CO-OCCURRENCE AND MORPHOLOGICAL ANALYSIS FOR COLON TISSUE BIOPSY CLASSIFICATION

Khalid Masood, Nasir Rajpoot, Hammad Qureshi

Department of Computer Science
University of Warwick
Coventry, CV4 7AL, UK

Kashif Rajpoot

Wolfson Medical Vision Lab
University of Oxford
UK

ABSTRACT

Diagnosis and cure of colon cancer can be improved by efficiently classifying the colon tissue cells from biopsy slides into normal and malignant classes. This paper presents the classification of hyperspectral colon tissue cells using morphology of gland nuclei of cells. The application of hyperspectral imaging techniques in medical image analysis is a new domain for researchers. The main advantage of using hyperspectral imaging is the increased spectral resolution and detailed subpixel information. The proposed classification algorithm is based on the subspace projection techniques. Support vector machine, with 3rd degree polynomial kernel, is employed in final set of experiments. Dimensionality reduction and tissue segmentation is achieved by Independent Component Analysis (ICA) and k-means clustering. Morphological features, which describe the shape, orientation and other geometrical attributes, are extracted in one set of experiments. Grey level co-occurrence matrices are also computed for the second set of experiments. For classification, kernel discriminant analysis (LDA) with co-occurrence features gives comparable classification accuracy to SVM using a gaussian kernel. The algorithm is tested on a limited set of samples containing ten biopsy slides and its applicability is demonstrated with the help of measures such as classification accuracy rate and the area under the convex hull of ROC curves.

1. INTRODUCTION

Colon cancer is a malignant disease of the large bowel. After lung and breast cancer, colorectal cancer (a combined term for colon and rectal cancer) is the most common cause of death for cancers in the Western world. The incidence of disease in England and Wales is about 30,000 cases/year, resulting in approximately 17,000 death/annum [11], and it has been estimated that at least half a million cases of colorectal cancer occur each year worldwide. It is caused by colonic polyps, an abnormal growth of tissue that projects in due course from the lining of the intestine or rectum, into colorectal cancer. These polyps are often benign and usually produce no symptoms. They may, however, cause painless rectal bleeding usually not apparent to the naked eye. The normal time for a polyp to reach 1 cm in diameter is five years or a little more. This 1 cm polyp will take around 5-10 years for the cancer to cause symptoms by which time it is frequently too late [10].

Diets low in fruits, less protein from vegetable sources, high age and family history are associated with increased risk of polyps. Persons smoking more than 20 cigarettes a day are 250 percent more likely to have polyps as opposed to nonsmokers who otherwise have the same risks. There is an association of cancer risk with meat, fat or protein consumption which appear to break down in the gut into cancer causing compounds called carcinogens [8]. Smoking cessation is important to decrease the likelihood of developing colon cancer. Dietary supplementation with 1500 mg of calcium or more a day is associated with a lower incidence of colon cancer. Weight reduction may be helpful in reducing the risk for colorectal cancer. Daily exercise reduces the likelihood of developing colon cancer. Turmeric, the spice which gives curry its distinctive yellow color, may also prevent colon cancer [5].

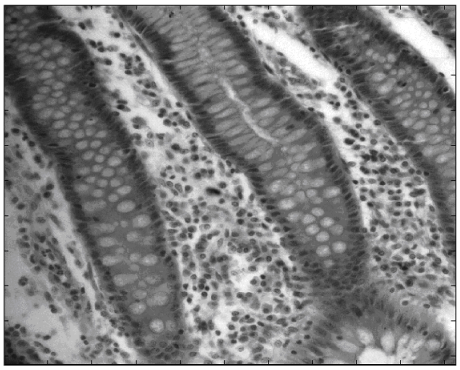
1.1. Hyperspectral Imaging

Hyperspectral imaging in laboratory experiments, is a non-contact sensing technique for obtaining both spectral and spatial information about a tissue sample. Hyperspectral imaging measures a spectrum for each pixel in an image. There are many types of spectroscopy which are being used to study the spectral signatures of individual cells and underlying tissue sections. In optical spectroscopy, which measures transmission through, or reflectance from, a sample by visible or near-infrared radiation at the same wavelength as the source, classification is done mostly by statistical measures [1].

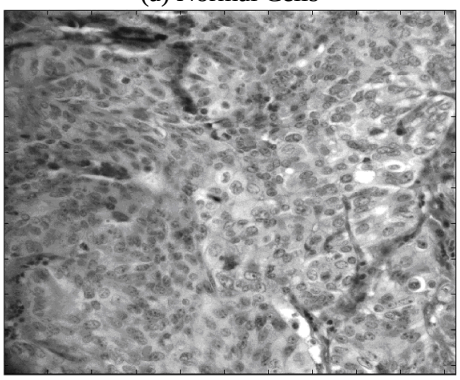
Hyperspectral images are normally produced by emission of spectra from imaging spectrometers. Spectroscopy is the study of light that is emitted by or reflected from

materials and its variation in energy with wavelength [9]. Spectrometers are used to make measurements of the light reflected from a test specimen. A prism in the centre of spectrometer splits this light into many different wavelength bands and the energy in each band is measured by detectors which are different for each band. By using large number of detectors (even a few thousand), spectrometers can make spectral measurements of bands as narrow as 0.01 micrometers over a wide wavelength range, typically at least 0.4 to 2.4 micrometers (visible through middle infrared wavelength ranges). Most approaches to analyse hyperspectral images concentrate on the spectral information in individual image cells, rather than spatial variations within individual bands or groups of bands. The statistical classification (clustering) methods often used with multispectral images can also be applied to hyperspectral images but may need to be adapted to handle high dimensionality.

Recent developments in hyperspectral imaging have enhanced the usefulness of the light microscope [3]. A standard epifluorescence microscope can be optically coupled to an imaging spectrograph, with output recorded by a CCD camera. Individual images are captured representing Y-wavelength planes, with the stage successively moved in the X direction, allowing an image cube to be constructed from the compilation of generated scan files. Hyperspectral imaging microscopy permits the capture and identification of different spectral signatures present in an optical field during a single-pass evaluation, including molecules with overlapping but distinct emission spectra. High resolution characteristics of hyperspectal imaging is reflected in two sample images in Figure 1 of colon tissue cells.



(a) Normal Cells



(b) Malignant Cells

Fig. 1. Colon Tissue Imagery

2. DIMENSIONALITY REDUCTION AND SUBSPACE PROJECTION

There is a large redundant information in the subbands of hyperspectral imagery. Independent component analysis (ICA) is used to discard the redundancy and extract the variance among different wavelengths of spectra. K-means clustering is used to help the dimensionality reduction procedure and to segregate the biopsy slide into its cellular components. Subspace projection is achieved with principal component analysis (PCA) and linear discriminant analysis (LDA). A brief introduction to the mathematical derivation of these methods is presented in the following subsections.

2.1. Independent Component Analysis (ICA)

The objective of Independent Component Analysis (ICA) is to perform a dimension reduction approach to achieve decorrelation between independent components [14]. Let us denote by $X = (x_1, x_2, \ldots, x_m)^T$ a zero-mean m-dimensional variable, and $S = (s_1, s_2, \ldots, s_n)^T$, $n < m$, is its linear transform with a constant matrix W [17]:

$$S = WX$$

Given X as observations, ICA aims to estimating W and S. The goal of ICA is to find a new variable S such that transformed components s_i are not only uncorrelated with each other, but also statistically as independent of each other as possible. An ICA algorithm consists of two parts, an objective function which measures the independence between components, entropy of each independent source or their higher order cumulants, and the second part is the optimisation method used to optimise the objective function. Higher order cumulants like kurtosis, and approximations of negentropy provide one-unit objective function. A decorrelation method is needed to prevent the objective function from converging to the same optimum for different independent components. Whitening or data sphering project the data onto its subspace as well as normalizing its v ariance.

2.2. K-Means Clustering

Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. This is done such that

patterns in the same cluster are alike and patterns belonging to two different clusters are different. The k-means method has been shown to be effective in producing good clustering results for many practical applications [2]. The aim of the k-means algorithm is to divide m points in n dimensions into k clusters so that the within-cluster sum of squared distance from the cluster centroids is minimised. The algorithm requires as input a matrix of m points in n dimensions and a matrix of k initial cluster centres in n dimensions. The number of clusters k is assumed to be fixed in k-means clustering. Let the k prototypes $(w_1, \ldots, w_k)$ be initialised to one of the m input patterns $(i_1, \ldots, i_m)$. Therefore;

$$w_j = i_l, j \in \{1, \ldots, k\}, l \in \{1, \ldots, m\}$$

The appropriate choice of k is problem and domain dependent and generally a user must try several values of k. The quality of the clustering is determined by the following error function:

$$E = \sum_{j=1}^{k} \sum_{i_l \in C_j} |i_l - w_j|^2$$

The direct implementation of k-means method is computationally very intensive.

2.3. Kernel Principal Component Analysis

PCA is a kind of linear transform, while Kernel PCA is a nonliner transform. The basic idea of KPCA [13] is based on the theory that by doing nonlinear mapping of the data points to a higher dimensional space, better features are obtained which is more natural and compact representation of the data. The computational complexity arising from the high dimensionality mapping is mitigated by using the kernel trick. Consider a nonlinear mapping $\Phi(.) : R^n \rightarrow R^f,\ f > n$ the space of n dimensional data points to some higher dimensional space R^f. So every point x_n is mapped to some point $\phi(x_f)$ in a higher dimensional space. After mapping, KPCA is nothing but linear PCA done on the points in the higher dimensional space. Denoting a $m \times m\ matrix\ K\ by$

$$K_{ij} = k(x_i, x_j) = \Phi(x_i).\Phi(x_j)$$

the kernel PCA problem becomes

$$m\lambda k\alpha = k^2\alpha \equiv m\lambda\alpha = k\alpha$$

where α denotes a column vector with entries $\alpha_1, \cdots, \alpha_m$. The projection vectors in R^f to a lower dimensional space spanned by the eigenvectors w^Φ is the nonlinear principal components corresponding to Φ:

$$w^\Phi.\Phi(x) = \sum_{i=1}^{m} \alpha_i(\Phi(x_i)).\Phi(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x)$$

hence the first m nonlinear principal components are extracted without the expensive operation of high dimensional projection of the data.

2.4. Kernel Linear Discriminant Analysis

Similar to KPCA, mapping is performed on the input space to the high dimensional feature space with linear properties [13]. In the new space, the problem is solved in a classical way in the transformed space using the kernel operators.Denoting the wihin-class and between-class matrices by S_w and S_b and applying FDA in kernel space, the solution of the equation below will give the eigenvalues and eigenvectors w ;

$$\lambda S_w^\Phi w^\Phi = S_B^\Phi w^\Phi$$

which can be obtained by

$$W_{OPT}^\Phi = argmax_{w^\Phi} \frac{|(W^\Phi)^T S_B^\Phi W^\Phi|}{|(W^\Phi)^T S_W^\Phi W^\Phi|}$$

$$= [w_1^\Phi, \cdots, w_m^\Phi]$$

. Consider a c-classs problem and let the r_{th} sample of class t and s_{th} sample of class u be x_{tr} and x_{us} respectively. The kernel function is defined as

$$(k_{rs})_{tu} = k(x_{tr}, x_{us}) = \Phi(x_{tr}).\Phi(x_{us})$$

. Let K be a $m \times m$ matrix defined by the elements $(K_{tu})_{u=1,\cdots,c}^{t=1,\cdots,c}$ where K_{tu} is a matrix composed of dot products in the feature space R^f,i.e., $K = (K_{tu})_{u=1,\cdots,c}^{t=1,\cdots,c}$ where $K_{tu} = (k_{rs})_{s=1,\cdots,l_u}^{r=1,\cdots,l_t}$. Note K_{tu} is a $l_t \times l_u$ matrix, and K is a $m \times m$ symmetric matrix. We can also define a matrix Z :

$$Z = (Z_t)_{t=1,\cdots,c}$$

where $Z_t$is a $m \times m$ block diagonal matrix. The between-class and within-class scatter matrices in a high dimensional feature space R^f are defined as;

$$S_B^\Phi = \sum_{c=1}^{C} l_i \mu_i^\Phi (\mu_i^\Phi)^T$$

$$S_w^\Phi = \sum_{i=1}^{C} \sum_{j=1}^{l_i} \Phi(x_i j)(\Phi(x_i j)^T$$

where μ_i^Φ is the mean of class i in R^f, l_i is the number of samples belonging to class i. From the theory of reproducing kernels, any solution$w^\Phi \in R^f$ must lie in the span of all training samples in R^f,i.e.

$$w^\Phi = \sum_{p=1}^{C} \sum_{q=1}^{l_p} \alpha_{pq} \Phi(x_{pq})$$

3. METHODOLOGY

The proposed classification algorithm consists of three modules as shown in Figure 2. Brief description of dimensionality reduction and feature extraction modules is given in the following sub-sections. Detailed description of the segmentation can be found in [12].

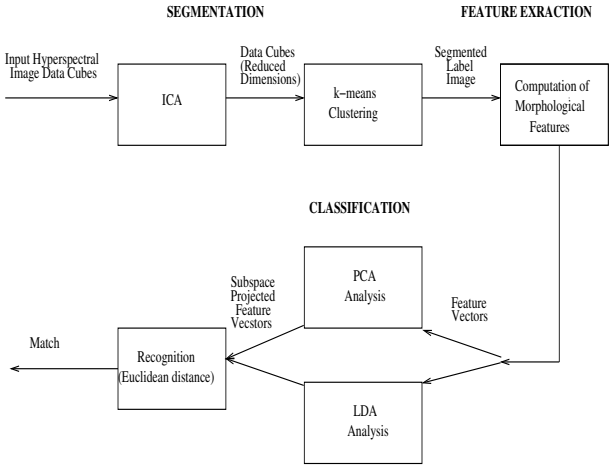
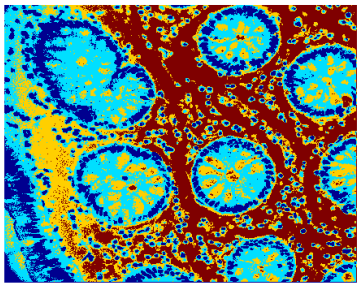
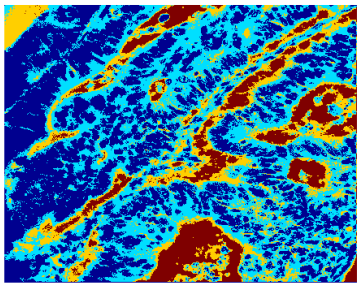


Fig. 2. Classification Algorithm Block Diagram



(a) Benign Cells



(b) Malignant Cells

Fig. 3. Segmentation Results

3.1. Segmentation

High dimensional data in the form of 3-D cubes is obtained using hyperspectral imaging. For efficient processing this data has to be dimensionally reduced. Dimensionality reduction involves two steps, extraction of statistically independent components using Independent Component Analysis (ICA) and colour segmentation using k-means clustering. Flexible ICA (FlexICA) [6], a fixed point algorithm for ICA, adopting a generalised Gaussian density, is used for data sphering (whitening) and achieves considerable dimensionality reduction. Data is distributed towards heavy-tailedness by the high-emphasis filters. The data with reduced dimensionality is then fed to k-means clustering algorithm for segmentation.

The hyperspectal data cube containing 28 subbands is segmented into four labeled parts. Each slide of the tissue cells is divided into four regions represented by four colours as shown in Figure 3. The four labeled parts are denoted by colours as dark blue for nuclei, light blue for cytoplasm, yellow for gland secretions and red for lamina propria.

3.2. Feature Extraction

3.2.1. Morphological Features

In order for the pattern recognition process to be tractable it is necessary to represent patterns into some mathematical or analytical model. The model should convert patterns into features or measurable values, which are condensed representations of the patterns, containing only salient information [7]. Features contain the characteristics of a pattern to make them comparable to standard templates making the pattern classification possible. The extraction of good features from these pattern models and the selection from them of the ones with the most discriminatory power are the basis for the success of the classification process. In this work morphological texture features, extracted from the segmented images of a hyperspectral data cube for a biopsy slide of colon tissue cells, are used for the classification of the tissue cells.

Morphological features, which describe the shape, size, orientation and other geometrical attributes of the cellular components, are extracted to discriminate between two classes of input data. The segmented image is first split into four binarised image in accordance with the four cellular components. In each binary image, the corresponding cellular components i.e. nuclei, cytoplasm, gland secretions and stroma of lamina propria have binary value equal to 1.

3.2.2. Co-occurnence Features

The co-coourrence approach is based on the grey level spatial dependence. Co-occurrence matrix is computed by second-order joint conditional probability density function $f(i, j|d, \theta)$. Each $f(i, j|d, \theta)$ is computed by counting all pairs of pixels separated by distance d having grey levels i and j, in the given direction θ. The angular displacement θ usually takes on the range of values from $\theta = 0, 45, 90, 135$ degrees. The co-occurrence matrix captures a significant amount of textural information. The diagonal values for a coarse texture are high while for a fine texture these diagonal values are scattered. To obtain rotataion invariant features the co-occurrence matrices obtained from the different directions are accumulated. The three set of attributes used in our experiments are Energy, Inertia and Local Homogeneity.

$$E = \sum_i \sum_j [f(i, j|d, \theta)]^2$$

$$I = \sum_i \sum_j [(i - j)^2 f(i, j|d, \theta)]$$

$$LH = \sum_i \sum_j \frac{f(i, j|d, \theta)}{1 + (i + j)^2}$$

4. EXPERIMENTS

4.1. Experimental Setup

The experimental setup consists of a CRI Nuance microscope and a CCD camera. Two different biopsy slides containing several microdots, where each microdot is from a distinct patient, is prepared. Then each slide is illuminated with a tuned light source (capable of emitting any combination of light frequencies in the range of 450-850 nm), followed by magnification to 400 X. Thus several images, each image using a different combination of light frequencies, are produced [4].

The first set of experiments is carried out with mixed training/test data. Each image is divided into 4096 patches of 16×16 dimensions per patch. Morphological operation is performed on the patches for extraction of feature vectors using different combinations of ten scalar morphological properties. The data (patches of all slides randomly mixed) is divided into training set (about one quarter of the patches) and test set (remaining three quarters of patches). In the other experiment, multiscale feature extraction is performed. Feature values are initially calculated for base patch size 16×16, patch size is then doubled and feature values are re-calculated. This process continues for at least upto five scales. Fusion of the features, for different scales, is done by simple concatenation. Thus largest feature vector for five scales and using ten morphological parameters has dimensions of 200 fetures values.

Classification Accuracy			
Method	**Features**	**Experiment**	**Accuracy (%)**
PCA	Morphological	Single scale	55
PCA	Morphological	Multiscale	75
LDA	Morphological	Single scale	65
LDA	Morphological	Multiscale	84
PCA	GLCM	LOO	70
LDA	GLCM	LOO	80
SVM (poly)	GLCM	LOO	90
SVM (gaussian)	GLCM	LOO	100

Table 1. Classification Results

4.2. Experiments with Leave one out data

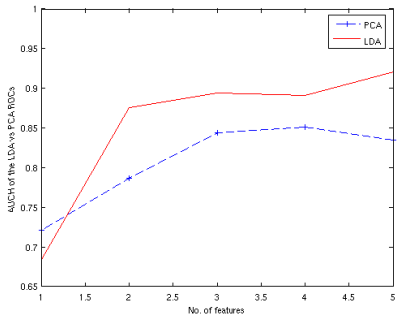
The second set of experiments are used with LOO (leave one out) settings and employing two different subsets. In the second setting, co-occurrence matrix is computed from the block size of 64x64 for each slide. Three co-occurrence features i.e. Angular second moment (Energy), variance and homogenity are calculated while pixel distance is varied from one pixel to two pixel values. Four directional features in the direction of 0, 45, 90 and 135 are concanated together so that feature vector with 24 dimensions is used in the classifiers. The last experiment is carried out with SVMs. Polynomial kernel of 3rd degree with parameters C=1 (cost of constrain violation), epsilon=0.001 (tolerance of termination criterion), and coefficient=0 is used in this exercise. Experiments with SVM using Gaussian kernel have the best classification accuracy of 100 percent for a threshold of 60 percent coorect patches.

4.3. Results

The classification accuracy in GLCM-LOO is about 90 percent and 9 slides on the whole are classified correctly with a threshold of 55 percent on the patches. Directly fed patches have a little less performance as compared to the co-occurrence features from the patches. As we have only limited number of input slides, so comparison is difficult on these set of experiments. Using new data with these setting will give better comparison for the classification accuracy.

5. CONCLUSIONS

In this paper, classification of colon tissue cells is achieved using the morphology of the glandular cells of the tissue region. There is an indication that the morphology of the cells, obtained from the hyperspectral analysis of biopsy slides, has strong discriminatory power. Regular structured cell shapes with some orientations are characteristics of normal cells, whereas irregular and deformed cell shapes represent



(b) AUCH of LDA vs PCA

Fig. 4. AUCH Performance Curves

malignant tissue. In morphological analysis, five features' subset is used to achieve 80 percent accuracy. In the second set of experiments with gray level co-occurrence matrix and using a feature vector of 24 dimensions, reasonable classification is performed even with simple classifiers like LDA. However, employing properly tuned Gaussian kernel with grid search method, accuracy level as good as 100 percent can be achieved.

Acknowledgments

The authors are greatly indebted to Gustave Davis, Mauro Maggioni and Ronald Coifman of the School of Medicine and the Applied Mathematics Department of Yale University for providing the data and for many fruitful discussions.

6. REFERENCES

[1] John Adams, M. Smith, and A. Gillespie. Imaging spectroscopy: Interpretation based on specral mixture analysis. *Remote Geochemical Analysis*, 1993.

[2] K. Alsabti, S. Ranka, and V. Singh. An efficient k-means clustering algorithm. *www.cise.ufl.edu.*, 1997.

[3] E. A. Cloutis. Hyperspectral geological remote sensing. *Evaluation of Analytical Techniques-International Journal of Remote Sensing*, 17:2215–2242, 1996.

[4] G. Davis, M. Maggioni, and R. Coifman et al. Spectral/spatial analysis of colon carcinoma. *Journal of Modern Pathology*, 2003.

[5] R. S. Houlston. Molecular pathology of of colorectal cancer. *Clinical Pathology*, 2001.

[6] A. Hyvarinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999.

[7] Anil Jain, Robert Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22, 2000.

[8] S. Kaster, S. Buckley, and T. Haseman. Colonoscopy and barium enema in the detection of colorectal cancer. *Gastrointestinal Endoscopy*, 1995.

[9] David Landgrebe. Hyperspectral image data analysis as a high dimensional signal processing problem. *IEEE Signal Processing magazine*, 2002.

[10] D. E. Mansell. Colon polyps & colon cancer. *American Cancer Society Textbook of Clinical Oncology*, 1991.

[11] Office of National Statistics. Cancer statistics: Registrations, england and wales. london. *HMSO*, 1999.

[12] K. M. Rajpoot and Nasir M Rajpoot. Hyperspectral colon tissue cell classification. *SPIE Medical Imaging (MI)*, 2004.

[13] N. Rajpoot and K. Masood. Human gait recognition with 3-d wavelets & kernel based subspace projections. *International Workshop on HAREM*, 2005.

[14] Kun Zhang and Lai-Wan Chan. Dimension reduction based on orthogonality-a decorrelation method in ica. *ICANN/ICONIP*, 2003.