A Precise and Automatic Gridding Approach to Noise-Affected and Distorted Microarray Images

Eleni Zacharia, Dimitris Maroulis, IEEE Member
Dept. of Informatics and telecommunications, University of Athens, Greece
eezacharia@gmail.com, dmaroulis@di.uoa.gr

Abstract

In this paper, a precise and fully-automatic approach to the determination of the grid alignment (Gridding) on microarray images is presented. The proposed approach is compared to state-of-the-art software programs and techniques. The conducted experiments demonstrate that it is very effective even when it is applied to noisy or distorted images as well as to images containing spots with various intensities.

1. Introduction

Over the last decades, cDNA microarrays have increasingly become one of the most indispensable tools in biological and medical research as they enable researchers to simultaneously monitor the expression levels of thousands of genes. The end product of a microarray experiment is a digital image which contains one or more distinct blocks, each one containing an equal number of spots.

The first crucial stage in microarray image processing is gridding, which is the process of segmenting a microarray image into numerous compartments, each containing one individual spot and background. Gridding however, remains, until nowadays, a challenging task. The main difficulty lies in the nature of microarray images [1]. More precisely, the quality of images is often degraded due to various sources of noise during image acquisition. Additionally, the locations of blocks or spots may slightly vary from image to image due to imperfections, such as rotations, and misalignments of the ideal rectangular grid. Finally, microarray images may contain low-intensity spots which are not clearly visible.

Due to the aforementioned nature of microarray images, existing software packages for gridding microarray images require different levels of human intervention in order to locate or correct the grid [2][3]. Moreover, the techniques that have been proposed to solve the rotation and misalignment problems [4][5] are not always adequately effective. As a result, they also require human intervention in order to correct the grid. This lack of automation in the gridding procedure is detrimental as human intervention affects the gene expression results.

In this paper, an original, precise and fully-automatic approach to gridding microarray images is presented. The proposed approach improves the one reported in [6] since it is noise-resistant and it can effectively cope with rotations, misalignments and local deformations of the ideal rectangular grid, while retaining the ability to be applied to images containing low-intensity spots. Experiments - over microarray images containing thousands of spots - showed that the proposed approach achieves an accuracy of more than 95% and it outperforms existing methods.

The rest of this paper is structured in three sections as follows: In Section 2, the proposed gridding method is presented. Experimental results are discussed in section 3 and concluding remarks are apposed in section 4.

2. Proposed approach

The proposed approach to gridding microarray images is divided into the following two main sub-stages: Firstly, the microarray image is segmented into blocks, and subsequently, each block of the microarray image is segmented into spots. In both sub-stages, a specific *Genetic Algorithm* (*GA*) - which optimally determines parallel and equidistant line-segments constituting the grid structure of the microarray image

- is used twice: Firstly, the *GA* determines optimally the "vertical" line-segments (i.e. L_{V1}, L_{V2}, L_{V3}, fig.1) and subsequently the *GA* determines optimally the "horizontal" line-segments (i.e. L_{H1}, L_{H2}, L_{H3}, fig.1).
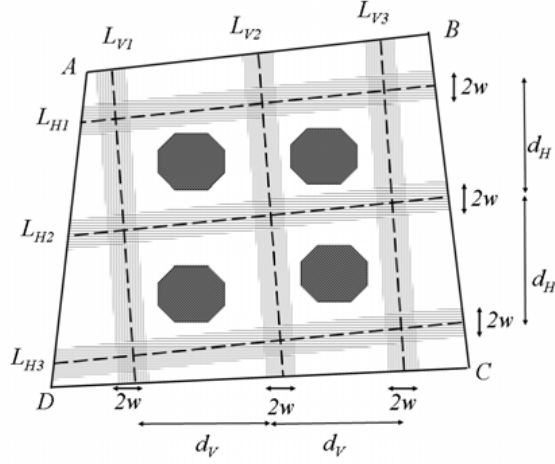

Fig. 1. Line-segments constituting the grid structure in a microarray image or block. In the former case, the polygons represent blocks while in the latter case, the polygons represent spots. The high-lighted areas on either sides of L_{V1}, …,L_{V3} and L_{H1}, …,L_{H3} denote the regions R_{V1},…,R_{V3} and R_{H1},…,R_{H3} correspondingly.

More precisely, the *GA* searches for the optimal line-segments as follows: Firstly, it creates an initial Population (*Pop₁*) of Chromosomes. Each Chromosome represents the "vertical" line-segments (i.e. L_{V1}, L_{V2}, L_{V3}, fig.1) or the "horizontal" line-segments (i.e. L_{H1}, L_{H2}, L_{H3}, fig.1). In the case when the Chromosome represents the "vertical" line-segments, it encodes the variables of the line-segment L_{V1} and the distance d_V (fig.1). In the case when Chromosome represents the "horizontal" line-segments it encodes the variables of the line-segment L_{H1} and the distance d_H (fig.1).

Subsequently, the Chromosomes constituting the *Pop₁* are evaluated using the Fitness Function. For this scope, we define the probability $P(L_i)$ of a line-segment L_i to be part of the grid by the following equation:

$$P(L_i) = f_B^{R_{Li}}(L_i) - f_S^{R_{Li}}(L_i) \qquad (1)$$

R_{Li} denotes the region of the image or block which contains those pixels whose distance from the line-segment L_i is less than a margin w (fig. 1). The real-valued function $f_B^{R_{Li}}(L_i)$ expresses the percentage of pixels of the region R_{Li} which belong to the background

while the real-valued function $f_S^{R_{Li}}(L_i)$ expresses the percentage of pixels of the region R_{Li} which belong to areas containing spots.

The Fitness Function $F(m)$ of a Chromosome m that encodes a possible solution to the particular optimization problem is defined by the following equation:

$$F(m) = \begin{cases} S_p(m) \cdot N(m), & \text{if } f_{LS}(m) \leq f_{Max} \\ S_p(m), & \text{otherwise} \end{cases} \qquad (2)$$

The real-valued function $S_p(m)$ denotes a total sum of the probabilities $P(L_i)$ of the line-segments L_i that are represented by the Chromosome m, and have a high probability $P(L_i)$ to be part of the grid. The real-valued function $f_{LS}(m)$ denotes the percentage of the line-segments L_i that are represented by the Chromosome m, and have a low probability $P(L_i)$ to be part of the grid. A high probability $P(L_i)$ is the one which is higher than a threshold P_{MAX} while a low one is the one which is lower than a threshold P_{LOW}.

Thereafter, the *GA* makes the Population *Pop₁* evolve into a new Population *Pop₂* using Genetic Operators. This Evolutionary Cycle from one Population to the next (*Pop₁* to *Pop₂*, *Pop₂* to *Pop₃* and so forth) continues until a maximum number of Populations is reached, for which the best Fitness Value has remained unchanged.

It should be noted that the line-segments - having the same direction and constituting the borders of blocks (or spots) - are ideally equidistant. However, this observation may not come true when rotations, misalignments and local deformations of the ideal rectangular grid exist. As a result, the determined line-segments may slightly vary from the optimal ones.

In order to tackle this problem, each line-segment L_i determined by the *GA* is replaced with a new one, L_i', on the condition that the following are valid:

$$L_i' \in R_{Li} \text{ and } P(L_i') - P(L_i) > T_p \qquad (3)$$

where T_p is a positive threshold.

3. Experimental Results

The proposed approach has been put to test using two sets of microarray images [7][8] containing thousands of spots. The Box-Cox transformation was applied as a pre-processing step, prior to gridding, in

order to adjust microarray spot intensities [9]. It should be noted that the gridding outcome of the proposed method is superimposed onto the original image directly, in order for the Box Cox transformation not to distort the subsequent data processing of the microarray images.

The accuracy of the proposed method was analyzed by means of a statistical analysis. In detail, applying the proposed approach in both datasets, 95.1% of spots were perfectly placed inside a compartment, 4.3% were very nearly gridded while only 0.6% of spots were gridded incorrectly.

Additionally, the aforementioned results of the proposed approach were compared to the ones reported by Blekas et. al [7]. The proposed method outperforms the method proposed by Blekas et. al as well as both the ScanAlyze and SpotFinder software programs.

Two gridding results of microarray sub-images are presented in figures 2, and 3. In the first figure, it is clear that the proposed method has efficiently located the grid structure even though the block contains low intensity spots and it is contaminated with noise. In the second figure, the proposed method has efficiently located the grid structure in a distorted sub-image.
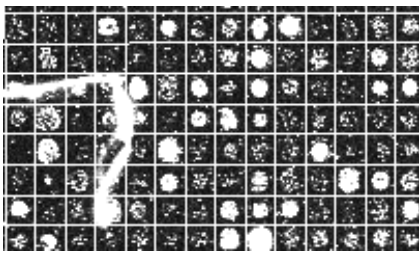

Fig. 2. A gridding result in an area of a microarray image containing low intensity spots and noise.

4. Conclusions

In this paper, an original method for the determination of the grid structure in a microarray image has been presented. The proposed method is fully-automatic. The experimental results demonstrate that the proposed method is efficient and it outperforms existing software programs and other state-of-the-art gridding techniques. It is also noise-resistant and it can be applied to distorted microarray images as well as to images containing spots of various intensities.
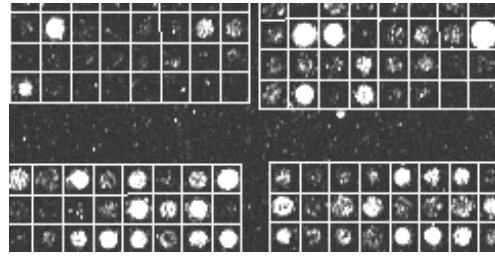

Fig. 3. A gridding result in an area of a distorted microarray image.

5. References

[1] M. Ceccarelli and G. Antoniol, "A Deformable Grid-Matching Approach for Microarray Images", IEEE Trans on Image Processing, vol. 15, no.10, Oct 2006.

[2] J. Buhler, T. Ideker, and D. Haynor, "Dapple: improved techniques for finding spots on DNA microarrays," UW CSE Technical Report UWTR 2000-08-05, pp. 1-12, 2000.

[3] N. Deng and H. Duan, "The Automatic Gridding Algorithm based on projection for Microarray Image", Proc. of the Int. Conf. on Intelligent Mechatronics and Automation, Chendu, China, August 2004, pp.254-257.

[4] P. Bajcsy, "Gridline: automatic grid alignment in DNA microarray scans," IEEE Trans. Image Processing, vol. 13, no. 1, pp. 15–25, Jan. 2004.

[5] J. Ho, W. L. Hwang, H. H. S. Lu, D. T. Lee, "Gridding Spot Centers of smoothly distorted microarray images," IEEE Trans. Image Processing, vol. 15, no. 2, pp. 342-353, Feb. 2006.

[6] E.Zacharia, D. Maroulis, "An Unsupervised and Fully-Automated Image Analysis Method for cDNA Microarrays," Proc. of the Int. Conf. on Computer-based Medical Systems, Maribor, Slovenia, June 2007, pp.389-394.

[7] K. Blekas, N. P. Galatsanos, A. Likas, and I. E. Lagaris, "Mixture Model Analysis of DNA Microarrray Images," IEEE Trans. Medical Imaging, vol. 24, no. 7, pp. 901-909, July 2005.

[8] D. Juric et al., "Differential Gene Expression Patterns and Interaction Networks in BCR-ABL-Positive and -Negative Adult Acute Lymphoblastic Leukemias," Journal of Clinical Oncology, vol. 25, no. 11, pp. 1341-1349, April 2007.

[9] C. T. Ekstrom, S. Bak, C. Kristensen and M. Rudemo, "Spot shape modelling and data transformations for microarrays", Bioinformatics, vol. 20, no. 14, pp. 2270-2278, Sep. 2004.