LEARNING MULTISPECTRAL TEXTURE FEATURES FOR CERVICAL CANCER DETECTION

Yanxi Liu, Tong Zhao and Jiayong Zhang

The Robotics Institute, Carnegie Mellon University,Pittsburgh, PA 15213
{yanxi,tzhao,zhangjy}@cs.cmu.edu

ABSTRACT

We present a bottom-up approach for automatic cancer cell detection in multispectral microscopic thin Pap smear images. Around 4,000 multispectral texture features are explored for cancer cell detection. Using two feature screening measures, the initial feature set is effectively reduced to a computationally manageable size. Based on pixel-level screening results, cancerous regions can thus be detected through a relatively simple procedure. Our experiments have demonstrated the potential of both multispectral and texture information to serve as valuable complementary cues to traditional detection methods.

1. INTRODUCTION

Finding abnormal cells in Pap smear images (Figure 1) is a "needle in a haystack" type of problem, which is tedious, labor-intensive and error-prone. It is therefore desirable to have an automatic screening tool such that human experts are only called for when complicated and subtle cases arise. Most research to date on automatic cervical screening tries

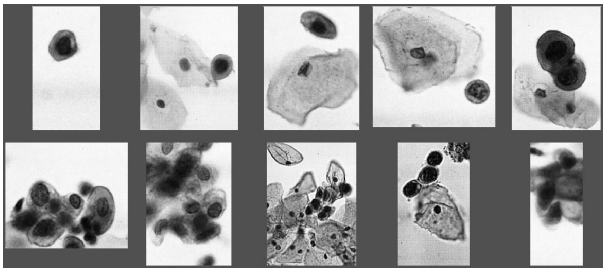


Fig. 1. A sample of Pap smear images.

to extract morphometric/photometric features at the cell level in accordance with "The Bethesda System" rules [1]. These methods usually depend on accurate segmentations between cytoplasm and nucleus. However, various factors such as the presence of blood, inflammatory cells or thick cell clumps make the segmentation rather difficult. Even with 100% segmentation accuracy, shape features may still fail to reflect the differences between normal and cancerous cells.

Recently, multispectral biomedical image analysis has received increasing attention. Using a micro-interferometric spectral imaging setup [2] in CMU, we have obtained a set of multispectral Pap smear images. The wavelengths range from 400 nm to 690 nm, evenly divided into 52 bands. In this paper, a bottom-up approach is proposed to automatically detect cancerous regions in such database. It takes advantage of the multispectral property and avoids segmentation difficulty. First, each cell pixel is evaluated as cancerous or normal based on various types of features extracted from multispectral bands. Then cancerous regions are detected through a relatively simple procedure.

Two essential problems exist in this scheme. First, we have to decide what kind of features to extract. This is often considered as an art and depends highly on the designer's experience. Here, we follow a least commitment philosophy such that subjective rejection of image features without concrete justification is minimized. Accordingly, four kinds of wavelet features plus statistical intensity features are extracted over different spectral bands, constituting a vector space with dimension near 4,000. Unfortunately, this space is prohibitive for most existing classifiers due to computational cost and poor generalization. Hence comes the second problem: how to remove those irrelevant and/or redundant features, and find a feature subset that is well balanced between performance and compactness. Considering the huge feature and sample complexity, we employ two simple screening measures (IG and AVR) to distill the most distinctive image features. Experiments on a multispectral image database demonstrated the effectiveness of our approach. The results also show that multispectral and texture information both provide valuable complementary cues for cancer detection from multispectral Pap smear images.

2. OVERVIEW OF OUR APPROACH

Figure 2 gives the flow chart of the proposed bottom-up method. In the preprocessing stage, we first segment cells

Gratefully acknowledge the support from NCI Unconventional Innovation Program, award # = N01-CO-07119, and Drs. D. Farkas, E. Wachman and Nath for providing the images used in this work.

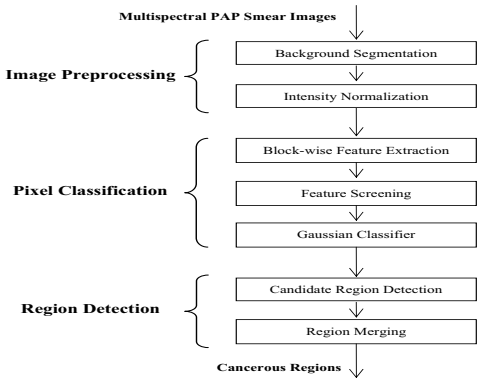


Fig. 2. The proposed bottom-up detection scheme.

from the background, which is much easier than traditional nuclei/cytoplasm segmentation. Then we normalize all band images by subtracting the spectral signature of the image background in order to suppress background noise. The next two stages will be explained in the following sections.

3. PIXEL LEVEL CLASSIFICATION

3.1. Block-wise Feature Extraction

For each pixel, we extract various types of image features in a fix-sized block around it in order to capture local texture information. This procedure is applied to every band in turn, resulting in a very high dimensional multispectral texture feature set. For computational efficiency we choose to use the following filtering methods.

Statistical Intensity Features We employ both basic and high order statistics to describe the statistical intensity properties of each block. Ten statistics are used: maximum, minimum, range, median, mean, standard deviation, energy, skewness, kurtosis and entropy. They are also computed on the blocks after the following wavelet transforms with tree-structured decomposition.

Orthogonal and Biorthogonal Wavelet Orthogonal and compactly supported wavelets, such as Daubechies, symlets and coiflets, have fast numerical algorithms, good time or space localization properties, but with poor regularity. We choose two asymmetric orthogonal wavelets: Daubechies 2 (low order) and Daubechies 16 (high order). Biorthogonal wavelets are symmetric with FIR filters.

Gabor Filters The 2D Gabor functions are defined as

$$g_{m\theta}(x,y) = \frac{a^{-m}}{2\pi\sigma_1\sigma_2}\exp[-\frac{1}{2}(\frac{x'^2}{\sigma_1^2}+\frac{y'^2}{\sigma_2^2})+2\pi iWx']$$

where $i=\sqrt{-1}, \sigma_1, \sigma_2, W$ are given, and $x' = a^{-m}(x\cos\theta + y\sin\theta), y' = a^{-m}(-x\sin\theta + y\cos\theta)$. A series of filters are generated with different choices of m and θ.

3.2. Feature Screening

A basic question in many image classification tasks is: what are good features for classification purpose? Although many feature subset selection methods have been proposed [3], few of them can be directly applied to domains with more than 100 dimensions. The huge feature dimension (near 4,000) and sample complexity (over 100,000) in this work make them computationally prohibitive. Therefore we employ sequentially two simple individual evaluation measures: Information Gain (IG) and Augmented Variance Ratio (AVR). Features are ranked according to these measures, and a subset is selected conservatively via a statistical significance test.

Information Gain Suppose data D is split by feature X into p partitions $D_1^X, \ldots, D_p^X$, and there are totally d classes $\omega_1, \ldots, \omega_d$. The information gain due to feature X is defined as the difference between the prior uncertainty and the expected posterior uncertainty using X:

$$IG(X) = I(D) - \Sigma_{j=1}^{p}\frac{|D_j^X|}{|D|}I(D_j^X),$$

where $I(D) = -\Sigma_{i=1}^{d}P_D(\omega_i)\log P_D(\omega_i)$ and $I(D_j^X) = -\Sigma_{i=1}^{d}P_{D_j^X}(\omega_i)\log P_{D_j^X}(\omega_i)$, $|D|$ is the number of instances in D and $P_D(\omega_i)$ are priors for data D. Feature X is chosen over feature Y if $IG(X) > IG(Y)$.

Augmented Variance Ratio The AVR for X is defined as

$$AVR(X) = \frac{S_b(X)}{\frac{1}{d}\sum_{i=1}^{d}\frac{S_i(X)}{\min_{i\neq j}(|M_i(X)-M_j(X)|)}}$$

where $S_b(X)$ is the between-class variance, $M_i(X)$ and $S_i(X)$ are the mean and within-class variance for ω_i respectively.

3.3. Regularized Gaussian Classifier

Let x be the n-dimensional feature vector associated with a pixel. For each class ω_c $(c=1,2)$, a quadratic discriminant function is defined as

$$g_c(x) = (x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c) + \log|\Sigma_c| - 2\log P_c$$

where μ_c, Σ_c and P_c are the mean, covariance matrix and *a priori* probability of ω_c respectively. Let $h(x) = g_1(x) - g_2(x)$, decisions will be made according to the following rule: $x \in \omega_1$ if $h(x) < 0$; otherwise $x \in \omega_2$.

In practice, the true distribution parameters in $h(x)$ are commonly substituted by the sample means and covariance matrices, which leads to the so-called maximum likelihood Gaussian classifier [4]. However, several problems exist with such maximum likelihood estimators when the feature

dimension is relatively high. Parameter estimation errors will induce large variance to the discriminant. In addition, the covariance matrix may be singular when sample size is small. To alleviate these problems, various regularization techniques are available in order to obtain robust covariance matrix estimate. Here we use a simple modified version of quadratic discriminant proposed by Kimura [5]:

$$g(x) = \sum_{i=1}^{k} \frac{1}{\lambda_i} \left[\varphi_i^T (x-\mu)\right]^2 + \sum_{i=k+1}^{n} \frac{1}{h^2} \left[\varphi_i^T (x-\mu)\right]^2 + \ln\left[h^{2(n-k)} \prod_{i=1}^{k} \lambda_i\right]$$

where λ_i denotes the i-th eigenvalue ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$) of Σ, and φ_i denotes the corresponding eigenvector. The basic idea behind the modification is to estimate only the k largest eigenvalues, and replace the remaining $(n-k)$ small ones by a constant h^2, which is chosen as the average of the $(k+1)$-th largest eigenvalues of all classes. The use of a Gaussian classifier is acceptable since the output of AVR is normally biased and we have used Neural Networks on the same data with similar performance.

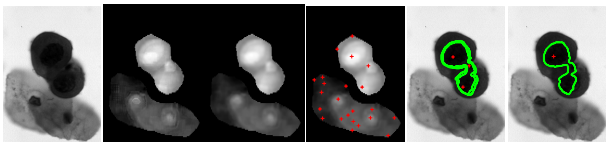
4. REGION LEVEL DETECTION

Based on the continuous output $h(x)$ of pixel classification, cancerous regions can be discovered by a relatively simple procedure (Figure 3). First, smooth $H = \{h(x)\}$ by a Gaussian filter. Second, find all the local maxima m_i in H, and their corresponding effective regions R_i, defined as the points immediately around m_i with values above a fixed fraction of $h(m_i)$. Third, for each R_i extract a geometric feature $G = C/L$, where C is the circumference of R_i, L is the distance from m_i to the boundary. Prune those R_i with $h(m_i)$ or G smaller than some thresholds, and generate the candidate region set. Finally, merge those candidate regions that are inter-connected. Note that we have combined the simple shape feature G into this bottom-up approach.

5. EXPERIMENTAL RESULTS

5.1. Dataset and Experimental Setup

The proposed bottom-up approach to cervical cancer detection in multispectral Pap smear images have been evaluated on a database containing 40 images (each with 52 bands), with a total of 149 cells (41 cancerous and 108 normal). First, all images are preprocessed to remove the background and normalize intensity. Then for each pixel to be classified, various image features are extracted in a 16×16 block around it over different bands, as described in section 3.1. Thus a very high dimensional multispectral texture feature vector is associated with each pixel. We collect a total of 156,732 sample vectors from all 40 images. Considering the fact that samples from the same image are often highly correlated, we always keep them as a whole when partitioning training and test sets in all the following experiments.



(a) (b) (c) (d) (e) (f)

Fig. 3. An example of cancerous region detection. (a) Band average image. (b) Scaled output H of pixel classification. (c) H smoothed by a Gaussian filter ($\sigma_x = \sigma_y = 1.5$). (d) Detected local maxima are marked as red crosses on H. (e) Contours of candidate regions are plotted on average-band image. (f) The final result after region merging.

5.2. Pixel Classification and Comparisons

To evaluate the pixel classification algorithm, 28 images are randomly chosen to make up the training set. By using the samples from these images, IG and AVR ranking scores are first computed, and those irrelevant features with low scores are removed from the feature set. Then a modified Gaussian classifier is trained based on the selected features. Finally, the classifier is applied to the samples from the remaining 12 images. By adjusting the decision threshold in the classifier, we record false positive rates (FPR) at different true positive rate (TPR) values (TPR = 1 - false negative rate), see Figure 4. This procedure is repeated 40 times, and the results are averaged. Besides pooling all four types of features (DB2, DB16, Bio2.2 and Gabor) in the original feature set, we have also done comparative experiments by employing only one type of them. Table 1 shows the feature dimensions before and after IG/AVR screening in a typical random partition test. Table 2 gives the test set results when TPR = 0.86, 0.90 and 0.95 respectively. It can be observed that DB16 and Gabor give better performance among the four types of features, which is consistent with the comparative study result of other researchers [6]. In addition, the combined version is always better or no less than any single type of features. Therefore, the combined scheme is used throughout the following experiments.

The motivation behind this work is the belief that both multispectral and texture information will help cervical cancer detection. To verify this point, we investigate the classification performance of three types of features with/without multispectral or texture information:

1. Block-wise texture features over all bands (MBT);
2. Texture features only on average band image (ABT);
3. The intensity of a single pixel across all bands (MBI).

Figure 4 shows their test set ROC curves. It is observed that MBT is consistently better than ABT (without multispectral info), and much better than MBI (without texture info). This indicates that both multispectral and texture information do

# Features	DB2	DB16	Bio2.2	Gabor	Combined
Original	800	800	900	1200	3700
After IG Screening	74	76	71	51	144
After AVR Screening	47	35	41	21	45

Table 1. Various dimensions before and after IG/AVR screening in a typical random test.

TPR	DB2	DB16	Bio2.2	Gabor	Combined
0.86	0.06 ± 0.08	0.06 ± 0.08	0.06 ± 0.07	0.07 ± 0.10	0.04 ± 0.07
0.90	0.10 ± 0.12	0.08 ± 0.11	0.09 ± 0.10	0.08 ± 0.12	0.07 ± 0.11
0.95	0.17 ± 0.18	0.12 ± 0.15	0.15 ± 0.15	0.11 ± 0.16	0.11 ± 0.15

Table 2. Comparison of pixel classification accuracy using different types of texture features. The False Positive Rates (FPR) at different True Positive Rate (TPR = 1 - false negative rate) values are given in the format of MEAN$\pm$ STD.

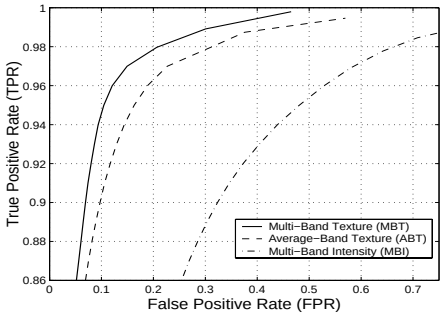


Fig. 4. Comparison of three pixel classification schemes with/without multispectral or texture information.

provide valuable cues to the detection problem.

5.3. Region Detection

As the number of available images is small, we evaluate the performance of the complete detection system using leave-one-out method. Each time 39 images are used to train the pixel classifier, and one image is reserved for test. Some typical detection results are shown in Figure 5. Among the 149 cells distributed in 40 images, one cancerous cell is missed (1- false negative rate = TPR $= 40/41 \approx 0.98$), and one normal cell is falsely detected (false positive rate FPR $= 1/108 \approx 0.01$).

6. CONCLUSION

We propose a bottom-up approach to cancer detection in multispectral Pap smear images. Around 4,000 multispectral texture features are extracted in a fix-sized block around each pixel. Two screening measures are sequentially employed, and the original feature set is effectively reduced to a computationally manageable size. Based on high-accuracy pixel classification, cancerous regions can be discovered by

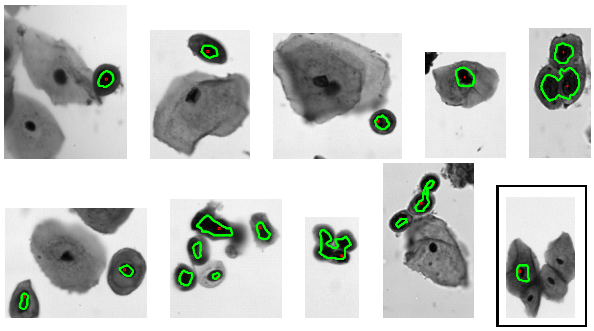


Fig. 5. Some detection results from the multi-band image database. All images contain one or more cancerous cells, except the last one (with black frame). The contours of the detected regions are shown on the average-band images.

a simple procedure. Experiments on a multispectral image database with 149 cells distributed over 40 images demonstrate the effectiveness of the approach. We have compared four types of wavelet transforms for pixel classification, and found that Daubechies 16 and Gabor give better performance. Three types of features with/without multispectral or texture information are also investigated. The result shows high potential for using both multispectral and texture information to provide complementary cues. We are currently investigating new feature selection algorithm, and are trying to combine this bottom-up approach efficiently with high-level shape analysis.

7. REFERENCES

[1] R. J. Kurman and D. Solomon, *The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses*, Springer-Verlag, New York, 1994.

[2] D. L. Farkas, B. Ballou, et al., "Optical image acquisition, analysis and processing for biomedical applications," *Springer Lecture Notes in Computer Science*, vol. 1311, pp. 663–671, 1997.

[3] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 245–271, 1997.

[4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, second edition, 2001.

[5] F. Kimura, K. Takashina, et al., "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Trans. on PAMI*, vol. 9, no. 1, pp. 149–153, 1987.

[6] W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation," *ICIP'95*, vol. 2, pp. 256–259, 1995.