An Improved Iterative Watershed and Morphological Transformation Techniques for Segmentation of Microarray Images

A.Sri Nagesh
Assistant Professor, CSE Department,
R.V.R.& J.C.College of Engineering,
Chowdavaram,Guntur.

Dr.G.P.S.Varma
Professor, IT Department,
S.R.K.R.Engineering College,
Bhimavaram.

Dr A Govardhan
Professor& Principal,CSE Department,
JNTUCEH, Jagtiyal ,
Hyderabad.

ABSTRACT

Microarrays are novel and dominant techniques that are being made use in the analysis of the expression level of DNA, with pharmacology, medical diagnosis, environmental engineering, and biological sciences being its current applications. Studies on microarray have shown that image processing techniques can considerably influence the precision of microarray data. A crucial issue identified in gene microarray data analysis is to perform accurate quantification of spot shapes and intensities of microarray image. Segmentation methods that have been employed in microarray analysis are a vital source of variability in microarray data that directly affects precision and the identification of differentially expressed genes. The effect of different segmentation methods on the variability of data derived from microarray images has been overlooked. This article proposes a methodology to investigate the accuracy of spot segmentation of a microarray image, using morphological image analysis techniques, watershed algorithm and iterative watershed algorithm. The input to the methodology is a microarray image, which is then subjected to spotted microarray image preprocessing and gridding. Subsequently, the resulting microarray sub grid is segmented using morphological operators, watershed algorithm and iterative watershed algorithm. Based on the precision of segmentation and its intensity profile, a formal investigation of the three segmentation algorithms employed (morphological operators, watershed algorithm and iterative watershed algorithm) is performed. The experimental results demonstrate the segmentation effectiveness of the proposed methodology and also the better of the three segmentation algorithms employed for segmentation.

General Terms

Machine Intelligence, Bio Informatics, Soft Computing, Retrieval, Medical Digital Image Processing, Medical Image Databases.

Keywords: Bioinformatics, Microarray, Genes, Spot Segmentation, Threshold, Fast Circular Cross Correlation, Morphological Operator, Morphological Filtering, Watershed Algorithm, Iterative Watershed Algorithm.

I. INTRODUCTION

In bioinformatics, microarrays have been emerging as vital technologies, proffering support in dealing with an extensive range of problems in medicine, health and environment, and drug development [1]. Microarray technology is being incorporated into fundamental biomedical research and is becoming an elementary molecular monitoring tool in clinical microbiology settings, particularly for diagnostic applications owing to the likelihood of multiplexing the simultaneous detection of a number of pathogens in a single reaction [2, 3, 4]. The utilization of microarrays in measuring gene expression levels in variable conditions offers biologists with a better understanding of gene functions, and has a plenty of applications in life sciences [5], [6]. A considerable number of clinical applications are on the basis of the development of DNA microarrays for the detection of specific genes or gene regions in pathogens, while others are on the basis of protein (ELISA microarrays) or other immobilized molecules [7]. Characteristic applications of microarrays include the quantification of RNA expression profiles of a system under varied experimental conditions, or expression profile comparisons of two systems under one or several conditions. DNA microarrays are an experimental technology for performing exploration of the genome (genotyping experiments). The technology proffers to biomedical investigators a simple tool for monitoring the expression levels of thousands of genes, under the same experimental conditions, and thus a simple way to identify and quantify gene expression levels for all genes in an organism [8]. Microarrays could be made use of to detect genes that are involved in specific diseases, by comparing gene expression in normal and abnormal cells.

On course of the biological experiment, the extraction of mRNA corresponding to two biological tissues of interest (i.e. normal and tumor) is performed. Every mRNA sample is reverse transcribed into complementary DNA (cDNA) copy and labeled with two distinct fluorescent dyes resulting in two fluorescence-tagged cDNA (red Cy5 and green Cy3) [9]. Subsequently, the probes on the cDNA array are mixed and cohybridized with the RNA samples. The resulting samples are then scanned to obtain a 16-bit gray-scale image for each dye. The relative profusion of a specific RNA type in the sample is the measure of the relative intensity of the dyes in each spot. The unequal distribution of probe material in the spot, making spots irregular in shape and size, is induced [10] by a plenty of factors like 1) the

hydrophobicity of the pretreated glass surface, 2) the humidity as the probe dries, and 3) the speed of drying. Every microarray experiment is likely to produce thousands of spots per image and typically with multiple images to be analyzed, the development of automated algorithms is a vital issue so as to be capable of processing the enormous quantity of information in huge image databases [11].

Image processing plays a significant role in the extraction and quantitative analysis of the relative profusion of mRNA on the microarray images. The three stages involved in processing of microarray images [12] are: firstly, gridding is carried out to preliminarily locate the spots and blocks in the images. Subsequently, every individual microarray spot is segmented into foreground and background by making use of the available gridding information. Lastly, intensity extraction, computes foreground fluorescence intensity and background intensities. As the accurateness of the resultant data is vital in posterior analysis, the above stages are quite significant [9]. Gene expression data analysis is the subsequent step of microarray analysis and it is accomplished with different statistical and/or unsupervised learning methods. In general, image segmentation refers to the process of partitioning the image into numerous regions, with every individual region having its own properties. Microarray image segmentation is essential to derive reliable information. Robust segmentation is significant for accurate classification of genes' expression and to extrapolate an assortment of spot quality measures. The objective of segmentation is to classify a pixel as either foreground inside a spot, or as background outside the spot.

Accurate spot segmentation is indispensable for quantifying the image intensity, since microarrays are likely to be affected by a high-level background noise level. Lossless spot segmentation and quantification of spot shape variability can be achieved by means of mathematical morphology methods. Before the spot features can be extracted from a microarray image, initially, we have to segment the microarray image, which is just an intensity image characterized by connected pixels of identical intensity values. Four chief approaches used for the segmentation of an intensity image are: a) Threshold techniques - based on the principle that all pixels whose value lie within a certain range belong to a particular class of hybrid levels, b) Boundary-based techniques [13], which keep track of rapidly changing pixel values at the boundary between two regions, c) Region-based methods [14]- based on comparing one pixel with its neighbor, and if, they all have identical values, they are classified to the same class (an important special case is the algorithm called Seeded Region Growing (SRG) and d) Hybrid techniques- a combination of boundary and region-based methods, and are very reliable in producing closed boundaries [15] (belongs to this class). There is a variety of software available to perform segmentation of spotted microarray images e.g. Spot [16], which uses Seeded Region Growing.

In this research, we have presented a methodology for microarray spot segmentation, to evaluate three common segmentation methods: morphological operators, watershed algorithm and iterative watershed algorithm, based on intensities of the microarray image. The proposed methodology is inputted with a microarray image and is composed of three distinct phases namely, 1) Preprocessing, 2) Gridding and 3) Segmentation use morphological operators, watershed algorithm and iterative watershed algorithm. Initially, a single sub grid of a microarray image is extracted, and is then preprocessed. Subsequently, segment spots are estimated using mean intensity calculation, a correlation function and a background removal method. Finally, the processed microarray sub grid is segmented by using the three segmentation methods employed (morphological operators, watershed algorithm and iterative watershed algorithm). The experimental result of the proposed methodology identifies the better of the three segmentation algorithms employed, based on image intensity and segmentation precision.

The rest of the paper is organized as follows. Section II presents a brief review of some recent significant researches in Microarray image segmentation. A concise description of the techniques utilized in the proposed methodology is given in section III. The proposed methodology for investigating the spot segmentation methods used in microarray image analysis is explained in section IV. Experimental results and analysis of the proposed methodology are discussed in Section V. Finally, concluding remarks are provided in Section VI.

II. REVIEW OF RELATED RESEARCHES:

Numerous researches have been proposed by researchers for the segmentation of Microarray Images. In this section, a brief review of some important contributions from the existing literature is presented.

Eisen *et al.* [17] have implemented fixed circle segmentation, a traditional technique that was first used in ScanAnalyze. It works by assigning the same diameter and shape (circle) to all spots. Fixed circle method is also provided options by GenePix [18] and ScanArray Express [19]. The adaptive circle segmentation technique was proposed to overcome the drawback of fixed circle segmentation and was used in GenePix, ScanAnalyze, ScanArray Express, Imagene, and Dapple [20]. Generally, image segmentation has been successfully achieved using the Seeded Region Growing (SRG) algorithm and has been currently introduced in the processing of microarray image [21]. Histogram-based approach is yet another successful technique that has been made use of in microarray image segmentation. A simple and intuitive idea is to make use of histograms to classify a pixel into either foreground or background. Chen *et al.* [22] have presented a method that utilizes a circular target mask to cover all the foreground pixels, and performs a threshold computation using Mann-Whitney test.

Roberto Hirata *et al.* [23] have presented a mathematical morphology based technique, segments the image into three classes of objects: 1) subarrays (i.e., set of grouped spots), 2) spot box (i.e., the rectangular neighborhood that contains a spot) and 3) spot (i.e., region of the image where there exists signal). Antonio P. G. Damiance *et al.* [24] have proposed a data clustering method for the segmentation of microarray images on the basis of dynamical system modeling. A network containing a number of interacting elements that receives attractions from other elements within a certain region has been employed in their proposed approach. Those attractions, computed by a predefined similarity measure, coerce the elements to converge to their

corresponding cluster centre. Furthermore, the gridding process has been omitted in their proposed model.

Jesus Angulo and Jean Serra [25] have presented an automatic non-supervised set of algorithms for speedy and precise extraction of spot data from DNA microarrays using morphological operators, which were robust to both intensity variation and artifacts. The approach could be summarized as follows. The comparison of the proposed algorithm with two packages: ScanAnalyze and Genepix, has portrayed the robustness and precision of the proposed algorithm. The design of morphological methods that facilitate the normalization and quantification of differential gene expression directly on the image has been their next objective.

Kashif I. Siddiqui *et al.* [26] have proposed a method that makes use of morphological image analysis techniques for segmenting the spots of a microarray image accurately, followed by the utilization of B-Splines for the quantification of the shapes of the segmented spots. The watershed segmentation and other morphological techniques for image analysis are being employed for spot extraction of a gene microarray image. Their method was robust to noise problems that are likely to lead to over segmentation. Also, their procedure necessitated very low computational requirements. The B-spline coefficients of the shape were attained from the detected boundaries of the extracted spots and were stored in a database for quantification of spot variations.

Jesus Angulo [27] has dealt with the development of model-based image processing algorithms to adaptively qualify/segment/quantify every spot in an image based on its morphology. They have also introduced a series of morphological models for spot intensities. The spot feature extraction and classification (without segmenting) is on the basis of performing a conversion from the spot image to polar coordinates and, after computing the radial/angular projections, the computation of granulometric curves and derived parameters from those projections. Moreover, spot contour segmentation can be solved by processing in polar coordinates, calculating the up/down minimal path, which is effortlessly attained with the generalized distance function. A regularized segmentation could be achieved with the proposed model-based technique by the controlling different elements of the algorithm.

An automated algorithm for gridding based on hierarchical refinement was presented by Yu Wang *et al.* [28] to perk up the efficiency, robustness and reproducibility of microarray data analysis. The techniques employed in the proposed algorithm include: morphological reconstruction, global and local rotation detection, non-parametric optimal thresholding and local fine-tuning without any human intervention. They have illustrated that their algorithm can identify and compensate for alignment and rotation problems by utilizing synthetic data and real microarray images of different sizes and with diverse degrees of rotation of subarrays, to attain reliable and robust results.

III. OVERVIEW OF THE TECHNIQUES UTILIZED IN THE PROPOSED METHODOLGY:

The significant techniques utilized in microarray image segmentation namely: Wiener Filter, Adaptive Thresholding, Morphological operators and watershed algorithm are detailed in the below subsections.

A. Wiener Filter

The wiener filter can be defined as a Mean Squared Error optimal stationary linear filter for images degraded by additive noise and blurring. It has to be presumed that the signal and noise processes are second-order stationary (in the random process sense), for the computation of wiener filtering [29]. The Wiener filtering is a linear estimation of the original image. The approach is based on a stochastic framework. The orthogonality principle implies that the Wiener filter in Fourier domain can be expressed as follows:

$$W(f_1, f_2) = \frac{H*(f_1, f_2)S_{xx}(f_1, f_2)}{|H(f_1, f_2)|^2 \ S_{xx}(f_1, f_2) + S_{\eta\eta}(f_1, f_2)} \qquad (1)$$

Where $S_{xx}(f_1, f_2)$, $S_{\eta\eta}(f_1, f_2)$ are respectively power spectra of the original image and the additive noise, and $H(f_1, f_2)$ is the blurring filter. It is simple to see that the Wiener filter has two separate parts, an inverse filtering part and a noise smoothing part. It not only performs the deconvolution by inverse filtering (high pass filtering) but also removes the noise with a compression operation (low pass filtering) [30].

B. Adaptive Threshold

The adaptive thresholding method [31] is on basis of analyzing the statistical parameters like arithmetic mean, geometrical mean and standard deviation of the sub-band coefficients. Some researchers have applied local adaptive thresholding scheme that binarizes and enhances poor quality and degraded document for locating significant textual information [32]. The input to a typical adaptive thresholding is a grayscale or color image and, in its simplest implementation, outputs a binary image symbolizing the segmentation. For each pixel in the image, a threshold has to be calculated. If the pixel value is below the threshold, it is set to the background value or else it takes the foreground value.

C. OTSU Threshold

The objective of thresholding is to recognize and extract the target from its background based on the distribution of gray levels or texture in image objects. It serves as an important technique for image segmentation. Otsu's method is made use of to achieve automated histogram shape-based image thresholding, [33] or, the reduction of a gray level image to a binary image. The algorithm works with the assumption that the input image for thresholding is composed of two classes of pixels (e.g. foreground and background), which then computes the optimum threshold by separating the two classes such that their combined

spread (intra-class variance) is minimal. In Otsu's method, we perform an exhaustive search for the threshold that minimizes the intra-class variance, termed as a weighted sum of variances of the two classes:

$$\sigma_\omega^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \qquad (2)$$

Weights ω_i represent the probabilities of the two classes separated by a threshold t and σ_i^2 variances of these classes. Otsu depicts that the minimization of the intra-class variance is the identical to the maximization of the inter-class variance [34]:

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \qquad (3)$$

that is expressed in terms of ω_i (class probabilities) and μ_i (class means) which in turn can be iteratively updated. The above idea results an effective algorithm.

Algorithm

a. Computation of the histogram and probabilities of every intensity level

b. Initialize $\omega_i(0)$ and $\mu_i(0)$

c. Step through all possible thresholds $t = 1 \cdots\cdots$ maximum intensity

    i. Update ω_i and μ_i

    ii. Compute $\sigma_b^2(t)$

d. Desired threshold corresponds to the maximum $\sigma_b^2(t)$.

D. Morphological Operators

In this section, we present a brief description of the fundamental principles, definitions and notations made use of in mathematical morphology [35], [36]. A branch of digital image processing and analysis is mathematical morphology that utilizes the concepts from algebra (set theory, complete lattices) and geometry (translation, distance, convexity). The application of the morphological operation on a binary image is defined as binary morphology. Let *f be* function that defines a grayscale image defined on $\Re^2 \to \Re$, B be a planar structuring element such that $B \subseteq \Re^2$ and ψ be an image operator that transforms a grayscale image *f* into another image in accordance with some specific task. An important morphological operator erosion (dilation) is defined as,

$$\psi_{erosion}(f) = \underset{(\varepsilon,\eta)\in B}{\Lambda} f(x+\varepsilon, y+\eta) = (f \ominus B) \qquad (4)$$

$$\psi_{dilation}(\mathrm{f}) = \underset{(\varepsilon,\eta)\in B}{\Lambda} f(x-\varepsilon, y-\eta) = (f \oplus \mathrm{B}) \qquad (5)$$

Erosion (dilation) replaces the value of the image f at a pixel $(x;y)$ by the infimum (supremum) of the values of f over a structuring element B (B' - reflection of B around the origin), which results in "shrinkage" ("expansion") of the image. Another extensively used and very important morphological operation is called structural opening (structural closing) and is defined as,

$$f \circ \mathrm{b} = (\mathrm{f} \ominus \mathrm{b}) \oplus \mathrm{b} \qquad (6)$$

$$f \bullet b = (f \oplus \mathrm{b}) \ominus \mathrm{b} \qquad (7)$$

and is made use of to undo the effect of erosion (dilation) by the application of the associated dilation (erosion) operation. Another valuable operator is the so-called area opening, which is utilized to remove area having grains, from the image based on a threshold level. Mathematically it is defined as,

$$\psi_{aopn}(f,a) = \underset{t\in\Re}{\Lambda}\{(x,y)\in \bigcup\{F_s(t){:}|\ F_s(t)|\geq a\}\} \quad (8)$$

where, $F_s(t)$ is a cross-section of the image intensity f and $F_s(t)|\ s = 1;2;3$ are grains of the cross-section $F(t)$ of the image f and a is the threshold level [37]. Conversely, area closing is used to fill in the holes in the image, whose area is smaller than a given value. It is important to note that openings (closings) are increasing, anti-extensive (extensive) and idempotent. They both are smoothing filters and are used for smoothing contours of an image, suppressing small islands and cutting narrow isthmuses. The amount of smoothening is determined by the size and shape of the structuring element employed.

A morphological operator ψ is said to be a morphological filter, if it is increasing and idempotent. The combination of different morphological filters also results in a morphological filter. Alternating filters are combination of closings and openings and are defined as,

$$\pi_k(f) = (f \circ kB) \bullet kB \qquad (9)$$

$$\rho_k(f) = (f \bullet kB) \circ kB \qquad (10)$$

Where kB represents $(k-1)$ dilations and k is the size of the filter. The top-hat filtering is defined on an intensity or binary image using a predefined neighborhood or structuring element. Top-hat filtering is the equivalent of subtracting the result of performing a morphological opening operation on the input image from the input image itself.

F. Watershed Algorithm

Watershed algorithm is an image processing algorithm that achieves segmentation by splitting an image into areas, on the basis of image topology. The length of the gradients is construed as elevation information. During the successive flooding of the grey value relief, watersheds with adjacent catchment basins are constructed. This flooding process is performed on the gradient image, i.e. the basins should emerge along the edges. In general, this leads to over-segmentation of the image, particularly for noisy image material, e.g. medical data. F. Meyer, in the early 90's, has introduced one of the most common watershed algorithms [39].

The watershed algorithm starts with a grayscale image F and a set M of markers with distinct labels (in our case, these will be the minima of F), and then, expands the set M as much as possible, while preserving the number of connected components of M:

1. Every neighbor x of every marked area is inserted into a hierarchical queue, based on the priority level corresponding to the grey level $F(x)$. It has to be noted that a point cannot be inserted twice in the queue;

2. A point x at the highest priority level is extracted from the hierarchical queue, i.e., the lowest grey level. If the neighborhood $\Gamma(x)$ of x is composed of only points with the same label, then is marked with this label, and its neighbors that are not yet marked are inserted into the hierarchical queue;

Step 2 has to be repeated until an empty hierarchical queue is met. The watershed lines set are the complement of the set of labeled points. It is to be noted that this algorithm does neither label nor propagate watershed pixels, which "stop" the flooding. Therefore, the watershed lines formed by Meyer's algorithm are generally thinner than lines formed by other watershed algorithms [38].

G. Iterative Watershed Algorithm:

Iterative watershed algorithm is an image segmentation algorithm devised for objects having severely irregular shape with inhomogeneous intensities. Watershed transform is one of the classical and effective methods in the field of segmentation [43]. But in practice, conventional watershed transformation typically produces over-segmentation due to numerous local minima and irregularities existing in real images. Iterative watershed algorithm can be seen as possible enhancement that defines a set of markers to mark those regions that require segmentation.

Overview

The proposed iterative watershed algorithm can be summarized as follows:

   I.   Initial watershed with the ridge constraint
        a.   Perform double-threshold seed detection.
        b.   Perform scale-space ridge detection.
        c.   Superimpose the ridge as the highest waterline and the seed as the lowest water-basin on the original image.
        d.   Apply watershed transform.
  II.   Iterative watershed using Bayes rule
        e.   Set blob probability and feature probability
        f.   Perform Bayes pseudo-blob classification.
        g.   Perform "winner-take-all" merge.
        h.   Repeat Step (e)-(g) until no blob needs to be merged or a maximum number of iterations are reached.

The segmentation accuracy of the iterative watershed depends on two main steps, 1) pseudo-blob classification rule and 2) pseudo-blob merge rule. The first step determines the over-segmented blob, named as pseudo-blob and the second step reallocates the pixels of a pseudo-blob into other real-blobs.

Pseudo-Blob Classification Rule

Let v_s be a feature vector extracted from image pixel at location $s = (x, y)$ and iteration time t. For a blob, the posterior probability of v_s being from a pseudo-blob b_1 or real-blob b_2 is given by

$$P(b_i \mid v_s, t) = \frac{P(v_s \mid b_i, t)P(b_i \mid t)}{P(v_s \mid t)}, i = 1,2 \tag{11}$$

Then, using the Bayes decision rule, the pixel is classified as belonging to a pseudo-blob if the feature vector satisfies,

$$P(b_1 \mid v_s, t) > P(b_2 \mid v_s, t) \tag{12}$$

Nothing that the feature vector v_s associated the pixel $s = (x, y)$ are either from a real-blob or a pseudo-blob, it follows that

$$P(v_s \mid t) = P(v_s \mid b_1, t)P(b_1 \mid t) + P(v_s \mid b_2, t)P(b_2 \mid t) \tag{13}$$

Taking all the pixels in a blob into account, and substituting (11) and (13) in (12), it becomes

$$2P(b_1 \mid t) > \sum_{s \in b_1} P(v_s \mid t) \Big/ \sum_{s \in b_1} P(v_s \mid b_1, t) \tag{14}$$

The prior probability of pseudo-blob at iteration t is updated recursively using equation (15),

$$P(b_1 \mid t) = \alpha P(b_1 \mid t-1) \; and \; \alpha = \left(\alpha_0 / P(b_1 \mid 0)\right)^{1/N} \tag{15}$$

Where, N is the maximum number of iterations, making $P(b_1 \mid N) = \alpha_0$ at the final iteration. For simplicity, in this work, we chose feature vector v_s as a binary ridge descriptor.

$$P(v_s \mid t) = \begin{cases} 1 & s = ridge \\ 0 & s \neq ridge \end{cases} \tag{16}$$

Pseudo-Blob Merge Rule

There are two fundamental ideas by which the pixels inside a pseudo-blob can be re-grouped, either pixel-by-pixel or all pixels as a whole. The grouping of all pixels as a whole has produced better results, and hence, the merge rule is based on "winner-take-all" principle.

- Label the pixels inside a pseudo-blob as blank.
- Re-watershed the images to let the surrounding blobs encroach the pseudo-blob pixels.
- Count the maximum occurrence label of the pseudo-blob pixels.
- Assign the maximum occurrence label to all pseudo-blob pixels.
-

IV. PERFORMANCE EVALUATION OF MICROARRAY SPOT SEGMENTATION

Image Segmentation is defined as the process of isolating objects in the image from the background i.e., partitioning the image into disjointed regions, such that each region is homogeneous with respect to some property. Consequently, spot segmentation can be defined as the process of extracting the appropriate homogenous spots and the noise background, having the desired homogeneity property, from a microarray image. In this section, the proposed approach for evaluating two important spot segmentation methods: Morphological operators and watershed algorithm is detailed. The processes involved in the approach proposed for the evaluation of the segmentation methods are depicted in the block diagram below (Figure 1). The raw microarray images used to evaluate the proposed methodology were collected from the database [40]. A number of preprocessing steps are initially applied to visibly locate and estimate the spots and the distance between two consecutive rows and columns of spots in a microarray. Once the spot space is estimated, we segment the using morphological operators and watershed algorithm. The presented spot segmentation methodology for microarray images consists of three steps:

1. Preprocessing.
2. Gridding.
3. Segmentation via Morphological operator, Watershed algorithm and Iterative Watershed algorithm.
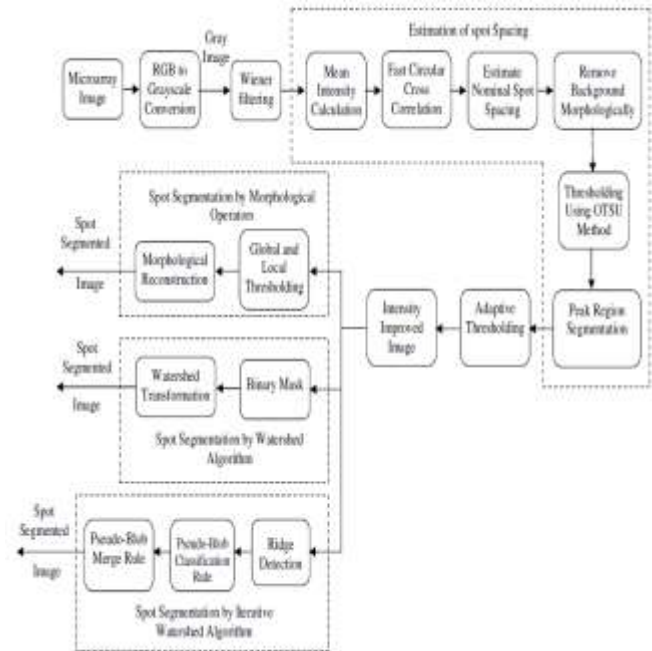

Figure: 1 Block diagram of Microarray image spot segmentation

A. Preprocessing:

The proposed research considers the original microarray images that consist of indexed images with an associated RGB (red-green-blue) color map by "R" and "G" for "red" and "green" with R corresponding to the Cy5 and G to Cy3 respectively. Each microarray image is composed of an enormous number of n sub microarrays and each sub microarray contains A columns and B rows of gene spots, where each spot corresponds to the location of a cDNA probe to which mRNA from the cells of interest have been bound. The microarray images are to be first converted into grayscale, so as to preprocess the images and address its spots. It is just a supplementary step as the original red and green channels can be used to extract the true intensities after the spots are located. With the aim of obtaining the grayscale images, the RGB color model is converted to a YIQ (luminance-hue saturation) model. The principal advantage of the YIQ model is that, it decouples luminance and chromaticity, codifying in different channels grayscale and color data. In order to attain the grayscale information, the I and Q components are set to zero; this is chiefly aimed at enhancing the spots position color plane at the pixel's location with 8 bits for each color. The Y component is attained through the weighted sum of the R, G and B channels, as described in Eq. (17) [34]:

$$Y = 0.299R + 0.587G + 0.114B \qquad (17)$$

The microarray image is composed of a set of gray levels and its value ranges from x_{min} to x_{max} of the image at point $x = (X, Y)$. The points and the spaces between genes

are denoted by the lower points and the zero values, which symbolize the background points and the points that contain noise. The microarray image converted from RGB to grayscale is bound to contain noises and artifacts. Hence, in order to remove those noises and artifacts and sharpen the image, the grayscale image is filtered with the aid of wiener filtering (see Eq. (1)).

B. Gridding:

Initially, to locate the regular grid of spots in the microarray image, we calculate the mean intensity for each column of the image. Using the mean intensity computed, we identify the centers of the spots and the gaps between the spots. In an ideal scenario, the spots in a microarray image would be periodically spaced consistently printed. But practically, the spots are likely to have different sizes and intensities, and hence the horizontal profile obtained will be irregular. Hence, the proposed methodology employs fast circular cross correlation to enhance the self similarity of the profile. The fast circular cross correlation algorithm uses Fast Fourier Transform (FFT) to compute the circular cross correlation of two periodic signal vectors. The circular cross-correlation of two signals x and y in C^N can be given as [41]

$$\hat{r}_{xy}(l) \overset{\Delta}{=} \frac{1}{N}(x * y)(l) \overset{\Delta}{=} \frac{1}{N}\sum_{n=0}^{N-1}\overline{x(n)}y(n+l), \quad l = 0,1,2,\mathsf{L}, N-1 \quad (18)$$

Where l is an integer variable, $\hat{r}_{xy}(l)$ is an estimator of the true cross-correlation, which is an assumed statistical property of the signal itself. The pseudo code for estimating the nominal spacing of the microarray spots is presented below,

Pseudo code:

$$Y \rightarrow Gray \text{ Image}$$

$$M \rightarrow Mean \text{ Intensity of Image}$$

$$FC_c \rightarrow Fast\ Circular\ Cross\ Correlation$$

$$P \rightarrow Peak\ Value$$

$$L_s \rightarrow Left \text{ slope}$$

$$R_s \rightarrow Right \text{ slope}$$

$$N_s \rightarrow Nominal \text{ Spacing}$$

$$M << \{x : p(x)\}$$

$$\text{Where } p(x) = \sum_{j=1}^{N}(Y_j)$$

$$d = FC_c(M)$$

$$L_s = diff\left(d(j_1, j_m)\right)$$

$$R_s = diff\left(d(i_1, i_m)\right)$$

$$if\ \left((L_s > 0 \,\&\, \&(R_s > 0)\right)\ then$$

$$P \leftarrow d$$

$$end\ if$$

$$N_s = round\left(Median(diff(P))\right)$$

The next step of gridding is to perform background removal. With the aid of the nominal spacing estimate, we can devise a filter to remove the background noise from the intensity profile. Once the background noises are removed, we have clean and anchored gaps between the peaks of the spotted microarray image. To segment the peak regions, we employ the Otsu's thresholding method (see section 3.3). The Otsu's thresholding method is a simple but effectual global automatic thresholding method, in which the threshold values are determined with the help of the statistical properties of the data. Subsequently, we extort the center of the peaks that correspond to the horizontal centers of the spots with the aid of blob analysis. Then, we determine the midpoints between the adjacent peaks, that yield the grid point locations of the microarray image. With the aid of aforesaid processes, we have estimated the spot space for vertical grid (vertical spacing).

Once the spot space for vertical grids are estimated, the next step is to estimate the spot space for the horizontal grid. To estimate the horizontal spot spacing, initially, we transpose the image and repeat all the aforesaid processes on the transposed image. Using the pairs of neighboring grid points (vertical grid and horizontal grid), the position and size coordinates of each spot are tabulated into ROI called Region Of Interest. Once the spot spaces are estimated and the position coordinates are determined, we can segment the spots from background by using an appropriate thresholding algorithm. In the proposed methodology, initially, we apply a single adaptive threshold to the whole image. We cannot achieve effective segmentation using a single threshold level because of the large differences in spot brightness. To improve the spot brightness, we transform the intensity values to logarithmic space that yields the spot segmented image with improved intensity.

C. SEGMENT SPOTS

All the aforesaid steps prepare the input microarray image for spot segmentation. Finally, the intensity improved microarray image after adaptive thresholding is segmented using morphological operators, watershed transformation and iterative watershed transformation. The application of the morphological operators (erosion, dilation opening and closing) achieves spot segmentation, but the silhouettes of the segmented spots will contain some pinholes. Hence, the pinholes in the segmented image will be filled with the aid of the morphological reconstruction algorithm. Spot segmentation of a microarray image achieved using morphological techniques was efficient for

image analysis and also insensitive to noise problems leading to over segmentation.

Also, on the same intensity improved microarray image, we employ the watershed transform. Prior to watershed transform, we first convert the intensity improved image to a binary image. Then, the watershed transform (see section 3.5) is applied to the binary image for spot segmentation. The resulting watershed lines will serve as the boundaries around spots of the microarray image. One advantage of watershed segmentation is that, we can extract and characterize noise background features since the watershed provides regions in the neighborhood of each spot.

The third segmentation considered for study, iterative watershed transformation is then applied to the same intensity improved microarray image. The iterative watershed algorithm [43] employed in the proposed study extends the original watershed algorithm in three aspects. (i) The controlled markers consist of seeds and boundaries and each selected by, 1) Seeds via an effective double-threshold approach and 2) Boundaries via ridge detection (ii) the detected boundaries need not to be closed, and hence, a seed need not correspond to one closed boundary. (iii) The over-segmentation dilemma is solved by means of iterative classification using Bayes classification rule and a merging procedure. The iterative watershed algorithm works well for microarray images without clear boundaries and homogenous intensities and also prevents over segmentation.

V. EXPERIMENTAL RESULTS AND ANALYSIS:

This section presents the experimental results of the proposed methodology and the formal investigation of the parameters considered for evaluating the two spot segmentation methods employed namely, 1) Morphological operators and 2) Watershed transformation. The proposed methodology has been evaluated on real microarray images collected from the publicly available database that belongs to Lymphoma/Leukemia Molecular Profiling Project Gateway [40]. The experimental results obtained for the proposed methodology are portrayed in Figure 2, 3, 4 & 5.
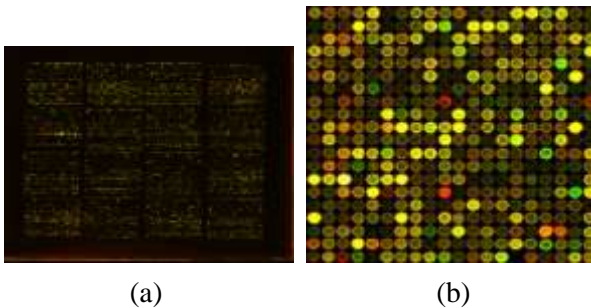

(a)   (b)

Figure: 2 Original Microarray Image (a) Microarray image (b) A subgrid of the microarray image
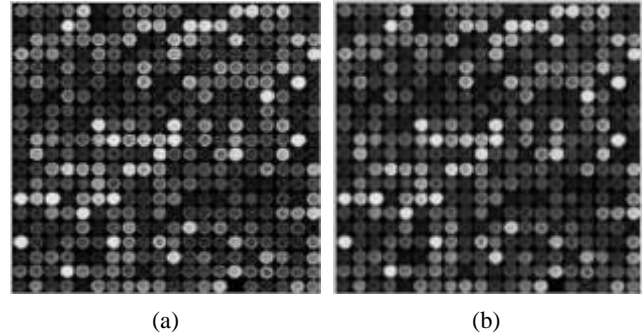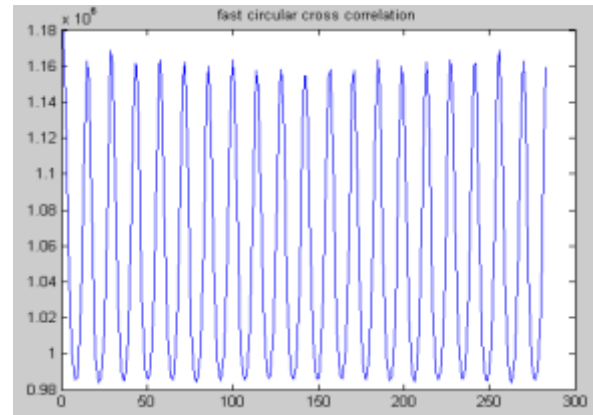

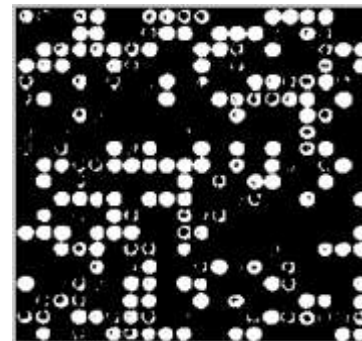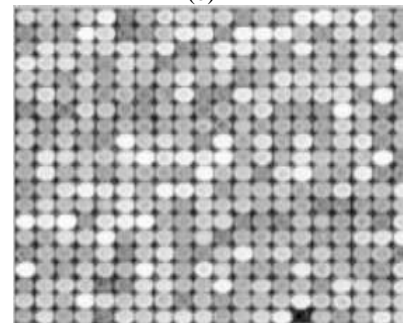(a)   (b)

Figure: 3 Preprocessing (a) Grayscale Converted Subgrid of the Microarray Image (b) Weiner Filtered Image
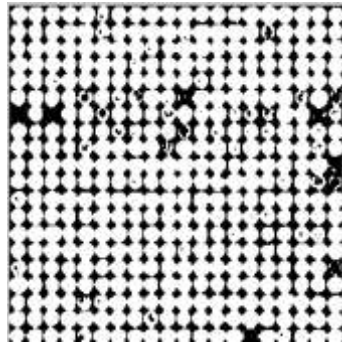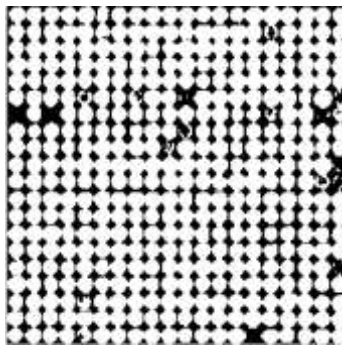

(a)

(b)

(c)

Figure: 4 Gridding (a) Application of Fast Circular Cross Correlation (b) Initially Segmented image By Adaptive Thresholding (c) Intensity Improved Image
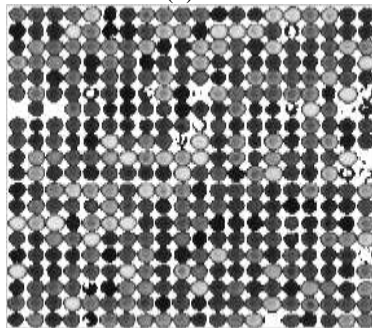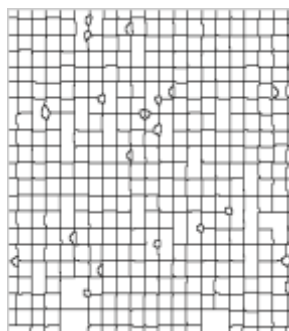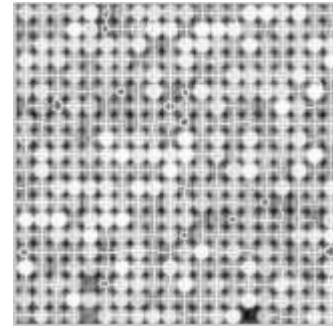
(a)

(b)

(c)

Figure: 5 Spot segmentation Using Morphological Operators (a) Spot segmentation by Combined Thresholding (b) Spot segmentation after morphological 'imfill' operator (c) Final Morphologically Segmented output
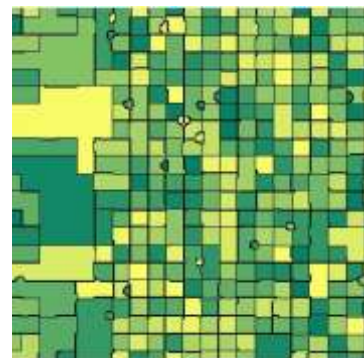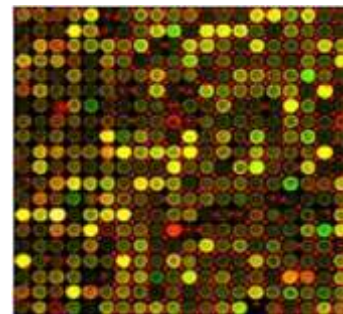

(a)

(b)

Figure: 6 (a) & (b) Spot Segmentation Using Watershed transform

(b)

Figure: 7 (a) & (b) Spot Segmentation Using Iterative Watershed transform

Here, we present the investigation to determine the better of the three spot segmentation algorithms employed (Morphological operators, Watershed algorithm and Iterative Watershed algorithm) in the proposed methodology, emphasizing the importance of segmentation algorithms in microarray image analysis. Regarding the importance of segmentation, a critical component in carrying out microarray experiments is the segmentation of images, following scanning. Works in the literature have portrayed significant differences in precision and the number of differentially expressed genes based on the employed segmentation algorithms. Differences amid segmentation methods could effect from differences in summarizing pixel-level data as pixel-to-pixel variability has been revealed to have vital effects on data quality [42].

In the proposed methodology, we specifically evaluate the segmentation effectiveness of the three segmentation methods on

the microarray image, based on the intensity levels of the segmented output. From the results, it could be deduced that spot segmentation of the microarray image has been achieved in an efficient manner using morphological techniques when compared to the watershed and iterative watershed transformation. The microarray image segmented using the morphological operators produced results with better intensity profile giving a better illustration of the segmented spots. The image clarity (better intensity levels and more spot segmentation) of the segmented output will facilitate gene expression evaluation, the subsequent step of microarray image analysis. Generally, the use of Mathematical Morphology as the paradigm for data analysis provides the following advantages. First, because of the strong mathematical foundation, it leads to reproducible results. Second, the choice of every parameter used in the method is motivated by physical measurements such as shape and dimensions. Moreover, it enables easy customization of the proposed methodology for each commercial microarray.

VI. CONCLUSION

In this paper, we have proposed a methodology for investigating the accuracy of spot segmentation of a microarray image, using morphological image analysis techniques, watershed algorithm and iterative watershed algorithm. The microarray image inputted has been subjected to spotted microarray image preprocessing and gridding. Subsequently, the resulting microarray sub grid has been segmented using morphological operators, watershed algorithm and iterative watershed algorithm. A formal investigation of the three segmentation algorithms employed (morphological operators, watershed algorithm and iterative watershed algorithm) has been conducted on the basis of the precision of segmentation and its intensity profile. The experimental results have illustrated the segmentation effectiveness of the proposed methodology and also the better of the three segmentation algorithms employed for segmentation. Moreover, it can be showed that morphological segmentation permits us to perform better shape and intensity analysis when compared with watershed and iterative watershed segmentation.

VII. REFERENCES

[1] Luis Rueda and Li Qin, "An Unsupervised Learning Scheme for DNA Microarray Image Spot Detection", Proc. of the First International Conference on Complex Medical Engineering, Takamatsu, Japan, pp. 996-1000, 2005.

[2] A. Loury and L. Bodrossy, "Highly parallel microbial diagnostics using oligonucleotide microarrays", Clinica Chimical Acta, Vol. 363, pp. 106-119, 2006.

[3] G. López-Campos, L. Garcia-Albert, F. Martín Sanchez and A., Garcia-Saiz, "Analysis and management of HIV peptide microarray experiments", Methods of Information in Medicine, Vol. 45, pp. 158-162, 2006.

[4] G. López-Campos, M. Coiras, J.P. Sánchez-Merino, M.R. López-Huertas, I. Spiteri, F. Martín-Sanchez, P. Pérez-Breña, "Oligonucleotide microarray design for detection and serotyping

of human respiratory adenoviruses by using a virtual amplicon retrieval software", Journal of Virological Methods, vol. 145, pp. 127-136, 2007.

[5] S. Draghici, "Data Analysis Tools for DNA Microarrays", Chapman & Hall/CRC, 2003.

[6] M. Schena, "Microarray Analysis", John Wiley & Sons, 2002.

[7] Lopez-Campos G, Lopez Alonso V, Martin-Sanchez F., "Addressing the Biomedical Informatics Needs of a Microarray Laboratory in a Clinical Microbiology Context", Stud Health Technol Inform., S.K. Andersen et al. (Eds.), IOS Press, Vol. 136, pp. 45-50, 2008.

[8] Antonis Daskalakis, Dionisis Cavouras, Panagiotis Bougioukos, Spiros Kostopoulos, Pantelis Georgiadis, Ioannis Kalatzis, George Kagadis, and George Nikiforidis, "Effective Quantification of Gene Expression Levels in Microarray Images Using a Spot-Adaptive Compound Clustering-Enhancement-Segmentation Scheme", O. Gervasi and M. Gavrilova (Eds.): ICCSA 2007, Vol. LNCS 4707, Part III, pp. 555 – 565, 2007.

[9] Nikolaos Giannakeas, Petros S. Karvelis, and Dimitrios I. Fotiadis, "A Classification-Based Segmentation of cDNA Microarray Images using Support Vector Machines", 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, pp. 875-878, August 20-24, 2008.

[10] Hedge, P., QI, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Earle-Hugues, J., Snesrud, E., Lee, N. and Quackenbush, J., "A concise guide to cDNA microarray analysis", Biotechniques, Vol. 29, pp. 548–562, 2000.

[11] Raphael Gottardo, Julian Besag, Matthew Stephens, Alejandro Murua, "Probabilistic segmentation and intensity estimation for microarray images", Biostatistics, Vol. 7, No. 1, pp. 85–99, 2006.

[12] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative motoring of gene expression patterns with a complementary DNA microarray," Science, vol. 270, pp. 467-470, 1995.

[13] L. S. Davis, "A survey of edge detection techniques", Comput. Graphics and Image Processing, vol. 4, pp. 179-205, 1975.

[14] S. W. Zucker, "Region growing: Childhood and adolescence", Comput. Graphics and Image Processing, vol. 5, pp. 382-399, 1976.

[15] S. Beucher, "The watershed transformation applied to image segmentation", 10th Conf. on Signal and Image Processing in Microscopy and Microanalysis, Cambridge, UK, 1991.

[16] Yee Hwa Yang, Michael Buckley, Sandrine Dudoit and Terry Speed, "Comparison of methods for image analysis on cDNA microarray data", University of California, Berkeley, Technical Report # 584, 2002.

[17] M. Eisen, "ScanAlyze User's Manual", M. Eisen, 1999.

[18] Axon Instruments, "Genepix 4000A: User's Manual", Axon Instruments Inc., 1999.

[19] GSI Lumonics, "QuantArray Analysis Software: Operator's Manual", 1999.

[20] J. Buhler, T. Ideker, and D. Haynor, "Dapple: Improved Techniques for Finding Sports on DNA Microarrays", Technical Report UWTR 2000-08-05, University of Washington, 2000.

[21] Y. Yang, M. Buckley, S. Dudoit, and T. Speed, "Comparison of Methods for Image Analysis on cDNA Microarray Data", Journal of Computational and Graphical Statistics, Vol. 11, pp. 108-136, 2002.

[22] Y. Chen, E. Dougherty, and M. Bittner, "Ratio-based Decision and the Quantitative Analysis of cDNA Microarray Images", Journal of Biomedical Optics, Vol. 2, pp. 364-374, 1997.

[23] Roberto Hirata, Jr., Junior Barrera, Ronaldo F. Hashimoto, Daniel O. Dantas and Gustavo H. Esteves, "Segmentation of Microarray Images by Mathematical Morphology", Real-Time Imaging, Vol. 8, No. 6, pp. 491-505, December 2002.

[24] Antonio P. G. Damiance, Jr., Liang Zhao, Andre C. P. L. F. Carvalho, "A dynamical model with adaptive pixel moving for microarray images segmentation", Real-Time Imaging, Special issue on imaging in bioinformatics: Part III, Vol. 10, No. 4, pp. 189 - 195, August 2004.

[25] Jesus Angulo and Jean Serra, "Automatic analysis of DNA microarray images using mathematical morphology", Bioinformatics, Vol. 19, No. 5, pp.553-562, 2003.

[26] Kashif I.Siddiqui, Alfred O. Hero and Matheen M. Siddiqui, "Mathematical Morphology applied to Spot Segmentation and Quantification of Gene Microarray Images", In: proceedings of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, vol.1, pp.926 - 930, 3-6 November 2002.

[27] Jesus Angulo, "Polar modelling and segmentation of genomic microarray spots using mathematical morphology", Image Analysis and Stereology, Vol. 27, No.2, pp. 107-124, June, 2008.

[28] Yu Wang, Marc Q. Ma, Kai Zhang, Frank Y. Shih, "A hierarchical refinement algorithm for fully automatic gridding in spotted DNA microarray image processing", Information Sciences: an International Journal, Vol. 177, No. 4, pp. 1123-1135, February 2007.

[29] Saeed V. Vaseghi, "Advanced signal processing and digital noise reduction (Paperback)", John Wiley & Sons Inc, pp. 416, July 1996.

[30] Dr. Rich Baraniak, Ramesh Neelamani, "Weiner Filtering", from http://www.owlnet.rice.edu/~elec539/Projects99/BACH/proj2/wiener.html

[31] D. Gnanadurai, and V. Sadasivam, "An Efficient Adaptive Thresholding Technique for Wavelet Based Image Denoising", International Journal of Signal Processing, Vol. 2, No. 2, pp. 114-119, 2006.

[32] Yahia S. Halabi, Zaid SA"SA, Faris Hamdan, Khaled Haj Yousef, "Modeling Adaptive Degraded Document Image Binarization and Optical Character System", European Journal of Scientific Research, Vol. 28, No.1, pp.14-32, 2009.

[33] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", Journal of Electronic Imaging, Vol. 13, No. 1, pp. 146–165, 2004.

[34] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Sys., Man., Cyber., Vol. 9, pp. 62 – 66, 1979.

[35] H. J. A. M. Heijmans, "Morphological Image Operators", Boston: Academic Press, 1994.

[36] Dougherty, E. R., "An Introduction to Morphological Image Processing", SPIE Press, Bellingham, 1992.

[37] J. Goutsias and S. Batman, "Morphological Methods for Biomedical Image Analysis", Handbook of Medical Imaging, SPIE Optical Engineering Press, Ed., M. Sonka and J. M. Fitzpatrick, Vol. 2, pp. 175-272, 2000.

[38] Laurent Najman and Michel Couprie, "Watershed algorithms and contrast preservation", Lecture Notes in Computer Science, Vol. 2886, pp.62-71, 2003.

[39] Meyer, F., "Topographic distance and watershed lines", Signal Processing, Special issue on Mathematical Morphology, Vol. 38, pp. 113-126, 1994.

[40] "Microarray Images", from http://llmpp.nih.gov/lymphoma/data/rawdata/

[41] Smith, J.O., "Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications", Second Edition, from http://ccrma-www.stanford.edu/~jos/st/Cross_Correlation.html

[42] Brown C.S., Goodwin, P.C. and Sorger, P.K., "Image metrics in the statistical analysis of DNA microarray data", Proc. Natl Acad. Sci. USA, Vol. 98, No. 16, pp. 8944–8949, 2001.

[43] Li Chen, Min Jiang, and JianXun Chen, "Image Segmentation Using Iterative Watersheding Plus Ridge Detection", 2009 IEEE International Conference on Image Processing, iro, Egypt • Saturday, November 7 - Tuesday, November 10, 2009.