

Classification of Potential Nuclei in Prostate Histology Images using Shape Manifold Learning

Muhammad Arif¹, Nasir Rajpoot²

¹ Pakistan Institute of Engineering and Applied Sciences, Nilore, Islamabad, Pakistan

² Department of Computer Science, University of Warwick, Coventry, UK

Email: {arif.nasir}@dcs.warwick.ac.uk

Abstract – The demanding step in the development of ancillary systems for the diagnosis of cancer and other diseases based on nuclear morphometry is the delineation of nuclei in the images of stained tissue sections. Various constituents of the tissue section such as cellular and extra-cellular elements, staining artefacts, debris of nuclei, and clusters of overlapping nuclei apart from the image acquisition noise to name a few contribute to in the complexity of the task. In this paper, we pose the problem of selection of nuclei in tissue section as classification of shapes using manifold learning on training images followed by out-of-sample extension for unknown test images. Experimental results demonstrate the effectiveness of the proposed algorithm

Index Terms – Manifold learning, diffusion maps, nuclear morphometry, out-of-sample extension.

I. INTRODUCTION

Histopathologists study slides of stained tissue sections under microscope for the diagnosis and prognosis of diseases, hormonal imbalances and other disorders. Significant inter- and intra-observer variability has been reported in the literature [1], understandably due to the element of subjectivity in the inference mechanism. Furthermore, the whole procedure is tedious and demands highly skilled experts. To overcome these problems and to detect minute changes otherwise undetectable by a human observer, computer vision and image processing communities in collaboration with histopathologists strive for the development of Computer-Assisted Diagnosis (CAD) systems in order to incorporate objectivity in the whole process of diagnosis. The advancements in the computational power and mass storage capabilities of computing machines and imaging devices augmented with their availability at throwaway prices have spurred the research in the realization of such systems. This is of obvious relevance to developed as well as developing societies.

A. Relevant Work

The focus of current research in the development of histology-based CAD systems is to analyze the slides of stained tissue sections at tissue level and at cellular level [2]. Abnormality of any kind is manifested with changes in the texture of the tissue accompanied with discernable changes in spatial arrangement of cells and their shapes. Co-occurrence matrix, run-length matrix and wavelets have

been used to detect the texture level changes [3]-[8]. The methods at cellular level detect changes in the shapes of the nuclei and their texture [9]-[12]. The quantitative descriptors based on nuclei include area, radius, major/minor axis, compactness, perimeter, fractal dimension and texture of the chromatin inside the nuclei, to name a few. Another class of techniques at cellular level uses approaches from graph theory to quantify spatial arrangement of cells to detect changes in histopathological sections [3], [13]. Techniques from machine learning such as k -NN, neural networks, support vector machines, fuzzy and genetic algorithms, kernel PCA, etc. have been used to train and classify normal and abnormal tissues using features extracted from above techniques.

Most shape-related methods in the development of histology-based CAD systems assume that the nuclei in a given tissue section are detected with ease using staining techniques. Although the most challenging task in this pursuit is to isolate sufficient number of nuclei for shape analysis from the microscopic image of the tissue section which is populated with nuclei, clusters of overlapping/touching nuclei, cut off or damaged nuclei resulting from tissue sectioning procedures, staining artefacts and cellular and extra-cellular elements. Image acquisition noise and nonconformity of various imaging devices to a common language also contribute in the complexity of this step. Automatic and semiautomatic application-dependent techniques have been proposed in the literature to address this problem.

An automatic method to segment breast and cervical cells using a parametric elliptical model for cells has been introduced in [14]. The parameters of the model are adjusted by minimizing a cost function. It has successfully segmented isolated and overlapping cells. In [15] compact Hough transform and maximum likelihood approach has been used to detect cell nuclei. Mathematical morphology has been used in [11] to segment nuclei and an analysis of their shapes. A method has been described in [16] for the segmentation of 2D and 3D cell nuclei which combines intensity information with edge strength and their shapes.

A semiautomatic method based on active contour models has been proposed in [17] to extract the shapes of the nuclei. Human interaction is required for the initialization of approximate boundaries of the nuclei. In [18] a semiauto-

matic algorithm has been presented to segment cells assuming their boundaries to be piecewise elliptic and end points of the curves are specified by the user. Gray-weighted distance transform has been used in [19] to segment cells in microscopic images. However, the algorithm requires manual intervention in that a user is required to mark two points per cell.

B. Main Contributions

In our previous work [20], we have shown that manifold learning techniques such as diffusion maps [28] efficiently arrange projections of feature vectors of various objects in the tissue section in a meaningful way in a much lower-dimensional space. The preliminary results presented in [20] were based on thresholding of the lower-dimensional coordinates from the *same* training image of the tissue section. The main contribution of this paper is the extension of the framework proposed in [20] such that it can perform the classification of *unseen* test images using the out-of-sample extension [31]. The segmentation algorithm (section II.A) employed in this work is much improved as compared to that of [20], as it incorporates bilateral filtering [22] instead of anisotropic diffusion [21] for stability reasons as well as for better preservation of edges around the nuclei. The proposed classification algorithm is computationally efficient: it just requires computation of the unseen objects' coordinates in low dimensional space without having to recalculate the eigenvectors of the similarity matrix followed by classification on the basis of predetermined threshold computed using training images. Experimental results on unseen test images demonstrate the effectiveness of the proposed algorithm.

C. Paper Organization:

The remainder of the paper is organized as follows. Section II describes the imaging modality and the proposed method. The pre-processing steps on images of histological sections are presented in Section II.A. Section II.B outlines the feature extraction from the pre-processed images. Details of manifold learning and out-of-sample extension parts of the method are provided in Sections II.C-E. Results are presented and discussed in section III, followed by conclusions.

II. MATERIALS AND METHODS

A block diagram of our proposed algorithm is shown in Figure 1. Input data are Haematoxylin and Eosin (H&E) stained microscopic color images of prostate tissue samples captured under similar conditions. A typical specimen of the stained tissue section used in our experimentation is shown in the Figure 2 (a).

A. Pre-processing

The images were converted into grey scale and pre-processed for noise removal using edge preserving smoothing filter before segmentation. Various options were avail-

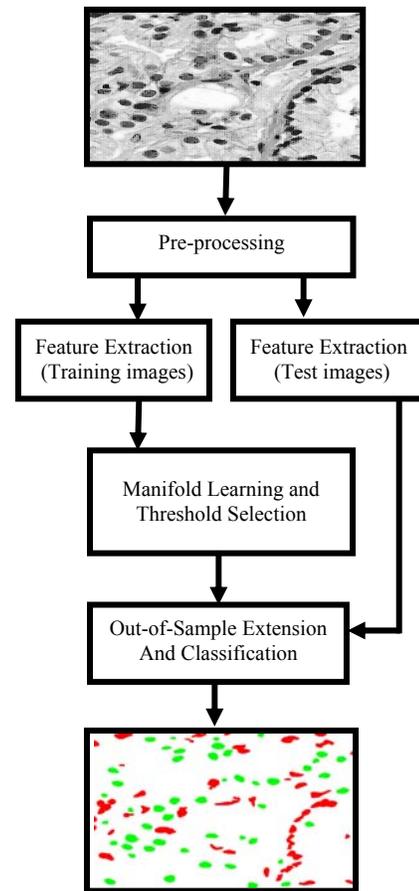


Figure 1 – Block Diagram of the Proposed Algorithm.

able for this purpose, such as anisotropic diffusion filtering [21], bilateral filtering [22], and mean shift algorithm [23]. In our previous work, we used anisotropic diffusion filter for smoothing followed by k -mean clustering algorithm for segmentation [20]. The diffusion methods are iterative as they involve solution of partial differential equations raising issues of instability and computational efficiency [22]. In this work, we use bilateral filtering for smoothing while preserving edges followed by k -mean clustering on grey scale images. Some of the advantages of bilateral filtering are its simplicity and being non-iterative in nature. The connection between bilateral filtering and anisotropic diffusion filtering has been explored in [24].

It is well known that linear FIR smoothing filtering, such as Gaussian filtering, blurs the edges because weights of the filter kernel depend on the proximity of pixels from neighbourhood center only. The weights of the kernel of a bilateral filter, on the other hand, depend on the proximity of the pixels from their neighbourhood center as well as the similarity in their intensity values. If $x = [x_1 \ x_2]^T$ is the coordinate vector of the neighbourhood center, then discrete

version of bilateral filtering for a window of size S is given by [24],

$$h(\mathbf{x}) = \frac{\sum_{j=-S}^S \sum_{k=-S}^S I(x_1 + j, x_2 + k) w(x_1, x_2, x_1 + j, x_2 + k)}{\sum_{j=-S}^S \sum_{k=-S}^S w(x_1, x_2, x_1 + j, x_2 + k)} \quad (1)$$

where $I(\mathbf{x})$ and $h(\mathbf{x})$ are the intensity values of the input and filtered images at location \mathbf{x} respectively.

It can be seen from the above equation that the weights w of the filter are a function of the location of the pixels being modified as well as their neighbourhood. The filter weights corresponding to pixels in the proximity of a neighbourhood center with coordinates \mathbf{x} are given by

$$w(\mathbf{x}, \boldsymbol{\zeta}) = \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\zeta}\|^2}{2\sigma_d^2}\right) \exp\left(-\frac{(I(\mathbf{x}) - I(\boldsymbol{\zeta}))^2}{2\sigma_r^2}\right) \quad (2)$$

where $\boldsymbol{\zeta} = [x_1 + j, x_2 + k]^T$ are coordinates of the proximity pixels and σ_d and σ_r are geometric spread and photometric spread, respectively. The larger the value of σ_d , the more blur in the filtered image. Similarly, pixel intensity values within the photometric spread σ_r are given higher weight than the ones further away [22].

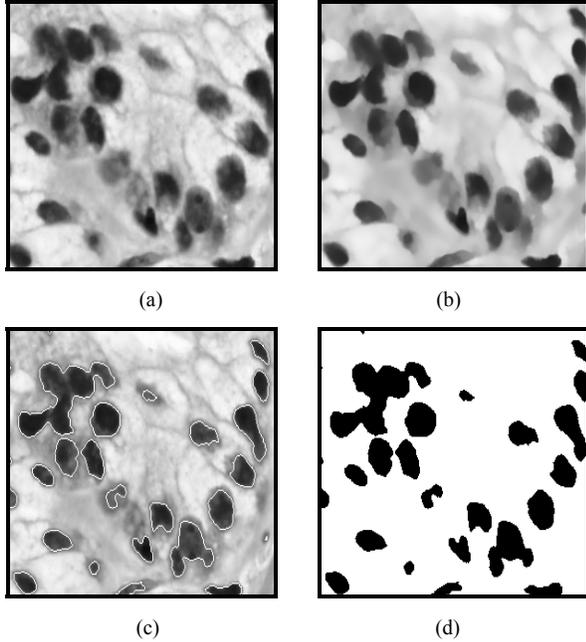


Figure 2- (a) A Typical Cropped Microscopic Image of Stained Tissue Section of Prostate Gland. (b) The Results of Bilateral Filtering. (c) Bounded Objects Represent Detected Objects by k -Mean Clustering. (d) Potential Nuclei after Morphological Operations on Binary Image.

The result of bilateral filtering applied on an image shown in Figure 2 (a) is shown in Figure 2 (b). Note that the

bilateral filter has successfully smoothed the chromatin texture inside the nuclei while preserving their shapes and edges.

Pixels associated with one of the clusters resulting from k -mean clustering algorithm applied on grey values of the filtered images belong to the potential nuclei. These pixels are delineated in Figure 2 (c). By turning the delineated pixels on while the rest off, we get a binary image. In order to remove the segmentation artefacts, morphological operation such as opening, filling etc. are performed on the binary image. The resulting image after these operations is shown in Figure 2 (d).

B. Feature Extraction

In order to learn shape manifolds, boundary points of various objects in the binary image of Figure 2 (d) are extracted and resampled to equal number N . For each object the centroidal distance vector $\mathbf{r} = \{r_1, r_2, \dots, r_N\}$ is computed using [25]

$$r_i = \sqrt{([x_i - x_c]^2 + [y_i - y_c]^2)} \quad (3)$$

where (x_i, y_i) denote the coordinates of the i th boundary point and (x_c, y_c) denote the centroid of the object.

The distance vector \mathbf{r} is transformed into frequency domain using FFT. Our feature vector \mathbf{f} for manifold learning algorithm is derived as follows,

$$\mathbf{f} = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{N/2}|}{|F_0|} \right]^T \quad (4)$$

where F_i denote the i th Fourier coefficient, with F_0 being the DC component. The feature vector \mathbf{f} is made rotation and scale invariant by dividing magnitudes of the Fourier coefficients by the DC component.

C. Manifold Learning

In a wide class of applications, the minimum number of parameters needed to account for the modes of variation (i.e., intrinsic dimensionality) of a given data set is much smaller than the given dimensionality of the feature vector. In order to decrease the complexity of the classification algorithm and/or to facilitate visualization, it is desirable to transform the higher dimensional data to a meaningful representation of the reduced dimensionality.

The dimensionality reduction techniques are classified into two categories: linear techniques and non-linear techniques. Linear techniques of dimensionality such as PCA and LDA assume that the data lie on or near a linear submanifold in high dimensional space. When the data lies on a nonlinear submanifold, techniques such as Locally Linear Embedding [26], Laplacian Eigen Maps [27], Diffusion Maps [28] etc., have been proposed in recent years. L.J.P. van der Maaten et al. [29] present a comprehensive review

and comparative study of various dimensionality reduction techniques.

We use Diffusion Maps for extracting intrinsic dimensionality of our data set. Like other spectral methods of dimensionality reduction, Diffusion maps also exploit the spectral properties (eigenvectors and eigenvalues) of a similarity matrix. The feature vectors are treated as nodes of a symmetric graph and their connectivity is reflected by a robust (i.e., with respect to noise) distance metric on the data set.

For a given feature vector set $\Omega = \{f_1, f_2, \dots, f_n\}$, the weight of edge between nodes f_i and f_j of the graph is computed using a Gaussian Kernel of width ε ,

$$w(f_i, f_j) = \exp\left(-\frac{\|f_i - f_j\|^2}{2\varepsilon}\right) \quad (5)$$

The weights of edges between nodes define a similarity matrix W whose (i,j) th element is given by $w(f_i, f_j)$. The similarity matrix is converted into Markov matrix P with its (i,j) th element p_{ij} given by

$$p_{ij} = \frac{w(f_i, f_j)}{d(f_i)} \quad (6)$$

where $d(f_i)$ denotes the degree of node f_i in the graph. It is defined as

$$d(f_i) = \sum_{z \in \Omega} w(f_i, z). \quad (7)$$

The (i,j) th element of P^t characterize the probability of going from node f_i to node f_j in t steps. If $\{\lambda_i\}$ is the sequence of eigenvalues of P such that $|\lambda_0| \geq |\lambda_1| \geq \dots$, and $\{v_i\}$ are the corresponding eigen vectors of P , then a mapping from the data set Ω to a low-dimensional Euclidean space \mathfrak{R}^m , where m is the dimensionality of the lower-dimensional space and can be a function of t , and is given by (see [28], [30] for details),

$$\Psi^t = \left\{ \lambda_1^t v_1, \lambda_2^t v_2, \dots, \lambda_m^t v_m \right\}^T \quad (8)$$

Dimensionality reduction occurs due to spectral fall-off and the time t of the random walk. Large-scale structures in the data set are captured with fewer diffusion coordinates for large values of t of the random walk.

The two-dimensional diffusion map for $(N/2-1)$ -dimensional feature vectors (with $N = 200$ boundary points) for a typical tissue section is shown in Figure 3.

In order to explore the usefulness of the embedding, various objects in the tissue section have been overlaid on their corresponding embeddings (the 3rd quadrant of Figure 3) in Figure 4. Note that almost all the regular shaped objects are

confined in this region implying that the low-dimensional embeddings have successfully captured the attributes of the data characterizing various shapes in the tissue sections.

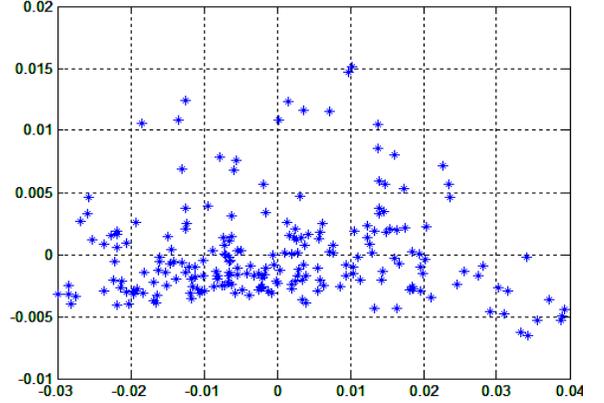


Figure 3- Diffusion Maps for a Typical Tissue Section used in our Experimentation.

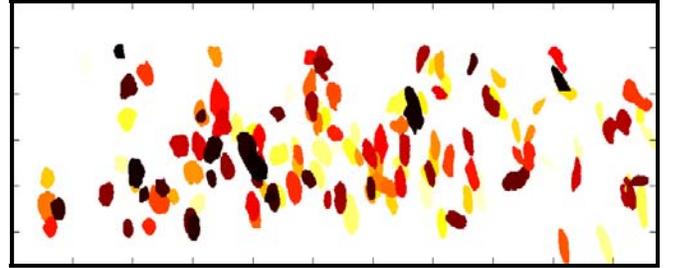


Figure 4- Objects in the Tissue Section overlaid on their Corresponding Coordinates for a Portion of Figure 3.

D. Out-of-Sample Extension

Given the embedding for the training data, we propose to use Nystrom's formula to embed unseen test data [31]-[32]. If the coordinates of the projection of a new feature vector f in low-dimensional space is given by the vector $[\lambda_1^t y_1, \lambda_2^t y_2, \dots, \lambda_m^t y_m]^T$, then y_k is computed using Nystrom's formula,

$$y_k = \frac{1}{\lambda_k} \sum_{i=1}^N v_{i,k} p(f, f_i) \quad (9)$$

where $v_{i,k}$ is the i th coordinate of k th eigen vector and λ_k is corresponding eigen value and $p(f, f_i)$ is the probability of going from node f to node f_i in one step and is computed using (6).

E. Classification

Appropriately thresholding the low dimensional space after embedding for a given training sample isolates the potential nuclei from rest of the objects in the tissue section. To classify objects in test samples, out-of-sample extension

is performed for the given feature vectors. The decision to classify objects in a test microscopic image of a tissue section as potential nuclei or non-nuclei is based on the thresholds obtained from the training data.

III. RESULTS AND DISCUSSION

A cropped version of a typical prostate histological section image used for training is shown in Figure 5 (a). Diffusion maps in 2D (Figure 3) were computed for the training image. By overlaying various objects on their respective coordinates, the region in which the regular-shaped nuclei lie was identified and values of the two thresholds (for two diffusion coordinates) were selected manually. Green objects in Figure 5 (b) show the potential nuclei detected by our algorithm after threshold selection.

The effectiveness of the classification using out-of-sample extension for two unseen test images is demonstrated in Figures 6 and 7. Again the green objects in these figures are potential nuclei detected by our algorithm. Some of the touching/overlapping nuclei (shown in red) in these images can be separated by post-processing procedures such as watershed algorithm and mathematical morphology.

We believe that these results are very encouraging. The next logical step in this study will be to compute diagnostically important features from the detected nuclei and use machine learning techniques to detect abnormalities.

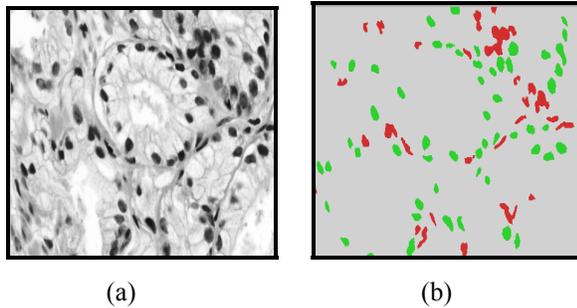


Figure 5- (a) A Typical **Training** Tissue Section of Prostate. (b) Nuclei Detected (in Green) by Thresholding the Diffusion Maps.

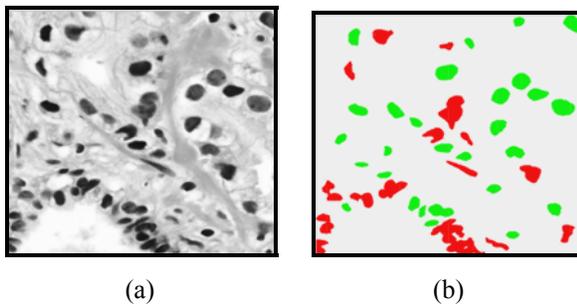


Figure 6- (a) An **Unseen** Test Image for Detection of Nuclei. (b) Nuclei Detected by Out-of-Sample Extension (in Green). Objects other than Nuclei are shown in Red.

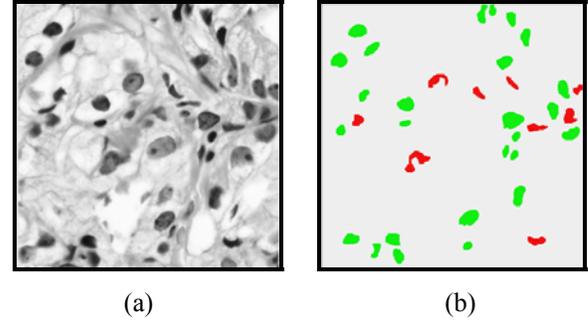


Figure 7- (a) Another **Unseen** Test Image used for the Detection of Nuclei. (b) Nuclei Detected by Out-of-Sample Extension (in Green). Objects other than Nuclei are shown in Red.

IV. CONCLUSIONS

This paper suggests an algorithm to classify constituents of a prostate tissue section. Diffusion maps have been used for training and Nystrom's formula was employed for embedding feature vectors of various objects in unseen histological images of prostate tissue sections. The experimental results demonstrate the potential of our method for use in automated or semi-automated histology-based CAD systems, which is the main aim of our future work.

ACKNOWLEDGEMENTS

The first author gratefully acknowledges the financial support provided by HEC, Govt. of Pakistan to carry out this research and the Department of Computer Science, University of Warwick, UK, for hosting his visit.

REFERENCES

- [1] P. Klinkhamer, G.P. Vooijs, and A. de Haan, "Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears," *Acta Cytology*, vol. 32, no. 6, pp. 794-800, Nov-Dec 1998.
- [2] C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: A systematic survey," Rensselaer Polytechnic Institute Technical Report # TR-05-09, 2005.
- [3] C. Demir, S. Humayun Gultekin, and B. Yener, "Learning the topological properties of brain tumors," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 2, no. 3, July-September 2005.
- [4] Q. Ji, J. Engel, and E. Craine, "Texture analysis for classification of cervix lesions," *IEEE Transactions on Medical Imaging*, vol. 19, no. 1, November 2000.
- [5] K. Masood, N. Rajpoot, K. Rajpoot et al., "Hyperspectral texture analysis for colon tissue biopsy classification," *International Symposium on Health Informatics and Bioinformatics*, 2007.
- [6] W. Christen-Barry and A. Partin, "Quantitative grading of tissue and nuclei in prostate cancer for prognosis prediction," *John-Hopkins APL Technical Digest*, vol. 18, no. 2, 1997.

- [7] H. Qureshi, N. Rajpoot, K. Masood et al., "Classification of meningiomas using discriminant wavelet packets and learning vector quantization," *In Proceedings of Medical Image Understanding and Analysis*, 2006.
- [8] K. Rajpoot, N. Rajpoot and M. Turner, "Hyperspectral colon tissue cell classification," *In Proceedings of SPIE Medical Imaging, San Diego, California (USA)*, 2004.
- [9] K. Rodenacker and E. Bengtsson, "A Feature set for cytometry on digitized microscopic images," *Analytical Cellular Pathology*, vol. 25 pp. 1-36, 2003, ISO Press.
- [10] W. H. Wolberg, W. N. Street, D. M. Heisey, and O. L. Mangasarian, "Computer-derived nuclear features distinguish malignant from benign breast cytology," *Hum. Pathology*, vol. 26, no. 7, pp. 792-796, July 1995.
- [11] J. P. Thiran and B. Macq, "Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tissue," *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 10, October 1996.
- [12] Yogesan, T. Jorgensen, F. albregtsen et al., "Entropy-based texture analysis of chromatin structure in advanced prostate cancer," *Cytometry*, vol. 24 pp. 268-276, 1996.
- [13] B. Chaudhuri, K. Rodenacker, and G. Burger, "Characterization and featurizing of histological section images," *Pattern Recognition Letters*, 7, pp. 245-262, 1988.
- [14] H.-S Wu, J. Barbra and J. Gil, "A parametric fitting algorithm for segmentation of cell images," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 3, March 1998.
- [15] T. Mouroutis, S. Roberts, and A. Bharath, "Robust cell nuclei segmentation using statistical modelling," *Bioimaging* no. 6, pp. 79-91, 1998.
- [16] C. Wahlby, I.-M. Sintron, F. Erlandsson et al., "Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections," *Journal of Microscopy*, vol. 215, pp. 67-76, July 2004.
- [17] W. Street, W. Wolberg and O. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis," *In IS&T/SPIE 1993 International Symposium on Electronic Imaging: Science and Technology*, vol. 1905, pp. 861-870, San Jose, California, 1993.
- [18] H.-S Wu and Barbra, "An efficient semiautomatic algorithm for cell contour extraction," *J. Microsc.* 179, pp. 270-179, 1995.
- [19] D. Bagget, M. Nakaya, M. McAuliffe et al., "Whole cell segmentation in solid tissue sections," *Cytometry Part A* 67A, pp. 137-143, 2005.
- [20] M. Arif, and N. Rajpoot, "Detection of nuclei by unsupervised manifold learning," in *Proceedings of Medical Image Understanding and Analysis (MIUA'2007)*, July 2007.
- [21] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629-639, July 1990.
- [22] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color Images," *In Proceedings International Conference on Computer Vision*, Bombay, India, 1998.
- [23] D. Comaniciu and P. Meer, "Mean Shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, May 2002.
- [24] D. Barash, "A fundamental relationship between bilateral filtering, adaptive smoothing and nonlinear diffusion equations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, June 2002.
- [25] H. Kauppinen, T. seppanen, and M. Pietikainen, "An experimental comparison of autoregressive and fourier-based descriptors in 2-D Shape classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, pp. 201-207, 1995.
- [26] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323-2326, 2000.
- [27] M. Bilkin and P. Niyogy, "Laplacian eigen maps for dimensionality reduction and data representation," *Neural Computation*, Vol. 6. No. 15, pp. 1373-1396, June 2003.
- [28] R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, Special Issue on diffusion maps and wavelets, Vol. 21, pp. 5-30, July 2006.
- [29] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik, "Dimensionality reduction: A comparative review," Preprint, 2007.
- [30] S. Lafon and A. Lee, "Diffusion Maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 9, September 2006.
- [31] Y. Bengio, J. Paiement, and P. Vincent, "Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering" D'epartement d'Informatique et Recherche Operationnelle, Universit'e de Montr'eal Montr'eal, Canada, Technical Report #1238, July 2005.
- [32] S. Lafon, Y. Keller, and R.R. Coifman, "Data fusion and multicue data matching by diffusion maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 11, November 2006.