

Automatic Detection of Molecular markers in Digital Images

A.R.S. Marçal, C.M.R. Caridade, P. Albuquerque, M.V. Mendes, F. Tavares

Abstract—Labeled molecular markers are an important tool in molecular biology. This work presents a method for the automatic identification of molecular markers in dot blot images. The method detects the location of markers in the image and their size. An experiment was made with 6 test images, which were used to produce an additional set of 222 images with various rotation, translation, contrast and noise levels. Over 7500 markers were identified automatically and compared with reference values obtained manually. The RMS error for the marker positioning in the original test images were between 1.1 and 3.8 pixels, which is about 1/10 of the typical radius (26 pixels). The method proposed was found to be almost insensitive to grid rotation and translation, and reasonably robust to image contrast changes and presence of noise.

I. INTRODUCTION

Labeled molecular markers are commonly used in molecular biology as probes to identify, by hybridization, the corresponding DNA template from a large number of template molecules. Therefore, a probe can only confirm the presence or absence of a specific locus. This technique is an important tool in molecular biology, particularly to evaluate simultaneously the expression of numerous genes, and for the identification of microorganisms for which DNA-signatures or taxa-specific loci have been validated as probes [1] [2].

Macroarrays designed for detection of different bacteria have already been proposed for a large variety of targets including bacteria pathogenic on potato [3], phytopathogenic *Pseudomonas* [1], *Lactobacillus* species [4], *Pythium* species [5] and *Aeromonas* spp. [6], among others. These assays are reported as being specific for the target bacteria with good inter-experiment consistency of results. When properly validated, macroarrays provide, at this time, a much better cost-benefit ratio for use in routine analysis than the much more expensive and elaborate microarray platforms.

In this work we used a macroarray prototype developed for detection of plant pathogenic bacteria [7]. Each molecular marker was blotted into a nylon membrane as a dot, to form a dot blot. A dot blot has a pre-defined grid of evenly spaced dots, for example 48 dots arranged in 8 rows and 6 columns,

A.R.S. Marçal is with Faculdade de Ciências, Univ. Porto, DMA, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal andre.marcal@fc.up.pt

C.M.R. Caridade is with Instituto Superior de Engenharia de Coimbra, R. Pedro Nunes, Qt. Nora, Coimbra, Portugal and Faculdade de Ciências, Univ. Porto, DMA, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal caridade@isec.pt

P. Albuquerque and F. Tavares are with IBMC – Instituto de Biologia Molecular e Celular, Univ. Porto and Faculdade de Ciências, Departamento de Botânica, Univ. Porto psilva@ibmc.up.pt
FTavares@ibmc.up.pt

M.V. Mendes is with IBMC – Instituto de Biologia Molecular e Celular, Univ. Porto mvm@ibmc.up.pt

as the images presented in figure 1.

The dot blot images were acquired with a GS-800 densito-

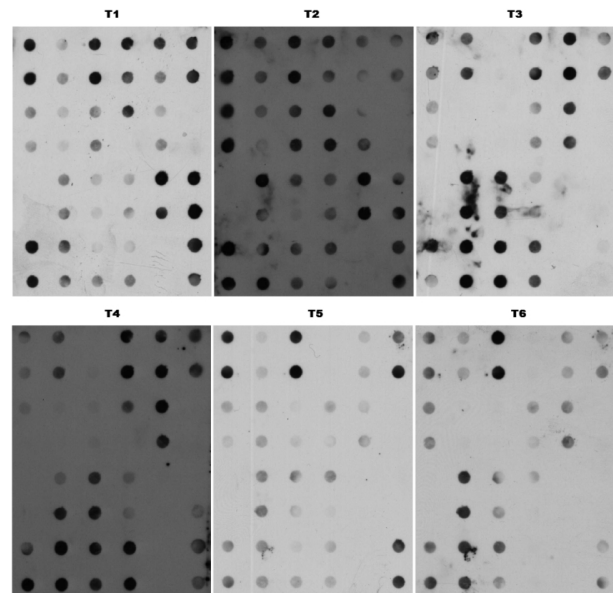


Fig. 1. Dot blot images with a grid of 8x6 Molecular Markers.

meter (Bio-rad, Hercules, CA), producing grayscale images with 1100 by 820 pixels (902000). An ideal positive dot will be a dark mark in a light gray background, whereas a negative dot will ideally be undistinguishable from the background. The gray level intensity of the background and the amount of noise varies considerably from image to image.

The test images used (*T1* to *T6*) are presented in figure 1. All have eight positive controls at the corners, in locations (1,1), (2,1), (1,6), (2,6), (7,1), (8,1), (7,6) and (8,6), and six negative controls located at (5,1), (6,1), (3,6), (4,6), (7,5) and (8,5). Although the positive controls are expected to result in a clear mark signature, this is not always the case (e.g. the bottom right corner marks of image *T3*). The gray intensity of the background is generally uniform within an image, but noise is present in some areas, well noticed in image *T3*. This type of noise can disturb the identification of nearby marks.

The analysis of dot blot images is currently based on the human identification of positive / negative dots, and interpretation / classification of the results. The process is time consuming, subjective and consequently erratic, thus the interest of developing a fully automatic system based on image processing and data classification techniques. The purpose of this work is to present a method for the automatic identification of molecular markers in dot blot images. The

robustness of the method to rotation, translation and noise is evaluated using the set of test images.

II. METHODOLOGY

The algorithm developed receives as input a digital image and assumes prior knowledge of the grid size (number of dots per line and column). The process can be divided in three stages: initial estimate of the grid orientation, image to grid matching, final identification of markers.

A. Initial estimate of the grid orientation

Initially, the identification of very dark dots in the grayscale image G is performed. A binary image (B) is obtained by thresholding G , using a threshold value obtained by the Otsu method [8]. As an illustration, figure 2 shows the result of this process applied to image $T1$ (figure 1 top left), where the threshold value obtained by the Otsu method was 0.49 (in the range 0 to 1). The next step is noise reduction, using the morphological operation opening (erosion followed by dilation) with a circular structuring element of 5 pixel radius [9]. Once the very small objects are removed from the binary image, morphological reconstruction, using the implementation on MATLAB [10], is used to fill holes in the remaining objects. The center of mass of each object is computed, and those objects whose radius is at least 15 pixels are selected as a marker. For the example of figure 2, 25 markers were obtained.

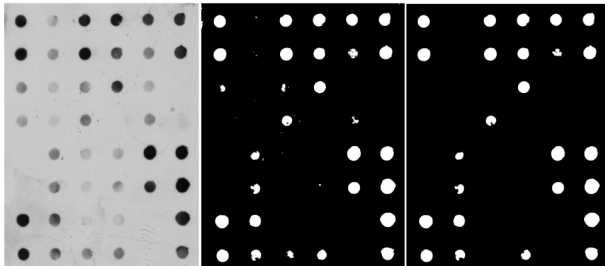


Fig. 2. Image $T1$ (left) with binary images obtained by thresholding (center) followed by noise removal (right).

The set of markers identified is used to establish the orientation of the grid. This is done by searching for the directions between all marker pairs. The assumption is that the two main directions of the grid are orthogonal. The angles between all pairs of markers are computed, reducing the results to the domain $[0, 90^\circ]$. For n markers, $n(n-1)/2$ directions are computed. Figure 3 illustrates the process of computing the direction between marker $M1$ and markers $M2, M3, M4$ and $M5$. In this case the results for $M1-M2$ and $M1-M3$ are the same, as the two directions are orthogonal. With all directions between marker pairs computed, the orientation of the grid (α) is obtained as the most frequent angle, using a bin search of 2 degrees. The original image is then rotated by α , resulting in a new version of the image with the grid nearly aligned with the image.

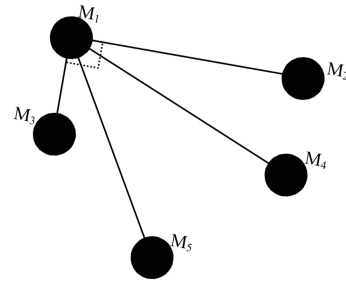


Fig. 3. Main direction search for the initial estimate of the grid orientation.

B. Image to grid matching

Once the initial correction of the image orientation is done, it is necessary to identify the correspondence between the markers detected by thresholding and the grid. The image is divided in sections, considering the size of the image and the number of grid elements (in columns and rows). Each marker is thus assigned a position in the grid. An ideal grid is then mapped into the image, considering the possibility of a rotation (α), a scale parameter (λ) and a translation (X_0 and Y_0). Figure 4 illustrates this process.

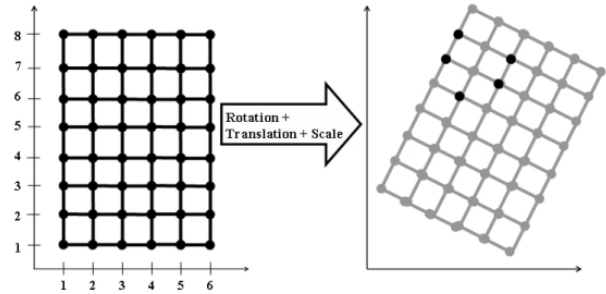


Fig. 4. Image to grid matching.

The position (X_i, Y_i) of a marker (i) in the digital image is obtained from the ideal grid co-ordinates (X'_i, Y'_i) using equation (1). The system of equations (1) is solved by the least square errors method, equations (2, 3, 4), where the 4 unknowns (matrix W) account for the translation, rotation and scale [11]. The observables (L) and coefficients (A) matrices have 2 lines for each tie point, obtained for each marker paired between the ideal grid and the binary image.

$$\begin{aligned} X_i &= X_0 - \lambda \sin \alpha Y'_i + \lambda \cos \alpha X'_i \\ Y_i &= Y_0 + \lambda \sin \alpha X'_i + \lambda \cos \alpha Y'_i \end{aligned} \quad (1)$$

$$L = AW \Leftrightarrow W = (A^T A)^{-1} A^T L \quad (2)$$

$$L = \begin{bmatrix} X_i \\ Y_i \\ \dots \end{bmatrix}, A = \begin{bmatrix} 1 & 0 & -Y'_i & X'_i \\ 0 & 1 & X'_i & Y'_i \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (3)$$

$$W = \begin{bmatrix} X_0 \\ Y_0 \\ \lambda \sin \alpha \\ \lambda \cos \alpha \end{bmatrix} \quad (4)$$

C. Final identification of markers

Once the image to grid mapping is established, a local thresholding is performed to detect the possible presence of a marker in each section of the image. The threshold value is again obtained by the Otsu method [8], and the noise removal procedure is the same as described in section 2.1. For the example with image *T1* (figure 2), only 25 markers were detected in the first step, enough to estimate the grid orientation, while the local approach detected all 41 markers visible in the image.

III. EXPERIMENTAL SETUP

In order to evaluate the method developed for the automatic identification of markers, an experiment was prepared using the 6 test images presented in figure 1. These images were all obtained with a grid of 8 by 6, but with different characteristics in terms of contrast and noise.

For each test image, the visible markers were identified manually, in terms of center location (x, y) and radius (r) . The positioning error (d) of a marker is computed as the Euclidean distance between the coordinates of its center obtained by manual and automatic processes. The average positioning error for an image d_{RMS} is computed as the Root Mean Square (RMS) of all marker errors.

Each test image was subjected to transformations (rotation, translation, contrast and noise) in order to evaluate the ability of the proposed method to identify the markers under different conditions. A total of 8 new images were obtained for each test image, by applying a rotation of an angle (θ) between -40° and 40° , with increment of 10° . The original test images were also subjected to 9 translations, of 50, 100, 200 in X, in Y, and in both X and Y. The contrast of the original test images was changed using a gamma function [9] with the following values for γ : 0.1, 0.2, 0.4, 0.6, 0.8, 2, 4, 6, 8. Two types of noise were also added to the test images: Salt & Pepper (f - fraction of contaminated pixels of 1.0%, 1.5%, 2.0%, 2.5%, 3.0%) and zero mean Gaussian (with σ values of 0.01, 0.02, 0.04, 0.06 and 0.10). A total of 222 images were tested (1+8+9+9+10 for each test image), with over 7500 markers identified and evaluated all together.

IV. RESULTS

A summary of the results for the 6 original test images is presented in table I. The table shows the number of markers identified in each image, and the number of markers with positioning error (d) below 1 pixel, between 1 and 2 pixels, and higher than 2 pixels. The maximum value of d and the RMS error for the whole image are also included. In the bottom part of table I two estimates of the marker radius are presented: the average radius of all marks identified manually

TABLE I
SUMMARY OF RESULTS FOR THE 6 ORIGINAL TEST IMAGES (SEE TEXT FOR DETAILS).

	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
#markers	41	40	30	31	35	29
with $d < 1$	25	3	3	4	6	5
with $1 \leq d \leq 2$	13	9	12	7	5	12
with $d > 2$	3	28	15	20	24	12
d_{max}	2.4	7.8	4.3	6.6	4.2	5.7
d_{RMS}	1.1	3.8	2.3	2.8	2.6	2.4
R_{man}	26.2	26.6	26.2	26.4	26.1	26.0
R_{auto}	26.3	26.1	26.7	26.3	25.2	25.4

TABLE II
AVERAGE RMS ERRORS FOR THE ROTATION EXPERIMENT.

Rotation	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
$\theta = -40^\circ$	1.1	3.5	2.4	2.3	2.5	2.5
$\theta = -30^\circ$	1.1	3.5	2.4	2.3	2.5	2.4
$\theta = -20^\circ$	1.1	3.6	2.4	2.3	2.5	2.4
$\theta = -10^\circ$	1.1	3.6	2.4	2.5	2.6	2.4
$\theta = 0^\circ$	1.1	3.8	2.3	2.8	2.6	2.4
$\theta = +10^\circ$	1.1	3.6	2.3	2.6	2.6	2.4
$\theta = +20^\circ$	1.1	3.6	2.4	2.3	2.5	2.4
$\theta = +30^\circ$	1.1	3.5	2.4	2.4	2.5	2.5
$\theta = +40^\circ$	1.1	3.5	2.4	2.3	2.5	2.5

(R_{man}), and the average radius computed by the automatic process (R_{auto}).

The RMS errors for the image rotation experiment are presented in table II. There are almost no changes in RMS error with the grid orientation, which shows that the method is robust to rotation. The test for translation also showed that the method is robust. In 3 images the RMS errors were the same for all 9 translations tested and for the other 3 there were only small variations (less than 0.03).

The results for the image contrast changes are presented in table III. In this case there are considerable differences in the RMS errors as the γ coefficient varies. For those images where the original had a good contrast between background and foreground (markers), the RMS errors remain low, close to the reference error of the original image, for all γ values tested, except for image *T3* that has considerable noise. For the two test images that were originally very dark (*T2* and *T4*), with low contrast between markers and background, there are some variability in the RMS error, and for very high values of γ the marker detection in fact fails.

The results for the experiment with Salt & Pepper and zero mean Gaussian noise are presented in table IV. Some images are nearly unaffected by the presence of noise (e.g. *T1* and *T5*) while other are greatly affected. The results presented in tables III and IV suggest that the mark detection is more affected by low contrast than by noise. However, the images were subjected to extreme changes in the contrast experiment, whereas only moderate noise intensity was tested.

TABLE III
AVERAGE RMS ERRORS FOR THE CONTRAST EXPERIMENT.

<i>Gamma factor</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
$\gamma = 0.1$	1.6	3.4	4.0	5.5	2.1	2.2
$\gamma = 0.2$	1.6	4.6	7.8	2.0	2.1	2.6
$\gamma = 0.4$	2.2	5.2	4.3	3.0	2.2	6.0
$\gamma = 0.6$	2.1	6.0	3.5	2.0	2.1	2.7
$\gamma = 0.8$	2.1	2.2	3.0	4.1	3.0	2.4
$\gamma = 1$	1.1	3.8	2.3	2.8	2.6	2.4
$\gamma = 2$	1.2	3.1	3.1	2.1	2.4	3.0
$\gamma = 4$	1.2	2.6	5.0	4.0	2.5	3.8
$\gamma = 6$	1.2	fail	2.0	fail	3.0	2.1
$\gamma = 8$	1.2	fail	2.1	fail	3.0	1.8

TABLE IV
AVERAGE RMS ERRORS FOR THE NOISE EXPERIMENT.

<i>Noise type, level</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	<i>T6</i>
S&P, $f = 1.0\%$	1.9	2.2	2.7	3.7	2.1	4.8
S&P, $f = 1.5\%$	1.8	2.2	2.6	3.4	2.1	4.9
S&P, $f = 2.0\%$	2.0	2.1	2.7	3.5	2.1	5.0
S&P, $f = 2.5\%$	1.7	2.3	3.4	6.8	2.2	4.9
S&P, $f = 3.0\%$	1.6	7.4	2.6	5.6	2.9	3.3
Gaussian, $\sigma = 0.01$	1.7	5.7	2.8	2.7	2.2	2.6
Gaussian, $\sigma = 0.02$	1.7	4.8	3.2	2.5	2.2	3.1
Gaussian, $\sigma = 0.04$	1.8	5.8	3.8	2.4	2.0	6.4
Gaussian, $\sigma = 0.06$	1.7	5.7	4.5	2.4	2.5	2.6
Gaussian, $\sigma = 0.10$	1.8	3.3	4.6	2.6	2.6	3.1

V. CONCLUSIONS

The proposed method for the automatic identification of molecular markers in dot blot images proved to be effective. The detection process was found to be robust to rotation and translation of the grid, and to variations in contrast and noise level in the image. All visible marks were detected in 218 out of 222 images tested. The only images (4) that the mark detection failed had very low contrast originally, and were subjected to a γ correction with high value (6 and 8).

The experiment with a manual identification of markers position and size showed that the marker size (radius) was correctly detected by the automatic system. The difference between the manual and automatic estimation of the marker radius was found to be less than 1 pixel (or less than 4%, as the typical value for the radius is 26 pixels), which can be considered a good result. The RMS error in positioning (in XY) the molecular marker centers was between 1.1 and 3.8 in the original images. The estimate of the location of the molecular marker centers was found to be robust to rotation and translation, and to vary within reasonable limits with changes in contrast and noise level.

The final goal of this work is to develop an automatic system to fully process the dot blot images. This includes the identification of the grid, with the position and size of the molecular markers, but also the extraction of a probability of being ON for each mark. The development of an automatic application for the analysis of dot blot images, will increase

the reliability of macroarrays used for bacteria detection, and is therefore an important contribution in diagnostic microbiology.

REFERENCES

- [1] J. Vieira, M.V. Mendes, P. Albuquerque, P. Moradas-Ferreira, and F. Tavares, "A novel approach for the identification of bacterial tax-specific molecular markers", *Letters in Applied Microbiology*, vol. 44, 2007, pp. 506-512.
- [2] P. Albuquerque, M.V. Mendes, C.L. Santos, Moradas-Ferreira, and F.P. Tavares, "DNA signature-based approaches for bacterial detection and identification", *Sci. Total Environ*, vol. (in press), 2009.
- [3] A. Fessehaie, S.H. De Boer, et al., "An oligonucleotide array for the identification and differentiation of bacteria pathogenic on potato", *Phytopathology*, vol. 93(3), 2003, pp. 262-269.
- [4] P. Poltronieri, O.F. D'Urso, et al., "DNA Arrays and Membrane Hybridization Methods for Screening of Six Lactobacillus Species Common in Food Products", *Food Analytical Methods*, vol. 1(3), 2008, pp. 171-180.
- [5] J.T. Tambong, A. Cock, et al., "Oligonucleotide array for identification and detection of Pythium species", *Applied and Environmental Microbiology*, vol. 72(4), 2006, pp. 2691-2706.
- [6] R. khushiramani, S.K. Girisha, et al., "Evaluation of a digoxigenin-labelled probe for detection of Aeromonas spp", *Letters in Applied Microbiology*, vol. 48(3), 2009, pp. 383-385.
- [7] P. Albuquerque, F. Tavares, P. Moradas-Ferreira, and M.V. Mendes, "Twenty molecular markers for an efficient detection of Ralstonia solanacearum", in *Proceedings of the 2nd FEMS Congress, 4-6 July 2006, Madrid, Spain*. FEMS, 2006.
- [8] N. Otsu, "A threshold selection method from gray-level histograms", *IEEE Trans. on Systems Man Cybernetics*, vol. 9(1), 2002, pp. 62-69.
- [9] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Prentice Hall, Upper Saddle River, New Jersey, 3rd edition, 2008.
- [10] The MathWorks, *Using Matlab, Version 6.5.*, The MathWorks, Inc. Natick, MA, 2002.
- [11] K. Kraus, *Photogrammetry - Fundamentals and Standard Processes*, Dummer, Bonn, 4th edition, 1993.