + \beta \|H - WX\|_F^2 + \lambda \|M - D^m X\|_F^2 \text{ s.t. } \forall i \|x_i\|_0 \leq T, \end{aligned} \quad (3)$$

where $X \in \mathbb{R}^{K \times N_p} = [x_1, \dots, x_{N_p}]$ is the matrix of sparse coefficients, where each column vector x_i is associated with a patch, p_i . A and W are linear transformation matrix of label-consistency and linear classifier matrix, respectively. α , β , and λ are scalar weights, which control the contribution of the corresponding terms in the dictionary learning process. To solve (3) using the efficient K-SVD method, the optimization problem is re-written as follows,

$$\begin{aligned} \langle D^p, D^m, A, W \rangle = \arg \min_{D^p, D^m, A, W} & \left\| \begin{pmatrix} P \\ \sqrt{\alpha} Q \end{pmatrix} - \begin{pmatrix} D^p \\ \sqrt{\beta} H \end{pmatrix} X \right\|_F^2 \\ & + \lambda \|M - D^m X\|_F^2 \text{ s.t. } \forall i \|x_i\|_0 \leq T. \end{aligned} \quad (4)$$

In (4), D^p , A , and W are vertically concatenated to form a new matrix D^{new} , and P , Q , and H are vertically concatenated to form a new matrix of patches P^{new} , forming a LC-KSVD learner. However, we do not include D^m in the newly constructed D^{new} in (4) for following reason. Due to the large portion of zeros in D^m for background patches, the concatenation of P^{new} and D^m would result in an imbalanced dictionary after atom normalization, finally leading to imbalanced dictionary learning process.

To solve (4), the optimization problem is iteratively updated, in which each iteration consists of two steps including *sparse coding* and *dictionary atoms update*. Specifically, in each iteration, we first utilize the orthogonal matching pursuit (OMP) [13] to find sparse vectors x_i for $i \in [1, \dots, N_p]$. Based on the K-SVD algorithm, atoms of D^{new} are updated one-by-one. After an atom, d_n^{new} , is being updated, its corresponding sparse coefficients are used to train a corresponding

atom d_n^m in D^m . The optimization method of SI-LC-KSVD is represented in algorithm 1, where X_k^R is the k^{th} row of the sparse matrix, X , and E_k^p is the reconstruction error of patches, Y^{new} , without considering the k^{th} atom of D^p . E_k^m is the reconstruction error of masks, M , without considering the k^{th} atom of D^m . \tilde{E}_k^p and \tilde{E}_k^m are reduced matrices of E_k^p and E_k^m , where atoms corresponding to zero elements in X_k^R are removed. \tilde{X}_k^R is also obtained by removing zero coefficients in X_k^R . As shown in algorithm 1, d_k^m 's are updated using the efficient K-SVD method [14], as it has been shown more effective training for D^m compared to the original K-SVD method [3].

Algorithm 1: Optimization of SI-LC-DDS

Input: $P, H, Q, M, D_{(0)}^p, D_{(0)}^m, A_{(0)}, W_{(0)}, \alpha, \beta, \lambda, N_{Itr}$;

Output: D^p, D^m, A, W ;

begin

Set $P^{new} = [P^T, \sqrt{\alpha}Q^T, \sqrt{\beta}H^T]^T$,
 $D^{new} = [D_{(0)}^{mT}, \sqrt{\alpha}A_{(0)}^T, \sqrt{\beta}W_{(0)}^T]^T$;

for $iter \in 1, \dots, N_{Itr}$ **do**

Sparse Coding $X = OMP(P^{new}, D^{new})$;

for $k \in 1, \dots, K$ **do**

Calculate: $E_k^p = (Y - \sum_{j \neq k} d_j^{new} X_j^R)$;

Hold non-zero entries of X_j^R : \tilde{E}_k^p ;

Calculate: $USV^T = SVD(\tilde{E}_k^p)$;

Update: $d_k^{new} = U(:, 1)$, $\tilde{X}_k^R = \Sigma(1, 1)V(:, 1)$;

Calculate: $E_k^m = (M - \sum_{j \neq k} d_j^m X_j)$;

Hold non-zero entries of X_j^R : \tilde{E}_k^m ;

Update: $d_k^m = \tilde{E}_k^m \times \tilde{X}_k^R$;

2.4. Dictionaries Initialization

The proposed optimization method of SI-LC-DDS in Algorithm 1 needs initial matrices including $D_{(0)}^p, D_{(0)}^m, A_{(0)}$, and $W_{(0)}$. In LC-KSVD, the K-SVD approach is used for training dictionaries for each class, which are then combined to form the initial dictionary. However, this is not efficient in the sense that the number of atoms in all classes are selected equal, whereas dictionary sizes should be proportional to the structural complexity (texture and shape information) of each class. In this paper, we introduce a new scheme to provide initial dictionaries $D_{(0)}^p$ and $D_{(0)}^m$, using an adaptive size dictionary learning method (DLENE) [15], which trains efficient dictionaries for each class. Then, we calculate $A_{(0)}$, and $W_{(0)}$ based on the explained method in [4].

Assume for each class, $cl \in [1, \dots, N_C]$, a set of training images, $\{I^{cl,l}\}$, and their corresponding binarized masks, $\{I_M^{cl,l}\}$, exist. Then patches from $\{I^{cl,l}\}$ and $\{I_M^{cl,l}\}$ are extracted. We define $P^{cl} = \emptyset$, $M^{cl} = \emptyset$, and $P^{bg} = \emptyset$ as a set of image patches in the cl^{th} class, a set of mask patches in the cl^{th} class, and background patches. Note, only the patch, p_i , whose corresponding mask, m_i , has at least a portion of

non-zero cells, $(\sum_{j=1}^{n_p} m_{i,j}) > th \times n_p$, is added to P^{cl} , meanwhile its corresponding mask is added to M^{cl} . A patch with $(\sum_{j=1}^{n_p} m_{i,j}) = 0$ is added to P^{bg} . After collecting patches, they are used to learn dictionaries of background and classes.

For each class, P^{cl} and M^{cl} are vertically concatenated to form $Y^{cl} = [\{P^{cl}\}^T, \{M^{cl}\}^T]^T$, and the DLENE method is used to train dictionaries with efficient number of atoms as,

$$[\{D^{cl,p}\}^T, \{D^{cl,m}\}^T]^T = DLENE(Y^{cl}, ANNZC_{des}, RMSE_{des}), \quad (5)$$

$$D^{bg} = DLENE(P^{bg}, ANNZC_{des}, RMSE_{des}), \quad (6)$$

where $ANNZC_{des}$ and $RMSE_{des}$ are desired average number of non-zero coefficients and desired root-mean-square-error [15]. To match dictionary learning of initialization and SI-LC-DDS processes, we set $ANNZC_{des} = T$. After finishing the process of learning the dictionaries, the initial dictionaries are formed as, $D_{(0)}^p = [D^{bg}, D^{1,p}, \dots, D^{N_C,p}]$ and $D_{(0)}^m = [\emptyset, D^{1,m}, \dots, D^{N_C,m}]$. Based on the sizes of learnt dictionaries of background and classes, the matrices Q and H are created as instructed in [4]. Finally, $A_{(0)}$ and $W_{(0)}$ are obtained as explained in [4, 16].

2.5. Object Detection and Segmentation

The learnt dictionaries are used to automatically detect and segment objects in input images. Given a query image, $I^{in} \in \mathbb{R}^{s_x \times s_y}$, patches are extracted, $P = [p_1, \dots, p_{N_p}] = ext(I^{in}, \kappa_{seg}, r)$, where κ_{seg} is a step size for segmentation process, and r is the same as r of the dictionary learning process. Since the intensity information is only available for a query image, sparse vectors are initially calculated based on image construction only,

$$X^e = arg \min_X \|P - D^p X\|_F \text{ s.t. } \|x_i\|_0 < T \quad \forall i, \quad (7)$$

where $X^e = [x_1, \dots, x_{N_p}]$ is the matrix of sparse vectors. Then, $H^e = W X^e$ and $M^e = D^m X^e$ are calculated, representing rough estimations of H and M of the input image, I^{in} . Each row of $H^e = [\{h_b^R g\}^T, \{h_1^R\}^T, \dots, \{h_{N_C}^R\}^T]^T$, represents the membership of patches to a class, where $h_{cl}^R \in \mathbb{R}^{1, N_p}$. We define $H_{cl}^e = v_{cl} h_{cl}^R$, where $v_{cl} \in \mathbb{R}^{n_p}$ is a vector with zero cells, except a single one at the location $(2r+2)r+1$, which corresponds to the center point of a patch of radius, r . Now, we transform data from the patch domain into the image domain by reconstructing $I^{M^e} = rec(M^e, \kappa_{seg}, r, s_x, s_y)$, and $I^{H,cl} = rec(H_{cl}^e, \kappa_{seg}, r, s_x, s_y)$. For each class, we calculate $I^{M^e,cl} = I^{M^e} \times I^{H,cl}$, where \times is the pixel-wised multiplying operator. $I^{M^e,cl}$ provides estimated segmentation of objects of the cl^{th} class in the input image. The process of generating $I^{M^e,cl}$ is called Class-Specific Segmentation.

Because $I^{M^e,cl}$ is obtained by intersecting mask and class information, it contains wealthier information compared to I^{M^e} and $I^{H,cl}$ s. We extract patches from $I^{M^e,cl}$ as $M^{rd,cl} = ext(I^{M^e,cl}, \kappa_{seg}, r) = [m_{cl}^1, \dots, m_{cl}^{N_p}]$, and then, we generate refined classes of patches as, $h_{cl}^{rd} = [h_{cl,1}, h_{cl,2}, \dots, h_{cl,N_p}]$, where $h_{cl,i} = \frac{(\sum_{j=1}^{n_p} m_{i,j}^{cl})}{n_p}$. Now,

H^{rd} is generated by vertically concatenating h_{cl}^{rd} s, as $H^{rd} = [\{h_b^R g\}^T, \{h_1^d\}^T, \dots, \{h_{N_C}^d\}^T]^T$. In this paper, the process of generating H^{rd} from $I^{M^e, cl}$ s is called Mask to Class Translation. Now, we feed H^{rd} into the sparse coding as,

$$X^e = \arg \min_X \|P' - D'X\|_F \text{ s.t. } \|x_i\|_0 < T \quad \forall i, \quad (8)$$

where $P' = [\{P\}^T, \{\sqrt{\beta}H^{rd}\}^T]^T$ and $D' = [\{D^p\}^T, \{\sqrt{\beta}W\}^T]^T$, and each column of D' is normalized. Then, $H^e = WX^e$ and $M^e = D^m X^e$ are calculated again, and the Class-Specific Segmentation process is repeated to find $I^{M^e, cl}$ s.

3. EXPERIMENTS AND DISCUSSIONS

To evaluate the performance of the proposed algorithm for object detection and segmentation, two classes of objects, including cars and motorbikes, from the Caltech-101 database [17], are used as target segmentation objects. We randomly selected 40 images of cars and 40 images of motorbikes to form our database. For each image, ground truth data are manually generated. For the proposed method of this paper, we set $r = 15$, $\kappa = 4$, $\kappa_{seg} = 2$, $\alpha = 10$, $\beta = 0.1$, and $N_{It} = 30$. For the initialization step, the DLENE's parameters are set as $ANNZC_{des} = 6$ and $RMSE_{des} = 0.01$.

Comparative experimentation is developed, where the proposed algorithm and the state-of-the-art methods, including D-KSVD [16] and LC-KSVD [4], are examined. Specifically, for each examined method, we perform three trainings (i.e. dictionary learning processes) as follows: (1) for car images only, (2) for motorbike images only, and (3) for both car and motorbike images. Thus, the first and second trainings generate two binary-class dictionaries, each corresponding to background and one object class, and the third training involves three classes including background and two object classes. For the first and second trainings, 20 images are randomly selected as training set and the rest of images are used as evaluation set. The segmentation accuracy of each method is evaluated in the first two binary-class cases. For the third training, 20 car images and 20 motorbike images are randomly selected as training set, and the rest of images form evaluation set. In this scenario, detection and segmentation accuracies of cars and motorbike classes for each method are measured. We use accuracy measure, ACC_{sg} , to calculate segmentation accuracy of methods[18]. To evaluate detection performance, we define N_{cr} and N_{mb} as the number of correct detection of cars and motorbikes, respectively. Then, detection accuracy is defined as $ACC_{dt} = \frac{N_{cd}}{N_T}$, where $N_{cd} = N_{cr} + N_{mb}$ and $N_T = 40$ are the number of correct detections and total number of evaluation images, respectively.

Table 1 demonstrates the detection accuracy of D-KSVD, LC-KSVD, and SI-LC-KSVD. The LC-KSVD leads the detection accuracy by 5%, compared to the proposed method. Table 2 shows the segmentation accuracy of the methods. Accordingly, SI-LC-KSVD provides higher segmentation accuracy for the three trainings. Figure 1 shows two examples of segmenting objects using the proposed method of this paper.

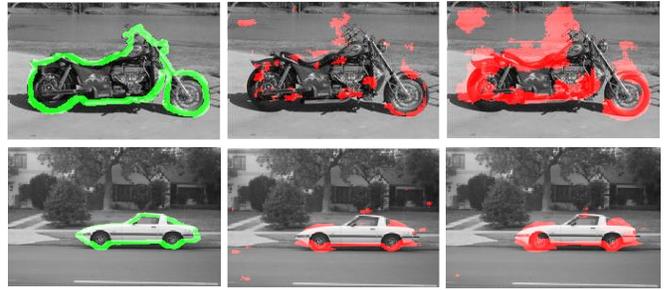


Fig. 1. Examples of segmentation results of a motorbike and a car. The left, middle, and right columns correspond to ground-truth, segmentation output at first iteration, and segmentation output at second iteration, respectively.

Table 1. Comparing detection accuracy of the proposed method with state-of-the-art.

	N_{cr}	N_{mb}	ACC
D-KSVD	19	6	0.63
LC-KSVD	18	14	0.80
SI-LC-KSVD	13	17	0.75

Table 2. Comparing segmentation accuracy, ACC_{sg} , of the proposed method with state-of-the-art for three cases: (a) only cars (Car), (b) only motorbikes (MotorB), and (c) both cars and motorbikes (Car&MotorB). Results are shown as $\mu \pm \sigma$ where μ and σ are mean and standard deviation, respectively.

	Car	MotorB	Car&MotorB
D-KSVD	0.859 ± 0.102	0.696 ± 0.082	0.814 ± 0.112
LC-KSVD	0.824 ± 0.145	0.728 ± 0.068	0.801 ± 0.103
SI-LC-KSVD	0.928 ± 0.025	0.816 ± 0.039	0.820 ± 0.138

In Figure 1, the green boundary in the left sub-figure shows the ground-truth, and the red boundary in middle and right sub-figures shows the automated segmentation result. Figure 1 shows the improved segmentation performance in the second iteration (the right sub-figure), compared to the segmentation performance of the first iteration (the middle figure).

4. CONCLUSIONS

In this paper, we introduced a new segmentation approach based on discriminative dictionary learning in sparse representation. By including prior shape information of target objects in the form of binarized masks in the proposed SI-LC-KSVD learning, the resulting dictionary facilitated subsequent object detection and segmentation. We also introduced a novel class-specific segmentation process, where estimated shapes were fed back into sparse coding recursively to refine object's detection/segmentation. In evaluation, the proposed segmentation method was compared to prior arts by detecting and segmenting two types of objects in images. Experimentation results suggested superiority of the proposed approach in object segmentation. As the selection of training patches in dictionary learning and proper parameter settings could potentially affect detection and segmentation performances, we will discuss these issues in our future works.

5. REFERENCES

- [1] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *IEEE Proceedings*, vol. 98, no. 6, pp. 1045–1057, 2010.
- [2] J. Weight, A. Yang, A. Genush, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. 31, no. 2, pp. 210 – 227, 2009.
- [3] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Process.*, vol. 54, no. 11, pp. 4311 – 4322, 2006.
- [4] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent k-svd: learning a discriminative dictionary for recognition," *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. 35, no. 11, pp. 2651 – 2664, 2013.
- [5] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," in *IEEE Conf. on Computer Vis. and Patt. Recog. (CVPR)*, 2011, pp. 1697–1704.
- [6] P. Du, Z. Xue, J. Li, and A. Plaza, "Learning discriminative sparse representations for hyperspectral image classification," *IEEE J. of Selected Topics in Signal Process.*, vol. 9, no. 6, pp. 1089–1104, 2015.
- [7] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Med. Image Anal.*, vol. 16, no. 7, pp. 1385 – 1396, 2012.
- [8] Y. Wang, X. Wang, C. Can, and C. Wan, "Image segmentation by sparse representation," in *IEEE Int. Conf. on Audio, Lang. and Image Process. (ICALIP)*, 2012, pp. 365 – 369.
- [9] T. Tong, R. Wolz, P. Coupe, J. V. Hajnal, and D. Rueckert, "Segmentation of mr images via discriminative dictionary learning and sparse coding: application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11 – 23, 2013.
- [10] Y. Guo, Y. Gao, Y. Shao, T. Price, A. Oto, and D. Shen, "Deformable segmentation of 3d mr prostate images via distributed discriminative dictionary and ensemble learning," *Med. phys.*, vol. 41, no. 7, pp. 072303, 2014.
- [11] T. Tong, R. Wolz, Z. Wang, Q. Gao, K. Misawa, M. Fujiwara, K. Mori, J. V. Hajnal, and D. Rueckert, "Discriminative dictionary learning for abdominal multi-organ segmentation," *Med. Image Anal.*, vol. 23, no. 1, pp. 92–104, 2015.
- [12] M. Marsousi and Konstantinos N. Plataniotis, "Binomial classification based on dlene features in sparse representation: Application in kidney detection in 3d ultrasound," in *IEEE Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 1017 – 1021.
- [13] Y. C. Pati, R. Rezaifar, and P. S. Krishnapras, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *IEEE Conf. In Signals, Sys. and Computers*, 1993, pp. 40–44.
- [14] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *CS Technion*, vol. 40, no. 8, pp. 1–15, 2008.
- [15] M. Marsousi, K. Abhari, P. Babyn, and J. Alirezaie, "An adaptive approach to learn overcomplete dictionaries with efficient numbers of elements," *IEEE Trans. on Signal Process.*, vol. 62, no. 12, pp. 3272–3283, 2014.
- [16] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," in *IEEE Conf. on Computer Vis. and Patt. Recog. (CVPR)*, 2010, pp. 2691–2698.
- [17] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vis. and Image Understand.*, vol. 106, no. 1, pp. 59–70, 2007.
- [18] M. Ganz, X. Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *IEEE Trans. Bio. Eng.*, vol. 59, no. 8, pp. 2144–2151, 2012.