A Computational Intelligence Approach to Unsupervised Microarray Image Gridding

G. Antoniol, M. Ceccarelli
Research Centre on Software Technologies - RCOST,
University of Sannio
Via Traiano, Benevento 82100 Italy
{antoniol,ceccarelli}@ieee.org

Keywords.Microarray Image, Markov Random Fields, non-linear optimization

Introduction

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 [4]: (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*. Most of available gridding approaches require human intervention, for example to specify some points in the spot grid or even to register individual spots. Automating this part of the process will allow high throughput analysis [1].
The paper reports a novel approach for the problem of automatic gridding in Microarray images. The method uses a two step process. First a regular rectangular grid is superimposed on the image by interpolating a set of guide spots, this is done by solving a non-linear optimization process with an evolutionary approach. Second, the interpolating grid is adapted, with Markov Chain Monte Carlo method, to local deformations. This is done by modeling the solution as a Markov Random Field with a Gibbs prior possibly containing first order cliques (1-clique). The algorithm is completely automatic and no human intervention is required, it efficiently accounts grid rotations and irregularities.

1.The grid model

The Bayesian framework is useful to deal with complex pattern recognition problems where prior knowledge and uncertainty about observed data [3], here we model the image gridding problem as a MAximum a Posteriori estimate (MAP). Let I the observed image, and $\mathcal{G} = \{\mathbf{g}_i : i = 1,...,n\}$ a grid represented as a set of image points coordinates, then we try to solve the following MAP problem:

$$\mathcal{G}^* = \arg\max_{\mathcal{G}} P(I \mid \mathcal{G}) \cdot P(\mathcal{G}) \tag{1}$$

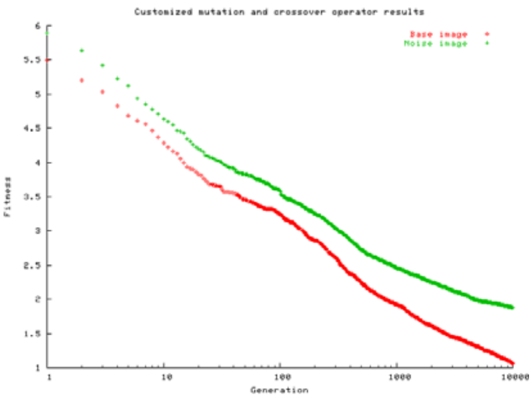
Where we model the prior term $P(\mathcal{G})$ as a Gaussiam Markov Random Field with mean a reference grid $\mathcal{T}$:

$$P(\mathcal{G}) \propto \exp\left[-\frac{1}{2} \sum_i (\mathbf{g}_i - \mathbf{t}_i)^T \Sigma_i^{-1} (\mathbf{g}_i - \mathbf{t}_i) \right] \tag{2}$$

the reference grid $\mathcal{T} = \{\mathbf{t}_1,...,\mathbf{t}_n\}$ is computed by interpolating a set of guide spots detected with a generalization of the Hough Transform (*OMT*) for circles [2]. Whereas the likelihood term $P(I \mid \mathcal{G})$ is modeled as $P(I \mid \mathcal{G}) \propto \exp[-\frac{1}{2}\sum_i (1 - OMT(\mathbf{g}_i))^2]$, where $OMT(\boldsymbol{x})$ represents the value of the OMT Transform at the point $\boldsymbol{x}$.

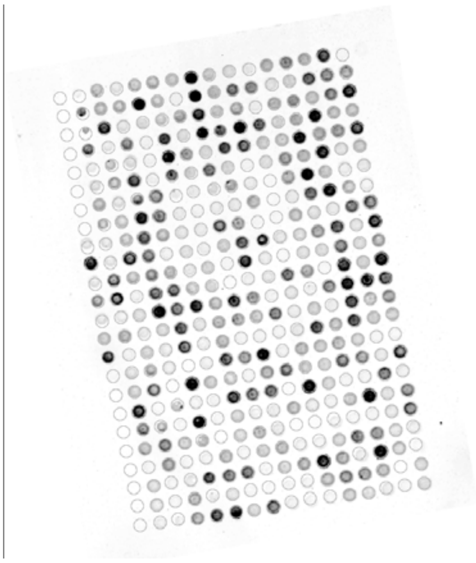
The main computational task of the algorithm consists into the search of the reference grid $\mathcal{T}$. Here we follow an heuristic search method based on a customized evolutionary algorithm. Given a set of microarray guide spots, automatically detected as the local maxima of the OMT, with center coordinates $\boldsymbol{c}_1,..., \boldsymbol{c}_k$ for any given grid, defined by a tuple <$\boldsymbol{x}_0$, $\boldsymbol{y}_0\alpha,\beta,\Delta x,\Delta y$> (respectively the coordinate of the upper left spot, the angle with the x and y axes of the grid directions and the grid spacing), we adopt as cost figure the maximum Euclidean distance (minimum over $\boldsymbol{c}_r$) from the nearest $\boldsymbol{t}_k$:

$$\min_{<x_0, y_0, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Delta x, \Delta y>} \{\max_i \min_j \| \mathbf{t}_j - \mathbf{c}_i \|\}$$



Being such criterion non linear and non differentiable, we applied metaheuristic search. Metaheuristic algorithms, such as genetic algorithms and simulated annealing, have been successfully applied to a number of engineering problems ranging from load balancing in the process industries, through electromagnetic system design, to aircraft control and aerodynamics.

We applied an algorithm inspired by stochastic hill climbing and genetic algorithm with elitism. The tuple $<\boldsymbol{x}_0, \boldsymbol{y}_0 \alpha, \beta, \Delta x, \Delta y>$ was represented as a vector of six real number; by means of an heuristic algorithm suitable ranges for the tuple parameters are defined for the initial population. Preliminary experiments revealed that parameters $\boldsymbol{x}_0$, $\boldsymbol{y}_0$ exhibited co-adaptation thus the crossover operator was reimplemented to avoid disrupting co-adapted solutions. Mutation operator was implemented via a normal distribution with zero mean and unitary variance. The algorithm has been tested over a range of computer generated images and on real microarray images. Preliminary algorithm accuracy assessment was carried out on generated images with and without superimposed noise. Different crossover, mutation probabilities as well as different selection schemas were compared. The figure shows the results obtained by 100 simulations with mutation probability 0.01, cross over probability 0.6 adopted selection criterion was the stochastic remainder sampling. The size of the initial population was fixed in 200 individuals and the number of generation 10000. As expected, accuracy on noise image was consistently lower with respect to non noise image. However, it is worth noting that when the number of generation increases the absolute maximum error drops below 2 pixels. The last figure reports the resulting grid obtained by minimizing (1) and using the interpolating grid generated by our method. It efficiently accounts for local deformations of the grid. At the time of writing we are quantitatively evaluating the accuracy of the method over a set of reference data.



References

[1] P. Bajcsy,’ Gridline: Automatic Grid Alignment in DNA Microarray Scans’, *IEEE Trans. Image Processing*, vol. 13, n. 1, pp. 15-25.

[2] M. Ceccarelli, A. Petrosino (2001), 'The Orientation Matching Approach to Circular Object Detection', in *Proceedings of IEEE International Conference on Image Processing* 2001, pp. 712-715, IEEE Press

[3] K. Hartelius, J. M. Carstensen, 'Bayesian Grid Matching’, *IEEE Trans. pn Pattern Analysis and Machine Intelligence*, vol. 25, n. 2, pp. 162-173.

[4] Y. H. Yang, M. Buckley, S. Dudoit, T. Speed (2002), 'Comparison of Methods for Image Analysis on cDNA Microarray Data', *Journal of Computational and Graphical Statistics*, vol. 11, pp. 108-136.