GROUP SPARSITY BASED CLASSIFICATION FOR CERVIGRAM SEGMENTATION

Yang Yu[⋆], *Junzhou Huang*[⋆], *Shaoting Zhang*[⋆], *Christophe Restif*[⋆], *Xiaolei Huang*[†], *Dimitris Metaxas*[⋆]

[⋆] CBIM, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
[†]Computer Science & Engineering Department, Lehigh University, Bethlehem, PA, USA

ABSTRACT

This paper presents an algorithm to classify pixels in uterine cervix images into two classes, namely normal and abnormal tissues, and simultaneously select relevant features, using group sparsity. Because of the large variations in image appearance due to changes of illumination, specular reflections and other visual noise, the two classes have a strong overlap in feature space, whether features are obtained from color or texture information. Using more features makes the classes more separable and increases the segmentation's quality, but also its complexity. However, the properties of these features have not been well investigated. In most cases, a group of features is selected prior to the segmentation process; features with minor contributions to the results are kept and add to the computational cost. We propose feature selection as a significant improvement in this problem. It provides a robust trade-off between segmentation quality and computational complexity. In this work we formulate the cervigram segmentation problem as a feature-selection-based classification method, and we introduce a regularization-based feature-selection algorithm to leverage both the sparsity and clustering properties of the features used. We implemented our method to automatically segment the biomarker AcetoWhite (AW) regions in a dataset of 200 images of the uterine cervix, for which manual segmentation is available. We compare the performance of several regularization-based feature-selection methods. The experimental results demonstrate that on this dataset, our proposed group-sparsity-based method gives overall better results in terms of sensitivity, specificity and sparsity.

Index Terms— segmentation, cervix image, biomarker AcetoWhite, feature selection, group sparsity, classification

1. INTRODUCTION

Cervical cancer is the second most common cancer affecting women worldwide and has a high mortality rate if not treated timely. Thus early detection and high quality screening are very important [18]. An effective approach to cervical cancer screening is based on the color changes of cervix tissues when exposed to acetic acid. Abnormal tissues usually turn white and look more opaque, which are termed Acetowhite (AW). Since the texture, size and location of AW regions have been shown to correlate with the pathologic grade of disease severity, analyzing them is a significant part of the diagnosis. In this work, we present an algorithm to automatically segment the AW regions in uterine cervix images. We use optical cervigrams acquired by cervicography using specially-designed cameras for the visual screening of the cervix. They were collected for the National Cancer Institute (NCI) Guanacaste project [8] and were digitized and maintained by the National Library of Medicine (NLM) and NCI, to study the correlation between visual features and the development of pre-cancerous lesions. Accurate segmentation of AW regions in cervigrams is useful for indexing and retrieval of the images and helpful for improving diagnosis. However, it is a challenging problem in computer vision because of large variations in image appearance, caused by illumination variations and specular reflections. As a result, the distributions of color and texture features of AW and non-AW regions show significant overlap.

Many classification methods have been applied to segment the AW regions, such as K-means clustering [14], Support Vector Machine (SVM) classifiers [7], supervised learning based segmentation [12, 13] and sparse representations [20, 21]. Shape priors have also been proposed [3]. Non-convex regions [2] and AW regions [4] can also be used via shape features. Their performances are promising, especially when increasing the number of features, using cervix color and texture [9]. However, the intrinsic diversity among images and the resulting overlap between feature distributions of different classes, make it difficult to train a single classifier that can perform tissue classification with low error on a large image set. Another potential solution is to use a Multiple Classifier System (MCS) [1], which trains a set of diverse classifiers that disagree on their predictions and effectively combines the predictions in order to reduce the classification error, such as voting, AdaBoost, bagging and STAPLE [16]. But as mentioned in [15], the classifier ensemble methods could improve performance only when the base classifiers are sufficiently good, with at least 50% sensitivity and specificity. This is an issue given that most base classifiers may have very poor performance.

The main reason for the poor performance of base classifiers is the feature-overlap problem. Since the samples that are overlapped in one feature space may be separated in other feature spaces, increasing the number of features can alleviate the problem. This is generally one of the most effective ways to improve the segmentation performance. However, one of the main shortcomings of existing work is that feature selection has not been well investigated. Features are often preselected from the beginning. These predefined features do not equally or positively contribute to the segmentation performance, while they still increase the computation complexity. Thus pruning features without adversely affecting the performance is an important task. Choosing the best base classifiers [15], by selecting a subset of the training dataset based on dissimilarity between probability functions, achieves great improvement over previous methods. For the task at hand however, the training set is expected to contain significantly different images from test sets, due to the variety of conditions of patients and imaging quality. As a result, it is preferable for cervigram image segmentation to use a more general feature-selection method, which would extract a subset of important features from the entire dataset.

In this paper we formulate this cervigram segmentation problem as a feature selection based classification problem, and present a group sparsity based method to solve it. Our method is based on two underlying observations. First, the samples could be classified using only a few features. Since more than enough features are used in our work and some of them are redundant, a sparse set of selected features can capture the class differences and increase classification speed. Second, features in color or texture spaces are often mutually dependent. They tend to simultaneously either have or not have an effect on the classification. Similar priors have already been used in other machine learning applications, such as image annotation [19] with good results. Thus, the above observed priors motivate us to use a group sparcity method [17], which is theoretically proven to improve classification performance under specific conditions list in [5], and achieves great success in many signal recovery applications.

Our work leverages the special properties of cervigram image features to improve segmentation performance. Several types of features are used and compared, and they are automatically pruned based on group and sparsity priors to improve cervigram image segmentation performance and accelerate speed. In Section 2 we detail the classification method. In Section 3 we present our experiments and compare the performance of different features as well as regularization-based feature-selection methods. We present our conclusions and future work in Section 4.

2. METHODOLOGY

Our segmentation algorithm includes two steps: training and classification. The training step selects features based on both sparsity and group clustering priors. These priors improve the model's robustness to noise. The method we use is a variation of solutions to regularization problems, and is inspired by the recently proposed group sparsity in the compressive sensing community. The classification step automatically prunes features and classifies each pixel of the input images.

Table 1. Details of the training and classification steps of the our method.

Training (computing weights)
Input: Feature matrix of training images, $F \in \mathbb{R}^{m \times p}$, where m is the number of training samples, and each row is a feature vector $f_i \in \mathbb{R}^p$, computed in a patch centered on that pixel; classification target vector $Y \in \mathbb{R}^n$, where $y_i = 1$ if pixel i is AW, and $y_i = -1$ if it is non-AW.
Computation:
Generate groups G_1 to G_m according to feature types.
Use Projected Gradient Method [11] (or other optimization solvers) to solve (4)
Output: weight vector $w \in \mathbb{R}^p$.

Classification
Input: Weight vector $w \in \mathbb{R}^p$; feature vector of all m pixels in the test image $T \in \mathbb{R}^{m \times p}$, where each row is a feature vector $t_i \in \mathbb{R}^p$.
Computation:
Prune columns of T and w according to the zero elements in w. The columns of T corresponding to zero elements are removed: the sizes of T and w are reduced.
Loop: Iterate over the rows of T
$y = t_i \cdot w$
If $y > 0$
The ith pixel is positive (*i.e.* in an AW region)
Else
The ith pixel is negative (*i.e.* in the background)
Output: The classification result for all pixels.

2.1. Problem Formulation

For the ith observation in the training set, the data y_i is used to indicate whether the sample is AW or not, and the feature vector $x_i = (x_{i1}, ..., x_{ip})$ is extracted from the predefined feature spaces to characterize the sample. We consider the feature selection step as a regression problem, where $Y = (y_1, ..., y_n) \in \mathbb{R}^n$ is the response and $X = (x_1, ..., x_n) \in \mathbb{R}^{n \times p}$ is the regressor. All the responses are generated from a sparse linear combination of the features adding a stochastic noise vector $\epsilon \in \mathbb{R}^n$: $Y = Xw + \epsilon$, where w is the weight vector for different features.

In our framework, the regressor y_i represents the classification target, where $y_i = 1$ if the ith pixel is in an AW region, and $y_i = -1$ if it is in a non-AW region. x_i is the feature vector for the ith pixel, obtained from a patch centered around that pixel. As an example, in the case where only each of the

Table 2. Comparison of color and texture features applied on this segmentation problem. Equal weights are used to combine them together.

Features	$\overline{p}$	σp	$\overline{q}$	σq	$\overline{DCS}$	σDCS
RGB (color)	0.502	0.050	0.803	0.037	0.515	0.072
HSV (color)	0.652	0.062	0.815	0.045	0.601	0.077
LAB (color)	0.484	0.053	0.834	0.037	0.503	0.067
Opponent (color)	0.502	0.050	0.813	0.036	0.515	0.072
Haar(texture)	0.403	0.061	0.724	0.045	0.435	0.062
All features	**0.682**	0.049	**0.822**	0.037	**0.653**	0.071

three RGB features are used, and the patch size is 5×5, we have $p = 3{\times}5{\times}5$ and x_i is a $1{\times}75$ vector. The actual features used in our experiments are reported in Section 3. While we should note here that more than enough features are used in the training step, which is more than the number of samples, thus, the linear system $Xw = Y$ is underdetermined.

2.2. Group Sparsity Based Classification

The direct loss function to calculate w is the least square estimate, which minimizes the residual sum of squared errors:

$$\hat{w} = \underset{w\in\mathbb{R}^p}{\arg\min}\ \|Xw - Y\|_2^2 \tag{1}$$

where the analytical solution of w can be represented as $(X^TX)^{-1}X^TY$. However, the matrix X^TX is singular since the linear system is underdetermined, making the model unstable. Ridge regularization is widely used to alleviate this problem, which can be written in the following format:

$$\hat{w} = \underset{w\in\mathbb{R}^p}{\arg\min}\ \left[\frac{1}{n}\|Xw - Y\|_2^2 + \lambda\|w\|_2\right] \tag{2}$$

where $\lambda\|w\|_2$ is a restriction on the L^2 norm of weight vector w, which increases the stability of the solution. But L^2 norm does not encourage sparsity, which means it will perform poorly when irrelevant features are present in X.

Lasso is an often-used method to fulfill the sparsity prior, which is different from ridge regularization only on the penalty term:

$$\hat{w} = \underset{w\in\mathbb{R}^p}{\arg\min}\ \left[\frac{1}{n}\|Xw - Y\|_2^2 + \lambda\|w\|_1\right] \tag{3}$$

L^1 norm is used by Lasso instead of L^2 norm in ridge regularization in the above function. The solutions produced can be as sparse as with L^0 regularization in some underdetermined systems. However, as shown in Section 3, using this prior alone can produce over-sparse solutions in our system, which also adversely affects the overall performance.

Features have a natural group structure based on the different kinds of features such as RGB and Haar, etc. The features in each group tend to have or not have effect since their magnitudes are not independent. Based on this observation, we add the group priority to this problem and reformulated it as:

$$\hat{w} = \underset{w\in\mathbb{R}^p}{\arg\min}\ \left[\frac{1}{n}\|Xw - Y\|_2^2 + \lambda\sum_{j=1}^{m}\|w_{G_j}\|_2\right] \tag{4}$$

where features are partitioned into m disjoint groups G_1, G_2, ..., G_m, and w_{G_j} denotes the vector in $\mathbb{R}^{G_j}$, which is identical to w in G_j. Both L^1 and L^2 norms are used in the term $\lambda\sum_{j=1}^{m}\|w_{G_j}\|_2$, where L^2 norm is used among the weights inside of the same group, while L^1 norm is used when summing the results between groups. As we known, L^2 norm does not induce parsimony, while L^1 norm does. So the weights are tend to be parsimony based on the group structure. The algorithm framework and optimization procedures are illustrated in Table 1. The weight w is produced during the training stage by leveraging the group sparsity in regularization techniques.

In the classification stage, if $w_j = 0$, the jth element in the feature vector x_i does not have a contribution, therefore the jth bin in the feature vectors can be pruned directly. Based on the pruned features and weights, the estimated classification target is obtained from the test image. Then a binary classifier is used to perform the classification. A target value above 0 classifies the corresponding pixel as AW, while a value below 0 puts the pixel as non-AW. The computation complexity for classification is $O(n_p n_f)$, where n_p is the number of pixels in the testing image and n_f is the number of features. Thus reducing the number of features decreases the computation complexity.

3. EXPERIMENTS

3.1. Experimental Settings

Our dataset consists of 200 cervigram images of diverse appearance from the NCI/NLM archive with corresponding multiple-expert boundary markings. 20 images are used for testing and validation and the remaining 180 ones are used for training. Four color spaces are tested, namely RGB, HSV, LAB and Opponent color space [10]. Haar features are used as the texture information. For each pixel, the feature vector

Table 3. Comparison of feature selection methods applied on this segmentation problem.

Methods	$\overline{p}$	σp	$\overline{q}$	σq	$\overline{DCS}$	σDCS	Sparsity
No regularization	0.601	0.035	0.663	0.031	0.478	0.102	0 %
L^1 regularization	0.668	0.051	0.671	0.093	0.511	0.122	76 %
L^2 regularization	0.675	0.060	0.704	0.069	0.533	0.118	2 %
Group sparsity (proposed)	0.693	0.053	0.725	0.185	0.617	0.123	60 %

consists of these color and texture features computed in a patch centered around that pixel. In addition to our proposed group sparsity method, we also implement the L^1 and L^2 regularization, as well as no regularization. The tuning parameter λ is chosen by cross validation. Note that non-regularization methods do not depend on the tuning parameter. All our implementations are done in Matlab R2009a.

3.2. Evaluation of Features

Table 2 compares the segmentation performance of different color and texture features. The mean and standard deviation of sensitivity p, specificity q and dice similarity coefficient (DSC) are reported. The performances of different features are dissimilar. HSV performs better than other color spaces, and color information achieves better results than texture. Generally, using single features does not perform well since the feature distributions of AW and non-AW regions are highly overlapped. Combining different types of features alleviates this problem. These experimental results confirm the earlier discussion on the importance of increasing the number of features.

3.3. Evaluation of Regularization Methods

Table 3 compares different regularization based feature selection methods applied on this cervigram segmentation problem. The sparsity is reported, which is defined as the percentage of zero elements. The comparisons include non-regularization, L^1, L^2 and group sparsity based regularization.

Least squares without regularization under-performs on this dataset since the linear system X can be singular or underdetermined. L^1, L^2 and group sparsity achieve better performance because the regularization terms improve the stability of the system. Group sparsity procudes the best results on this dataset since it leverages both sparsity and clustering priors and the clustering prior contributes to the stability of the model. The solution obtained with L^1 regularization is more sparse than the one of the group sparsity. The reason is that L^1 method can arbitrarily prune features, while group sparsity has structure constraints. Generally group sparsity achieves a good tradeoff between performance and sparsity. Using this method can decrease the computational complexity without adversely affecting the performance.

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a group-sparsity-based feature-selection method to automatically segment cervigram images and simultaneously select relevant features. This algorithm leverages both sparsity and clustering priors to prune the features. Compared to other regularization methods, it shows higher overall performance in this context, in terms of sensitivity and specificity. Since an increasing number of features are used in variant kinds of segmentation applications, this algorithm may also be applicable for other segmentation tasks. Future works will improve the feature selection algorithm by considering structured sparsity [6], and combine our method with smoothness-enforcing segmentation algorithms. The structure sparsity focuses on more general structured information, which is not limited to a fixed group structure. The features in cervigram images have strong local correlations inside each group. Taking advantage of this property may further improve the performance of our classifier. Meanwhile, our feature selection methods could also be embedded in other more general two-label segmentation methods that enforce local consistency of pixel labels, such as graph-cuts or level sets.

5. REFERENCES

[1] Y. Artan and X. Huang. Combining multiple 2v-svm classifiers for tissue segmentation. In *ISBI*, pages 488–491, 2008.

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

[3] S. Gordon, S. Lotenberg, and H. Greenspan. Shape priors for segmentation of the cervix region within uterine cervix images. In *SPIE, Medical Imaging*, 2008.

[4] H. Greenspan, S. Gordon, G. Zimmerman, S. Lotenberg, J. Jeronimo, S. Antani, and R. Long. Automatic detection of anatomical landmarks in uterine cervix images. *IEEE Trans. Med. Imag.*, 28(3):454–468, 2009.

[5] J. Huang and T. Zhang. The benefit of group sparsity. In *Technical Report arXiv:0901.2962, Rutgers University*, 2009.

[6] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. In *ICML*, pages 417–424, 2009.

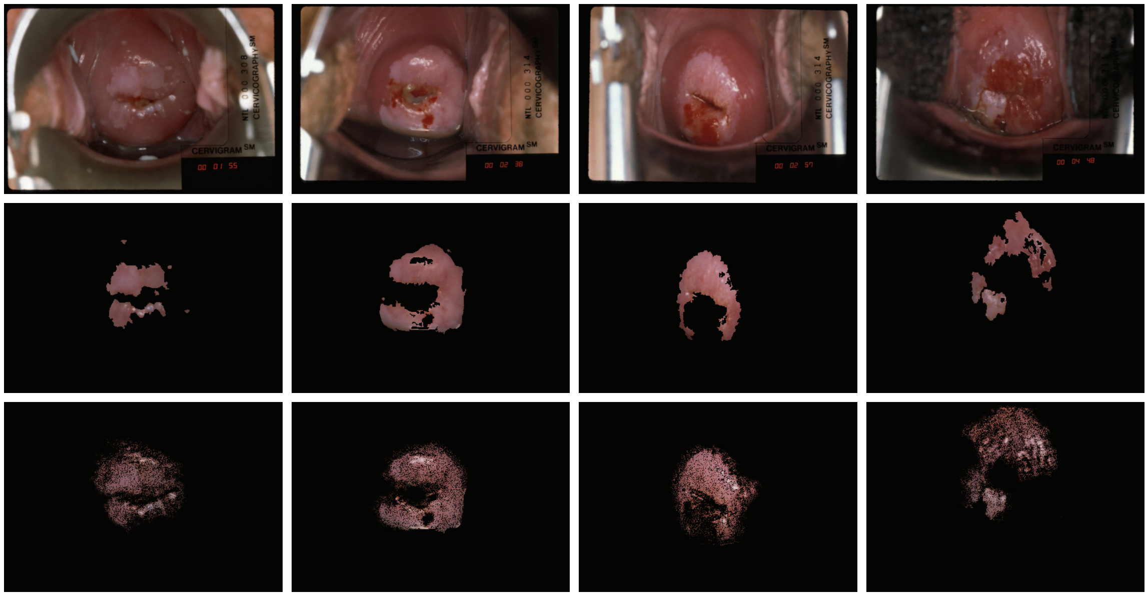

Fig. 1. Segmentation results using our algorithm. Top row: original images. Middle row: manually segmented results. Bottom row: results from our group sparsity method.

[7] 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*, 2008.

[8] 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*, pages 16–25, 2006.

[9] Q. Ji, J. Engel, , and E. Craine. Texture analysis for classification of cervix lesions. *IEEE Trans. Med. Imag.*, 19(11):1144–1149, 2000.

[10] K. Sande, T., and C. Snoek. Evaluating color descriptors for object and scene recognition. *PAMI*, 99(1), 2009.

[11] M. Schmidt, K. Murphy, G. Fung, and R. Rosales. Structure learning in random fields for heart motion abnormality detection. In *CVPR*, 2008.

[12] F. Schroff, A. Criminisi, and A. Zisserman. Single-histogram class models for image segmentation. In *ICVGIP*, 2006.

[13] 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.

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

[15] 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.

[16] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (staple): An algorithmfor the validation of image segmentation. In *IEEE Trans. on Med. Imag.*, pages 903–921, 2004.

[17] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B*, 68:49–67, 2006.

[18] J. Zhang and Y. Liu. Cervical cancer detection using svm based feature screening. In *MICCAI*, pages 873 – 880, 2004.

[19] S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. Metaxas. Automatic image annotation using group sparsity. In *CVPR*, pages 3312 –3319, 2010.

[20] S. Zhang, J. Huang, D. Metaxas, W. Wang, and X. Huang. Discriminative sparse representations for cervigram image segmentation. In *ISBI*, pages 133 – 136, 2010.

[21] S. Zhang, J. Huang, W. Wang, X. Huang, and D. Metaxas. Cervigram image segmentation based on reconstructive sparse representations. In *SPIE, Medical Imaging*, 2010.