Cervigram Image Segmentation Based On Reconstructive Sparse Representations

Shaoting Zhang[1], Junzhou Huang[1], Wei Wang[2], Xiaolei Huang[2], Dimitris Metaxas[1]

[1]CBIM, Rutgers University
110 Frelinghuysen Road Piscataway, NJ 08854, USA
[2]Department of Computer Science and Engineering
Lehigh University, Bethlehem, PA 18015, USA

ABSTRACT

We proposed an approach based on reconstructive sparse representations to segment tissues in optical images of the uterine cervix. Because of large variations in image appearance caused by the changing of the illumination and specular reflection, the color and texture features in optical images often overlap with each other and are not linearly separable. By leveraging sparse representations the data can be transformed to higher dimensions with sparse constraints and become more separated. K-SVD algorithm is employed to find sparse representations and corresponding dictionaries. The data can be reconstructed from its sparse representations and positive and/or negative dictionaries. Classification can be achieved based on comparing the reconstructive errors. In the experiments we applied our method to automatically segment the biomarker AcetoWhite (AW) regions in an archive of 60,000 images of the uterine cervix. Compared with other general methods, our approach showed lower space and time complexity and higher sensitivity.

Keywords: segmentation, cervix image, biomarker AcetoWhite, reconstructive errors, sparse representation, K-SVD, OMP

1. INTRODUCTION

Segmentation of different regions of medical images can assist doctors to analyze them. Area information from segmentation is important in many clinical cases. In this work, we propose an approach to automatically segment the biomarker AcetoWhite (AW) regions in an archive of 60,000 images of the uterine cervix. These images are optical cervigram images acquired by Cervicography using specially-designed cameras for visual screening of the cervix (Figure 1). They were collected from the NCI Guanacaste project[1] for the study of visual features correlated to the development of precancerous lesions. The most important observation in a cervigram image is the AW region, which is caused by whitening of potentially malignant regions of the cervix epitheliuem, following application of acetic acid to the cervix surface. Since the texture, size and location of AW regions have been shown to correlate with the pathologic grade of disease severity, accurate identification and segmentation of AW regions in cervigrams have significant implications for diagnosis and grading of cervical lesions. However, accurate tissue segmentation in cervigrams is a challenging problem because of large variations in the image appearance caused by the changing of illumination and specular reflection in pathology. As a result, the color and texture features in optical images often overlap with each other and are not linearly separable when training samples are larger than a certain level (Figure 2).

1.1 Related work

Previous work on cervigram segmentation has reported limited success using K-means clustering,[2] Gaussian Mixture Models,[3] Support Vector Machine (SVM) classifiers.[4] Shape priors are also proposed.[5] Although such priors are applicable to cervix boundary, it does not work well with AW since AW regions may have arbitrary shapes. Supervised learning based segmentation,[6,7] holds promise, especially with increasing number of features. However, because of the intrinsic diversity between images and the overlap between feature distributions of

{shaoting, jzhuang, dnm}@cs.rutgers.edu, {wew305, xih206}@lehigh.edu

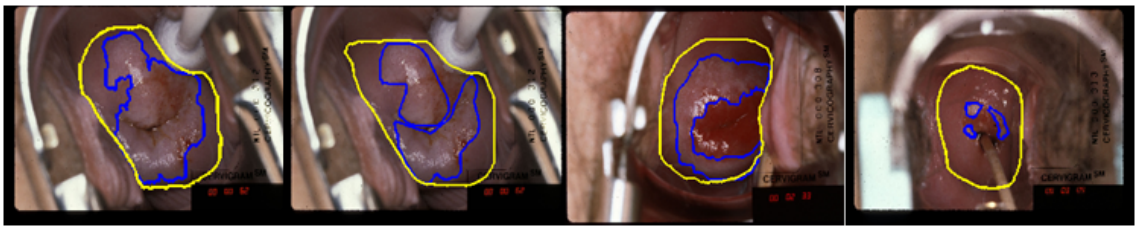


Figure 1. Examples of digitized cervicographic images (i.e. cervigrams) of the uterine cervix, created by the National Library of Medicine (NLM) and the National Cancer Institute (NCI). Ground truth boundary markings by 20 medical experts. Our work is aimed for automatically segmenting the biomarker AcetoWhite (AW) regions, which indicates clinical significance.

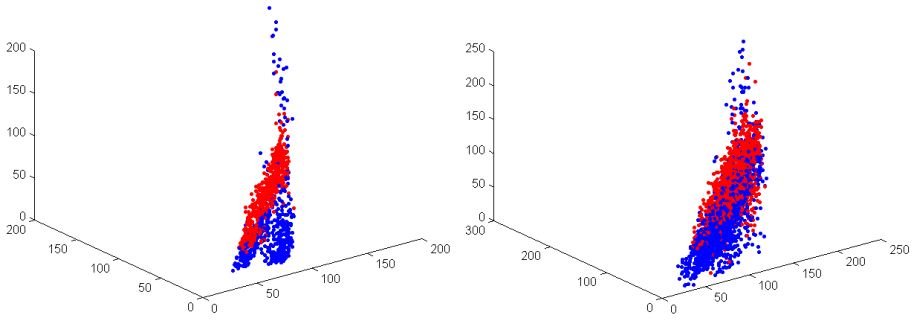


Figure 2. Color distribution of AW and non-AW regions. Red means AW and blue means non-AW. The left one comes from one sample. The right one comes from one hundred samples.

different classes, it is difficult to learn a single classifier that can perform tissue classification with low error rate for a large image set. Our empirical evaluation shows that overfitting is a serious problem when training a single (SVM) classifier using all training samples the average sensitivity of the classifier is relatively low. A potential solution is to use a Multiple Classifier System (MCS),[8] which trains a set of diverse classifiers that disagree on their predictions and effectively combines the predictions in order to reduce classification error. Voting, AdaBoost, bagging and STAPLE[9] can be employed. A necessary condition for the above ensemble methods is that all the base classifiers should provide sufficiently good performance, usually 50% or higher sensitivity and specificity in order to support the ensemble. However, there may be large variance in base classifier performance in our case. Some classifiers commonly have lower sensitivity than 50%. Wang and Huang[10] proposed a method to find the best base classifier based on distance guided selection, which achieves state-of-the-art results in a subset of the archive.

In our method we focus on finding a single classifier by transforming the data to a higher dimension with sparse constraints. Then the data can be more separated using sparse representations. This classifier is potentially useful for MCS since the sensitivity and specificity are always larger than 50% in our experiments. Finding the sparse representation typically consists of the sparse coding and codebook update. Greedy algorithms such as matching pursuit (MP)[11] and orthogonal matching pursuit (OMP)[12] can be employed for finding sparse coefficients (coding). Extensive study of these algorithms shows that if the sought solution is sparse enough, these greedy techniques can obtain the optimal solution.[13] When the sparse data has group clustering trend,[14] AdaDGS[15] can be employed to further improve the performance. To update codebook, method of optimal direction (MOD)[16] and K-SVD[17] are two effective approaches. Although both of them result in similar results, we prefer K-SVD because of its better convergence rate. After finding positive and negative dictionaries from training images and computing sparse coefficients from testing images, reconstructive errors can be obtained and compared. The pixel can be assigned to the class with lower errors. Our main contributions are the following: 1) we introduce the reconstructive sparse representations and K-SVD algorithms to the medical imaging community, which are originated from the compressive sensing field; 2) we apply this theory to solve the challenging cervigram image segmentation problem and achieve improved performance. Details of this method is explained in Section

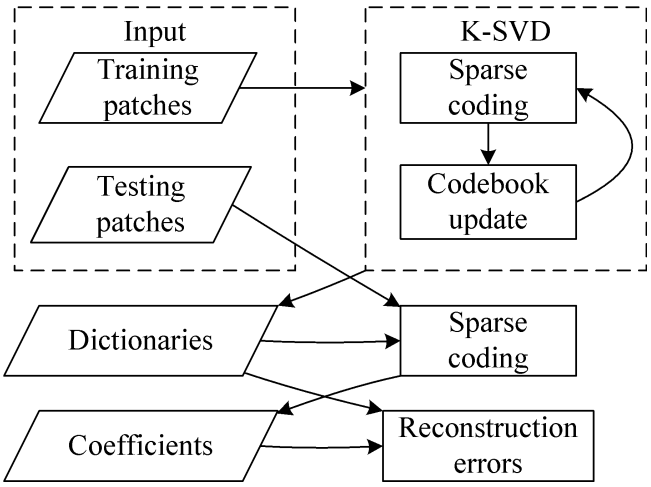


Figure 3. The algorithm framework. The left column represents the data. The right column represents algorithms including K-SVD.

2 and experiments are shown in Section 3.

2. METHODOLOGY

2.1 Framework

Figure 3 illustrates the algorithm framework. In the training stage, ground truth is manually obtained by clinical experts. Patches on the ground truth are labeled as positive ones, while the others are negative ones. These patches are fed into K-SVD to generate positive and negative dictionaries. Then using these two dictionaries, the sparse coding step is applied on patches extracted from testing images to compute two sets of sparse coefficients. From the coefficients and corresponding dictionaries, reconstructive errors are calculated and compared for classification. An alternative way is to classify the sparse coefficients (using SVM) instead of classifying the original data, which is also tested in Section 3. Details of the reconstructive sparse representation are discussed in Section 2.2 and Section 2.3.

2.2 Learning sparse dictionaries

Reconstructive sparse representation is used here to classify image patches. The objective of sparse representation is to find D and X by minimizing the following equation:

$$\min_{D,X}\{\|Y - DX\|_F^2\} \; subject\ to\ \forall i, \|x_i\|_0 \leq L \tag{1}$$

Where Y represents signals (image patches here), D is the overcomplete dictionary, X is the sparse coefficients, $\|\cdot\|_0$ is the l^0 norm counting the nonzero entries of a vector, $\|\cdot\|_F$ is the Frobenius norm. Denote y_i as the ith column of Y, x_i as the ith column of X, then y_i and x_i are the ith signal vector and coefficient vector respectively, with dimensionality $D \in \mathbb{R}^{n\times k}$, $y_i \in \mathbb{R}^n$ and $x_i \in \mathbb{R}^k$.

K-SVD algorithm starts from a random D and X obtained from the sparse coding stage. The sparse coding stage is based on pursuit algorithms to find the sparse coefficient x_i for each signal y_i. OMP is employed in this stage. OMP is an iterative greedy algorithm that selects at each step the dictionary element that best correlates with the residual part of the signal. Then it produces a new approximation by projecting the signal onto those elements already selected.[13] The algorithm framework of OMP is listed in the Table 1.

Table 1. The framework of the OMP algorithm.

Input: dictionary $D \in \mathbb{R}^{n \times k}$, input data $y_i \in \mathbb{R}^n$
Output: coefficients $x_i \in \mathbb{R}^k$
$\Gamma = \emptyset$
Loop: iter=$1, ..., L$

- Select the atom which most reduces the objective

$$\arg\min_j \left\{ \min_{x'} |y_i - D_{\Gamma \cup \{j\}} x'|_2^2 \right\} \tag{2}$$

- Update the active set: $\Gamma \leftarrow \Gamma \cup \{j\}$
- Update the residual using orthogonal projection

$$r \leftarrow \left(I - D_\Gamma (D_\Gamma^T D_\Gamma)^{-1} D_\Gamma^T\right) y_i \tag{3}$$

- Update the coefficients

$$x_\Gamma = (D_\Gamma^T D_\Gamma)^{-1} D_\Gamma^T y_i \tag{4}$$

In the codebook update stage K-SVD employs a similar approach as K-Means to update D and X iteratively. In each iteration D and X are fixed except only one column d_i and the coefficients corresponding to d_i (ith row in X), denoted as x_T^i. The Equation 1 can be rewritten as

$$\left\| Y - \sum_{j=1}^{k} d_j x_T^j \right\|_F^2 = \left\| \left(Y - \sum_{j \neq i} d_j x_T^j \right) - d_i x_T^i \right\|_F^2 \tag{5}$$

$$= \left\| E_i - d_i x_T^i \right\|_F^2 \tag{6}$$

We need to minimize the difference between E_i and $d_i x_T^i$ with fixed E_i, by finding alternative d_i and x_T^i. Since SVD finds the closest rank-1 matrix that approximates E_i, it can be used to minimize the Equation 5. Assume $E_i = U \Sigma V^T$, d_i is updated as the first column of U, which is the eigenvector corresponding to the largest eigenvalue. x_T^i is updated as the first column of V multiplied by $\Sigma(1,1)$.

However, the updated x_T^i may not be sparse anymore. The solution is logical and easy. We just discard the zero entries corresponding to the old x_T^i. The detail algorithms of K-SVD are listed in the table 2.

2.3 Reconstructive errors

Using K-SVD algorithm we can obtain two dictionaries for positive patches and negative patches separately, denoted as D_+ and D_- respectively. The simplest strategy to use dictionaries for discrimination is to compare the errors of a new patch y reconstructed by D_+ and D_- and choose the smaller one as its type, as shown in equation 9.

$$type = \underset{i=+,-}{arg\ min} \{\|y - D_i x\|_2^2\}\ subject\ to\ \|x\|_0 \leq L \tag{9}$$

The potential problem of this method is that the dictionaries are trained separately. That is to say, positive/negative dictionary only depends on positive/negative patches, so it attempts to reconstruct better for

Table 2. The framework of the K-SVD algorithm.

Input: dictionary $D \in \mathbb{R}^{n \times k}$, input data $y_i \in \mathbb{R}^n$ and coefficients $x_i \in \mathbb{R}^k$
Output: D and X
Loop: Repeat until convergence

- *Sparse coding:*
  use OMP to compute coefficient x_i for each signal y_i, to minimize

$$\min_{x_i} \{\|y_i - Dx_i\|_2^2\} \; subject \; to \; \|x_i\|_0 \leq L \tag{7}$$

- *Codebook update:*
  for $i = 1, 2, ..., k$, update each column d_i in D and also x_T^i (ith row)
  - Find the group using d_i ($x_T^i \neq 0$), denoted as ω_i
  - Compute error matrix E_i as in equation 5
  - Restrict E_i by choosing columns corresponding to ω_i. The resized error is denoted as E_i^R
  - Apply SVD and obtain

$$E_i^R = U \Sigma V^T \tag{8}$$

    Update d_i as the first column of U. Update nonzero elements in x_T^i as the first column of V multiplied by $\Sigma(1,1)$

positive/negative patches but not worse for negative/positive patches. Discriminative methods can be considered to alleviate this problem.[18] However, the tuning parameters of the discriminative system are very sensitive and it can only converge within small intervals. In our case, reconstructive method works relatively well. Discriminative method with this application is left to future investigation.

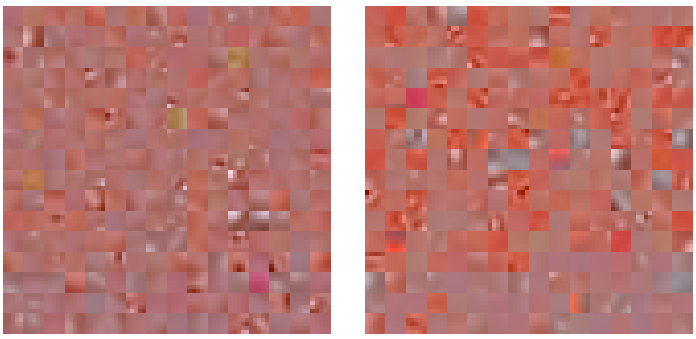
An alternative way is to classify the sparse coefficients x instead of y. x from training images can be fed into SVM or other classifier. The intuition is that x is in higher dimension with sparse constants and can be more separated. Both of these two approaches are tested in Section 3.

2.4 Tracing regions

Since there is no shape information considered, the resulting areas are usually disconnected. Inspired by the edge linking stage of Canny edge detector, similar procedure can also be applied on this application. Equation 9 can be rewritten as:

$$error = \|y - D_- x\|_2^2 - \|y - D_+ x\|_2^2 \tag{10}$$

When $error < 0$, the testing data is assigned to the negative samples. Otherwise it is positive. However, due to noise, there may be positive instances below the threshold (0). Thus similar to the Canny edge detector, two thresholds T_1 and T_2 ($T_1 > T_2$) can be predefined. In the first pass, $T_1 = 0$ is used as the threshold and classification is performed. This procedure is the same as Section 2.2. In the second pass, $T_2 < 0$ is set as the new threshold. The *error*s of neighboring points of the first results are checked, and the points with $error > T_2$ are merged into the positive samples. With ideal thresholds, the disconnectivity problem can be alleviated in a certain level. However, the value of T_2 highly depends on the application and currently is found by cross validation and brute force. Starting from 0, T_2 is decreased by a small step each time. The sensitivity and specificity are computed in each step. The parameters causing the best performance are chosen. More sophisticated approaches are left for future investigations.



(a) AW dictionaries (b) non-AW dictionaries

Figure 4. positive (AW) and negative (non-AW) dictionaries trained from K-SVD. Each dictionary is displayed as 16 by 16 cells (256 cells). Each cell comes from the column of the dictionary and is reshaped as a 5 by 5 patch. Some patches from different dictionaries are quite similar. Note that these patches do not directly represent image patches, since the columns are normalized in HSV color space.

Table 3. Performance comparison between 4 classifiers, measured by the mean of sensitivity and specificity.

	Sensitivity	**Specificity**
SVM with image patches	50.24%	63.59%
Nearest Neighbor	55.62%	69.18%
Compare reconstructive errors	62.71%	75.85%
SVM with sparse coefficients	61.46%	76.37%

2.5 Implementation details

Cervigram images from the NCI/NLM archive with multiple-expert boundary markings are available for training and validation purposes. 100 images of diverse appearance were selected for training and testing. To maximally mix the samples, 10 image is used for testing and validation and the remaining 90 ones are used for training. The mean sensitivity and specificity are reported. Different color spaces including RGB, HSV and Lab are tested. HSV is chosen since it is slightly better. Other color spaces, texture and appearance information can also be considered. Each patch is a 5 by 5 square centered in the pixel and concatenated H,S,V information into single vectors (75 by 1, $n = 75$). We choose the sparse factor $L = 6$ and dictionaries of size $k = 256$. Although there are many choices for these values, they are not arbitrary. They need to satisfy the constraints mentioned in[19] to guarantee convergence. In each image 1,000 patches are randomly selected from both AW and non-AW regions, 500 for each. Overall 90,000 patches are generated from the training images. 60 iterations of K-SVD are performed. The positive and negative dictionaries represent AW and non-AW regions respectively. Each column of dictionaries are reshaped as 5 by 5 patches, and they are displayed together in Figure 4. Some patches from different dictionaries are similar, proving that classification is a challenging task. Utilizing sparse representations can alleviate this problem.

3. RESULTS

The method was implemented in Matlab R2009a and tested on a 2.40 GHz Intel Core2 Quad computer with 8G RAM. It was compared with SVM and k nearest neighbors. SVM failed to handle 90,000 patches since it would consume most memories and couldn't converge. Thus the data for SVM was down sampled. Instead of feeding image patches into SVM, we also trained SVM using sparse coefficients. Nearest neighbor method was also time and space consuming because of the large training set. K-SVD was more efficient with 5 seconds for each iteration and less than 1GB RAM because of its sparsity.

Table 3 shows the results of different classifiers measured by sensitivity and specificity. Figure 5 visualizes the segmentation results of a specific image. Since the distributions of image patches were overlapped (Figure 2),

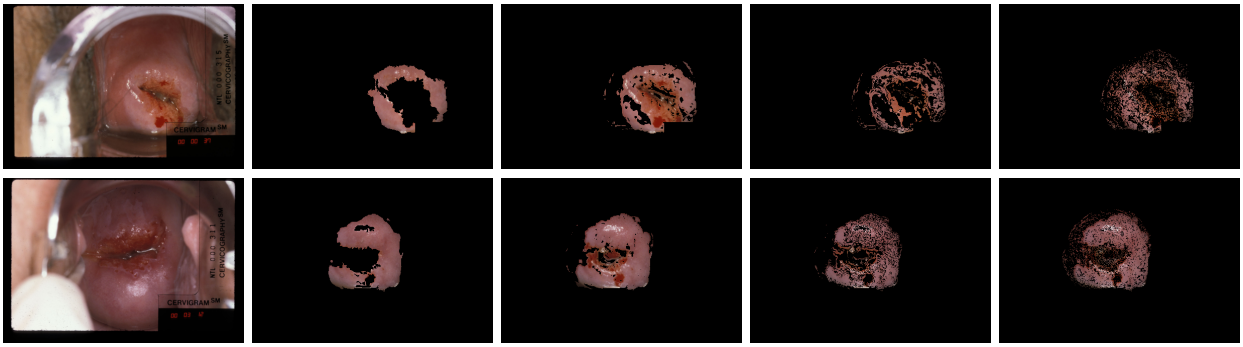


Figure 5. Results of a specific image. From left to right: original image; ground truth; SVM with image patches; nearest neighbor; reconstructive errors (equation 9).

SVM with image patches underperformed the classification task. Using sparse coefficients in higher dimension, SVM performed better since the data was more separated. Although comparing reconstructive errors also achieved good results, it still had many noises from non-AW regions. The reason is that the K-SVD algorithm is not discriminative and there is no shape information considered.

From the comparisons in Table 3 and Figure 5 we can find that sparse representations are very useful in this challenging problem.

4. CONCLUSIONS

In this paper we proposed classifiers based on reconstructive sparse representations to segment tissues in optical images of the uterine cervix. Our method was compared with some typical learning methods and worked better in this challenging case. In the future we would like to put efforts on three directions. 1) We will develop discriminative sparse representations instead of reconstructive ones. The dictionaries should perform better for its own class and also worse for other classes. Current discriminative methods highly depend on the parameters and are not stable. It can be further improved. 2) We will add our method into "best base classifier"[10] to improve the sensitivity and specificity of the whole system. 3) Combining sparse representations with SVM is also interesting.

5. ACKNOWLEDGEMENTS

The authors would like to thank the Communications Engineering Branch, National Library of Medicine-NIH, and the Hormonal and Reproductive Epidemiology Branch, National Cancer Institute-NIH, for providing the data and support of this work. The authors are thankful to Hongsheng Li (Lehigh), Tian Shen (Lehigh), Xiaoxu Wang (Rutgers), Ming Jin (Rutgers), Yuchi Huang (Rutgers) and Yang Yu (Rutgers) for stimulating discussions.

REFERENCES

[1] Jeronimo, J., Long, L., Neve, L., Bopf, M., Antani, S., and Schiffman, M., "Digital tools for collecting data from cervigrams for research and training in colposcopy," in [*Colposcopy, J Low Gen Tract Disease*], 16–25 (2006).

[2] Tulpule, B., Hernes, D., Srinivasan, Y., Yang, S., Mitra, S., Sriraja, Y., Nutter, B., Phillips, B., Long, L., and Ferris, D., "A probabilistic approach to segmentation and classification of neoplasia in uterine cervix images using color and geometric features," in [*SPIE, Medical Imaging: Image Processing*], 995–1003 (2005).

[3] Gordon, S., Zimmerman, G., Long, R., Antani, S., Jeronimo, J., and Greenspan, H., "Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams," in [*SPIE, Medical Imaging: Image Processing*], 2037–2045 (2006).

[4] Huang, X., Wang, W., Xue, Z., Antani, S., Long, L. R., and Jeronimo, J., "Tissue classification using cluster features for lesion detection in digital cervigrams," in [*SPIE, Medical Imaging: Image Processing*], (2008).

[5] Gordon, S., Lotenberg, S., and Greenspan, H., "Shape priors for segmentation of the cervix region within uterine cervix images," in [*SPIE, Medical Imaging: Image Processing*], (2008).

[6] Shotton, J., J.Winn, Rother, C., and Criminisi, A., "Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in [*ECCV*], (2006).

[7] Schroff, F., Criminisi, A., and Zisserman, A., "Singlehistogram class models for image segmentation," in [*ICVGIP*], (2006).

[8] Artan, Y. and Huang, X., "Combining multiple 2v-svm classifiers for tissue segmentation," in [*ISBI*], 488–491 (2008).

[9] Warfield, S., Zou, K., and Wells, W., "Simultaneous truth and performance level estimation (staple): An algorithmfor the validation of image segmentation," in [*IEEE Trans. on Medical Imaging*], 903–921 (2004).

[10] Wang, W. and Huang, X., "Distance guided selection of the best base classifier in an ensemble with application to cervigram image segmentation," in [*MMBIA*], (2009).

[11] Mallat, S. and Zhang, Z., "Matching pursuits with time-frequency dictionaries," in [*IEEE Trans. Signal Process*], 3397–3415 (1993).

[12] Chen, S., Billings, S. A., and Luo, W., "Orthogonal least squares methods and their application to non-linear system identification," in [*Int'l. J. Contr*], 1873–1896 (1989).

[13] Tropp, J. A., "Greed is good: Algorithmic results for sparse approximation," in [*IEEE Trans. Inf. Theory*], 2231–2242 (2004).

[14] Huang, J. and Zhang, T., "The benefit of group sparsity," in [*Technical Report arXiv:0901.2962, Rutgers University*], (2009).

[15] Huang, J., Huang, X., and Metaxas, D., "Learning with dynamic group sparsity," *ICCV* (2009).

[16] Engan, K., Aase, S. O., and Hakon-Husoy, J. H., "Method of optimal directions for frame design," in [*IEEE Int. Conf. Acoust., Speech, Signal Process*], 2443–2446 (1999).

[17] Aharon, M., Elad, M., , and Bruckstein, A., "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," in [*IEEE Trans. Signal Process*], 4311–4322 (2006).

[18] Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A., "Discriminative learned dictionaries for local image analysis," in [*CVPR*], 4311–4322 (2008).

[19] Starck, J. L., Elad, M., and Donoho, D., "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. on Image Processing* **14**, 1570–1582 (2004).