

MICROARRAY IMAGE ADDRESSING BASED ON THE RADON TRANSFORM

G. Antoniol, M. Ceccarelli,

A. Petrosino

Reserch Centre on Software Technologies-RCOST,
University of Sannio, Benevento, Italy.

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 at 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.

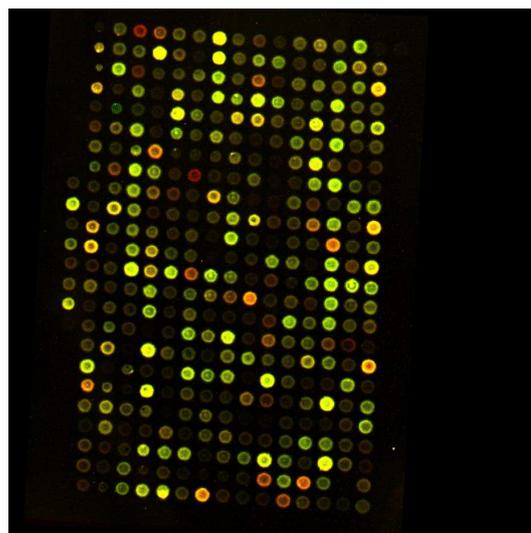


Fig. 1. Distorted Microarray Image.

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 rectangular grid model has also been successfully exploited in [10] together with heuristic search procedure, where several restrictions (alignment with x and y axes) over the grid are imposed without any further adaptation, however computing times may be a problem in the adopted search. Our method improves the approach reported in [10] by allowing arbitrary grid alignments with respect to the image axes and by refining the interpolating grid for dealing with local deformations. Our approach has some similarities with the heuristic method reported [3] where a similar idea of a prior image filtering step is exploited. However, here we show how the steps of angular optimization and parameter optimization, which are performed by heuristic search in [3], can be efficiently implemented by exploiting the properties of the Radon Transform. The idea of using guide spots and the interpolation has also been used in [10], and for arbitrary grid rotations in [2].

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, \mathcal{T} described by six parameters: $\langle x_0, y_0, \alpha, \beta, \Delta x, \Delta y \rangle$, where x_0 and y_0 are the coordinates of the upper left spot, α and β are the angles between the the grid directions and the x axis, and Δx and Δy are the grid spacings in both directions. Let M and N , the number of spots along the grid directions, the grid points i, j , $i = 1, \dots, M$, $j = 1, \dots, N$ 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, $OM(u, v)$ represents the evidence that the point (u, v) be the center of a circular object.

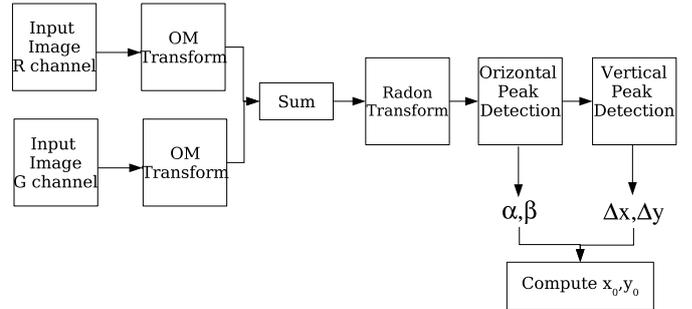


Fig. 2. Overall scheme of the *GridMatching* step.

We can compute the grid parameters α and β , the angles of the principal grid directions with respect to the x axis, as anticipated, by using the Radon Transform (RT) of the transformed image. RT consists of mapping from the image space to the projection space. In the case of images, the 2D RT of an image $I(x, y)$ is formulated as:

$$\mathcal{R}\{I\}(s, \phi) = \int I(x, y) \delta(x \cos \phi + y \sin \phi - s) dx dy \quad (1)$$

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:

$$\Gamma(\phi) = \int_s \mathcal{R}\{OMT\{I\}\}(s, \phi) ds \quad (2)$$

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 Δx and

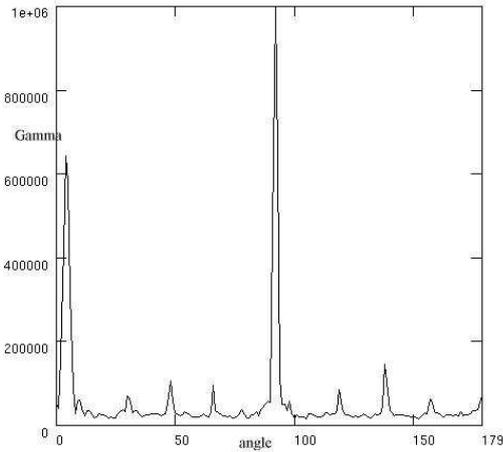


Fig. 3. The function $\Gamma(\phi)$

Δy 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 α and β . Therefore, if we project the OMT image along one of them, for example α , 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 α are organized as a sequence of rows which are Δy apart. Therefore, the projection of the OMT image along α direction will eventually produce a profile having a sequence of peaks which are Δy apart. Once the RT has been computed, it will be enough to extract the columns corresponding to the directions α and β . 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 Δx is the average distance between successive peaks in $\mathcal{R}\{OMT\{I\}\}(\cdot, \beta)$, whereas the estimate of Δy is the average distance between successive peaks in $\mathcal{R}\{OMT\{I\}\}(\cdot, \alpha)$.

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

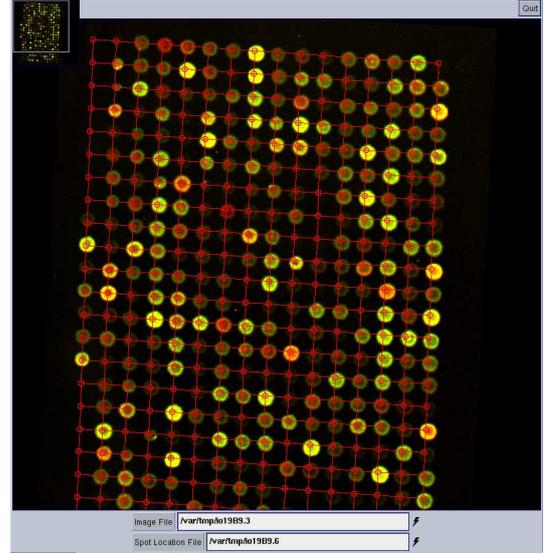


Fig. 4. The result of the *GridMatching* step; the output grid, in this case, has parameters $\langle x_0, y_0, \alpha, \beta, \Delta x, \Delta y \rangle = \langle 121.35, 30.87, 0.0698, 1.6057, 35.00, 35.00 \rangle$.

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 σ_1 in both directions, of course σ_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 α and β are random variables with $\alpha \sim N(0, \sigma_2)$ and $\beta \sim N(90^\circ, \sigma_2)$ and σ_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 and $\Delta x, \Delta y$, we do not consider particular alterations of these parameters for the quantitative analysis. For each combination of σ_1 and σ_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 σ_1 and σ_2 .

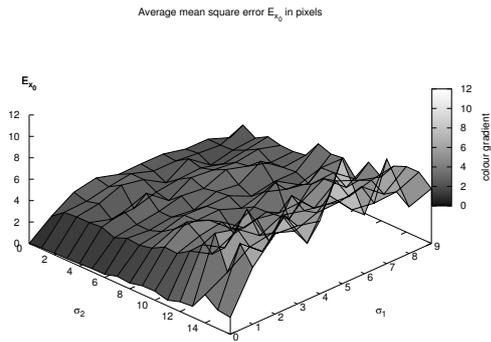


Fig. 5. The mean square error (in pixels) between the true X_0 grid parameter and x_0 computed by the GridMatching module.

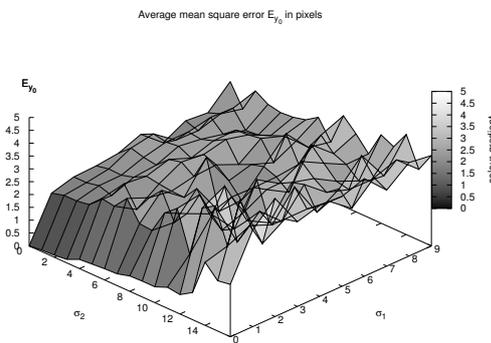


Fig. 6. The mean square error (in pixels) between the true Y_0 grid parameter and y_0 computed by the GridMatching module.

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 Δ_x, Δ_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, α and β 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 is 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 that the minimum spot is as the half of the maximum spot radius. It is worth noting that the accuracy of this step affects 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

- [1] Angulo, J., Serra, J., (2003), 'Automatic Analysis of DNA Microarray Images Using Mathematical Morphology', *Bioinformatics*, **19** (5), pp. 553-562.
- [2] Antoniol, G., Ceccarelli, M. (2004), 'A Markov Random Field Approach to Microarray Image Gridding', in *Proceedings of IEEE International Conference on Pattern Recognition 2004*, IEEE CS press.
- [3] Bajcsy, P. (2004), 'Gridline: Automatic Grid Alignment in DNA Microarray Scans', *IEEE Transactions on Image Processing*, **13** (1), pp. 15-25.
- [4] Ceccarelli, M., Petrosino, A. (2001), 'The Orientation Matching Approach to Circular Object Detection', in *Proceedings of IEEE International Conference on Image Processing 2001*, pp. 712-715, IEEE Press
- [5] Eisen, M., B., Brouwn, P. O. (1999), 'DNA arrays for analysis of gene expression', *Methods in Enzymology*, **303**, 1999.
- [6] Geman, G., Geman, D. (1984), 'Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images', *IEEE Transactions on PAMI*, **6**, pp. 721-742.
- [7] Jain, A., N., Tokuyasu, T., Snijders, A., Segreaves, R., Albertso, D., Pinkel, D. (2003), 'Fully Automatic Quantification of Microarray Image Data', *Genome Research*, **12**, pp. 325-332.
- [8] Hartelius, K., Carstensen, J., M. (2003), 'Bayesian Grid Matching', *IEEE Transactions on PAMI*, **25** (2), pp. 162-173.
- [9] Katzer, M., Kummert, F., Sagerer, G. (2003), 'A Markov Random Field Model of Microarray Gridding', *Proceed. of SAC 2003*, ACM
- [10] Liew A., W., Yan, H., Yang, M. (2003), 'Robust Adaptive Spot Segmentation of DNA Microarray Images', *Pattern Recognition*, **36**, pp. 1251-1254.
- [11] Yang, Y., H., M. Buckley, M., S. Dudoit, T. Speed (2002), 'Comparison of Methods for Image Analysis on cDNA Microarray Data', *Journal of Computational and Graphical Statistics*, **11**, pp. 108-136.