Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 53(67), Fascicola 2, 2008

Microarray Image Processing Using Harris Corner Detector Method

Raul Măluţan[1,3], Aurélien Bricier[2], Monica Borda[1], Pedro Gómez Vilda[3]

Abstract – The raw data from a DNA microarray experiment is a scanned image of a microarray slide that has been subjected to an experimental process. The analysis of data involves an image processing task and has substantial impact on the accuracy and effectiveness of the subsequent gene expression and identification analysis. To extract data from the microarray, a spot localization has to be done. The current work proposes a method for microarray image gridding using a classical image processing algorithm, the Harris corner detector, with slight modifications.
Keywords: microarray, gridding, corner detector

I. INTRODUCTION

Microarrays have now been recognized as a powerful and cost-effective tool for large scale analysis of gene expression. One issue of the technique is the structure of the resulting data. The raw data of an experiment is a scanned image of a microarray, which had been subjected to an experimental process before. This image has to be converted into quantitative data, which is basically an image processing task. Conversion has to be done carefully, since the quality of the resulting data has substantial impact on the accuracy and effectiveness of the subsequent gene expression and identification analysis. There are many recent publications that deal with microarray image processing [1], [2], [3], [4] so it seems like the problem is not solved yet and there is still a need of re-search in this field.

According to [5] the processing of microarray images can generally be separated into three tasks. The first step called gridding or addressing, assigns coordinates to each spot. In the second step, namely segmentation, pixels are classified either as foreground (representing the DNA spots) or as background (some recently developed methods also consider a third group called artifact). The third step calculates intensity values of each spot and estimates background values. In addition, the above mentioned steps can be preceded by different preprocessing methods and the last step can be extended to deliver some quality measures.

The current paper uses a traditional image processing method, known as Harris corner detector, with the aim of proposing a new approach in microarray image processing.

II. MICROARRAY IMAGE PROCESSING

The basic structure of a microarray image can be seen in Fig. 1. The image contains the array which in turn contains a number of spot groups. Each spot group in turn contains a number of spots. Independent of the desired result, spot centers or target masks, positions of the spots have to be estimated somehow. This can be done by gridding.

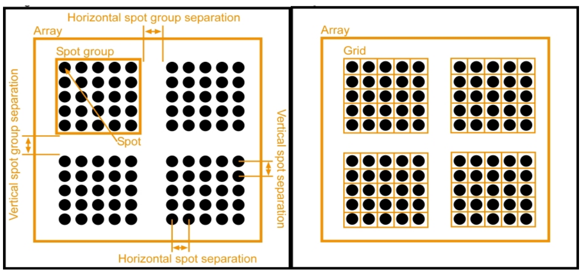

Fig. 1. Microarray image gridding (taken from [10])

First approaches to gridding were done by manual grid placement of the user [7]. Nowadays, it is commonly accepted that addressing algorithms should work without user interaction. This is not just because manual addressing is tedious and time-consuming but also because the addressing step can be repeated easily in the case of full automation. The same logic applies to the number of free parameters in an automatic addressing algorithm. It is desirable to have as few free parameters as possible to avoid time-consuming parameter testing.

After successful gridding, the approximate positions of the spots are known, simplifying the process of segmentation that allows the classification of pixels as foreground, *i.e.* belonging to a spot, or as background.

After having obtained a segmentation mask it follows the intensity extraction step that deals with the process

[1] Technical University of Cluj Napoca, Communication Department, 26-28 George Baritiu St., Room 364, 400027, Cluj-Napoca, Romania
e-mail: raul.malutan@com.utcluj.ro
[2] ENSEIRB, University of Bordeaux I, Telecommunications Department, 1, av. du Dr. Albert Schweitzer, F-33402 Talence Cedex, France
[3] Universidad Politécnica de Madrid, Departamento DATSI, Campus de Montegancedo s/n, 28660, Boadilla del Monte, Madrid, Spain

of conversion of image data to a few numbers that represent each spot.
Widely used methods for gridding are based on the computation of vertical and horizontal image profiles, localization of the spots being performed on these profiles using a sliding window to find local minimum of energy [8]. It is obvious that this method cannot be used on microarray images containing large missing spot areas or important high-energy artifacts inducing unusable profiles. That is why the consideration of a criterion less dependent of the pixel intensity gives the possibility to overcome this type of problem. To provide a good robustness, we propose in this work the use of Harris corner detector algorithm with a modified Harris measure.

III. HARRIS CORNER DETECTOR

The Harris corner detector algorithm [6] relies on a central principle: at a corner, the image intensity will change largely in multiple directions. This can alternatively be formulated by examining the changes of intensity due to shifts in a local window. Around a corner point, the image intensity will change greatly when the window is shifted in an arbitrary direction. Following this intuition and through a clever decomposition, the Harris detector uses the second moment matrix as the basis of it's corner decisions. The autocorrelation matrix *A* has values closely related to the derivates of image intensity:

$$A = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2(\mathrm{x}) & I_x I_y(\mathrm{x}) \\ I_x I_y(\mathrm{x}) & I_y^2(\mathrm{x}) \end{bmatrix} \tag{1}$$

where I_x and I_y are the respective derivates of pixel intensity in the x and y direction. The off-diagonal entries are the product of I_x and I_y, while the diagonal entries are the square of the respective derivates. The weighting function *w(x,y)* can be uniform, but is more typically an isotropic, circular Gaussian:

$$w(x,y) = \frac{1}{2\pi\sigma} e^{\left(-\frac{x^2+y^2}{2\sigma}\right)} \tag{2}$$

that acts to average in a local region while weighting those values near the center more heavily.
As it turns out, the matrix *A* describes the shape of the autocorrelation measure as due to shifts in window location. Thus, if we let λ_1 and λ_2 be the eigenvalues of *A*, then these values will provide a quantitative description of how the autocorrelation measure changes in space: it's principal curvatures. As Harris and Stephens point out [2] the *A* matrix centered on corner points will have two large positive eigenvalues. This is why they use the Harris measure based on the trace and determinant:

$$\begin{aligned} R &= \det(A) - \alpha \operatorname{trace}^2(A) \\ &= \lambda_1\lambda_2 - \alpha(\lambda_1+\lambda_2)^2 \end{aligned} \tag{3}$$

where α is a constant. Corner points have large, positive eigenvalues and would thus have a large Harris measure. Thus, corner points are identified as local maxima of the Harris measure that are above a specified threshold.
A corner, or in general an interest point, is characterized by a large variation of S in all direction of the vector. By analyzing the eigenvalues of A, this characterization can be expressed in the following way: A should have two "large" eigenvalues for an interest point.
Considering the Harris measure, a modified measure was proposed to test it on the current image processing. The relation for this new measure is:

$$R' = \frac{\lambda_1\lambda_2}{(\lambda_1+\lambda_2+\varepsilon)} \tag{4}$$

where ε is a constant. The result of the corner detection in a microarray image once applying (4) can be seen on Fig. 2.

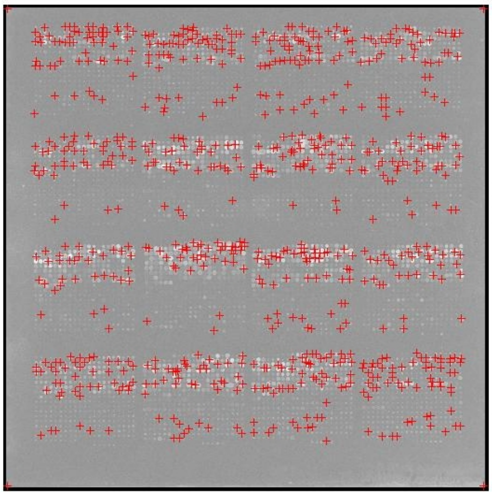

Fig. 2. Corner detection on a microarray image

III. MICROARRAY GRIDDING AND IDEXATION

Once the Harris corner detection was applied on the microarray image, the spot group localization, or global gridding, is done using a sliding window which identifies the areas that separates the spot groups as to be the region with no corners followed by a region with an important quantity of corner. Applying this on the horizontal and vertical dimension, the global gridding is performed, Fig. 3.
After the position of all spot groups from the image where determined, the spots from within each group must be localized. To speed up the process the localization of the spots within a spot group, known as the local gridding, was done on a single spot group which was cropped from the image.

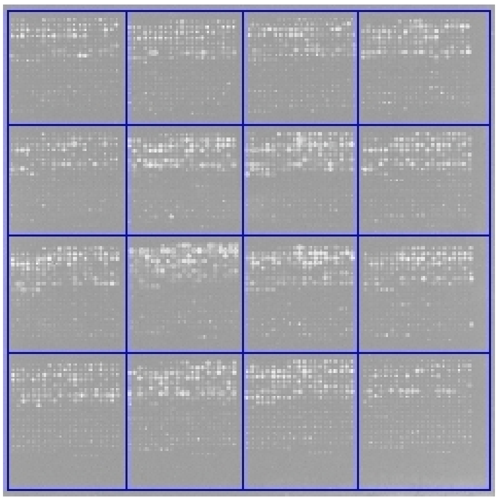

Fig. 3. Global gridding on microarray image

The local gridding current methods are based on energetic profiles, so it induces the same issues as global gridding. Once again, this process is based on a method less dependent to energy. Thereby, before searching for local minimum of energy, a color segmentation is applied to the group spot image using a model containing the mean of the two channels μ, and their covariance Σ. To produce the segmentation mask the Mahalanobis distance is computed, and a threshold is used to determine if a pixel belongs to a spot or not.

$$d_{maha}(x,y)=\sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)} \qquad (5)$$

Then a morphomathematic process is applied and the energetic profiles of the mask are calculated and the use of the Hill Climbing algorithm, described in [9], allows an estimation of the separation positions. Searching the original image on neighborhoods of previously determined separations, with the result of a rigid local gridding, Fig. 4, was done with the purpose of improving the accuracy of this method.

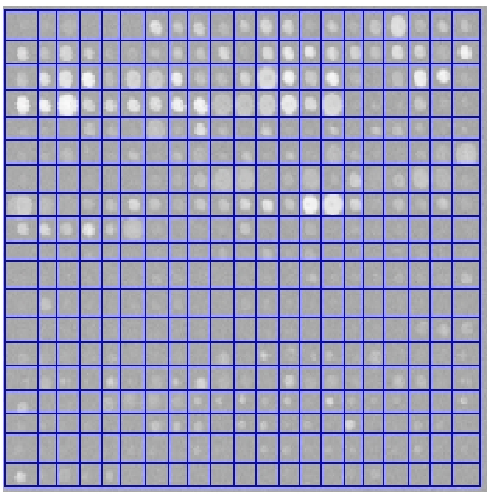

Fig. 4. Rigid local gridding

Next, to consider a spot-by-spot approach and to improve the accuracy of localization of the spots, for each spot is performed a positioning step. Starting from the previous localization, each spot is included in a rectangle, and on each side neighborhood of this rectangle a minimization of the energy is performed. This way a flexible local gridding is obtained, Fig. 5, with a localization of each spot independently of its neighborhoods.
This boxing allows a more accurate localization of the spot, is less sensitive to light deformation, maximize the energy per pixel on the spot area and decrease the surface to proceed in the sequel of the extraction process.
Having a box for each spot can be useful to perform an indexation step based on the computation of the box center, as its shown in Fig. 6. The coordinates of this point characterize the spot and are directly linked, in the index array, to the number of its spot group, its line and its column. This information corresponds to the information required in the MIAME database [11].

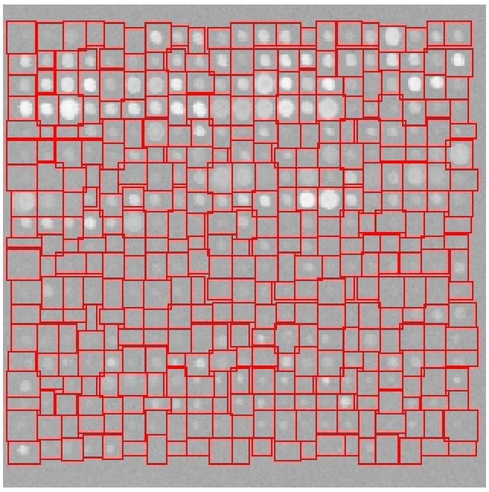

Fig. 5. Flexible local gridding

IV. CONCLUSIONS

The global and local gridding was tested on Apoa1 microarray image database [12]. By testing 19 images with the proposed method, only for two images were found misplaced spots, resulting an accuracy of 99,19%, while when applying the methods described in [10] in all cases were determined misplaced spots and the accuracy was only 95,58%.
The method applied in the current work aims at increasing the accuracy of the spot localization and the insensibility to variable image quality by working with parameter free algorithms.

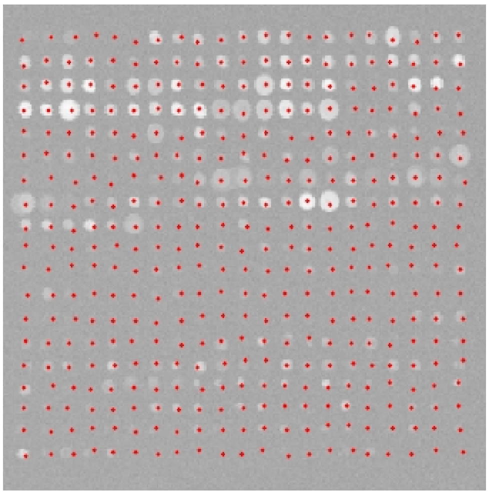

Fig. 6. Spot center localization

V. ACKNOWLEDGEMENTS

This research is being carried out under grant No. 527/2007 and No. 332/2007 from UEFISCSU.

REFERENCES

[1] P. Bajcsy, "Gridline: Automatic grid alignment in dna microarray scans", *IEEE Transactions on Image Processing*, vol. 13(1), 2004, pp. 15 – 25.
[2] R. Beare and M. Buckley, *Spot: cDNA Microarray Image Analysis Users Guide*, 2004.
[3] J. Angulo and J. Serra, "Automatic analysis of dna mi-croarray images using mathematical morphology", *Bioinformatics*, vol. 19(5), 2003, pp. 553 – 562.
[4] M. Katzer, F. Kummert, and G. Sagerer, "Methods for automatic microarray image segmentation", *IEEE Transactions on Nanobioscience*, 2003, vol. 2(4), 2003, pp. 203 – 204.
[5] Y. H. Yang, M. J. Buckley, S. Dudoit, and T. P. Speed, *Comparison of methods for image analysis on cDNA microarray data*, Technical Report 584, Department of statistics - University of California - Berkeley, 200.
[6] C. Harris, M.J. Stephens, "A combined corner and edge detector", *Alvey Vision Conference*, 1988, pp. 147 – 152.
[7] M. Eisen, *ScanAlyze User Manual*, 1999.
[8] K. Blekas, N. P. Galatsanos, A. Likas, I. E. Lagaris, "Mixture model analysis of DNA microarray images", *IEEE Transactions on Medical Imaging*, vol. 24(7), 2005, pp. 901 – 909.
[8] V. Vidyadharan, *Automatic gridding of microarray images*, MSc thesis, University of Windsor, 2004.
[10] B. Stetter, *Gene expression by automatic detection of hy-bridization spots in cDNA microarray images*, MSc thesis. Technical University of Munich, 2006
[11] Version 1.1 Draft 6 of the MIAME specification, http://www.mged.org/Workgroups/MIAME/miame_1.1.html
[12] Apoa1 data., http://www.stat.berkeley.edu/users/terry/zarray/Html/apodata.html. 2001.