

Sub-grid and Spot Detection in DNA Microarray Images Using Optimal Multi-level Thresholding

Iman Rezaeian and Luis Rueda

School of Computer Science, University of Windsor
401 Sunset Ave., Windsor, ON, N9B3P4, Canada
{rezaeia, lrueda}@uwindsor.ca

Abstract. The analysis of DNA microarray images is a crucial step in gene expression analysis, since any errors in early stages are propagated in future steps in the analysis. When processing the underlying images, accurately separating the sub-grids and spots is of extreme importance for subsequent steps that include segmentation, quantification, normalization and clustering. We propose a fully automatic approach that first detects the sub-grids given the entire microarray image, and then detects the locations of the spots in each sub-grid. The approach first detects and corrects rotations in the images by an affine transformation, followed by a polynomial-time optimal multi-level thresholding algorithm to find the positions of the sub-grids and spots. Additionally, a new validity index is proposed in order to find the correct number of sub-grids in the microarray image, and the correct number of spots in each sub-grid. Extensive experiments on real-life microarray images show that the method performs these tasks automatically and with a high degree of accuracy.

Keywords: Microarray image gridding; image analysis; multi level thresholding.

1 Introduction

Microarrays are one of the most important technologies used in molecular biology to massively explore the abilities of the genes to express themselves into proteins and other molecular machines responsible for different functions in an organism. These expressions are monitored in cells and organisms under specific conditions, and are present in many applications in medical diagnosis, pharmacology, disease treatment, just to mention a few. We consider DNA microarrays, which are produced on a slide, typically, in two channels. Scanning the slides at a very high resolution produces images composed of sub-grids of spots. Image processing and analysis are two important aspects of microarrays, since the aim of the whole experimental procedure is to obtain meaningful biological conclusions, which depends on the accuracy of the different stages, mainly those at the beginning of the process. The first task is gridding, which if done correctly, helps substantially improve the efficiency of the subsequent steps that include segmentation, quantification, normalization and data mining. When producing DNA microarrays, many parameters are specified, such as the number and size of spots, number of sub-grids, and even their exact location. However, many physicochemical factors

produce noise, misalignment, and even deformations in the sub-grid template that it is virtually impossible to know the exact location of the spots after the scanning is performed, at least with the current technology. Roughly speaking, gridding consists of determining the spot locations in a microarray image (typically, in a sub-grid). The gridding process requires the knowledge of the sub-grids in advance in order to proceed.

Many approaches have been proposed for sub-gridding and spot detection. The Markov random field (MRF) is a well known approach that applies specific constraints and heuristic criteria [11]. Another gridding method is mathematical morphology, which represents the image as a function and applies erosion operators and morphological filters, helping remove peaks and ridges from the topological surface of the images [7]. A method for detecting spot locations based on a Bayesian model has been recently proposed, and uses a deformable template model to fit the grid of spots in such a template using a posterior probability model which learns its parameters by means of a simulated-annealing-based algorithm [1,6]. Another method for finding spot locations uses a hill-climbing approach to maximize the energy, seen as the intensities of the spots which are fit to different probabilistic models [10]. Fitting the image to a mixture of Gaussians is another technique that has been applied to gridding microarray images by considering radial and perspective distortions [5]. A Radon-transform-based method that separates the sub-grids in a DNA microarray image has been proposed in [8]. Other approaches for DNA microarray gridding include the following. A gridding method that performs sub-gridding and spot detection is the one proposed in [13], which performs a series of steps including rotation detection based on a simple method that compares the running sum of the topmost and bottommost parts of the image. This method, which detects rotation angles with respect to one of the axes, either x or y , has not been tested on images having regions with high noise (e.g. bottommost $\frac{1}{3}$ of the image is quite noisy). Another method for gridding DNA microarray images uses an evolutionary algorithm to separate sub-grids and detect the positions of the spots [4]. The approach is based on a genetic algorithm that discovers parallel and equidistant line segments that compose the grid structure. Using maximum margin is another method for automatic gridding of DNA microarray images based on the maximization of the margin between the rows and columns of the spots [2]. In another approach, properties of planar (2D) grids are addressed from a mathematical point of view and an algorithm for recognizing distorted grids with perspective transformations is presented [5]. The approach involves recognizing parameters of affinely distorted grids by fitting Gaussian mixture models to grid spectrums, rebuilding the grid structures via a generating iteration based on the acquired parameters, and eliminating nonlinear effects caused by perspective transformations with the median of infinite lines from local structures.

In this paper, we propose a fully automatic approach that first detects the sub-grids given the entire microarray image, and then detects the locations of the spots in each sub-grid. The method proposed here uses an optimal multi-level thresholding algorithm to find the positions of the sub-grids in the image and the positions of the spots in each sub-grid. Additionally, a new validity index is proposed in order to find the correct number of sub-grids in the microarray image, and the correct number of spots in each sub-grid.

2 The Proposed Gridding Method

A DNA microarray image typically contains a number of sub-grids and each sub-grid contains a number of spots arranged in rows and columns. The aim is to perform a two-stage process in such a way that the sub-grid locations are found in first stage, and then spot locations within a sub-grid can be found in the second stage. Consider an image (matrix) $A = \{a_{ij}\}$, $i = 1, \dots, n$ and $j = 1, \dots, m$, where $a_{ij} \in \mathbb{Z}^+$ and a_{ij} represents the intensity of pixel (i,j) (usually, a_{ij} is in the range [0..65,535] in a TIFF image). The aim of gridding is to obtain a matrix G (grid) where $G = \{g_{ij}\}$, $i = 1, \dots, n$ and $j = 1, \dots, m$, $g_{ij} = 0$ or $g_{ij} = 1$ (a binary image), with 0 meaning that g_{ij} belongs to a grid separator, and 1 meaning the pixel is inside a spot region. This image could be thought of as a “free-form” grid. However, in order to restrict our definition to a rectangular grid, our aim is to obtain vectors \mathbf{v} and \mathbf{h} , $\mathbf{v} = [v_1, \dots, v_m]^t$, $\mathbf{h} = [h_1, \dots, h_n]^t$, where $v_i \in [1, m]$ and $h_j \in [1, n]$. Each vertical and horizontal vectors are used to separate sub-grids and spots.

The sub-grids in a microarray image are detected by applying the Radon transform as a preprocessing phase and then using optimal multilevel thresholding in the next stage. By combining optimal multilevel thresholding and the β index (Eq. 13), the correct number of thresholds (sub-grids or spots) can be found. Figure 1 depicts the process of finding the sub-grids in a microarray image. The input for the Radon transform process is a microarray image and the output of the whole process is the location (and partitioning) of the sub-grids. Analogously, the locations of the spots in each sub-grid are found by using optimal multilevel thresholding combined with the proposed β index to find the best number of rows and columns of spots. The input for this process is a sub-grid (already extracted from the sub-gridding step) and the output is given by the partitioning of the sub-grid into spots (spot regions).

We apply the Radon transform as a preprocessing step (to the raw images) in order to detect and correct rotations, if any, in the whole image or in a sub-grid. Rotations of an image can be seen in two different directions, with respect to the x and y axes. The aim is find two independent angles of rotation for an affine transformation, and for this the Radon transform is applied. Details on the Radon transform and how to use it in correcting rotations can be found in [8].

3 Optimal Multilevel Thresholding

Multilevel thresholding is one of the most widely-used techniques in image processing, including segmentation, classification and object discrimination. Given a histogram with frequencies or probabilities for each bin, the aim of multilevel thresholding is to divide the histogram into a number of groups (or classes) of contiguous bins in such a way that a criterion is optimized. In microarray image gridding, we compute the vertical (or horizontal) running sum of pixel intensities, obtaining a histogram in which each bin represents one column (or row respectively), and the running sum of intensities correspond to the frequency of that bin. The frequencies are then normalized in order to be considered as probabilities. Figure 2 depicts a typical DNA microarray image (AT-20387-ch2, see its description in section 5) that contains 12×4 sub-grids, along with

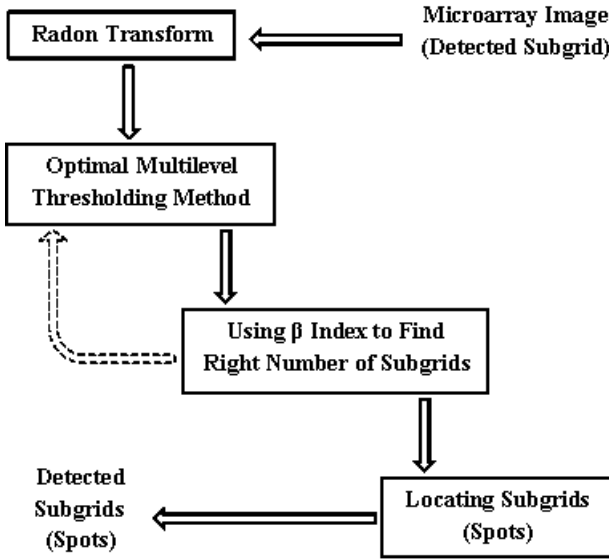


Fig. 1. Schematic representation of the process for finding sub-grids (spots) in microarray images (detected sub-grids)

the corresponding histograms representing the horizontal and vertical running sums. Each histogram is then processed (see below) to obtain the optimal thresholding that will determine the locations of the lines separating the sub-grids. Analogously, we apply the same method to each sub-grid to obtain the corresponding lines separating the spot regions.

Consider a histogram H , an ordered set $\{1, 2, \dots, n - 1, n\}$, where the i th value corresponds to the i th bin and has a probability, p_i . Given an image, $A = \{a_{ij}\}$, H can be obtained by means of the horizontal (vertical) running sum as follows: $p_i = \sum_{j=1}^m a_{ij}$ ($p_j = \sum_{i=1}^n a_{ij}$). We also consider a threshold set T , defined as an ordered set $T = \{t_0, t_1, \dots, t_k, t_{k+1}\}$, where $0 = t_0 < t_1 < \dots < t_k < t_{k+1} = n$ and $t_i \in \{0\} \cup H$. The problem of multilevel thresholding consists of finding a threshold set, T^* , in such a way that a function $f : H^k \times [0, 1]^n \rightarrow \mathbb{R}^+$ is maximized/minimized. Using this threshold set, H is divided into $k + 1$ classes: $\zeta_1 = \{1, 2, \dots, t_1\}$, $\zeta_2 = \{t_1 + 1, t_1 + 2, \dots, t_2\}$, \dots , $\zeta_k = \{t_{k-1} + 1, t_{k-1} + 2, \dots, t_k\}$, $\zeta_{k+1} = \{t_k + 1, t_k + 2, \dots, n\}$. The most important criteria for multilevel thresholding are the following [9]:

Between class variance:

$$\Psi_{BC}(T) = \sum_{j=1}^{k+1} \omega_j \mu_j^2 \tag{1}$$

where $\omega_j = \sum_{i=t_{j-1}+1}^{t_j} p_i$, $\mu_j = \frac{1}{\omega_j} \sum_{i=t_{j-1}+1}^{t_j} i p_i$;

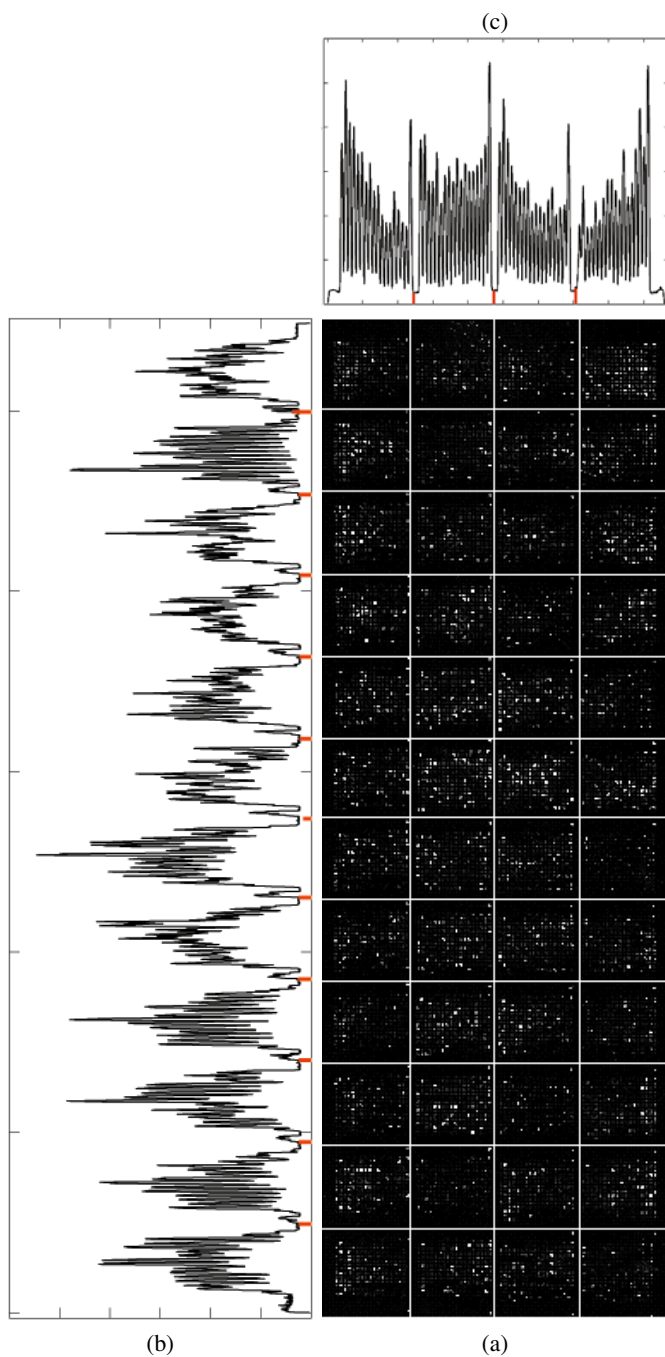


Fig. 2. (a) detected sub-grids in AT-20387-ch2 microarray image, (b) vertical histogram and detected valleys correspond to vertical lines, (c) horizontal histogram and detected valleys correspond to horizontal lines

Entropy-based:

$$\Psi_H(T) = \sum_{j=1}^{k+1} H_j \tag{2}$$

where $H_j = -\sum_{i=t_{j-1}+1}^{t_j} \frac{p_i}{\omega_j} \log \frac{p_i}{\omega_j}$;

Minimum error:

$$\Psi_{ME}(T) = 1 + 2 \sum_{j=1}^{k+1} \omega_j (\log \sigma_j - \log \omega_j) \tag{3}$$

where $\sigma_j^2 = \sum_{i=t_{j-1}+1}^{t_j} \frac{p_i(i-\mu_j)^2}{\omega_j}$.

A dynamic programming algorithm for *optimal* multilevel thresholding was proposed in our previous work [9], which is an extension for irregularly sampled histograms. For this, the criterion has to be decomposed as a sum of terms as follows:

$$\Psi(T_{0,m}) = \Psi(\{t_0, t_1, \dots, t_m\}) \triangleq \sum_{j=1}^m \psi_{t_{j-1}+1, t_j}, \tag{4}$$

where $1 \leq m \leq k + 1$ and the function $\psi_{l,r}$, where $l \leq r$, is a real, positive function of p_l, p_{l+1}, \dots, p_r , $\psi_{l,r} : H^2 \times [0, 1]^{l-r+1} \rightarrow \mathbb{R}^+ \cup \{0\}$. If $m = 0$, then $\Psi(\{t_0\}) = \psi_{t_0, t_0} = \psi_{0,0} = 0$. The thresholding algorithm can be found in [9]. In the algorithm, a table C is filled in, where $C(t_j, j)$ contains the optimal solution for $T_{0,j} = t_0, t_1, \dots, t_j$, $\Psi^*(T_{0,j})$, which is found from $\min\{t_j\} \leq t_j \leq \max\{t_j\}$. Another table, $D(t_j, j)$, contains the value of t_{j-1} for which $\Psi^*(T_{0,j})$ is optimal. The worst-case time complexity of the algorithm has been shown to be $\Theta(kn^2)$.

To implement the between-class variance criterion, the function $\Psi_{BC}(T)$ is expressed as: $\Psi_{BC}(T) = \sum_{j=1}^{k+1} \omega_j \mu_j^2 = \sum_{j=1}^{k+1} \psi_{t_{j-1}+1, t_j}$, where $\psi_{t_j+1, t_{j+1}} = \omega_j \mu_j^2$. We consider the temporary variables, a and b , which are computed as follows:

$$a \leftarrow p_{t_{j-1}+1} + \sum_{i=t_{j-1}+2}^{t_j} p_i, \quad \text{and} \tag{5}$$

$$b \leftarrow (t_{j-1} + 1)p_{t_{j-1}+1} + \sum_{i=t_{j-1}+2}^{t_j} ip_i. \tag{6}$$

Since from (5) and (6), a and b are known, then $\psi_{t_{j-1}+2, t_j}$, for the next step, can be re-computed as follows in $\Theta(1)$:

$$a \leftarrow a - p_{t_{j-1}+1}, \tag{7}$$

$$b \leftarrow b - (t_{j-1} + 1)p_{t_{j-1}+1}, \text{ and} \tag{8}$$

$$\psi_{t_{j-1}+2, t_j} \leftarrow \frac{b^2}{a}. \tag{9}$$

Similar decomposition allows that the minimum error and entropy-based criteria be recomputed in $\Theta(1)$ [9].

4 Automatic Detection of the Number of Sub-grids and Spots

Finding the correct number of sub-grids in a microarray image and number of spots in each sub-grid is one of the most important phases in sub-grid and spot detection. For this, we resort on validity indices used for clustering. By analyzing the traditional indices for clustering validity, we found that combining these indices with one of our measures, we propose a new index of validity for this specific problem. Initially, we considered four clustering validity indices (cf. [12]) in the context of the partitioning obtained by the multilevel thresholding method. We found that the best is the I index, which is defined as follows:

$$I(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^2, \quad (10)$$

where, $E_K = \sum_{i=1}^K \sum_{k=1}^{n_i} p_k \|k - z_i\|$, $D_K = \underbrace{\max}_{i,j=1}^K \|z_i - z_j\|$, n is the total number of points in data set, and z_k is the center of the k th cluster. To find the best number of thresholds, we perform an exhaustive search on all possible values of K , from 2 to \sqrt{n} [3]. The value of K for which $I(K)$ is maximal is considered to be correct number of clusters. We must note that the complexity of the algorithm remains $\Theta(kn^2)$, since the index I is computed for all values of K as the optimal thresholding algorithm fills in table C .

Based on this index and another measurement (see Eq. (11) below), we propose a new index for finding the correct number of sub-grids or spots. We consider the average value of the thresholds in a histogram, which is computed as follows:

$$A(K) = \frac{1}{K} \sum_{i=1}^K f(t_i), \quad (11)$$

where t_i is the i th threshold found by optimal multilevel thresholding and $f(t_i)$ is its respective value in the histogram.

The proposed index computes the value of A for different numbers of thresholds, K . Then, the best number of thresholds K^* can be found as follows:

$$K^* = \arg \min_{1 \leq K \leq \delta} \left\{ \frac{1}{K} \sum_{i=1}^K f(t_i) \right\} \quad (12)$$

where δ is the maximum number of thresholds and equals to \sqrt{n} [3]. Based on our experimental studies, the best results were obtained from a combination of our proposed index (11) and the I index (10) is as follows:

$$\beta(K) = \frac{I(K)}{A(K)} \quad (13)$$

For maximizing $I(K)$ and minimizing $A(K)$, the value of $\beta(K)$ must be maximized. Thus, the best number of thresholds K^* based on the β index is given by:

$$K^* = \arg \max_{1 \leq K \leq \delta} \beta(K) = \arg \max_{1 \leq K \leq \delta} \frac{\left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^2}{\frac{1}{K} \sum_{i=1}^K f(t_i)} \quad (14)$$

5 Experimental Results

For the experiments, two different kinds of cDNA microarray images have been used. The images have been selected from different sources, and have different scanning resolutions, in order to study the flexibility of the proposed method to detect sub-grids and spots with different sizes and features. The first set of images has been drawn from the Stanford Microarray Database (SMD), and corresponds to a study of the global transcriptional factors for hormone treatment of *Arabidopsis thaliana*¹ samples. Ten images were selected for testing the proposed method, and they correspond to channels 1 and 2 for experiments IDs 20385, 20387, 20391, 20392 and 20395. The images have been named using AT (which stands for *Arabidopsis thaliana*), followed by the experiment ID, and the channel number (1 or 2). The images have a resolution of 1910×5550 pixels and are in TIFF format. The spot resolution is 24×24 pixels per spot. Also, each image contains 48 sub-grids, arranged in 12 rows and 4 columns. The second test suite consists of a set of images from Gene Expression Omnibus (GEO) and corresponds to an Atlantic salmon macrophage study² samples. Eight images were selected for testing the proposed method, and they correspond to channels 1 and 2 for experiments IDs GSM16101, GSM16389 and GSM16391 and also channels 1 of GSM15898 and channels 2 of GSM15898. The images have been named using GSM followed by the experiment ID, and the channel number (1 or 2). The images have a resolution of 1900×5500 pixels and are in TIFF format. The spot resolution is 12×12 pixels per spot. Also each image contains 48 sub-grids, arranged in 12 rows and 4 columns.

To assess the performance of the detection method, we consider the following. We call *false positive* (FP) a grid line that separates an area into two different sub-areas and at least one of them does not contain a spot or a sub-grid. Similarly, a *false negative* (FN) occurs when two adjacent areas containing spots or sub-grids are not separated by a grid line. True positives (TP) and true negatives (TN) are obtained by the corresponding differences between the total number of cases minus FP or FN, respectively. Considering N as the total number of grid lines in the image, accuracy is calculated as $\frac{(TP+TN)}{N}$.

We have used the between-class variance as the thresholding criteria, since it is the one that delivers the best results. Table 1 shows the results of applying the proposed method for sub-grid and spot detection on the Thaliana dataset. All the sub-grids in each image are detected accurately, and also spot locations in each sub-grid can be detected efficiently with an average accuracy of 96.2% for this dataset. The same sets of experiments were repeated for the GEO dataset and the results are shown in Table 2. The sub-grids in each microarray image are accurately detected with a 100% accuracy and the spot locations in each sub-grid are detected efficiently with an average performance of 96% for this dataset. As shown in Tables 1 and 2, for all of images, in the sub-grid detection phase, the false negative and false positive rates are both 0%, yielding an accuracy of 100%. This means the proposed method works perfectly in sub-grid

¹ The images can be downloaded from smd.stanford.edu, by searching “Hormone treatment” as category and “Transcription factors” as subcategory.

² The images can be downloaded from ncbi.nlm.nih.gov, by selecting “GEO Datasets” as category and searching the name of images.

Table 1. Accuracy results of detected sub-grids and spots for each image in the Thaliana dataset and their respective FP and FN rates

Image	Sub-grid Detection			Spot Detection		
	False Negative	False Positive	Accuracy	False Negative	False Positive	Accuracy
AT-20385-CH1	0.0%	0.0%	100%	6.5%	0.4%	93.1%
AT-20385-CH2	0.0%	0.0%	100%	3.3%	1.5%	95.4%
AT-20387-CH1	0.0%	0.0%	100%	7.4%	0.5%	92.1%
AT-20387-CH2	0.0%	0.0%	100%	0.0%	0.6%	99.4%
AT-20391-CH1	0.0%	0.0%	100%	0.0%	1.2%	98.8%
AT-20391-CH2	0.0%	0.0%	100%	3.7%	1.3%	98.8%
AT-20392-CH1	0.0%	0.0%	100%	0.7%	1.0%	95.0%
AT-20392-CH2	0.0%	0.0%	100%	3.1%	1.3%	98.3%
AT-20395-CH1	0.0%	0.0%	100%	6.5%	0.4%	95.6%
AT-20395-CH2	0.0%	0.0%	100%	6.5%	0.4%	95.7%

Table 2. Accuracy results of detected sub-grids and spots for each image in the GEO dataset and their respective FP and FN rates

Image	Sub-grid Detection			Spot Detection		
	False Negative	False Positive	Accuracy	False Negative	False Positive	Accuracy
GSM15898-CH1	0.0%	0.0%	100%	3.2%	0.1%	96.7%
GSM15899-CH2	0.0%	0.0%	100%	3.2%	0.2%	96.6%
GSM16101-CH1	0.0%	0.0%	100%	3.0%	0.0%	97.0%
GSM16101-CH2	0.0%	0.0%	100%	3.1%	0.0%	96.9%
GSM16389-CH1	0.0%	0.0%	100%	5.8%	0.0%	94.2%
GSM16389-CH2	0.0%	0.0%	100%	3.1%	0.0%	96.9%
GSM16391-CH1	0.0%	0.0%	100%	6.7%	0.0%	93.3%
GSM16391-CH2	0.0%	0.0%	100%	3.6%	0.0%	96.4%

detection. Additionally, in the spot detection phase, accuracy of the proposed method is very high, being above 96% in both cases.

One of the reasons for the slightly lower accuracy in spot detection is that the distance between spots is smaller than the distance between sub-grids. In both datasets, there are approximately eight pixels between adjacent spots, and approximately 30 pixels horizontally and 100 pixels vertically between sub-grids in the Thaliana dataset, and 200 pixels between sub-grids in the GEO dataset. Another possible reason for this behavior is that the number of pixels in each sub-grid is far lower than a microarray image (around $\frac{1}{50}$). Thus, existing noise affects the spot detection phase much more than the noise present in the sub-grid extraction stage. It is important to highlight, however, that the sub-grid detection process is not affected by the presence of noise at all.

We have also performed a visual analysis in order to obtain a different perspective of our results. Figure 3 shows the detected sub-grids from the AT-20387-ch2 image (left) and the detected spots in one of sub-grids (right). As shown in the figure, the proposed method finely detects the sub-grids location at first, and in the next stage, each sub-grid

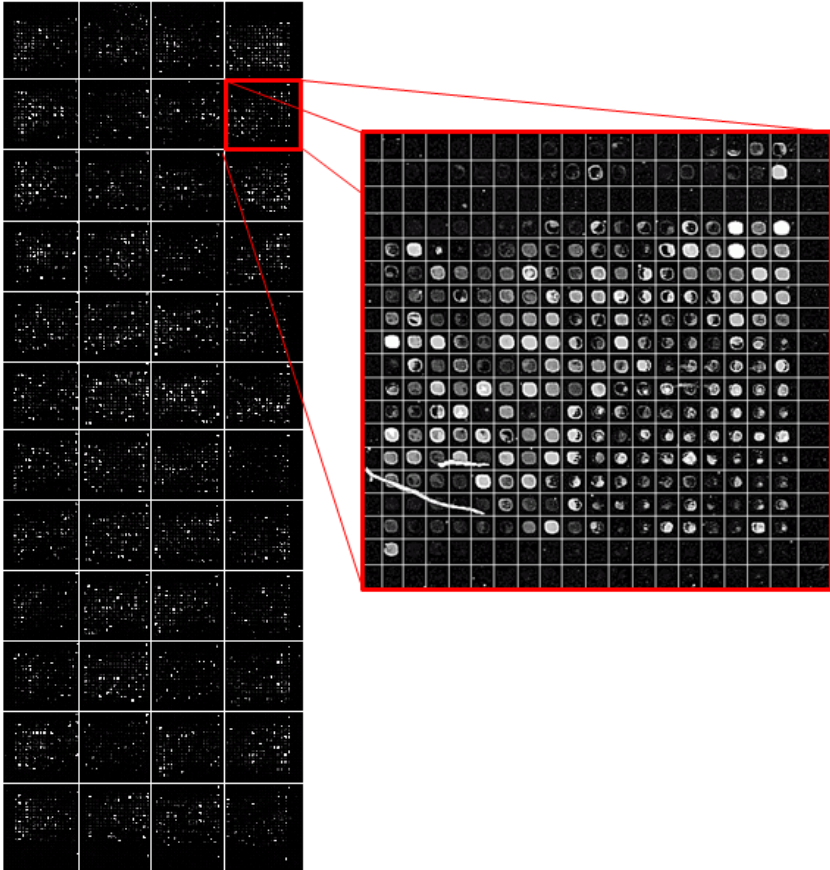


Fig. 3. Detected sub-grids in AT-20387-ch2 microarray image (left) and detected spots in one of sub-grids (right)

is divided nicely into the corresponding spots with the same method. The robustness of the proposed method is so high that spots in sub-grids can be detected very well in noisy conditions such as those observable in the selected sub-grid in Figure 3. The ability to detect sub-grids and spots in different microarray images with different resolutions and spacing (12×12 pixels for each spot in GEO dataset vs 24×24 pixels for each spot in SMD dataset) is another important feature of the proposed method.

To visually analyze the efficiency of the proposed method to automatically detect the correct number of spots and sub-grids, we show in Figure 4 a plot for the indices against the number of classes (sub-grids) for AT-20387-ch2. Sub-figures (a), (b) and (c) represent the values of the index functions for the horizontal lines for the I index, A index and β index respectively, while (d), (e) and (f) contain the plots of the indices for the vertical separating lines. We observe that it could be rather difficult to find the correct number of classes (sub-grids) using solo the I index or the A index,

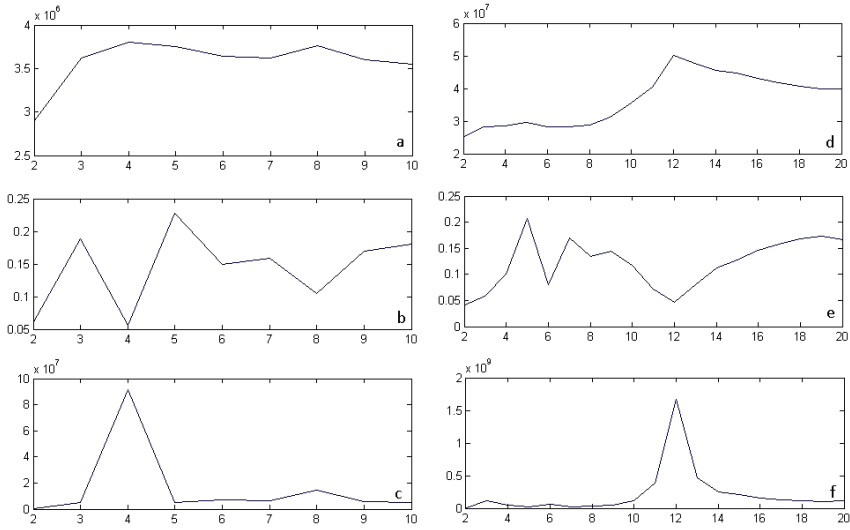


Fig. 4. Plots of the index functions for the AT-20387-ch2 microarray image. (a),(d): I index; (b),(e): A index; (c),(f): β index. The plots on the left correspond to the horizontal lines, while the ones on the right correspond to the vertical lines. the x axis corresponds to the number of classes (sub-grids).

while the β index clearly reveals the correct number of horizontal and vertical sub-grids by producing an almost flat curve with a pronounced peak at 4 and 12 respectively. For example, it is clearly observable in Figure 4 (a) that the I index could miss the correct number of sub-grids, 4, by showing two peaks (local and global maxima).

6 Conclusions

A novel method for separating sub-grids and spot centers in cDNA microarray images has been proposed. The method performs three main steps involving the Radon transform for detecting rotations with respect to the x and y axes, the use of polynomial-time optimal multilevel thresholding to find the positions of the lines separating sub-grids and spots, and a new index for detecting the correct number of sub-grids and spots. The proposed method has been tested on real-life, high-resolution microarray images drawn from two sources, the SMD and GEO. The results show that (i) the rotations are effectively detected and corrected by affine transformations, (ii) the sub-grids are accurately detected in all cases, even in abnormal conditions such as extremely noisy areas present in the images, (iii) the spots in each sub-grid are accurately detected using the same method, and (iv) because of using an optimal and parameterless algorithm for detecting threshold locations, this method can be used for microarray images with different features, and also for images with various spot sizes and configurations effectively.

References

1. Ceccarelli, B., Antoniol, G.: A Deformable Grid-matching Approach for Microarray Images. *IEEE Transactions on Image Processing* 15(10), 3178–3188 (2006)
2. Bariamis, D., Maroulis, D., Iakovidis, D.: M^3G : Maximum Margin Microarray Gridding. *BMC Bioinformatics* 11, 49 (2010)
3. Duda, R., Hart, P., Stork, D.: *Pattern Classification*, 2nd edn. John Wiley and Sons, Inc., New York (2000)
4. Zacharia, E., Maroulis, D.: Micoarray image gridding via an evolutionary algorithm. In: *IEEE International Conference on Image Processing*, pp. 1444–1447 (2008)
5. Qi, F., Luo, Y., Hu, D.: Recognition of perspectively distorted planar grids. *Pattern Recognition Letters* 27(14), 1725–1731 (2006)
6. Antoniol, G., Ceccarelli, M.: A Markov Random Field Approach to Microarray Image Gridding. In: *Proc. of the 17th International Conference on Pattern Recognition*, pp. 550–553 (2004)
7. Angulo, J., Serra, J.: Automatic Analysis of DNA Microarray Images Using Mathematical Morphology. *Bioinformatics* 19(5), 553–562 (2003)
8. Rueda, L.: Sub-grid Detection in DNA Microarray Images. In: *Proceedings of the IEEE Pacific-RIM Symposium on Image and Video Technology*, pp. 248–259 (2007)
9. Rueda, L.: An Efficient Algorithm for Optimal Multilevel Thresholding of Irregularly Sampled Histograms. In: *Proceedings of the 7th International Workshop on Statistical Pattern Recognition*, pp. 612–621 (2008)
10. Rueda, L., Vidyadharan, V.: A Hill-climbing Approach for Automatic Gridding of cDNA Microarray Images. *IEEE Transactions on Computational Biology and Bioinformatics* 3(1), 72–83 (2006)
11. Katzer, M., Kummer, F., Sagerer, G.: A Markov Random Field Model of Microarray Gridding. In: *Proceeding of the 2003 ACM Symposium on Applied Computing*, pp. 72–77 (2003)
12. Maulik, U., Bandyopadhyay, S.: Performance Evaluation of Some Clustering Algorithms and Validity Indices. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 24(12), 1650–1655 (2002)
13. Wang, Y., Ma, M., Zhang, K., Shih, F.: A Hierarchical Refinement Algorithm for Fully Automatic Gridding in Spotted DNA Microarray Image Processing. *Information Sciences* 177(4), 1123–1135 (2007)