MICROARRAY IMAGE ADDRESSING BASED ON THE RADON TRANSFORM

G. Antoniol, M. Ceccarelli,
Research Centre on Software Technologies-RCOST, University of Sannio, Benevento, Italy.

A. Petrosino
Department of Applied Sciences University "Parthenope", Naples, Italy

ABSTRACT

A fundamental step of Microarray image analysis is the detection of the grid structure for the accurate localization of each spot, representing the state of a given gene in a particular experimental condition. This step is known as gridding or microarray addressing. Most of the available microarray gridding approaches require human intervention; for example, to specify landmarks, some points in the spot grid, or even to precisely locate individual spots. Automating this part of the process can allow high throughput analysis [11].

This paper is aimed towards the development fully automated procedures for the problem of automatic microarray gridding. Indeed, many of the automatic gridding approaches are based on two phases, the first aimed at the generation of an hypothesis consisting into a regular interpolating grid, whereas the second performs an adaptation of the hypothesis. Here we show that the first step can efficiently be accomplished by using the the Radon Transform, whereas the second step could be modeled by an iterative posterior maximization procedure [2].

1. INTRODUCTION

DNA microarrays [5] technology has a large impact in many application areas, such as diagnostic human diseases and treatments (determination of risk factors, monitoring disease stage and treatment progress, etc.), agricultural development (plant biotechnology) or quantification of genetically modified organisms, drug discovery and design. In cDNA microarrays, a set of genetic DNA probes (from several hundreds to some thousands) are spotted on a slide. Two populations of mRNA, tagged with fluorescent dyes, are then hybridized with the slide spots, and finally the slide is read with a scanner. The outlined process produces two images, one for each mRNA population, each of which varies in intensity according to the level of hybridization represented as the quantity of fluorescent dye contained in each spot. The image of figure 1 is an example of a scanned microarray image where one of the images is put in the green channel, and the other in the red channel.

Fig. 1. Distorted Microarray Image.

Image analysis is an essential aspect of microarray experiment; measures over the scanned image can substantially affect successive steps such as clustering and identification of differentially expressed genes. Scanned microarray image processing has three main tasks [11]: (i) gridding, which is the process of assigning the coordinates to the spots, (ii) segmentation, it allows the separation between foreground and background pixels, and (iii) intensity extraction, it consists in the computation of average foreground and background intensities for each spot of the array. Most of available gridding approaches require human intervention, for example, to specify some points in the spot grid or even to precisely register individual spots. Automating this part of the process will allow high throughput analysis. Therefore, this paper focuses on the development of an automated procedure for the problem of automatic microarray gridding.

Some efforts to help automatic microarray data processing have recently emerged in literature. However, most of them impose different kinds of restrictions and are based on stringent assumptions. For example, the approaches in [1][7] require that grid rows and columns are strictly aligned.
with the \( x \) and \( y \) image axes. Other approaches, such as [8] and [9], rely on the Bayesian paradigm to deal with uncertainty and noise. In particular, the approach presented in [9] describes a second order prior for microarray gridding, whereas [8] presents a general approach to the grid matching and image warping problems. Here we show that the computation of a reference regular grid which can be accomplished via the Radon Transform of a filtered microarray image. This reference grid can then be used as prior within a Bayesian Framework [2]. In particular, potential starting spot locations are computed according to the Orientation Matching Transform [4]. The use of a rectangula- 

2. GRID MATCHING

The aim of the paper is the definition of a general procedure for the first step of many microarray gridding operations. This consists into the computation of a reference grid which is then adjusted in successive steps. We call this step Grid Matching. It computes an initial grid guess consisting into a regular grid, \( T \) described by six parameters: \( < x_0, y_0, \alpha, \beta, \Delta x, \Delta y > \), where \( x_0 \) and \( y_0 \) are the coordinates of the upper left spot, \( \alpha \) and \( \beta \) are the angles between the grid directions and the \( x \) axis, and \( \Delta x \) and \( \Delta y \) are the grid spacings in both directions. Let \( M \) and \( N \), the number of spots along the grid directions, the grid points \( i, j \), \( i = 1, ..., M \), \( j = 1, ..., N \) have coordinates:

\[
x_{ij} = x_0 + i\Delta y \sin(\beta) + j\Delta x \sin(\alpha)
y_{ij} = y_0 + i\Delta y \cos(\beta) + j\Delta x \cos(\alpha)
\]

The main step of the GridMatching phase consists of the computing of the Radon Transform (RT) of the filtered images. By analyzing the peak of the RT, we can compute all the six parameters of the grid.

The steps of the GridMatching phase are sketched in figure 2. Each channel of the input image is filtered by the Orientation Matching Transform [4] (OMT) having as parameters just the minimum and maximum spot radii. OMT is an extension of the Hough Transform for circles, and has several advantages: it is a correlation-based transform, it allows to treat in the same manner a set of radius in a wide range, it can be tailored to recognize clear spots on dark background and vice versa or both. Given an image \( I(x, y) \), its OMT, \( OMT(u, v) \) represents the evidence that the point \( (u, v) \) be the center of a circular object.

\[
\mathcal{R}\{I\}(s, \phi) = \int I(x, y) \delta(x \cos \phi + y \sin \phi - s) dx dy
\]

In order to find the principal directions of the grid, we consider just the direction of the projection by integrating the space variable \( s \) in the transform. This will allow one to select the directions having the maximum score corresponding to the angles along which we have the maximum number of aligned spot centers. In particular, our algorithm computes the two main peaks of the function:

\[
\Gamma(\phi) = \int_s \mathcal{R}\{OMT\{I\}\}(s, \phi) ds
\]

Our peak detection algorithm computes the two highest local maxima of a smoothed version of \( \Gamma(\phi) \). The reason lies in the fact that to detect the directions along which the spot locations are maximally aligned, \( \mathcal{R}\{I\}(s, \phi) \) should have been scored many times for several values of \( s \). Following the above image example, we report in figure 3 the behavior of the function \( \Gamma(\phi) \) having two distinct peaks corresponding to 4 and 92 degrees, which we select as the principal orientations of the grid.

Once the principal orientations of the grid have been computed, we are able to determine the parameters \( \Delta x \) and...
\( \Delta y \) of the reference grid. This is done again by using the properties of the RT. We actually know that the possible spot centers are maximally aligned along directions \( \alpha \) and \( \beta \). Therefore, if we project the OMT image along one of them, for example \( \alpha \), we obtain a sequence of values which correspond to the spot alignments which are parallel to this direction. Fortunately, we know that the spots form a grid, and the alignments parallel to \( \alpha \) are organized as a sequence of rows which are \( \Delta y \) apart. Therefore, the projection of the OMT image along \( \alpha \) direction will eventually produce a profile having a sequence of peaks which are \( \Delta y \) apart. Once the RT has been computed, it will be enough to extract the columns corresponding to the directions \( \alpha \) and \( \beta \). It is worth noting that this method is a generalization of the algorithm proposed in [7]. Indeed, the use of the RT allows one to develop a framework able to deal with a wide range of directions, on the contrary, the original approach reported in [7] required the grid to be perfectly aligned with the \( x \) and \( y \) axes. Our estimate of \( \Delta x \) is the average distance between successive peaks in \( \mathcal{R}\{OMT[I]\}(. \alpha) \), whereas the estimate of \( \Delta y \) is the average distance between successive peaks in \( \mathcal{R}\{OMT[I]\}(. \beta) \).

Finally, the computation of \( x_0 \) and \( y_0 \) is a straightforward step. It is enough to project back the coordinates of the center of spots and take the minimum of attained values. A plot of the produced result is reported in figure 4.

3. RESULTS

Due to the difficulty of having labeled grid images, and in order to perform an accurate quantitative analysis, as first experiment we investigate the capability of the proposed method over synthetic images representing microarray grids with artificially inserted deformations.

Therefore, for each spot location, the grid generation module randomly selects a radius in the given interval, and then alters the center coordinates according to the adopted noise model. In our experiments, we assume a Gaussian distribution of the spot location noise and a Gaussian distribution of the deviation of the grid directions with respect to the horizontal and vertical directions of the image. In this way we can quantitatively evaluate the robustness of the proposed method with respect to grid noise and deformations. Given a set of reference grid parameters, we run the experiments by generating synthetic images containing two kinds of noise. The first refers to the spot centers and, as already explained, is modeled by a Gaussian function with zero mean and variance \( \sigma_1 \) in both directions, of course \( \sigma_1 \) is measured in pixels.

The second kind of artificially induced distortion refers to the deviation of the grid principal directions from the \( x \) and \( y \) image axes, in this case the angles \( \alpha \) and \( \beta \) are random variables with \( \alpha \sim N(0, \sigma_2) \) and \( \beta \sim N(90^\circ, \sigma_2) \) and \( \sigma_2 \) is measured in degrees. Since the results of the algorithm do not depend on the specific values of the other parameters, such as \( x_0, y_0, \Delta x, \Delta y \), we do not consider particular alterations of these parameters for the quantitative analysis. For each combination of \( \sigma_1 \) and \( \sigma_2 \), we generated a set of 100 images and recorded the matching grids computed by GridMatching. As a figure of merit, we adopt the Mean Square Error (MSE) between the true solution and the solution attained by the algorithm over all the 100 images for each value of \( \sigma_1 \) and \( \sigma_2 \).

Fig. 3. The function \( \Gamma(\phi) \)

Fig. 4. The result of the GridMatching step; the output grid, in this case, has parameters \( < x_0, y_0, \alpha, \beta, \Delta x, \Delta y >= < 
121.35, 30.87, 0.0698, 1.6057, 35.00, 35.00 > \).
The $x_0, y_0$ mean errors are reported in Figure 5 and 6 respectively. $E_{x_0}$ refers to the total mean square error on the $x_0$ parameter, and analogously for the $y_0$ errors. We did not report the error for the $\Delta_x$, $\Delta_y$ parameters since in all cases the algorithm computed the exact values; this is also due to the fact that we imposed such values to be integer valued. Finally, $\alpha$ and $\beta$ estimated values always attained an error below $10^{-4}$ and thus are not reported in the paper. From these results we see that the GridMatching step if very efficient at detecting the grid parameter even in the presence of random deviation which are of the same order of the spot radius, since we used an interval between 4 and 8 pixels. This means the the minimum spot is as the half of the maximum spot radius. It is worth noting that the accuracy of this step affect the overall performance, and makes the step of grid refinement particularly simple as most of the spot centers are already near the grid nodes. The final paper reports a set of experiments over real microarray images, whereas more information and source code can be downloaded from http://www.scoda.unisannio.it.

4. REFERENCES


