DISCRIMINATIVE SPARSE REPRESENTATIONS FOR CERVIGRAM IMAGE SEGMENTATION

Shaoting Zhang, Junzhou Huang, Dimitris Metaxas

CBIM, Rutgers University
Computer Science Department

Wei Wang, Xiaolei Huang

Lehigh University
Computer Science Department

ABSTRACT

This paper presents an algorithm using discriminative sparse representations to segment tissues in optical images of the uterine cervix. Because of the large variations in the image appearance caused by the changing of illumination and specular reflection, the different classes of color and texture features in optical images are often overlapped with each other. Using sparse representations they can be transformed to higher dimension with sparse constraints and become more linearly separated. Different from the previous reconstructive sparse representation, the discriminative method considers positive and negative samples simultaneously, which means that these generated dictionaries can be discriminative and perform better for their own classes but worse for the others. New 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 used our method to automatically segment the biomarker AcetoWhite (AW) regions in an archive of the uterine cervix. Compared with the other general methods including SVM, nearest neighbor and reconstructive sparse representations, our approach showed higher sensitivity and specificity.

Index Terms— segmentation, cervix image, biomarker AcetoWhite, discriminative sparse representation, reconstructive errors, classification

1. INTRODUCTION

Segmentation of different regions of medical images can assist doctors in making proper diagnosis. For example, the area information and landmarks from segmentation is important in many clinical cases [1]. In this work, we proposed an approach to automatically segment the biomarker AcetoWhite (AW) regions in an image archive of the uterine cervix. These images are optical cervigram images acquired by Cervicography using specially-designed cameras for visual screening of the cervix, and they were collected from the NCI Guanacaste project [2] 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 image appearance caused by the changing of illumination and specular reflection in pathology. As a result, the color and texture features in optical images are often overlapped with each other and not linearly separable (Fig. 1).

Previous work on cervigram segmentation has reported limited success using K-means clustering, Gaussian Mixture Models [3], Support Vector Machine (SVM) classifiers [4]. Shape priors are proposed [5], and non-convex regions are also solved [6]. Although such shape information is applicable to cervix boundary, it could not work well with AW since AW regions may have arbitrary shapes. Supervised learning based segmentation [7][8] holds promise, especially with increasing number of features. However, because of the intrinsic diversity between images and the overlap between feature distributions of different classes, it is difficult to learn a single classifier that can perform tissue classification with low error for a large image set. Using reconstructive sparse representations [9] can alleviate the overlapping problem, but this approach does not consider both positive (AW) and negative (non-AW) samples together when generating each dictionary. Thus the reconstructed dictionaries may induce noises and not separate different classes well. Another potential solution is to use a Multiple Classifier System (MCS) [10], 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 [11] 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. Some classifiers commonly have low sensitivity (30% or lower). Wang [12][13] proposed a method to find the best base classifier based on distance guided selection, which achieves good results in a subset of the archive.

In this paper we focus on solving the problem by trans-

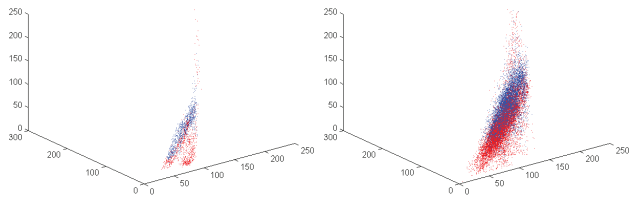

Fig. 1. Color distribution of AW and non-AW regions. Red represents AW and blue represents non-AW. The left one denotes one sample. The right one denotes one hundred samples.

forming the data to higher dimension with sparse constraints. Then the data can be more separated using sparse representations. Different from the reconstructive sparse representations [9], our approach considers all samples together for each dictionary. The generated dictionary deals with different classes discriminatively and performs better for its own class and worse for other classes. Thus the method is named as discriminative sparse representations [14]. The same as the reconstructive method, finding the discriminative sparse representation also consists of sparse coding and codebook update (Section 2.2). After finding positive and negative dictionaries from training images and computing sparse coefficients from testing images, reconstructive error of each pixel is computed and compared. The pixel is assigned to the class with lower errors. Details of this method is explained in Section 2 and experiments are shown in Section 3.

2. METHODOLOGY

2.1. Framework

Fig. 2 illustrates the algorithm framework. In the training stage, ground truth is obtained by clinical experts. 5 by 5 image patches on the ground truth are labeled as positive ones, while the others are negative ones. These patches are fed into general sparse models such as K-SVD to generate corresponding dictionaries. Then these reconstructive dictionaries are employed by the discriminative sparse model to calculate discriminative dictionaries. Both of the sparse models consist of the sparse coding and codebook update stages. Then using these two dictionaries, the sparse coding step is applied on patches extracted from testing images to compute two sets of sparse coefficients. Based on the coefficients and corresponding dictionaries, reconstructive errors are calculated and compared for classification. Details of the discriminative sparse model will be discussed in Section 2.2.

2.2. Discriminative learned dictionaries

The objective of general sparse representation is to find D and X by minimizing the following equation,

$$\underset{D,X}{min}\{\|Y - DX\|_F^2\} \; subject\; to\; \forall i, \|x_i\|_0 \leq L, \quad (1)$$

where Y represents signals (image patches here), D is the overcomplete dictionary, X is the sparse coefficients, $\|\cdot\|_0$ is l^0 norm counting the nonzero entries of a vector and $\|\cdot\|_F$

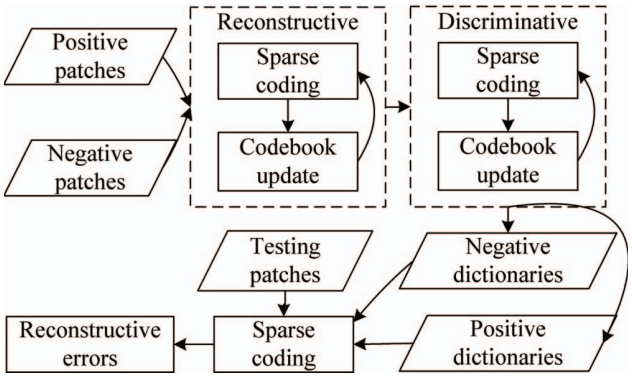

Fig. 2. The algorithm framework. Rectangles denote algorithm, and diamonds represent data.

is 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 $x_i \in \mathbb{R}^k$. Given a signal y_i and a dictionary D, the approximation of the sparse coefficient x_i can be found using greedy algorithms [15]. Then the signal can be reconstructed from D and x_i. The residual is represented as

$$R(y_i, D) = \|y_i - Dx_i\|_2^2. \quad (2)$$

Using K-SVD algorithm [16] 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, which is shown in (3).

$$type = \underset{i=+,-}{arg\; min}\{\|y - D_i x\|_2^2\} \; subject\; to\; \|x\|_0 \leq L. \quad (3)$$

It is also called reconstructive sparse method, which has been used to segment cervigram images [9]. In this method, positive and negative dictionaries are trained from positive and negative data separately. For example, positive dictionary only depends on positive data, so it attempts to reconstruct better for positive ones but not worse for negative ones, which may induce noises. The discriminative sparse method solves this problem by leveraging the logistic function,

$$C_i^\lambda(f_+, f_-) = log\left(\sum_{j\in\{+,-\}} e^{-\lambda(f_j - f_i)}\right), i \in \{+,-\}, \quad (4)$$

which is close to zero when the input f_i is smaller. Increasing the value of λ provides a higher penalty cost for each misclassification, which means more discriminative the model is. Instead of minimizing function 1, the discriminative model solves the following one,

$$\underset{D_{j,j\in\{+,-\}}}{min} \sum_{i\in\{+,-\}} C_i^\lambda(\{R(y_l, D_j)\}) + \lambda\gamma R(y_l, D_i). \quad (5)$$

The first half represents discriminative errors, and the second half denotes reconstructive ones. γ controls the trade-off

Input: Two dictionaries $D_i \in \mathbb{R}^{n\times k}$, $i \in \{+,-\}$ calculated from K-SVD, M input data $y_l \in \mathbb{R}^n$ and coefficients $x \in \mathbb{R}^{k\times MN}$
Output: D_i and X
Loop: Repeat until convergence
For $i = +,-$, for $j = 1 \dots k$, update d, the j-th column of D_i, and sparse coefficients

- Select the set of patches using d, denote as $\omega \leftarrow \{l \in 1 \dots M | x_{li}[j] \neq 0\}$
- For each patch l in ω, compute residual $r_l = y_l - D_i x_{li}$
- Compute weight w_l according to (7) for all patches
- Compute the new d by calculating the eigenvector of (6)
- Compute the new coefficients x_{li}
- Update D_i and X using d and x_{li}

Table 1. The algorithm of the discriminative sparse model.

between reconstruction and discrimination. The higher value of γ, the more reconstructive this model is. The optimization procedure is similar to the ideas of K-SVD. When updating d, the j-th atom of the dictionary i, it allows its corresponding coefficients to change. The method can be formalized as an eigenvalue problem, where new d is the eigenvector corresponding to the largest eigenvalue of

$$\sum_{p\in\{+,-\}} \sum_{l\in S_p\cap\omega} w_l (r_l + x_{li}[j]d)(r_l + x_{li}[j]d)^T, \quad (6)$$

where ω is the set of patches that uses d, the j-th column of D_i. r_l is the residual of patch l, $r_l = y_l - D_i x_{li}$. S_p means classification type p, positive or negative. The weight w_l is calculated from the logistic function,

$$w_l = \frac{\partial C_i^\lambda}{\partial f_p}\left(\{R(y_l, D_j)\}_{j\in\{+,-\}}\right) + \lambda\gamma 1_p(i), \quad (7)$$

where $1_p(i)$ is equal to 1 if $i = p$ and 0 otherwise. The algorithm is listed in Table 1.

2.3. 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 to the area. (3) can be rewritten as

$$error = \|y - D_- x\|_2^2 - \|y - D_+ x\|_2^2. \quad (8)$$

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 Canny edge detector, two thresholds $T_1 > T_2$ 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. However, the value of T_2 highly depends on the application and is found by brute force currently. Starting from 0, T_2 is decreased in a small step each time. The sensitivity and specificity are computed in each step. The parameters causing the best performance are chosen. More sophisticated approach is left for future investigation.

	Sensitivity	**Specificity**
SVM (RBF)	50.24%	63.59%
Nearest Neighbor	55.62%	69.18%
Reconstructive	62.71%	75.85%
Discriminative	71.15%	81.67%

Table 2. Performance comparison between 4 classifiers, in terms of sensitivity and specificity.

3. EXPERIMENTS

The method was implemented in Matlab R2009a and tested on a 2.40 GHz Intel Core2 Quad computer with 8G RAM. Cervigram images from the NCI/NLM archive with multiple-expert boundary markings were 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 were reported. Different color spaces including RGB, HSV and Lab were tested. HSV was chosen since it was slightly better. Each patch was a 5 by 5 square centered in the pixel and was concatenated H, S and V information into single vectors (75 by 1, $n = 75$). Gaussian mask or Laplacian filter can be applied on each patch as a preprocessing. We chose the sparse factor $L = 6$ and dictionaries of size $k = 256$. 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. 10 iterations of K-SVD and 30 iterations of the discriminative model were performed.

The values of λ and γ are also critical for the stability and efficiency [14]. To automatically choose the value of λ that gives the best result in term of classification performance, we can change the value of λ in an ascending series. Starting from a low value for λ, it is possible to check every few iterations whether increasing this value provides a better classification rate. For γ one has to set it large enough to ensure the stability of the scheme, but as small as possible to enforce the discriminative power. We can vary the value of γ by building a descending series and updating the same way as for λ, gives stability to the algorithm.

Our method was compared with SVM, nearest neighbor and reconstructive sparse representations. SVM failed to handle so many 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. Sparse representation was more efficient with 20 seconds for each iteration and less than 1GB RAM because of its sparsity. Some qualitative results are presented

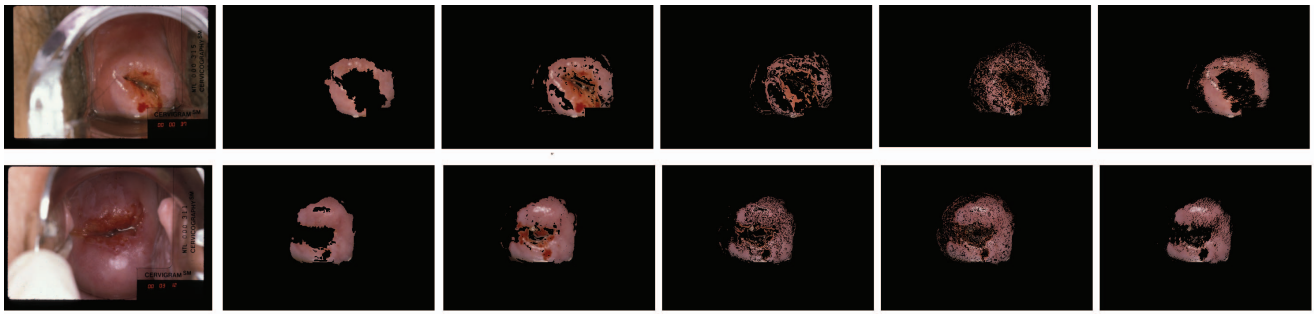

Fig. 3. Results of different algorithm applied on two images. 1st column: original image; 2nd column: ground truth; 3rd column: SVM with radial basis function kernel; 4th column: 1 nearest neighbor; 5th column: reconstructive sparse representations: 6th column: discriminative sparse representations.

in Fig. 3. Table 2 shows the results of different classifiers measured by sensitivity and specificity. SVM with radial basis kernels generally performs better than linear and polynomial kernels. Thus it is chosen in this comparison experiment. Since the distributions of image patches are highly overlapped (Fig. 1), SVM still underperforms the classification task. The reconstructive sparse model achieves good results, but it still has many noises from non-AW regions. The reason is that the K-SVD algorithm is not discriminative and there is no region information considered. The positive dictionary only attempts to reconstruct better for positive ones but not worse for negative ones. Thus there are false alarms in non-AW regions. The discriminative sparse model can alleviate this problem by considering both positive and negative patches together.

4. CONCLUSIONS

In this paper we proposed a classifier based on the discriminative sparse representations to segment tissues in optical images of the uterine cervix. Different from the reconstructive method, discriminative method considers all samples simultaneously and discriminatively for each dictionary. Thus the generated dictionaries perform better. Our method was compared with the reconstructive sparse representations and other general learning methods, and showed higher sensitivity and specificity. In the future, we would like to combine SVM and sparse representations together, to alleviate the problem caused by the sensitivity of parameters.

5. REFERENCES

[1] G. Zimmerman, S. Gordon, and H. Greenspan, "Automatic landmark detection in uterine cervix images for indexing in a content-retrieval system," in *ISBI*, 2006, pp. 1348–1351.

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

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

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

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

[6] S. Gordon and H. Greenspan, "Segmentation of non-convex regions within uterine cervix images," in *ISBI*, 2007.

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

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

[9] S. Zhang, J. Huang, W. Wang, X. Huang, and D. Metaxas, "Cervigram image segmentation based on reconstructive sparse representations," in *SPIE, Medical Imaging: Image Processing*, 2010.

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

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

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

[13] W. Wang, X. Huang, Y. Zhu, D. Lopresti, L. R. Long, S. Antani, Z. Xue, and G. Thoma, "A classifier ensemble based on performance level estimation," in *ISBI*, 2009.

[14] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *CVPR*, 2008, pp. 1–8.

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

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