Available online at www.sciencedirect.com

Image and Vision Computing 25 (2007) 155–163

www.elsevier.com/locate/imavis

Microarray image gridding with stochastic search based approaches

Giuliano Antoniol, Michele Ceccarelli *

Research Center on Software Technologies-RCOST, University of Sannio, Via Traiano, 82100 Benevento, Italy

Received 22 April 2004; received in revised form 22 July 2005; accepted 31 January 2006

Abstract

The paper reports a novel approach for the problem of automatic gridding in Microarray images. Such problem often requires human intervention; therefore, the development of automated procedures is a fundamental issue for large-scale functional genomic experiments involving many microarray images. Our 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 a stochastic search producing the best interpolating grid parameterized by a six values vector. Second, the interpolating grid is adapted, with a Markov Chain Monte Carlo method, to local deformations. This is done by modeling the solution 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 arbitrary grid rotations, irregularities and various spot sizes.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Markov random fields; Gridding; Genetic algorithm; Microarray

1. Introduction

DNA microarrays [5] allow the simultaneous monitoring of the expression levels for thousands of genes. This 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 GMOs, 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 measure gene expression level is represented in terms of ratio between samples stained with different dyes.

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 [17]: (i) *gridding*, which is the process of assigning the coordinates to the spots; (ii) *segmentation*, it allows the separation between foreground and background pixels; (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 individually adjust 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 gridding. Automated segmentation is another problem, which has also been studied by several authors [3,12]. Indeed, the image processing stage in microarray experiments is often the source of strong variations in the whole experiments and affects all subsequent analysis steps [1], for example the experiment reported in [12] shows that image analysis by manual gridding performed by different researchers leads to large discrepancies in the gene expression level values from the same array.

The problem of automatic gridding is complicated by the fact that microarray images are usually highly contaminated with noise and artifacts of the wet lab processes. Often rotations, misalignment and local deformations of the ideally rectangular grid can occur. There is a high need of methods for microarray gridding, which are robust and flexible at the same time.

---

* Corresponding author. Tel.: +39 08 24305131.

E-mail addresses: antoniol@ieee.org (G. Antoniol), ceccarelli@unisannio.it (M. Ceccarelli).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.01.023

Some efforts to help automatic microarray data processing have recently emerged in literature [2,8,9,15,17]. However, most of them impose different kind of restrictions and are based on strong assumptions. For example, the approaches in [2,10] require that grid rows and columns are strictly aligned with the x and y axes. Other approaches such as [8,9] 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 problem. Here, we adopt a prior, possibly, containing just 1-cliques, this can simplify, from a computational point of view, the search of the maximum a posteriori solution.

The adopted prior requires the previous computation of a reference *regular* grid, which is generated by interpolating a set of guide spots located by using the Orientation Matching Transform (OMT) [4]. Interpolating a set of guide spots has also been exploited in [14], by imposing several restriction over the grid, and without any further adaptation. Here, we perform the grid interpolation by solving a highly non-linear optimization problem via stochastic search approaches. This allows to efficiently deal with any grid rotations and shifting. The final result is obtained via a Maximum A Posteriori (MAP) grid obtained by a Markov Chain Monte Carlo approach. The adopted probabilistic model allows to deal with local deformations of the interpolating grid.

Our method improves the approach reported in [14] by allowing arbitrary grid alignments with respect to the image axes and by refining the interpolating grid for dealing with local deformations. It also simplifies the general model in [8] by adopting a simple prior based on 1-cliques.

The paper is organized as follows, Section 2 reports the prior model with the grid interpolation step, whereas in Section 3, we present out the Bayesian approach. A set of experiments on computer generated and real images is reported in Section 5.

2. The prior

Here, we approach the problem of microarray image addressing within a Bayesian framework [6]. Indeed, Bayesian methods are pervasive in all areas of Computer Vision, they are useful to encompass into a unique model the data consistency constraint, the model of observation and all our a priori assumption and knowledge about the solution of our problem. The general Bayesian framework codifies how to proceed with data analysis and inference in presence of uncertainty [11]. Given the data D and the model $\mathcal{M}$, the Bayes principle suggest to take decisions on the basis of the posterior probability $p(\mathcal{M}|D)$, which in turn is proportional to the product between the likelihood term, representing the observation model, $P(D|\mathcal{M})$ and the *prior* term, representing our knowledge, $P(D)$. For the case of microarray image gridding, we have an observed datum, the input image, I, which is the raw visual representation of an ideal grid, $\mathcal{G}$ consisting of a sequence of spot locations with a well-defined organization. The Maximum A Posteriori (MAP) grid estimate consists into searching the most likely grid given the observed image.

This paper assumes a notation similar to the one adopted in [8]; let us define the list of node locations $\mathcal{G} = \{\mathbf{g}_i, i = 1,\ldots,n\}$, where $\mathbf{g}_i$ is the vector of image coordinates of the ith node. Let $\mathcal{T} = \{\mathbf{t}_i, i = 1,\ldots,n\}$ be the reference grid computed as reported in Section 2.1, then the joint distribution of $\mathcal{G}$ is modeled by a Gaussian Markov random field:

$$P(\mathcal{G}) \propto \exp\left[-\frac{1}{2}\sum_i (g_i - t_i)^{\mathrm{T}} \sum_i^{-1} (g_i - t_i)\right] \tag{1}$$

Therefore, the grid location of the ith point is assumed to be Gaussian distributed with mean $\mathbf{t}_i$, the computation of a rectangular reference grid with nodes $\mathbf{t}_i$ is the main part of our approach. This is performed in a two-step procedure. In the first step, guide spots are located via OMT. The second phase defines a grid interpolating the identified spots via metaheuristic search (i.e. stochastic hill climbing and genetic algorithms). The technique of using a set of guide spot has been also exploited in [14].

Guide spots are located by using the orientation matching transform for circular object detection. OMT [4] 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, $\mathrm{OM}(u, v)$ represents the evidence that the point (u, v) be the center of a circular object. The local maxima of $\mathrm{OM}(u, v)$, which are above a given threshold are considered as centers of the guide spots. In our experiment, we assumed a threshold of 0.8.

2.1. The reference grid

The detected guide spots computed in the previous step are used to generate the rectangular grid, $\mathcal{T}$, which best interpolates their centers. The problem to determine the optimal rectangular grid $\mathcal{T}$ was modeled as a coordinate transformation from a given coordinate system (X, Y) to a new one (X', Y'). Guide spots coordinates are expressed into the (X, Y) axes, and produced with respect to the hardware and software acquiring the microarray image. The new coordinate system (X', Y') attempts to minimize a large fraction of distortions introduced by different microarray fabrication imprecision and errors. Ideally, in the new coordinate system spots are exactly located in position $i\Delta x, j\Delta y$; $i = 0, 1,\ldots, N$; $j = 0, 1,\ldots, M$; this also means that any cost function should take into consideration the displacement of the microarray spots from the ideal grid (i.e. the points $i\Delta x, j\Delta y$; $i = 0, 1,\ldots, N$; $j = 0, 1,\ldots, M$).

As shown in Fig. 1, in this paper, we assume that (X', Y') distortion was mainly due to a roto-translation of the X, Y axes. The new origin is located in x_0, y_0, and axes underwent independent rotations. In other words, the new coordinate system is identified by a tuple of six parameters: x_0, y_0 (the coordinates of the lower left point of the grid), α and β (i.e. the angles of the grid axes with the x direction), Δx and Δy (i.e. the grid spacing).

Download English Version:

https://daneshyari.com/en/article/527703

Download Persian Version:

https://daneshyari.com/article/527703

Daneshyari.com