Annals of the BMVA Vol. 2008, No. 4, pp 1–16 (2008)

Spatial Analysis for Colon Biopsy Classification from Hyperspectral Imagery

Khalid Masood and Nasir Rajpoot

Department of Computer Science, University of Warwick
⟨khalid,nasir@dcs.warwick.ac.uk⟩

Abstract

Automatic classification of histology images, the objective of our research, is aimed at supporting the pathologists in their diagnosis. In this paper, we present a comparative study between 3D spectral/spatial analysis (SSA) and 2D spatial analysis (SA) for the classification of colon biopsy samples from their hyperspectral images. Classification results suggest that textural analysis of 2D bands can achieve comparable performance to 3D spectral/spatial analysis.

1 Introduction

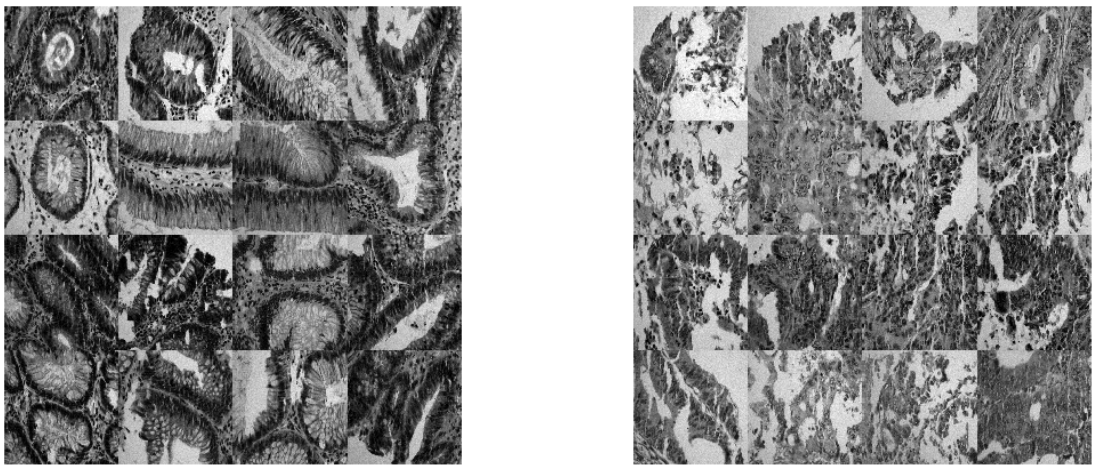
Colon cancer is the second most common cause of death from cancer in the Western and industrialised world Houlston [2001]. Early diagnosis of colon cancer can lead to an effective treatment. Regular screening is, therefore, very useful and particularly recommended for people aged over 50 years Ferrandez and Disario [2003]. Routine screening also means an increased workload for the pathologists, and hence the need for development of robust automated techniques to assist the manual diagnosis by pathologists. Automated techniques are also necessary because there is a significant observational difference in the diagnosis of pathologists – due to subjectivity, fatigue or tiredness – which can potentially result in different treatment plans Buhmann et al. [2007]. With the development of a computerised classification algorithm, we are making an attempt to assist pathologists in their diagnosis so as to benefit from automated classification.

Computational histopathology deals with the automated analysis of images from histology sections of tissue biopsy samples Rajpoot et al. [2008]. In most cases, the task is to label a slide or grade different stages of cancer. One particular case is the labelling of a biopsy slide as normal or malignant, depending upon the textural contents or the overall macro-architecture of the biopsy image. A few researchers have proposed solutions for the automatic classification of colon biopsy slide images. Todman *et al.*Todman et al. [2001] proposed

that metrics based on the responses of receptive field operators model the orientational selectivity of neurons. These operators are found in the early visual pathway and are capable of discriminating between images of normal, dysplastic (transitional) and cancerous samples. Their results, however, do not reflect the true applicability of their method due to the limited size of dataset. The approach in Maggioni et al. [2006] uses hyperspectral images of colon biopsy slides. The proposed algorithm is based on spectral analysis to discriminate between normal and cancerous biopsies of the colon tissue. Tissue segmentation is supervised, in the sense that different points in a biopsy image are labelled manually. The proposed algorithm automatically assigns a class to each labelled location by using Laplacian eigenmaps. In each biopsy slide, 40 to 60 gland nuclei are obtained after segmentation. If a pre-determined fraction of nuclei is malignant, then the whole slide is classified as cancerous. Hidović-Rowe *et al.*Hidović-Rowe et al. [2006] have presented an analysis of the colon tissue reflectance spectra. Using their model, they extracted changes in parameters like blood volume fraction, haemoglobin saturation, the size of the collagen fibres and layer thickness, which are characteristic of the onset of colon cancer. Despite the promise of some of the above methods, the required accuracy for automated systems is still not achieved. Most of these approaches need large refinements to turn them into an applicable robust automated system.

In this paper, we address the following two questions: (*a*) Does the textural analysis on a single spectral band achieve comparable performance to the existing spatial analysis on 3D hyperspectral data, and (*b*) Does textural information alone suffice to produce a reasonable classification accuracy? Hyperspectral imaging is widely used in remote-sensing applications due to its highly specific characterisation of different kinds of land cover. In Boucheron et al. [2007], a comparison is performed between multispectral and RGB data for nuclei classification of breast tissue. The authors' experiments involved multispectral images with 29 bands from a total of 58 H&E stained biopsy samples. Using SVM classifiers, they conclude that multispectral bands do not contain much more discriminatory spectral information than do the RGB bands for nuclei classification of breast tissue. The study in Boucheron et al. [2007], however, is limited to the classification of nuclei in breast histology images. In this paper, we present a study based on the comparison of two approaches: 3D spectral/spatial analysis (SSA) and 2D spatial analysis (SA). This study compares the results of textural analysis on single hyperspectral band to 3D spectral spatial analysis performed by Rajpoot & Rajpoot Rajpoot and Rajpoot [2004]. First, we employ a relatively recent development in subspace projection, namely 2D principal component analysis (2DPCA). This approach uses a single spectral band (2D image) to help answer the second question above as to how important it is to present a pattern in a characteristic feature vector form. In SSA, as established in the previous studies Masood and Rajpoot [2007], Rajpoot and Rajpoot [2004], morphological features are extracted, while in SA, circular local binary pattern (CLBP) features are used to test the efficacy of textural analysis. To save computational cost, we adopt an innovative feature selection algorithm for the selection of morphological and textural features. We investigate whether similar performance can be achieved with textural analysis on a carefully selected spectral band to those analyses which use 3D hyperspectral data Maggioni et al. [2006], Rajpoot and Rajpoot [2004]. If this is possible, a computationally efficient algorithm can be used for the classification of colon biopsy samples.

The remainder of this paper is organised as follows. Section 2 describes the hyperspectral data in SSA and introduces the various steps of SSA and SA. A novel feature selection algorithm is also described in this section. In subsection 2.4, 2DPCA and SVM are briefly described while details of PCA and LDA can be found in Zhao et al. [1996] and Belhumeur



(a) Benign tissue samples (b) Malignant tissue samples

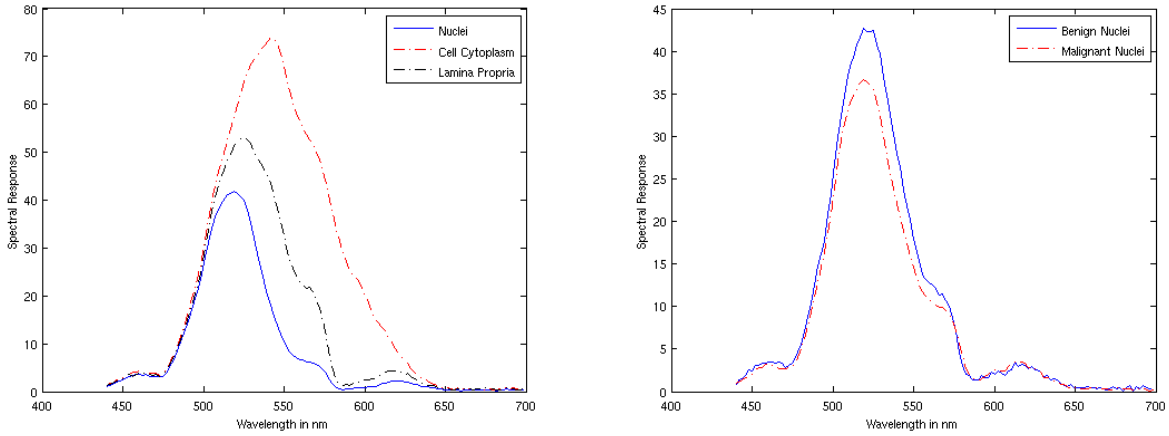
Figure 1: 70th spectral band (588 *nm*) for colon tissue biopsy samples at 40X magnification using Hematoxylin & Eosin ($H\&E$) staining.

et al. [1997], respectively. In Section 3, experimental results are presented and the two approaches are compared. Finally, conclusions are drawn from this study and a future road map is presented in Section 4.

2 Materials and Methods

Our experimental dataset consists of hyperspectral cubes of colon biopsy samples on a tissue micro-array. Figure 1 shows the 70th spectral band from the available 128 spectral bands of the dataset. Biopsy samples are taken from 32 different patients and are fixed in formalin. Tissues are embedded in paraffin and cut with a microtome. Hematoxylin & Eosin ($H\&E$) are used for staining. Typically, the nuclei stain dark blue, while the cytoplasm stains pink or orange. The hyperspectral imaging setup consists of a tuned light source based on a digital mirror device which transmits any combination of light frequencies Davis et al. [2003]. The tuned light source generates 128 wavelengths within the visual range of $440 - 700$ *nm* with a resolution of 2 *nm*. A Nikon Biophot microscope with a CCD camera is used for 40X magnification. Hyperspectral pictures of tissues obtained with a CCD camera are captured by a computer and a single data cube is collected for each biopsy. Each image cube has a spatial resolution of $491 \times 652 \times 128$ pixels.

Figure 2(a) shows the spectral responses of nuclei, glandular cytoplasm and lamina propria of a benign tissue using all 128 spectral bands. In Figure 2(b), spectral responses of benign and malignant nuclei are presented. It can be observed from this figure that the spread of benign and malignant nuclei is not very different: in fact the spectra of benign nuclei have higher peaks than those of malignant nuclei. It can be said that hyperspectral data is able to discriminate between different tissue parts but the spectral response for the two nuclei is not distinguishable. This motivates us to perform spatial analysis on this data



(a) Benign tissue components

(b) Benign and Malignant tissue nuclei

Figure 2: Spectral response of hyperspectral images.

for achieving reasonable classification results.

A block diagram of the two analyses is presented in Figure 3. In SSA, dimensionality reduction and segmentation are pre-requisite as suggested in Rajpoot and Rajpoot [2004]. ICA and *k*-means clustering are used for this purpose. Segmentation is followed by feature extraction and selection and different classifiers are used in the final stage. While in SA, feature extraction and selection are employed on a selected spectral band. Textural analysis using circular local binary patterns (CLBP) features is followed by same set of classifiers which are used in SSA.

2.1 SSA

In case of SSA, the first step is to reduce the dimensionality of hyperspectral data to cater for any redundant information in the bands of hyperspectral imagery. Independent component analysis (ICA) is used to discard the redundancy among different wavelengths of the spectra. *k*-means clustering is used to segregate the biopsy image into its tissue components. A brief introduction to ICA and *k*-means clustering is presented in the following subsections.

2.1.1 Independent Component Analysis (ICA)

The objective of ICA is to perform dimensionality reduction to achieve decorrelation between independent components Hyvarinen [1999]. Let $X = (x_1, x_2, \ldots, x_n)^T$ be a zero-mean n-dimensional variable and $S = (s_1, s_2, \ldots, s_m)^T$, $m < n$, be its linear transform with an unmixing matrix W given by:

$$S = WX \tag{1}$$

Given X as observations, ICA aims to find the linear transformation so that the components S are as independent as possible. ICA is a higher order method that seeks linear projections, not necessarily orthogonal to each other, as in PCA. For hyperspectral data, PCA is not suitable for dimensionality reduction as the data is highly correlated Rajpoot and Rajpoot [2004].

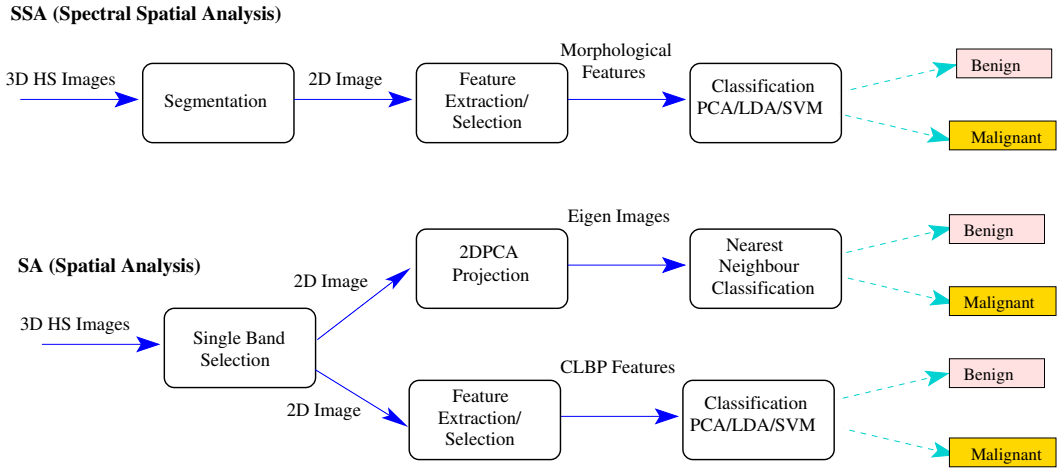


Figure 3: Block diagram of SSA & SA.

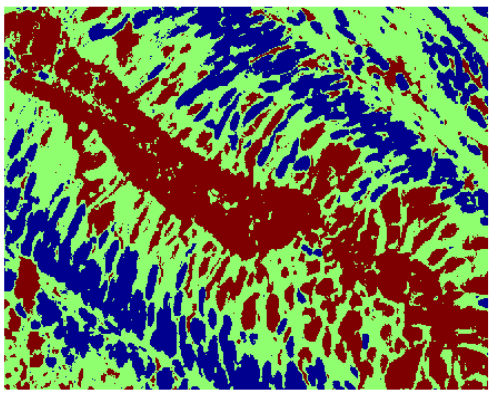
2.1.2 *k*-Means Clustering

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 different clusters are different. The *k*-means method has been shown to be effective in producing good clustering results for many practical applications Alsabti et al. [1998]. The aim of the *k*-means algorithm is to divide N points in m dimensions into k clusters so that the within-cluster sum of squared distance from the cluster centroids is minimised. The *k*-means algorithm is an iterative procedure which uses the centroid, the average location of a particular cluster, to discover structure in the data.

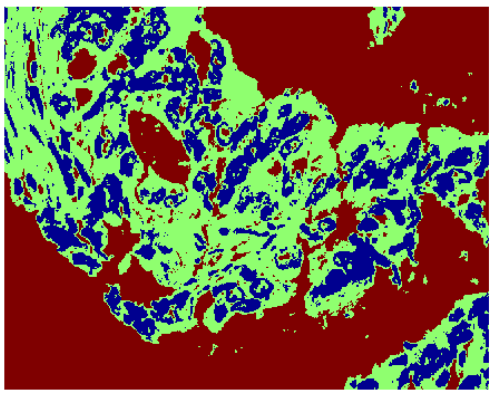
The goal of ICA is to recover components, which are as statistically independent from each other as possible. Considering the morphology of the tissue, we extracted three tissue components using ICA. Applying *k*-means clustering with $k = 3$, we get segmentation of the tissue consisting of nuclei (dark blue), glandular cytoplasm (green) and lamina propria (dark red), as shown in Figure 4.

2.1.3 Morphological Features

The feature vector in our SSA consists of morphological parameters, as suggested in Rajpoot and Rajpoot [2004]. Morphological features, which describe the shape, size, orientation and other geometrical attributes of the cell components, are extracted from the segmented image. At the end of the segmentation phase, we get three images for three different components of the tissue. From these images, a binary image containing only the cell nuclei is selected and feature extraction is applied on it. Our feature vector contains 11 morphological features: Euler number, Area, Extent, Convex area, Orientation, Solidity, Eccentricity, Equivalent diameter, Perimeter, Major Axis length and Minor axis length constitute the feature vector. As explained later in Section 2.3, the feature selection algorithm selects two feature sets, $Morph_1$ with 5 features (Euler number, Area, Extent, Orientation and Eccentricity) and $Morph_2$ with 3 features (Euler number, Area and Extent). Euler number, the key feature of morphological feature vector, was found to be the most discriminant feature in a separate set of experi-



(a) Benign tissue



(b) Malignant tissue

Figure 4: Segmentation using ICA and k-means clustering. Three different colors correspond to labels for three tissue parts: nuclei (dark blue), cytoplasm (light green) and lamina propria (dark red).

ments. It is equal to the number of objects (a set of foreground pixels) minus the number of holes (a set of background pixels) in those objects Bishnu et al. [2001]. It is defined as:

$$E = N(P - B) \tag{2}$$

where E is the Euler number, P is the set of foreground pixels and B is the number of holes. N is the number of connected components in a 4-connected or 8-connected neighbourhood.

The remaining four features of our $Morph_1$ feature set are: Area, Extent, Orientation and Eccentricity. Area estimates the total area of objects in the binary image and it is equal to the total number of pixels with a value of 1. The last three parameters are related to the spherical measure of a pattern and determine its shape. Extent is the radius of the smallest circle centered at one of the points on a closed shape which contains all the points in that shape. Orientation is related to the angle of the closed shape with the horizontal axis. Eccentricity of an ellipse tells about the flatness of an ellipse. An ellipse is defined as follows:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{3}$$

where x and y are set of the points, a is the half-width of the ellipse and b is its half-height. Eccentricity (ecc) depends on the ratio of width and height of the ellipse and is defined as follows:

$$ecc = \sqrt{1 - \left(\frac{b}{a}\right)^2} \tag{4}$$

In case of a circle, the eccentricity is zero. These features represent the pattern as an elliptical or circular shape which is true in our case as in most cases, the shape of the gland is almost elliptical.

2.2 SA

In our earlier work Masood and Rajpoot [2007], 70th spectral band (588 *nm*) showed greater classification accuracy than other spectral bands. So this band is selected and the feature extraction algorithm using circular local binary pattern (CLBP) feature vector is applied on it. The key component of SA is the feature extraction algorithm based on CLBP features.

2.2.1 Textural features using CLBP

Texture is based on two factors: texture patterns and strength of the patterns Ojala et al. [2002]. Texture can thus be regarded as a two-dimensional phenomenon characterised by two attributes, spatial structure (pattern) and contrast (the strength of the pattern). Texture in digital images can be described as variations in the captured intensities or colour. Although in general there is no information on the cause of variations, differences in image pixel intensities provide a practical means of analysing the properties of objects. There is no formal definition for texture but it can be categorised with attributes such as repetitive and directional patterns, high contrast and low contrast patterns and coarse or fine grained complex textures. Other features for texture are coarseness, edge and orientation.

The basic LBP operator is a gray scale invariant pattern measure characterising the texture in images Maenpaa et al. [2000]. In this method, texture is defined using local patterns on a pixel level. Each pixel is labelled with the code of the texture primitive that best matches the local neighbourhood. First, the centre pixel value is taken as a threshold and n neighbours of the pixel are selected. Then, each neighbouring pixel is given a weight based on its position and these weights are multiplied by the threshold values to generate basic LBP code. In the circular LBP (which is used in our algorithm), symmetric neighbours in a circle are used for a particular radius. Normally 8, 12 and 16 neighbourhoods are used. By taking 8 neighbours, 59 parameters represent the local variations in a region, whereas 12 and 16 neighbours result in 133 and 243 parameters, respectively. By using the highly unifying approach of LBP, texture in biopsy samples distinguishes benign tissue samples from malignant samples.

2.3 Feature Selection

In order for a pattern recognition system to be tractable, it is necessary to represent the given patterns in some mathematical or analytical model Jain et al. [2000]. The model should convert patterns into features or measurable values, which are condensed representations of the patterns, containing only salient information. Feature selection is an important task in machine learning, pattern recognition and data mining. It leads to saving in computational cost, since some of the features are discarded and the selected features may retain their original physical interpretation and may give a better understanding of the original data. In addition, the retrieved features may improve the classification accuracy by providing better discrimination ability.

Our feature selection algorithm is based on three measures related to the quality of clustering. These measures determine the separability and compactness of the clusters. First, Classification Scatter Index (CSI) C_o is a measure of compactness of clusters formed by a set

of features S. Lee et al. [2007]:

$$C_o = \sum_{j=1}^{m} p_j \sigma_j \tag{5}$$

where p_j is the prior probability of jth class, σ_j is the standard deviation of jth class and m denotes the number of clusters, two in our case. We scale the values of C_o between 0 and 1, with 1 representing the most compact clustering, as follows:

$$C = e^{-aC_o^2} \tag{6}$$

where a is a positive constant.

Second, the Rand index is a measure of similarity between two different clusters Hubert and Arabie [1985]. It ranges from zero when the two clusters are not similar to unity when the clusters are exactly the same. The adjusted Rand index provides a standardised measure such that its expected value is zero when the partitions are selected at random, and unity, when the partitions match completely. Suppose there are two partitions G_1 and G_2 of the same data having g_1 and g_2 clusters, respectively. The Rand index computes the proportion of n by 2 objects, n being the total number of points that belong to both the partitions. It is defined as follows:

$$R = \frac{\binom{n}{2} + \sum_{i=1}^{g_1}\sum_{j=1}^{g_2} n_{ij}^2 - \frac{1}{2}\sum_{i=1}^{g_1}\left[\sum_{j=1}^{g_2} n_{ij}\right]^2 - \frac{1}{2}\sum_{j=1}^{g_2}\left[\sum_{i=1}^{g_1} n_{ij}\right]^2}{\binom{n}{2}} \tag{7}$$

where n_{ij} is the number of points in cluster i of G_1 that also belong to cluster j of G_2.

Third, the Silhouette index is a measure for each point in the clustering as to how similar that point is to points in its own cluster and to points in other clusters Aranganayagi and Thangavel [2007]. Its value ranges between -1 to 1. It is defined as follows:

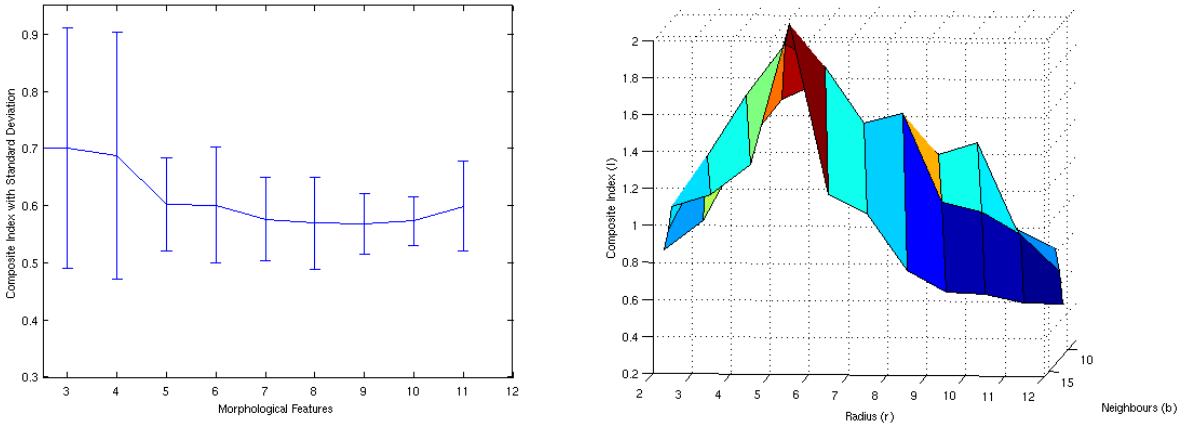
$$S(i) = \frac{min[D_b(i,k)] - D_w(i)}{max\left[D_w(i), min[D_b(i,k)]\right]} \tag{8}$$

where $D_w(i)$ is the average distance from the i-th point to the other points in its own cluster, while $D_b(i,k)$ is the average distance from the i-th point to points in another cluster k (it is the minimum average dissimilarity between the i-th element and any other cluster not containing the element). The silhouette coefficient S is the average of the coefficients S_i for all the points.

Taking the three indices described above, we compute a composite index I as the weighted average of C, R and S. The weights are calculated based on correlation coefficients of the indices with the clustering accuracy achieved by the k-means algorithm. Composite index determines the basis for optimum feature selection and is defined as follows:

$$I = r_c C + r_r R + r_s S \tag{9}$$

where r_c, r_r and r_s are correlation coefficients for C, R and S respectively. In SSA, a feature selection algorithm is employed for 9 different combinations of morphological parameters. The two best combinations, 3 and 5 features, are selected for classification experiments as shown in Figure 5(a). In Figure 5(a), Composite scatter index and its standard deviation is



(a) Composite Index for Morphological Features (b) Composite Index surface for CLBP Features

Figure 5: Composite Scatter Index.

plotted for various combinations of morphological parameters. We did not select the second best combination of the parameters as its standard deviation is quite high. In SA, CLBP(r, b) features are computed for 33 different combinations of radius r and number of neighbours b, with $r \in \{2, 3 \ldots, 12\}$ and $b \in \{8, 12, 16\}$. Figure 5(b) shows the surface obtained for I after sweeping through the ranges of r and b. We pick the best features sets that give the two highest values of composite indices, namely CLBP$(5, 8)$ and CLBP$(5, 12)$.

2.4 Classification Methods

2.4.1 Two-dimensional Principal Component Analysis (2DPCA)

Two-dimensional principal component analysis (2DPCA) Yang et al. [2004] is a recently proposed subspace projection method that has been successfully applied to face recognition. In conventional PCA, the image matrix is transformed into 1D vectors for covariance matrix, but in 2DPCA there is no need for transformation as the image covariance matrix is constructed directly from the training image matrices. This results in gaining two advantages over conventional PCA: the covariance matrix is evaluated accurately and it takes less computational time as the size of the covariance matrix is small.

For a given image I to be projected onto a subspace using a projection vector X, the following linear transformation is used:

$$P = IX \tag{10}$$

For an effective projection, the total scatter of the projected samples should be maximised as characterised by the trace of the covariance matrix of the projected feature vectors. To evaluate the covariance matrix C, image samples are not transformed into vectors as in the case of 1DPCA, instead original images are used. The covariance matrix C can be computed as follows:

$$C = \frac{1}{M} \sum_{i=1}^{M} (I_i - I_{av})^T (I_i - I_{av}) \tag{11}$$

where M denotes the number of training samples, I_i and I_{av} denote the ith training image and the average of all training images, respectively.

Projections of 2DPCA are matrices as opposed to vectors in case of PCA. For each image sample I, feature vectors are extracted by projection of k components onto the sample I:

$$P_k = IX_k, \;\; k = 1, 2, \ldots, d \tag{12}$$

where X_k are eigenvectors of the covariance matrix C corresponding to the d largest eigenvalues. A feature matrix $A = [P_1, \ldots, P_d]$ is built for each training image. For classification, the feature matrix of the test image is compared with feature matrices of training images and based on the sums of the Euclidean distances between corresponding rows, the test image is assigned the label of its nearest neighbour. 2DPCA is based on image matrix, so it is easy and simple to extract image features and component vectors are more accurate.

2.4.2 Support Vector Machine (SVM)

The support vector machine (SVM) introduces a nonlinear mapping from the input space to an implicit high-dimensional feature space, where the nonlinear boundary between the patterns in the input space is linearised Vapnik [1995]. The kernel function in SVM provides the possibility to avoid explicit evaluation of mapping. There is a wide range of available kernel functions but polynomial, Gaussian and sigmoidal kernel are used more often than others. The choice of kernel depends on the data and its clustering. A Gaussian kernel is used in most cases because of its superior performance. It is defined as follows:

$$K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2} + C \tag{13}$$

where parameter C is constant for a given data and does not affect the tuning of the SVM. The other parameter for the Gaussian kernel is the width γ of the basis function. Its selection depends on the input data and a variety of selection methods are used including grid search and Newton bisection methods Rajpoot and Rajpoot [2004]. Once a kernel is tuned properly, classification of the test data can be performed accurately.

3 Experimental Results and Discussion

Our dataset consists of 32 hyperspectral cubes of colon biopsy samples, taken from uniquely different patients, of which 24 samples are used for training and the remaining 8 samples comprise the testing set. Four-fold cross-validation is performed by repeating this process four times, varying the composition of the training and test datasets each time.

The first set of experiments is for SSA and six experiments are performed altogether using PCA, LDA and SVM. A morphological feature vector is used in these classifiers. For the first three experiments, the feature vector consists of five parameters while three features are used for the next three experiments. A selection of principal components for PCA is done empirically. Different numbers of components are tried, and based on the classification accuracy, two to four eigen components are selected. If there are two classes only, then LDA uses only one principal component for projection of the data samples. This results in poor performance of LDA as compared to PCA. To avoid this, we have used a modular LDA approach Huang et al. [2002]. The idea is to virtually increase the number of classes

and introduce sub-classes in the training set. This is possible only if the within-class scatter is large and there are enough training samples, which holds true in our case. Hence we divide our two main classes into four sub-classes. The first two sub-classes represent benign samples and the remaining two classes represent malignant samples. For SVM, a Gaussian kernel is used and it is tuned for different values of the kernel width parameter γ. The Newton bisection method is used for tuning. This is an iterative algorithm. For a range of γ, classification experiments are performed. The best γ is picked and subsequent iterations are based around this value. Finally, when we reach the convergence point for a threshold at which the value of γ is not changed, that γ represents a properly tuned kernel. Once a kernel is tuned by training, the same parameters are used in the testing stage.

Experimentation in SA starts with 2DPCA. No feature extraction process is involved and pixel values in each image represent features. Different numbers of eigen components are tested and finally 10 eigen components are selected based on their classification accuracy. For next six experiments, CLBP$(5,12)$ and CLBP$(5,8)$ are used as explained in the Feature selection algorithm. Experiments in SA are carried out using the same set of classifiers which were used in SSA. SVM is re-tuned as the feature vector using CLBP is larger in size than the feature vector of SSA.

Table 1 presents the classification accuracies and ROC performance measures for different classifiers. From this table, we observe that in two cases, classification accuracy achieved by LDA is even less than PCA. As mentioned above, we have used modular LDA to improve its performance but even that could not compensate its decreased performance as compared to PCA. One obvious reason is that the number of eigen components used by LDA is comparatively less than that for PCA.

ROC curves for SVMs of SSA and SA are presented in Figure 6. The maximum area under the convex hull (AUCH) of 0.9 is achieved by SVM using CLBP$(5,8)$. AUCH of 0.87 is achieved by SVM using CLBP$(5,12)$. AUCH of SVM using $Morph_1$ is 0.75, while SVM using $Morph_2$ has AUCH of 0.78. Various performance measures are used to evaluate the performances of different classifiers. These performance measures are derived from the counts such as true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) Fawcett [2006]. If an instance is positive and it is classified as positive, it is defined as true positive. If the instance is negative and it is classified as positive, it is false positive. While, a negative instance classified as negative is true negative (TN) and if it is classified as positive, it is called false negative (FN). Accuracy (Acc), Sensitivity (Sen) or Recall, and Specificity ($Spec$) are defined as follows:

$$Acc = \frac{TP+TN}{P+N}, Sen = \frac{TP}{TP+FN}, Spec = \frac{TN}{TN+FP}. \tag{14}$$

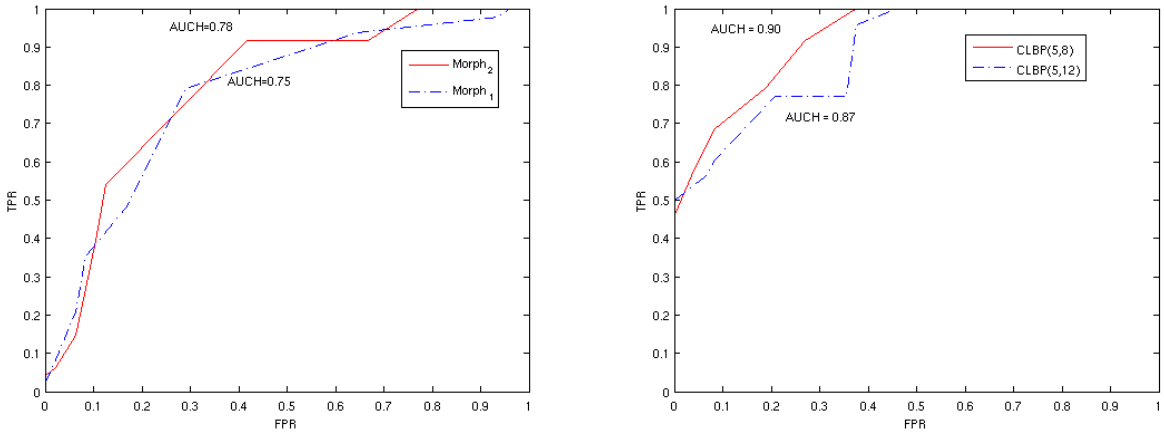
Positive Predictive Value (PPV) or Precision and Negative Predictive value (NPV) are defined as follows:

$$PPV = \frac{TP}{TP+FP}, NPV = \frac{TN}{TN+FN}. \tag{15}$$

The final measure, F-measure (F-meas), is the weighted harmonic mean between Recall and Precision. It is given by:

$$\text{F-meas} = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}}. \tag{16}$$

From Table 1, we observe that SVM yields the best overall performance in case of CLBP$(5,8)$ whereas PCA leads the table in case of CLBP$(5,8)$ with SVM closely behind. It can also be



(a) SVM in SSA (b) SVM in SA

Figure 6: Receiever operating characteristics (ROC) curves, based on true positive rate (TPR) and false positive rate (FPR), and area under convex hull (AUCH) of the ROC curves.

observed that the highest sensitivity (87.5%), F-measure (90.3%) and NPV (88.2%) values are achieved by SVM with CLBP$(5,8)$. Specificity and PPV are highest (100%) for PCA using the same feature set. These results are in close agreement with our feature selection method, which returns 3 features morphological set and CLBP$(5,8)$ as the best feature sets. The power of nonlinear classifier SVM is harnessed in case of CLBP$(5,12)$, our second best feature set, whereas PCA and SVM perform equally well in terms of the overall accuracy in case of CLBP$(5,8)$ due to PCA using a relatively large number of components and employing the best feature set.

Observations in Table 1 help us to analyse the situation and find answers to the questions posed in section 1. First of all, it is clear that textural features in SA achieve better classification performance than all classifiers of SSA. Even using only a single spectral band, in comparison to the segmented image obtained from 128 spectral bands, textural features outperform morphological features of SSA. Original images in 2DPCA may contain some textural variance but that alone cannot be used for discrimination purposes. CLBP features represent textural variance in a manner that yields the highest classification accuracy. It can be said that the gland patterns having tubular shape in most cases are represented efficiently by CLBP features. It is also observed that there is a non-linear boundary between the two classes and the Gaussian kernel SVM efficiently maps this boundary.

4 Conclusions and Future Work

We have made a comparison between spectral spatial analysis (SSA) and spatial analysis (SA) only for the classification of colon tissue samples. Our results indicate that textural analysis on a single band can achieve comparable classification accuracy to 3D spectral spatial analysis. In another study, Boucheron *et al.* Boucheron et al. [2007] claimed that hyperspectral bands do not contain additional discriminatory information but in our case, we suggest that by performing textural analysis on a single band, comparable classification results can

Category	Feature	Method	n	k	Acc	Sen	Spe	PPV	NPV	F-meas	Ave
SSA	$Morph_1$	PCA	5	4	71.9	75	68.7	70.6	73.3	72.7	72
		LDA	5	3	71.9	75	68.7	70.1	73.3	72.7	71.9
		SVM	5	-	84.4	75	**93.7**	**92.3**	78.9	82.7	84.5
	$Morph_2$	PCA	3	2	81.3	**87.5**	75	77.8	85.7	82.3	81.6
		LDA	3	1	78.2	**87.5**	68.7	70.6	73.3	72.7	75.2
		SVM	3	-	**87.5**	**87.5**	87.5	87.5	**87.5**	**87.5**	**87.5**
		2DPCA	32	10	65.6	37.5	93.7	85.7	60	52.2	65.8
SA	$CLBP(5,12)$	PCA	135	20	78.1	68.7	87.5	84.6	73.7	75.8	78.1
		LDA	135	3	81.4	68.7	93.7	91.7	75	78.5	81.5
		SVM	135	-	87.5	81.3	93.7	92.8	83.3	86.7	87.6
	$CLBP(5,8)$	PCA	59	20	**90.6**	81.2	**100**	**100**	84.2	89.6	**90.9**
		LDA	59	3	84.4	81.2	87.5	86.6	82.3	83.8	84.3
		SVM	59	-	**90.6**	**87.5**	93.7	93.3	**88.2**	**90.3**	**90.6**

Table 1: Performance measures in % for SSA and SA: n and k denote number of features and eigen components respectively. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F-measure and average of all these measures are denoted by Acc, Sen, Spe, PPV, NPV, F-meas and Ave, respectively. Bold entries indicate maximum values in each category.

be achieved. However, it may be noted that the SSA approach used here may not be at the highest level of sophistication, since it involves simple segmentation and classification using morphological features only, similar to Rajpoot and Rajpoot [2004]. The outstanding performance of the Gaussian kernel SVM holds both for SSA as well as for SA.

In case of SSA, there is room for detailed spectral analysis. At present, classification is based on spatial analysis and spectral analysis is used only for the segmentation phase. For instance, the spectral signal can be decomposed with wavelets for the classification. Another alternative is use of various stochastic methods, such as Markov chain Monte Carlo (MCMC), for extraction of the shape of tissue glands. Once the glands are extracted, classification can be based on shape parameters of the extracted glands.

5 Acknowledgements

The authors would like to gratefully acknowledge Professor Gus Davis (Yale) for sharing his pathology expertise and Dr. Mauro Maggioni (Duke) and Professor Ronald Coifman (Yale) for sharing the 3D hyperspectral imagery data. We would also like to thank Kashif Rajpoot (Oxford) for his help with the 3D analysis.

References

K. Alsabti, S. Ranka, and V. Singh. An efficient *k*-means clustering algorithm. *Proceedings First Workshop on High-Performance Data Mining*, 10(1), 1998.

S. Aranganayagi and K. Thangavel. Clustering categorical data using silhouette coefficient

as a relocating measure. *Proceedings IEEE Conference on Computational Intelligence and Multimedia Applications*, 2:13–17, 2007.

P. Belhumeur, J. Hesponha, and D. Kriegman. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

A. Bishnu, B. Bhattacharya, M. Kundu, C. Murthy, and T. Acharya. A pipeline architecture for computing the Euler number of a binary image. *International Conference on Image Processing (ICIP)*, 3:310–313, 2001.

L. Boucheron, Z. Bi, N. Harvey, B. Manjunath, and D. Rimm. Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery. *BMC Cell Biology*, 8(Suppl. 1):S8, 2007.

S. Buhmann, P. Herzog, J. Liang, M. Wolf, M. Saiganicoff, C. Kirchhoff, M. Reiser, and C. Becker. Clinical evaluation of a computer-aided diagnosis (CAD) prototype for the detection of pulmonary embolism. *Academic Radiology*, 14(6):651–658, 2007.

G. Davis, M. Maggioni, R. Coifman, R. Levinson, and D. Rimm. Spectral/spatial analysis of colon carcinoma. *Modern Pathology*, 16:3320–3321A, 2003.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

A. Ferrandez and J.A. Disario. Colorectal cancer: Screening and surveillance for high-risk individuals. *Expert Review of Anticancer Therapy*, 3(6):851–862, 2003.

D. Hidović-Rowe, E. Claridge, and T. Ismail. Analysis of multispectral images of the colon to reveal histological changes characteristic of cancer. *Proceedings Medical Image Understanding and Analysis (MIUA)*, pages 66–70, 2006.

R.S. Houlston. Molecular pathology of colorectal cancer. *Journal of Clinical Pathology*, 54: 206–214, 2001.

R. Huang, Q. Liu, and S. Ma. Solving the small sample size problem of LDA. *Proceedings International Conference on Pattern Recognition*, 3, 2002.

L. Hubert and P. Arabie. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

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

A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:441–458, 2000.

T. Maenpaa, T. Ojala, M. Pietikainen, and M. Sariano. Robust texture classification by subsets of local binary patterns. *Proc. 15th International Conference on Pattern Recognition*, pages 947–950, 2000.

M. Maggioni, G. Davis, F. Warner, F. Geshwind, A. Coppi, R. DeVerse, and R. Coifman. Hyperspectral microscopic analysis of normal, benign and carcinoma microarray tissue sections. *SPIE, Optical Biopsy VI*, 6091(1):60910I, 2006.

K. Masood and N. Rajpoot. Classification of colon biopsy samples by spatial analysis of a single spectral band from its hyperspectral cube. *Proceedings Medical Image Understanding and Analysis (MIUA)*, pages 42–48, 2007.

T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray scale and rotation invariant texture analysis with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

K. Rajpoot and N. Rajpoot. SVM optimization for hyperspectral colon tissue cell classification. *Proceedings Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 829–837, 2004.

N. Rajpoot, T. Nattkemper, and M. Gurcan. Special session on computational histopathology: Advances and new challenges. *Proceedings International Symposium on Biomedical Imaging (ISBI)*, 2008.

J. Kim S. Lee, K. Kim, J. Park, S. Park, and W. Moon. Optimal custering of kinetic patterns on malignant breast lesions: Comparison between *k*-means clustering and three-time-points method in dynamic contrast-enhanced MRI. *Proceedings 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2089–2093, 2007.

A. Todman, R. Naguib, and M. Bennett. Visual characteristics of colon images. *Proceedings Medical Image Understanding and Analysis (MIUA)*, pages 161–164, 2001.

V. Vapnik. The Nature of Statistical Learning Theory. *Springer-Verlag, New York*, 1995.

J. Yang, D. Zhang, and A. Frangi. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:131–137, 2004.

Y. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. *Proceedings 3rd International Conference on Automatic Face and Gesture recognition*, pages 336–341, 1996.