International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Robust Segmentation and Classification of Histopathological Image

Jincy Wilson[1], Vipin V. R.[2]

[1]P.G Scholor, Department of ECE, T K M Institute of Technology, Kollam, India

[2]Assistant Professor, Department of ECE, T K M Institute of Technology, Kollam, India

Abstract: ***Cancer is an uncontrolled cell growth caused due to abnormal cell division in any region of the body. When cancer is diagnosed at an early stage the treatment is simpler and effective. There are different types of cancer and each is classified by the type of cell that is initially affected. In this paper, early diagnosis of colon cancer is considered. In the proposed method, segmentation of tissue is done by graph cut, after the pre-processing of the histopathological image. Next, the features are extracted from segmented image, by structural and statistical methods. In structural approach, intensity histogram based features such as mean, variance etc. are computed. In statistical feature extraction, gray level co-occurrence matrix are used to find out the relation between pixels and thus energy, entropy, contrast etc. are measured. These extracted features are given to Support Vector Machine (SVM) classifier to classify them as cancerous and noncancerous. This system can be used in many real time problems like bioinformatics also.***

Keywords: Cancer, Histopathology, Structural, Statistical, Graph cut, Gray Level Co-occurrence Matrix, Support Vector Machine.

1. Introduction

Cancer is the one of the most important health problems that threat the human life. The likelihood of curing cancer increases with its early diagnosis and correct grading. Medical experts take advantage of numerous medical imaging techniques such as Magnetic Resonance Imaging (MRI), Computer Aided Tomography (CAT), Mammography, Colonoscopy, Ultrasound Imaging for cancer screening. Although these methods provide effective diagnosis tools for screening and early detection of tumors, they may not be helpful in determining their malignancy level. Moreover, these methods are not used as the gold standard and biopsy examination is still necessary to reach the final decision.

The main aim is to investigate robust and accurate image analysis algorithms for computer-assisted interpretation of histopathology imagery. Different image processing techniques will be applied for segmentation, image texture classification, cell type identification or classification to deriving quantitative measurements of disease features from histological images. This technique automatically determine whether a disease is present within analyzed samples or not and also help to decide the different grades or severity of disease.

Colon tissues are the examples of such tissues. They are formed of nuclear, luminal, and stromal tissue components. Nuclei of the epithelial cells of colon tissue are lined up around its luminal components and form glandular structures of the colon. Stromal tissue components are distributed between these glandular regions. This hierarchical organization of its nuclear, luminal, and stromal components reflects the major characteristics of the colon tissue. Figure 1 shows the histopathological tissue image of a colon tissue section.

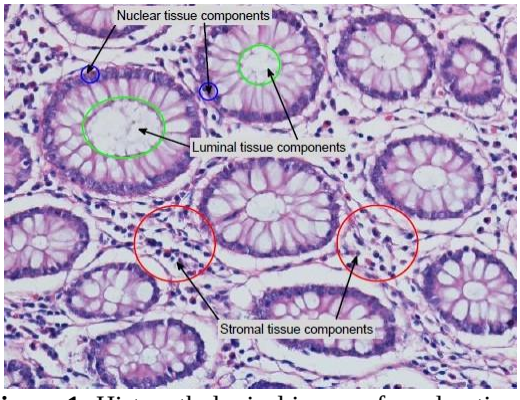

Figure 1: Histopathological image of a colon tissue

Histopathological tissue examination for cancer diagnosis and grading is the one of the most important medical practices. That experience observer variability. The major reasons behind the qualitative measures are done by the visual assessment. To deviate these problems, develop objective and mathematical analysis. Second order statistical feature is texture feature. Textural method avoids difficulties that related to correct localization of tissue cells. Here abnormalities can be modeled by textural changes in observed tissue from normal tissues. Texture of the feature including intensity histograms [2], co-occurance matrices[3],[4], multiwavelet coefficients[5], run-length matrices[6] etc. For data classification, a decision making based on set of features. In this paper, structural and statistical feature are extracted from the GLCM [7] and histogram based intensity [8] value. Then, a support vector classifier (SVM)[9] is used to classify the image into cancerous and non cancerous.

The organization of this paper as follows. In section II explain the pre processing, segmentation, feature extraction and SVM classifier. Section III explain the results and finally conclude in section IV

2. Methodology

Tissue graph segmentation technique is used to detect cancer. The cancer diagnosis and grading processes is to obtain quantitative information about the characteristics of tissues, and help reducing the subjectivity of pathologists. Therefore, proposed methods define mathematical representations of tissues with the use of textural features and structural features of histopathological images. These methods yield accurate results for diagnosis and grading of cancer.

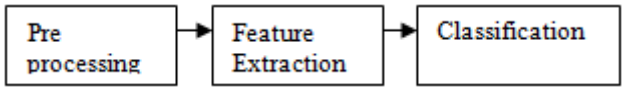

Figure 2: Block Diagram

The basic block diagram for the tissue image classification is as shown in the figure 2. The input image is Pre-processing for removing the noise and the contrast enhancement. Then segment the enhanced image by graph cut. In feature extraction stage structural and statistical features can be extracted. In classification stage, extracted features are used to classify image into two groups: cancerous and non cancerous with the help of Support Vector Machine (SVM) classifier.

2.1 Histopathological Images

Histopathology is the microscopic examination of biological tissues to observe the appearance of diseased cells and tissues in very fine detail. Typical cell is 10-20 µm in diameter. Cells are color less and translucent. So Variety of stains that provide sufficient contrast to make those features visible. Haematoxylin and Eosin stain is frequently used to examine thin section of tissue and which separates cell nuclei, cytoplasm and connective tissue. Hematoxylin stains cell nuclei blue, whereas Eosin stains cytoplasm and connective tissue pink. As an input stained colon images are used.

2.2 Pre-Processing

During the image collection, imaging devices are quite often interfered by various noise sources. Impulse noise degrades the biomedical image details such as edges, contours and textures. Reduce noise in signal or image by using wiener filter. When the image is blurred by a known low pass filter, it is possible to recover the image by inverse filtering. But inverse filtering is very sensitive to additive noise. The Wiener filtering executes an optimal trade-off between inverse filtering and noise smoothing. It removes the additive noise and inverts the blurring simultaneously. The wiener filter minimizes the mean square error between the estimated random process and the desired process. It minimizes the overall mean square error in the process of inverse filtering and noise smoothing. The Wiener filtering is a linear estimation of the original image. Improve contrast of the image by Adaptive Histogram Equalization. It is an excellent contrast enhancement method for both neural and medical images Image enhancement is among the simplest and most appealing areas of digital image processing. It is used to highlight certain features of interest in an image for increase the contrast of an image. There by improvement in quality of these degraded images can be achieved by using application of enhancement techniques like Adaptive Histogram Equalization method. This is an extension to traditional Histogram Equalization technique. It enhances the contrast of images by transforming the values in the intensity image. Unlike Histogram Equalization *histeq*, it operates on small data regions (tiles), rather than the entire image. Each tile's contrast is enhanced, so that the histogram of the output region approximately matches the specified histogram. The neighboring tiles are then combined using bilinear interpolation in order to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited in order to avoid amplifying the noise which might be present in the image.

2.3 Graph Cut Segmentation

Segmentation is an important part of image analysis. It refers to the process of partitioning an image into multiple segments. Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyse.

An undirected graph G = < V, E >, is defined as a set of nodes (vertices V) and a set of undirected edges (E) that connect these nodes. An example is shown in Figure 3. Each edge e ϵ E in the graph is assigned a nonnegative weight (cost) W_e. There are also two special nodes called terminals. A cut is a subset of edges C ⊂ Esuch that the terminals become separated on the induced graph G(C) = < V, E C >. It is normal in combinatorial optimization to define the cost of a cut as the sum of the costs of the edges that it severs.

$$|C| = \sum_{e \in C} W_e.$$

Graph cut formalism is well suited for segmentation of images. In fact, it is completely appropriate for N dimensional volumes.

Each pixel or voxels in an image as a node in a graph and added two terminal nodes connected to every pixels, (named S and T). The edges can represent any neighborhood relationship between the pixels. A cut partitions the nodes in the graph. As illustrated in Figure 2(c-d), this partitioning corresponds to a segmentation of an underlying image or volume. A minimum cost cut generates a segmentation that is optimal in terms of properties that are built into the edge weights. This technique is based on a well-known combinatorial optimization fact that a globally minimum cut of a graph with two terminals can be computed efficiently in low-order polynomial time.

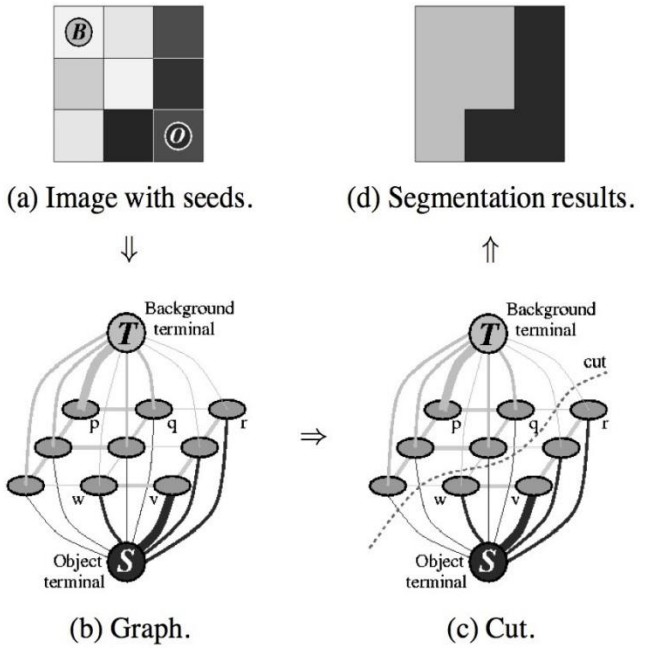

Figure 3: Graph cut segmentation

Segmentation Technique Assume that O and B denote the subsets of pixels marked as "object" and "background" seeds. Naturally, the subsets O ⊂ P and B ⊂ P are such that O ∩ B = ∅. The general flow is shown in Figure 3. From the image (Figure 2(a)) create a graph with two terminals shown in Figure 3(b). The edge weights reflect the parameters in the regional and the boundary terms of the cost function, as well as the known positions of seeds in the image. The next step is to compute the globally optimal minimum cut in Figure 3(c) separating two terminals. This cut gives a segmentation shown in Figure 3(d) of the original image. In the simplistic example of Figure 3, the image is divided into exactly one "object" and one "background" regions. In general, segmentation method generates binary segmentation with arbitrary topological properties.

2.4 Feature extraction

In pattern recognition and in image processing, feature extraction is the special form of dimensionality reduction. It involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Structural and statistical features are used for extraction feature.

- Texture feature

Texture feature classification using grey level co-occurrence matrices (GLCMs). The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. Grey Level Coocurrence Matrix (GLCM) method is a way of extracting second order statistical texture features. A GLCM is a matrix where the number of columns and rows is equal to the number of gray levels, G, in the image. The matrix element Q(i, j | x, y)is the relative frequency with which two pixels, separated by a pixel distance (x, y), occur within a given neighborhood, one with intensity i and the other with intensity j. The matrix element Q(i, j | d, Θ) which contains the second order statistical probability values for changes between gray levels i and j at a particular displacement distance d and at a particular angle(Θ). Using a large number of intensity levels G implies storing a lot of temporary data, i.e. a G x G matrix for each combination of (x, y) or (d, Θ). Due to their large dimensionality, the GLCMs are very sensitive to the size of the texture samples on which they are estimated. Thus, the number of gray levels is often reduced. From GLCM contrast, energy, correlation, homogeneity etc. can be computed, by using the following notation:

G is the number of gray levels used.
μ is the mean value of Q.
μ_x, μ_y, σ_x and σ_y are the means and standard deviations of Q_x and Q_y.
$Q_x(i)$ is the i^{th} entry in the marginal-probability matrix obtained by summing the rows of Q(i,j).

Contrast
This statistic measures the spatial frequency of an image and is difference moment of GLCM. It is the difference between the highest and the lowest values of a contiguous set of pixels.

$$\text{Contrast} = \sum_{i=0}^{G-1}(i-j)^2 \sum_{i=1}^{G}\sum_{j=1}^{G} Q(i,j). \quad (2)$$

Energy
This statistic is also called Uniformity or Angular second moment. It measures the textural uniformity that is pixel pair repetitions. It detects disorders in textures. Energy reaches a maximum value equal to one. High energy values occur when the gray level distribution has a constant or periodic form. Energy has a normalized range. The GLCM of less homogeneous image will have large number of small entries.

$$\text{Energy} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} Q(i,j)^2. \quad (3)$$

Correlation
The correlation feature is a measure of gray tone linear dependencies in the image.

$$\text{Correlation} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{(i,j)\,Q(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$$

Homogeneity
This statistic is also called as Inverse Difference Moment. It measures image homogeneity as it assumes larger values for smaller gray tone differences in pair elements.

$$\text{Homogeneity} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{1}{1+(i-j)^2} Q(i,j).$$

Structural features
Structural features describes pixel level characteristics of images are obtained from histogram based features. To this end, gray level or color histograms of intensity values and densitometric features are employed. The properties of these features such as mean, standard deviation, kurtosis, and skewness are computed to obtain first order statistical information about the texture of tissues. Let z be a random variable denoting image gray levels and P(z_i), i = 0,1,2,3,.L-1, be the corresponding normalized histogram, where L is the number of distinct gray levels.

Mean

$$\text{Mean, } m = \sum_{i=0}^{L-1} z_i P(z_i)$$

Variance

The variance gives the amount of gray level fluctuations from the mean gray level value

Variance, $\mu_2(z) = \sum_{i=0}^{L-1}(z_i - m)^2 P(z_i)$.

Skewness

Skewness is a measure of the asymmetry of the gray levels around the sample mean. If skewness is negative the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right.

Skewness, $\mu_3(z) = \sum_{i=0}^{L-1}(z_i - m)^3 P(z_i)$.

Kurtosis

Kurtosis is a measure of how outlier-prone a distribution. It describes the shape of the tail of the histogram.

Kurtosis, $\mu_4(z) = \sum_{i=0}^{L-1}(z_i - m)^4 P(z_i)$.

2.5. Support Vector Machine

For an effective classification method, it is important to accurately classify not only the observed data but also the unknown data. Thus, it is necessary to select the most appropriate boundary that will optimally separate the unpredicted data. Numerous parametric models that are based on the probability density estimations of classes are proposed for performing this task. Support vector machines provide a nonparametric classification method that solves an optimization problem. The support vector machine is a kernel-based supervised learning method. It is used for classification and regression. The SVM algorithm was proposed as a linear classifier. This algorithm mainly aims to partition data points of two classes in n dimensional space with an n - 1 dimensional hyperplane. ie, there may be more than one hyperplane separating the data points. The equation for the hyperplane is given by

$$y = w^T x + b \quad (4)$$

where w is known as the weight vector and b is the bias. The optimal hyperplane can be represented in an infinite number of different ways by scaling of w and b. As a matter of convention, among all the possible representations of the hyperplane, the one chosen is

$$|w^T x + b| = 1 \quad (5)$$

where x symbolizes the training examples closest to the hyperplane. In general, the training examples that are closest to the hyperplane are called support vectors. This representation is known as the canonical hyperplane. The distance of the hyperplane to the nearest point of the two classes is given by

$$M = \frac{2}{||w||} \quad (6)$$

The fundamental idea of the algorithm is to find the separating hyperplane that maximizes the margin, which is the distance of the nearest data points in both data sets to the separating hyperplane.

3. Results

In the experiment, stained tissue samples are used. An example tissue image shown in figure 4.

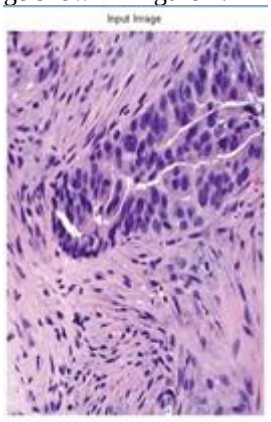

Figure 4: A sample of colon tissue image

In pre-processing stage, tissue image is converted to gray scale then noise removal and contrast enhancement is done. Noise removal is done by wiener filtering shown in figure 5.

The Wiener filter can be used to filter out the noise from the corrupted signal to provide an estimate of the underlying signal of interest. The Wiener filter is based on a statistical approach, and a more statistical account of the theory is given in the minimum mean-square error (MMSE). Segment the main objects from an image using a segmentation method

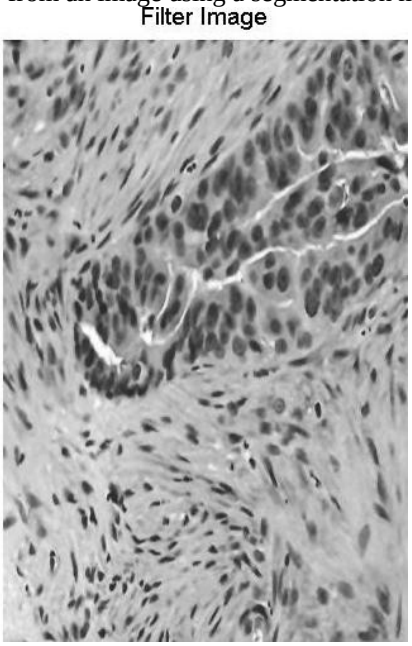

Figure 5: Wiener filtered image

based on graph cuts shown in figure 6. The segmentation helps to simplify or change the representation of an image into something that is more meaningful and easier to analyze.

Figure 6: Graph cut segmentation

Feature Extraction is used to analyze the object or image, extract the most prominent features to represent the various class of image. The main purpose of feature extraction is to reduce the original data by measuring certain property or feature. In this work texture features are used. Here histogram based and GLCM based features are extracted. In the training stage these features are taken. Table I shows the feature extracted value. According to this features further classification is done.

Support Vector Machine classifier is used to classify the data. It is important to accurately classify not only the observed data but also the unknown data.SVM are an example of a linear two-class classifier. The data for a two class learning problem consists of objects labeled with one of two labels corresponding to the two classes; for convenience assume the labels are +1 (Cancer) or -1 (Normal). Figure 7 and figure 8 shows the classified output of cancerous and normal(Non Cancerous) image.

4. Conclusion

This paper introduce a robust segmentation and classification for automated cancer diagnosis and grading are gaining importance in medicine as they provide objective

Table 1: Extracted feature values

Image type	Contrast	Correlation	Energy	Homogeneity	Mean	Std Deviatn	Skewness	Kurtosis
Normal	6.030351059	0.205656056	0.034172278	0.500340939	149.9150716	70.07811499	-0.423651764	-1.068483308
Normal	6.738834899	0.122134577	0.020761841	0.46003795	139.7948438	70.05069203	-0.171900124	-1.097046092
Normal	6.995851389	0.165459663	0.023319069	0.470375715	140.6822917	72.77985663	-0.217225787	-1.203797462
Normal	7.790068841	0.112491648	0.018657577	0.439952596	134.5055078	74.47465924	-0.083322997	-1.229540373
Normal	5.611129561	0.402236928	0.052105767	0.542062993	166.2448372	76.7485374	-0.451227687	-1.112564211
Cancerous	5.827849271	0.138721631	0.023543888	0.474918074	135.2728646	65.69193021	0.033501385	-1.010531689
Cancerous	5.594377357	0.178427446	0.024299679	0.482573194	130.59	66.33883719	0.131987554	-0.957856238
Cancerous	7.269538007	0.160835423	0.019261046	0.448155045	133.8816797	74.14185353	-0.042803891	-1.198612145
Cancerous	5.708425753	0.144703697	0.028238552	0.480152944	144.6452734	65.12949572	-0.420985194	-0.920535318
Cancerous	7.675329391	0.034484605	0.019697175	0.427792352	131.2605339	71.21036228	-0.077500889	-1.164420021
Cancerous	7.052548083	0.084944427	0.020710485	0.443627643	133.8761784	70.20559377	-0.122408323	-1.145904752
Cancerous	7.464976965	0.072204864	0.019494916	0.433999686	129.7835872	71.4665036	-0.039135637	-1.144767362

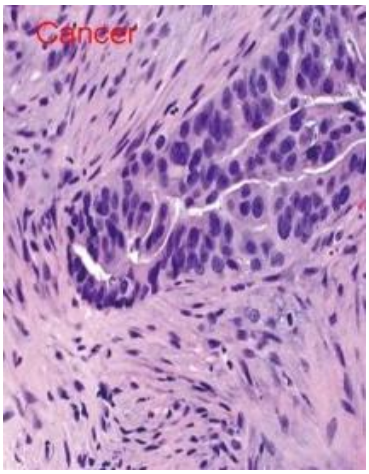

Figure 7: Cancerous image

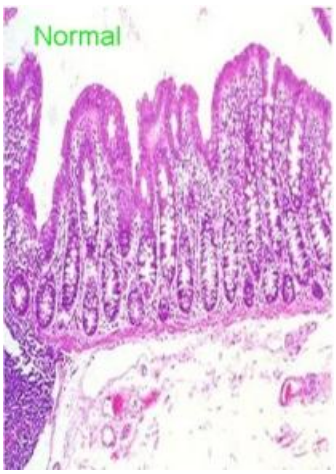

Figure 8:.Non cancerous image

mathematical measures. Graph cut segmentation technique is used for segmentation of tissue image. From the segmented image structural and statistical features are extracted to train the Support Vector Machine. The same features of the test image can be given to the trained classifier to detect the tissue image is cancerous or not. Once the training is over the system automatically able to detect cancer. In future enhancement histopathological image has been conducted for various cancer detection and grading applications, like prostate, breast, kidney and lung. Using different segmentation method, feature extraction and classification techniques can be used to classify or analyze the histopathology images and also reduces the computation time.

References

[1] Erdem Ozdemir and Cigdem Gunduz-Demir ”A hybrid classification model for digital pathology using structural and statistical pattern recognition” IEEE Trans. On Medical Imaging, vol.32,no.2,pp. 474-483,Feb. 2013.
[2] A.Tabesh, M.Teverovskiv, H.Y Pang, V.P Kumar, D. Verbel, A. Kotsianti, and O.Saidi”Multifeature prostatecancer diagnosis and Gleason grading of histological images,” IEEE Trans. On Medical Imaging, vol.26,no.10,pp. 1366-1378,Oct. 2007.
[3] A. N. Esgiar, R. N. G. Naguib, B. S Sharif.,M.K. Bennett, and A.Murray ”Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa,” IEEE T. Inf. Technol. Biomed., vol.2,no.3,pp. 197-203,Sep. 1998.
[4] S. Doyle, M. Feldman, J. Tomaszewski, and A. Madabhushi ”A boosted Bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies,” IEEE Trans. Biomed. Eng., 2011, in press. DOI: 10.1109/TBME.2010.2053540.
[5] Kourosh Jafari-Khouzani* and Hamid Soltanian-Zadeh, ”Multiwavelet Grading of Pathological Images of Prostate” IEEE Trans. On Biomed. Eng, vol.50,no.6,pp. 697-704,Jun. 2003.
[6] Akif Burak Tosun and Cigdem Gunduz-Demir ”Graph Run-Length Matrices for Histopathological Image Segmentation” IEEE Trans. On med.imag, vol.30,no.3,pp. 721-732,Mar. 2011.
[7] Branislav M and Allan Hanbury”Supervised texture detection in images”[online]Available:http://ai.stanford.edu/micusik/Papers/MicusikHanbury-CAIP2005.pdf
[8] Takumi Kobayashi”BoF meets HOG: Feature Extraction based on Histograms of Oriented p.d.f Gradients for Image Classification” [online]Available:https://staffa.ist.go.jp/takumi.kobayashi/publication/2013.pdf.
[9] Asa Ben-Hur and Jason Weston ”A User’s Guide to Support Vector Machines” pp. 1-16.
[10] P. Scheunders” A genetic c-means clustering algorithm applied to color image quantization,” IRecognit vol.30,no.6,pp. 859-866,1997
[11] Ali Tabesh, Mikhail Teverovskiy, Ho-Yuen Pang, Vinay P. Kumar, David Verbel,Angeliki Kotsianti, and Olivier Saidi ”Multifeature Prostate Cancer Diagnosis and Gleason Grading of Histological Images” IEEE Trans. on medical imaging, vol.26,no.10,pp. 1366-1378,Oct. 2007.
[12] Anil K. Jain, Robert P.W. Duin, and Jianchang Mao ” Statistical Pattern Recognition: A Review” IEEE Trans. On Pattern Analysis And Machine Intelligence vol.22,no.1,pp. 4-37,Jan. 2000.
[13] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, M. Feldman, and J. Tomaszeweski ” Automated grading of prostate cancer using architectural and textural image features” in Proc. Biomed. Imag pp.1284-1287, From Nano Macro, 2007.
[14] D.Altunbay, C.Cigar and Cigdem Gundez-Demir ”Color graphs for automated cancer diagnosis and grading” IEEE Trans. On Biomed. Eng, vol.57,no.3,pp. 665-674,Mar. 2010.
[15] Gonzalez and Woods,”Digital Image Processing” Prentice-Hall, Upper Saddle River, New Jersey 07458
[16] Fritz Albregtsen ”Statistical Texture Measures Computed from Gray Level Coocurrence Matrices”Image Processing Laboratory Department of Informatics University of Oslo pp. 1-14,Nov.2008
[17] Yuri Boykov and Gareth Funka-Lea ”Graph Cuts and Efficient N-D Image Segmentation”International Journal of Computer Vision vol.57,no.3,pp. 109-131,May. 2003.