Abstract

Microarray is an important tool and powerful technique that is used to analyse the expression of DNA in organisms for large scale gene sequences and gene expressions. Microarray technology allows massively parallel, high throughput profiling of gene expression in a single hybridization experiment. Processing of microarray images provides the input for further analysis of the extracted microarray data. It includes the following stages:

- Gridding is the process of identifying the area within an image that contain a single spot, subgrid and row, column within that subgrid.
- Segmentation is the process of grouping the pixels in a spot with similar features (this step results in the separation of foreground and background pixels).
- Intensity extraction calculates red and green foreground fluorescence intensity pairs.

Preprocessing

A major factor that complicates the task of image analysis and data mining is that microarray experiments involve a number of error-prone steps (occurring during fabrication, target labeling and hybridization), which induce noise on the resulting images. Microarray images are also corrupted by irregularities in the shape, size, and position of the spot. Most of the methods proposed by researchers have either considered high SNR (signal-to-noise ratio) images or various assumptions on factors such as type of thresholding used, parametric assumptions and decomposition levels. Also, some of the methods have discussed only impulse, Gaussian noise and fluorescent noise.

In this research work, methods are developed, for low SNR images and estimate many other types of noises other than Gaussian, impulse and fluorescent noises. Denoising of microarray image is an essential and challenging task in the pre-processing step of microarray image analysis.
Here, we have proposed two novel techniques for image restoration.

1. In the first technique a restoration system model is developed, that estimates the type of noise (standard noise) and appropriate filter is used to denoise the image. If the input image consists of mixture of noise sources then bilateral filter is used to denoise the noisy image. After applying the filtering techniques the denoised image becomes blurred. Hence Blind Deconvolution technique is used, which estimates PSF repeatedly over iterations.

2. In the second technique Mathematical Morphology is used to denoise the image. Two transforms Top-hat (erosion followed by dilation) and Bottom-hat (dilation followed by erosion) are used for restoration, further structuring element regulates the extent of thickening. To remove the insignificant spots, adaptive threshold concept is used.

**Gridding**

Existing gridding methods pay little attention to pre-processing of noisy microarray images and focus mainly on spot localization and spot segmentation. The aim of the present study is to propose a method that can deal with rotations, misalignments, and local deformations of the ideal rectangular grid. It is also noise-resistant and efficient even under adverse conditions such as the appearance of various spot sizes or the absence of spots.

We propose here three techniques for automatic gridding of noisy microarray images.

1. In the first technique, the microarray image is skew corrected, noise removed using adaptive thresholds computed on various segments, spatial topology of spots detected, gridding lines are positioned and finally grids are refined.

2. In the second technique, microarray image is skew corrected, noise removed using adaptive thresholds computed on various segments, spatial topology of spots detected, bounding boxes are fitted for each spot and thus facilitating the subsequent step of segmentation.

3. In the third technique, a novel approach for automatic gridding of noisy microarray images is presented. The projection profiles of the binarized image are obtained, noise removed using morphological operations. The unduly non uniform distance between grid lines in noisy microarray images are corrected using coefficient of variation (CV) of the successive differences in the gridding locations.
Abstract

Segmentation
The outcome of grid alignment is separation of spots in the microarray image. The next task is to segment these spots. This is nothing but calculation of foreground intensity, spot size and shape. Not all works have addressed low SNR, a common phenomena with microarray images. In this research work, we have developed methods to take care of this prevalent condition.

We propose here two techniques for automatic segmentation of noisy microarray images

1. Extended K-means algorithm, takes advantage of the spatial information around each spot of interest, calculating the of the found foreground F through existing K-means algorithm. Subsequently the method analyses the neighbors at the boundary of F carefully, before making the final decision of including some of the noisy pixels to F. Also the proposed work makes no assumption on the shape and the size of the spot.

2. Refinement of K-means algorithm, takes into account, intensity of pixels, which is the output of K-means method, as well as spatial information as in Multifeature method. Pixels that are neighbors to foreground are analyzed carefully from the two angles: intensity and spatial neighborhood, before making final decision.

Quantification
Quantification is nothing but ideal combinations of appropriate preprocessing, gridding, segmentation methods for a given image. Thus, it is not only important to assess the performance of each analysis stage independently, as it has been done in most previous studies (i.e. whether gridding or spot boundary detection is accurate or not) but also the performance of all processing steps as a whole in terms of reproducibility and validity in computing gene expression levels. All twelve frameworks using proposed restoration, gridding and segmentation methods are shown to be superior than existing contemporary methods through extensive experimentation. We have used the available metrics of performance evaluation namely Mean Absolute Error(MAE), Coefficient of Variation(CV), Jeffrey’s measure of divergence, Sum of Square Difference(SSD), Box Plot, MA plot, Scatter plot, concordance correlation and coefficient of determination. We have also proposed a new metric of quantification evaluation using Confidence Interval estimation.