ORIGINAL ARTICLE

Low-complexity PDE-based approach for automatic microarray image processing

Bogdan Belean · Romulus Terebes · Adrian Bot

Abstract Microarray image processing is known as a valuable tool for gene expression estimation, a crucial step in understanding biological processes within living organisms. Automation and reliability are open subjects in microarray image processing, where grid alignment and spot segmentation are essential processes that can influence the quality of gene expression information. The paper proposes a novel partial differential equation (PDE)-based approach for fully automatic grid alignment in case of microarray images. Our approach can handle image distortions and performs grid alignment using the vertical and horizontal luminance function profiles. These profiles are evolved using a hyperbolic shock filter PDE and then refined using the autocorrelation function. The results are compared with the ones delivered by state-of-the-art approaches for grid alignment in terms of accuracy and computational complexity. Using the same PDE formalism and curve fitting, automatic spot segmentation is achieved and visual results are presented. Considering microarray images with different spots layouts, reliable results in terms of accuracy and reduced computational complexity are achieved, compared with existing software platforms and state-of-the-art methods for microarray image processing.

Keywords Gene expression · Microarray · Image processing

B. Belean (✉) · A. Bot
Department of Mass Spectrometry, Chromatography and Applied
Physics, National Institute for Research and Development
of Isotopic and Molecular Technologies, Cluj-Napoca, Romania
e-mail: bogdan.belean@itim-cj.ro

B. Belean · R. Terebes
Department of Communications, Technical University
of Cluj-Napoca, Cluj-Napoca, Romania

1 Introduction

Gene expression identification for fully sequenced genomes leads to the understanding of biological processes within any living organisms. The informational pathway in gene expression is as follows: The protein-coding information is transmitted from nucleus to cytoplasm by an intermediate molecule called mRNA. The information encoded in this molecule is translated into functional gene products known as proteins.

Measurement of gene expression provides clues about regulatory mechanisms, biochemical pathways and broader cellular function. Gene expression levels reside in the quantity of the mRNA sample found in each cell. Microarray technology is known as a valuable tool used to identify genes in biological sequences and to determine their functionality and their expression levels under different conditions. The conduct of a microarray experiment starts with the transformation of the mRNA samples into complementary DNA (cDNA), to prevent the degradation of mRNA molecules. Further on, samples from two sources (cDNA from a target sample and cDNA from a reference sample) are labeled with two different fluorescent markers (cyanine 3—Cy3 and cyanine 5—Cy5, respectively), attached and hybridized on to a solid surface array—glass slide, obtaining a collection of DNA spots, known as a DNA microarray. After hybridization, the array is scanned using two light sources with different wavelengths for each marker (red and green), to determine the amount of labeled sample bound to each spot through hybridization process. The light sources induce fluorescence in the spots which is captured by a scanner, and a composite microarray image is produced, indicating the expression levels of each gene (spot) in both samples [34]. Thus, by estimating gene expression levels, microarray technology compares genes from normal

cells with abnormal or treated cells, determining and providing information for understanding the genes involved in different diseases [13].

Recent research developed several microarray image processing techniques, specific for cDNA microarray gene expression levels estimation. The classical flow of processing a microarray image is generally separated in the following tasks: preprocessing, for improving image quality and enhancing weakly expressed spots, addressing and segmentation. The addressing step, known also as gridding or grid alignment, associates logical coordinates to each spot of the image, whereas segmentation classifies pixels either as foreground, representing the DNA spots, either as background. In the last step, spot intensity values and background intensities corresponding to each spot are computed and, based on the results, gene expression levels are estimated and used for further interpretation.

In terms of automation, as reported in [6], microarray spot detection techniques can be classified as manual, semi-automated or fully automated. In case of currently available software like ImaGene (Biodiscovery, Inc.), GenePix Pro (Molecular Devices, Inc.), ScanAlyze and SpotFinder, the procedure of grid alignment is performed interactively by the user, using a template-based approach which requires various parameters adjustments [5]. Automatic template adjustments were introduced by GenePix, Feature Extraction Software and QuantArray [6, 36], but even so, if grid geometry deviation is increased, the methods are not efficient [5]. Nevertheless, for each type of microarray technology, different template definitions are necessary; thus, the methods are not fully automated. Regarding reliability, grid alignment methods should be able to determine spots with various shape and size in the presence of noise and artifacts introduced by microarray slide printing and by the hybridization process of the target material.

Various image processing techniques for automated grid alignment and spot detection were proposed for handling all the above issues. Mathematical morphology, a valuable tool for analyzing geometric structures, was used by Wang et al. [37] and Angulo et al. [4] for sub-grid detection and grid alignment. A hill-climbing approach was proposed in [33], where spot positions are determined using different probabilistic models for spot intensities distribution. Blekas et al. [10] proposed a Gaussian mixture model for accurate grid alignment, which lacks in terms of automation due to the prior requirements of the number of spots on rows and columns. A novel approach based on genetic algorithms was used by Zaharia et al. [38] for automatic grid alignment, providing better results than the Blekas et al. method, mainly due to its robustness to noise and accidental image rotations. As reported in [7, 8], automatic gridding for microarray images can be performed also using an SVM-based approach, by maximizing the margin between consecutive rows or columns of spots. A fully automated gridding method for microarray images has been also proposed by Rueda et al. [32]; the method uses optimal multilevel thresholding followed by a refinement procedure to find the positions of the sub-grids in the image and the positions of the spots in each detected sub-grid.

Spot segmentation within microarray images addresses the classification of the foreground and background pixels in the target regions. Ideally, each microarray spot has the shape of a circle with a constant diameter for all spots. Unfortunately, the scanning process introduces distortions leading to variable sizes, variable contours (sickle donut or interrupted shapes) or spatial artifacts. Various image processing techniques, classified in spatial and distributional methods [11], are proposed to deal with the mentioned segmentation-related problems. Adaptive pixel clustering techniques were used by [11, 17, 26, 30] for the segmentation of microarray spots having variable contours. Alternate methods, such as the snake fisher model or 3D spot modeling, were used in [21] and, respectively, [39] for spot segmentation. Markov random field modeling, used in [14, 23], combines both observed intensity and spatial information for spot segmentation. Image projections were successfully used both in foreground separation within microarray images [6] and in X-ray image segmentation [16]. A computational method based on geometric measures [40] and the growing concentric hexagon algorithm [18] were also proposed for classifying background and foreground pixels.

We propose a novel approach for automatic grid alignment using shock filters applied on the vertical and horizontal luminance function profiles. Our approach first detects grid positions and performs spot segmentation taking into account the detected rows and columns of spots. The novel contributions of the proposed approach are the following:

- fully automatic grid alignment using a PDE-based approach;
- microarray spot segmentation;
- lower computational complexity compared to state-of-the-art approaches;
- accuracy and robustness to noise and artifacts.

In order to validate the proposed approach for automatic grid alignment and spot segmentation, we compare our results with the ones obtained by dedicated software packages and state-of-the-art approaches on images drawn from public microarray databases.

1.1 Methods

Typically, the microarray scanning process delivers $16 \times$ bits grayscale images, TIFF format, in which spot

fluorescence levels are captured. Depending on the used microarray technology, the resulted microarray image contains one or more sub-grids, with each sub-grid consisting in a two-dimensional array of spots. Image processing techniques were used in order to determine the spot locations within each sub-grid, the spot sizes, the spot intensities and background intensity information, delivering them as raw data parameters.

The proposed methods for automatic microarray image analysis are detailed further on.

1.2 Preprocessing

A common characteristic of microarray images delivered by existing scanners is the low expression levels of microarray spots. Thus, microarray image analysis starts with a logarithm point-wise transformation used to enhance weekly expressed spots. Consider the input microarray image $I = \{p_{i,j}\}$, with $p_{i,j}$ is the pixel intensity value from row i and column j. The logarithm-transformed image is given by $I_L = \left\{ \log_2\left(p_{i,j} + 1\right) \right\}$, with $p_{i,j} = \overline{0 \dots 2^n - 1}$, where n is the number of bits for luminance function representation. We employ also a normalization step in order to insure that the intensity histogram fits the full dynamic range of the n—bits microarray image. Using a contrast stretching procedure, the normalized I_L pixel intensity values are mapped into new values $I_S = \left\{ p'_{i,j} \right\}$, such that the new $p'_{i,j}$ values are saturated considering 1 % of low and high intensities of I_L. Accidental microarray image rotation introduced by the scanning process is detected and eliminated using the Radon transform, as reported in [12, 32].

1.3 Morphological filtering and microarray sub-grid detection

In case of multiple groups of spots within the same microarray image, the sub-grid detection step estimates the location of each group. First, image vertical (VP) and horizontal profiles (HP) were computed by determining the mean pixel intensity value along rows and columns, respectively. The periodic features of these one-dimensional profiles were estimated using autocorrelation, which reveals the microarray spot size d. Using a disk structuring element of diameter d, morphological operators were applied on the preprocessed microarray image aiming the separation of microarray spot groups. Thus, after a morphological closing operator using a structuring element with a diameter d, the microarray spots within the same group were fused together while the borders of the spot groups are preserved. Once again the image vertical (VP') and horizontal profiles (HP') computation was performed and, by estimating profiles periodicity using autocorrelation, the number of spots groups on each row and column was determined together with spot group dimensions. Taking into account the minimum values of the VP' and HP' profiles, together with the number of spot groups and their sizes, the positions of the sub-grid were automatically detected. It is to be mentioned that, prior to image closing, a top-hat filtering was applied on the microarray image in order to minimize the contribution of the background variations to the image horizontal and vertical profiles computation, which can lead to mismatched spot group detection in case of a strongly varying background. The detected spot groups, represented as rectangles enclosing all spots within the same group, are illustrated in the Sect. 3.

1.4 Automatic grid alignment

Grid alignment in case of microarray image processing aims to determine each microarray spot location by registering a set of horizontal and vertical lines which describe a two-dimensional array of spots. The existing software platforms for microarray image analysis together with current research impose two approaches for grid alignment: template-based and, respectively, data-driven approaches. In first case, a pre-defined template is overlaid on the microarray image and it is adjusted in order to match the spots in the microarray image of interest. The second approach, considered in our method, is based on image profile analysis and image processing techniques for automatically determining the microarray grid position.

Each microarray spot group localized by the sub-grid detection was first smoothed using a 2D convolution with a Gaussian kernel in order to reduce the noise influence on image projections. Further on, the profiles for each spot group of size M(width) $\times$ N(height) were computed as described by (1) and (2), where VP and HP represent the vertical and horizontal profiles, respectively. We denote by $p_{i,j}$ the pixel intensity value from row i and column j within the microarray image.

$$VP(i) = \frac{1}{N} \sum_{j=0}^{N-1} p_{i,j} \tag{1}$$

$$HP(j) = \frac{1}{M} \sum_{i=0}^{M-1} p_{i,j} \tag{2}$$

The hyperbolic partial differential equation describing the one-dimensional (1D) shock filter was used for automatic grid alignment. The equation:

$$\frac{\partial U}{\partial t} + F(U_{xx})|U_x| = 0 \tag{3}$$

was proposed by Osher and Rudin in [29], aiming blurry edge enhancement. The initial condition of the shock filter

is $U(x, 0) = U_0(x)$, with the operator F fulfilling the following conditions: $F(0) = 0$ and $F(s) \cdot sign(s) \geq 0$. (Note: U_x and U_{xx} represent, respectively, the first- and the second-order derivatives). By choosing $F(s) = -sign(s)$, we obtain the classical shock filter Eq. (4):

$$U_t = -sign(U_{xx})|U_x| \qquad (4)$$

The discrete scheme described by (5) is used to apply the 1D shock filters on both horizontal and vertical microarray image profiles denoted by P:

$$P(i)^{n+1} = P(i)^n - \Delta t \cdot |DP(i)^n| \cdot sign(D^2 P(i)^n), \qquad (5)$$

with:

$$\begin{aligned} DP(i)^n &= m(\Delta_+ P(i)^n, \Delta_- P(i)^n) \\ D^2 P(i)^n &= (\Delta_+ \Delta_- P(i)^n) \end{aligned} \qquad (6)$$

$m(x, y)$ the "minmod" function:

$$m(x, y) = [sign(x) + sign(y)] \cdot \min(|x|, |y|) \qquad (7)$$

and

$$\Delta_\pm = \pm(P(i \pm 1) - P(i)) \qquad (8)$$

By applying shock filters, image profiles evolve as described by the Eq. (4); an example can be seen in Fig. 1a where a section of the original profile and its corresponding result after applying shock filtering are illustrated. The thin continuous line represents the microarray image profile which evolves in the direction pointed by the arrows. The final result is represented by the thick continuous line, whereas dotted lines representing intermediate steps in the evolution of the image profiles. The shock filter creates strong discontinuities at the inflexion points; thus, based on the resulted profiles, grid alignment was performed. Pairs of perpendicular lines were drawn over the picture as shown in Fig. 1b, considering the inflexion points on both the horizontal and vertical image profiles. The resulted grid is constructed as illustrated in Fig. 1c, with the grid lines locations computed as average positions between the coordinates $(2i, 2i + 1)$ and $(2j, 2j + 1)$ of the vertical and, respectively, horizontal lines.

For images containing one group of spots, Fig. 1c, the increased number of spots in a single column or line determines strong discontinuities at inflexion points within the image profiles Fig. 1a, which determine the correct grid lines. On such images, delivered by Agilent microarray scanners and included in our first data set, accurate results can be obtained using the previously described approach. Images containing multiple groups of spots having weakly expressed spots within a spot group (Fig. 1d) can lead to the detection of multiple inflexion points (Fig. 1d, e). For accurately handling these situations, we employed an

autocorrelation function-based refinement procedure. Let $P(i)$ with $i = 1…N$ be the image profile for a spot group of size N after applying the shock filters and d the profile periodicity determined using autocorrelation, as described by equation:

$$ac(t) = \sum_{i=1}^{N} P(i)P(i - t) \qquad (9)$$

In Fig. 1e, a section of the image profile is illustrated for the marked region with weekly expressed spots from Fig. 1d. The strong discontinuities determined using the shock-filtered $P(i)$ profile, which delineate each line of spots, are denoted by the positions $(2j - 1, 2j)$ and represented with dashed lines. The dotted lines over the $P(i)$ profile, denoted by the (e_1, e_2) pairs, represent the grid lines positions in case of multiple inflexion points are involved in delineating lines of spots. These pairs (e_1, e_2) were determined as follows: Considering the nearest computed discontinuity $(2j - 1, 2j)$ and the profile periodicity d, the closest inflexion points to the estimated $(2j - 1 + d, 2j + d)$ positions represent the (e_1, e_2) pairs.

1.5 Spot segmentation

Once grid alignment is performed, the subsequent task is to separate each spot from its local background within the overall microarray grid. Let n and m be the number of spots on horizontal and vertical axis within the microarray image. As illustrated in Fig. 2a, $I_{\text{Line}}(i)$ and $I_{\text{Col}}(j)$ with $i = 1…n$, and $j = 1…m$ represent microarray sub-images determined in the grid alignment process, each containing one line and, respectively, one column of microarray spots. For the I_{Line} and I_{Col} sub-images, the horizontal image profile $HP(I_{\text{Line}})$ and the vertical image profile $VP(I_{\text{Col}})$ were computed. The PDE formalism was used once again, this time as a preprocessing step for spot segmentation. We applied shock filtering on the horizontal profiles and vertical profiles for each sub-image representing spot lines and spot columns, respectively. The resulted profiles, represented by dotted lines in Fig. 2b, mark the pairs of inflexion points (A, B) and (A', B'), corresponding to the beginning and the end of each microarray spot, within each spot line and spot column. Thus, as shown in Fig. 2b, the pairs (A, A') and (B, B') define a rectangular area enclosing each microarray spot (i, j). The final step for the background segmentation is the determination of the ellipse E with the maximum area, inscribed in the rectangular area associated with a microarray spot. One can argue that the rectangular area could be rotated and not aligned on vertical and horizontal axes. Our method does not handle these cases; nevertheless, the circular nature of the microarray spots makes unnecessary to consider rotated rectangles areas. Various approaches of
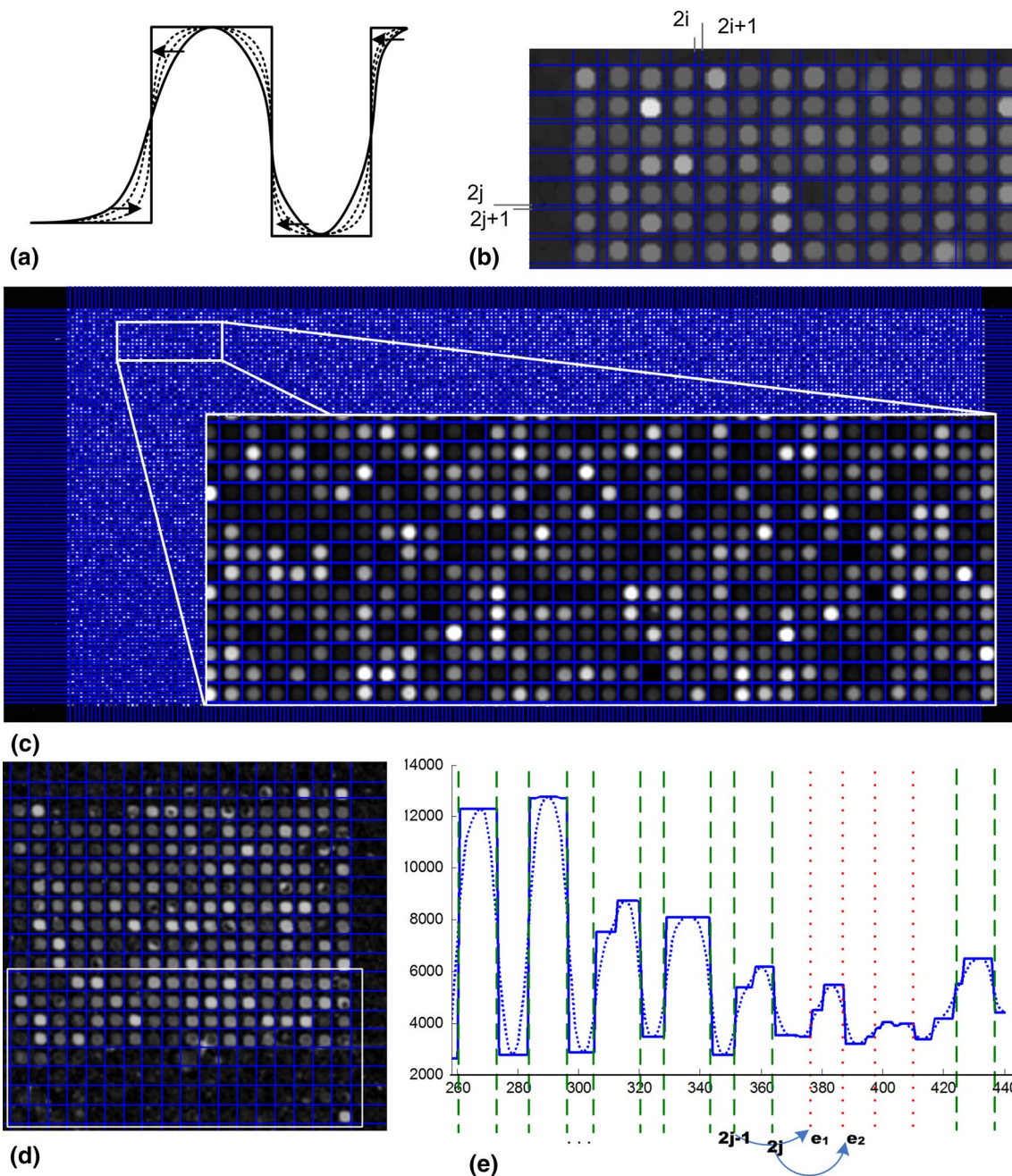
Fig. 1 Microarray grid alignment. **a** Image profile evolution based on partially differential equations, **b** detected microarray grid lines based on the determined inflexion points, **c** overview of the detected grid in case of real-life microarray image, **d** microarray spot group with weekly expressed spots, **e** grid lines determined by the autocorrelation-based refinement procedure

shock filter formulation such as [27] are available, where image denoising is also taken into account. Without loss of generality, our approach can also be used in cases where circular or elliptic shape detection is necessary [15, 31]. Visual results of the proposed spot segmentation approach together with its computational complexity estimation are included in the Sect. 2.

1.6 Output measures

The results obtained by our novel method for grid alignment and spot segmentation were compared with state-of-the-art results, and the results delivered by existing software platforms. The compute and output measures used to validate our results are presented in the subsequent.
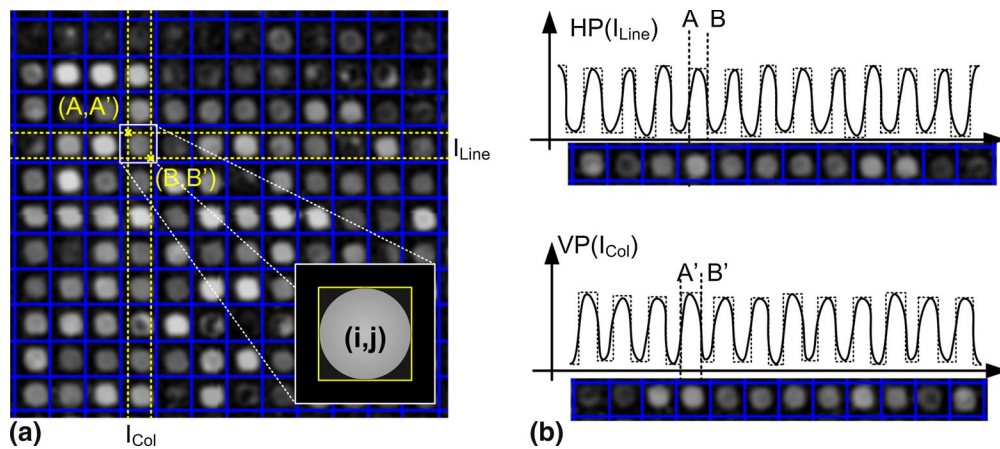
Fig. 2 Microarray spots segmentation. **a** Selection of I_{line} and I_{col} sub-images containing each one line and one column of spots, respectively, corresponding to the marked microarray spot (i, j), **b** the hori- zontal and vertical shock-filtered profiles $HP(I_{line})$ and $VP(I_{col})$ for the I_{line} and I_{col}, respectively, together with the pairs (AA') and (BB') which define the rectangular area enclosing the microarray spot (i, j)

A classical quality measure for the grid alignment step is the percentage of grid lines which separates spots correctly, marginally or incorrectly [7, 32]. The mass centers, mean intensity and the coefficient of determination of each spot, proposed by Agilent Feature Extraction and GenePix Pro software platforms, can be used also to validate both the grid alignment and the spot segmentation results. A detailed comparison of our results with the ones computed with competing methods on images drawn from public databases, using all the aforementioned quantitative measures, is presented in the next section.

2 Results

In order to assess the accuracy and reliability of the proposed methods for automatic microarray image processing, we use the GEO (Gene Expression Omnibus) and the SMD (Stanford Microarray Database) public databases. The proposed method is applied on microarray images with different scanning resolutions and different spot layouts. Two types of images are considered and analyzed: images containing multiple sub-grids (groups of spots) and images having a single grid. Our first data set consists of 10 6,100 × 2,160 pixels size microarray images delivered by Agilent Scanners, having only one spot group containing 22,575 spots. Three types of images, corresponding to both Cy2 and Cy5 channels, can be distinguished within this set: images which correspond to a study of liver treatments on "*mus musculus*" known as house mouse; images corresponding to sesame oil treatment for blood pressure on "*mus musculus*"; and images expressing proteasome inhibitor treatment for breast cancer. The second data set extracted from the SMD database contains 10

5,550 × 1,910 pixels size microarray images with 48 spot groups, 324 spots per each group. The data set represents a study of the global transcriptional factors for hormone treatment of "*Arabidopsis thaliana*" samples. Our third data set includes a set of 8 images from the GEO database, corresponding to an "*Atlantic salmon head kidney*" study. Each microarray image of 5,897 × 2,170 pixels size contains 48 spots groups with 182 spots per each group. For the last-mentioned two data sets, the same images as the ones referenced in [32] were selected, an extended data set being obtained.

In case of state-of-the-art approaches proposed in [7, 32], quality assessment of the image processing techniques for automatic microarray grid alignment is performed using the resulted gridded image. Thus, [7] evaluates each spot as being marginally, perfectly and incorrectly gridded, depending on the percentage of its pixels contained in the detected grid, whereas [32] determines the percentage of grid lines that separate spots marginally, perfectly or incorrect. To our knowledge, the results reported in [32] were not compared with the results drawn from the public databases. The information regarding the microarray spot segmentation introduced by our proposed PDE-based approach allowed us to perform an in-depth comparison of our results with the results delivered by existing software packages, on images from the GEO and SMD databases.

In case of our *first data set*, grid alignment was accurately performed, each spot residing in its detected grid cell. In our previous work [9], for the same data set, we determined spot locations as the center of each grid cell. The supplementary spot segmentation information allows us to perform an in-depth comparison of our results with those included in the GEO database. As described in [1], the Feature Extraction software package locates each spot

i by computing its center of mass on the scanned microarray image, denoted by pairs (XG_i, YG_i). Let $I_{i,j}$ be the pixel intensities for each spot enclosed in a squared area of size $m \times n$ defined by our proposed approach. The pairs (X_i, Y_i) representing the obtained mass center for each spot were computed as shown in Eq. (10).

$$X_i = \frac{1}{\sum I_{i,j}} \sum_{j=0}^{m} \left(\sum_{i=0}^{n} i \cdot I_{i,j} \right) \quad if \; I_{i,j} \subset E \qquad (10)$$

The mean euclidian distance d between the mass centers included in the GEO database and the ones obtained by our approach, together with the standard deviation, were estimated in case of each microarray image. In Table 1, results of the aforementioned comparison are listed on a subset of ten microarray images. For the whole data set, all spots reside inside the determined grid cells. For the whole population of spots, we obtained a mean distance d of 0.226 pixels, whereas the standard deviation of the mean distance over the whole population of spots was 0.254. Moreover, for 98.89 % of the total number of spots within the first data set, the distance between computed mass center and the mass centers determined by Feature Extraction was less than 1 pixel. We show in Fig. 3 visual results of the proposed segmentation method for the case of GSM102718; these results are consistent with the computed quality measures.

In case of the *second data set*, the grid alignment step was evaluated using the same quality measures as the ones from [7, 32]. Table 2 shows comparative results between our approach and the method described in [32], for images from the SMD database. The average accuracy for the whole data set, considering the percentage of grid lines that separates spot perfectly, was 99.39, 1.3 % higher than the state-of-the-arts methods.

Failure to detect some spot regions due to the extremely contaminated microarray image is reported in [32] for the AT-20392-ch1 microarray image. An example for such a situation is shown on the sub-image from Fig. 4c, the marked group of spots not being detected by the grid alignment method proposed in [10]. For the same microarray image, the results obtained by our proposed approach can be visualized in Fig. 4c, e. Despite the presence of large bright artifacts due to the slide printing process, grid alignment was correctly performed even in the vicinity of the artifacts. Moreover, spot segmentation was also performed as shown in Fig. 4e, where the elliptic shape inscribed in the rectangle area represents the microarray spot segmented from the background. For evaluating the segmentation accuracy, we use the images included in our second data set and the same experimental protocol used to quantify results obtained by the GenePix platform in [20]. For a better understanding, we present in the subsequent a brief description of the qualitative measures, parameters and results delivered by the GenePix platform. We assume the reference sample is labeled with Cy3, and the test sample

Table 1 Results of the proposed method in case of the first data set

Experiment ID	Mean euclidian distance d	% $d < 1$	Standard deviation σ
GSM102718	0.170	99.514	0.154
GSM102721	0.212	99.478	0.302
GSM102722	0.194	99.112	0.201
GSM135598	0.248	99.018	0.226
GSM135599	0.243	99.509	0.165
GSM135560	0.248	97.883	0.315
GSM207313	0.360	97.554	0.529
GSM207315	0.052	99.670	0.094
GSM207316	0.274	98.924	0.298
GSM207320	0.263	98.214	0.256

The mean Euclidian distance (in pixels) between the determined mass center and the ones drawn from public database for the microarray spots enclosed within each image from the first data set

is labeled with Cy5. The dynamic range for the fluorescent intensity measurements for both Cy3 and Cy5 channels of the digitized microarray image is I from 0 to 65,535 gray levels. The microarray spot characteristics are given by the following parameters: spot location (x, y), median intensity values of all pixels representing each spot in case of both Cy3 and Cy5 channels ($cy3\_med$, $cy5\_med$), median intensity of background intensity values of all pixels that fall within the local background of each spot from both Cy3 and Cy5 channels (b_cy3_med, b_cy5_med). Let $I_{cy3}(i)$ and $I_{cy5}(i)$ denote the fluorescence intensity levels for each pixel i within the rectangle area enclosing each microarray spot from the Cy3 and, respectively, Cy5 channels. The rectangle area includes both local background and microarray spot pixels. Considering the aforementioned spot characteristics, three different ratios quantities are computed by GenePix for evaluating whether changes in the intensity levels of the reference and test samples are significant: The *Ratio of Medians* computes the ratio of the background corrected median intensity values from the whole spot; the *Median of Ratios* calculates a median value of the ratios I_{Cy3}/I_{Cy5} intensity for each pixel within a microarray spot; and the *Regression Ratio*(Rgn R) represents an independent measure defined by the slope of the least-squares best fit regression line of the fluorescence intensity values for each pixel against each other [e.g., $I_{Cy3}(i)$ versus $I_{Cy5}(i)$]. The regression ratio indicates individual feature quality. Considering the regression pixels used to calculate the Rgn R, the coefficient of determination (Rgn R^2) for the least-squares regression fit of a microarray spot is defined as the square of the correlation coefficient and ranges values between 0 and 1 [20]. In general, higher Rgn R^2 values indicate higher spot quality and tight correlations between the different ratio measurements (ratio of medians, median
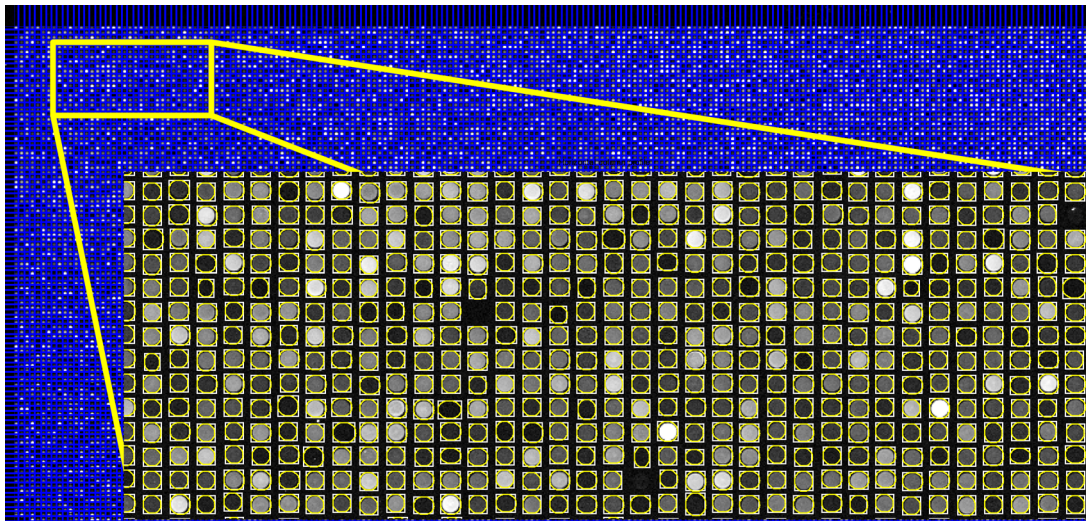
Fig. 3 Visual results of the spot segmentation method in case of GSM102718 microarray image

Table 2 Results of the proposed grid alignment and segmentation in case of the second data set

Experiment ID	Grid alignment						Segmentation (Rgn R^2) PDE versus GenePix		
	Incorrectly (%)		Marginally (%)		Perfectly (%)				
	Omtg	PDE	Omtg	PDE	Omtg	PDE	r value	Av. Diff.	SD
AT-20385-ch1	4.30	0.23	0.46	0.11	95.24	99.66	0.958	0.059	0.069
AT-20385-ch2	2.83	0.34	0.09	0.23	97.08	99.43			
AT-20387-ch1	2.90	0.46	0.14	0.57	96.96	98.97	0.915	0.102	0.092
AT-20387-ch2	0.52	0.20	0.11	0.48	99.37	99.32			
AT-20391-ch1	0.64	0.23	0.17	0.34	99.19	99.45	0.928	0.093	0.088
AT-20391-ch2	0.32	0.23	0.26	0.34	99.42	99.43			
AT-20392-ch1	4.10	0.37	0.33	0.23	95.57	99.40	0.973	0.041	0.062
AT-20392-ch2	0.21	0.37	0.25	0.57	99.54	99.06			
AT-20395-ch1	0.41	0.34	0.12	0.11	99.47	99.56	0.951	0.059	0.077
AT-20395-ch2	0.98	0.23	0.31	0.11	98.71	99.66			

The accuracy of the proposed grid alignment method for each image within the second data set determined by the percentage of spots perfectly, marginally or incorrectly aligned; the accuracy of the proposed segmentation approach determined by the Pearson correlation coefficient, average difference and the standard deviation between the computed coefficients of determination and the ones drawn from the SMD database

of ratios and regression ratio). For validating our approach, we correlated the coefficients of determination computed by our approach with the ones included in the SMD database for the entire population of microarray spots within each microarray image. Let Rgn $R^2_{PDE}(i)$ be the obtained coefficient of determination for the spot i using our approach and Rgn $R^2_{SMD}(i)$ be the coefficient of determination drawn from SMD database for the same spot i. The correlation coefficient together with the mean difference between our results and the public SMD results is described by Eqs. (11) and (12), respectively. Moreover, considering the whole population of spots within a microarray image, the

dispersion of the difference between the two coefficients of determination from the average is given by the computed standard deviation SD (see Table 2—segmentation section).

$$r \text{ value} = r\left(\text{Rgn } R^2_{\text{SMD}}, \text{ Rgn } R^2_{\text{PDE}}\right) \tag{11}$$

Average difference
$$= \text{mean}(\text{abs}(\text{Rgn } R^2_{\text{SMD}}(i) - \text{Rgn } R^2_{\text{PDE}}(i))) \tag{12}$$

An average Pearson correlation coefficient of 0.945 together with an average difference of 0.070 and an average
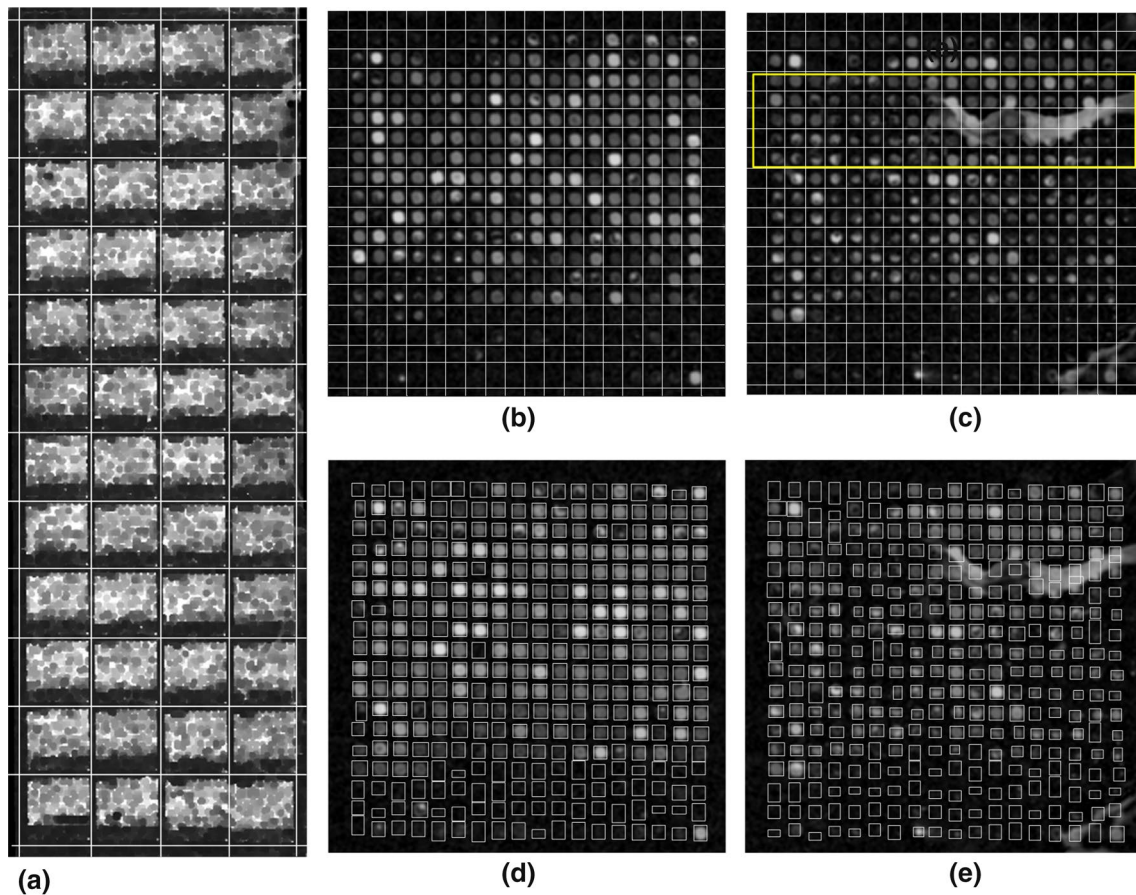
Fig. 4 Visual results in case of our grid alignment and segmentation methods in case of AT-20392-ch1 image in the presence of slide printing artifacts. **a** Spots group detection, **b** grid alignment for a selected spot group with weekly expressed spots, **c** grid alignment for a selected spot group containing large artifacts, **d**, **e** spot segmentation within the aforementioned spot groups

Table 3 Results of the proposed grid alignment and segmentation in case of the third data set

Experiment ID	Grid alignment						Segmentation ($\bar{I}_{spot}$) PDE versus GEO results		
	Incorrectly (%)		Marginally (%)		Perfectly (%)				
	Omtg	PDE	Omtg	PDE	Omtg	PDE	r value	Av. Diff.	SD
GSM15898-ch1	0.58	0.32	0.16	0	99.26	99.68	0.990	0.047	0.056
GSM15898-ch2	1.00	0.16	0.21	0.32	98.79	99.52	0.989	0.049	0.068
GSM16101-ch1	0.00	0.16	0.32	0.32	99.68	99.52	0.972	0.036	0.063
GSM16101-ch2	1.57	0.48	0.06	0.14	98.37	99.38	0.964	0.045	0.051
GSM16389-ch1	0.79	0.32	0.12	0.29	99.09	99.39	0.943	0.087	0.084
GSM16389-ch2	0.57	0.30	0.04	0.16	99.39	99.54	0.985	0.058	0.062
GSM16391-ch1	0.00	0.14	0.24	0.32	99.76	99.54	0.984	0.048	0.059
GSM16391-ch2	0.14	0.16	0.13	0	99.73	99.84	0.992	0.042	0.054

The accuracy of the proposed grid alignment method for each image within the third data set determined by the percentage of spots perfectly, marginally or incorrectly aligned; the accuracy of the proposed segmentation approach determined by the Pearson correlation coefficient, average difference and the standard deviation between the computed spots mean intensities values and the ones drawn from the GEO database

standard deviation of 0.077 was obtained by comparing the coefficients of determination obtained by our approach, and the ones delivered by GenePix.

For the third data set, we include in Table 3 quantitative results of the proposed grid alignment, showing an improvement compared to the OMTG approach presented

Table 4 Computational complexity of the proposed approach for grid alignment and segmentation

Autocorr. [2, 5, 35]	Hill Climbing [33]	Mathematic Morphology [4, 37]	SVM [7, 8]	OMTG [32]	PDE
Grid alignement					
$O(M + N)$	$O(M^2N^2)$	$O(2S_eMN)$	$O(MN(M + k))$	$O(t_sN^2)$	$O(2MN + p(M + N))$
Pixel Clustering [11]	Watershed transform [3]	Markov random field modeling [14, 22]	Snake fisher model [21, 41]	PDE	
Segmentation					
$O(rkMN)$	$O(S_eMN)$	$O(mMN\log(M^2N^2/m))$	$O(n^2MN)$		$O(6MN + (p + d)(\alpha M + \beta N))$

The proposed PDE-based approaches for grid alignment and segmentation are compared with the state-of-the-art methods in terms of computational complexity

in [32]. Regarding spot segmentation, our results were compared with the ones drawn from the GEO database. Taking into account that the coefficients of determination are not included within the GEO data repository, the mean intensity of each microarray spot i, denoted by $\bar{I}_{\text{spot}}(i)$, was used for evaluation. By mean spot intensity $\bar{I}_{\text{spot}}(i)$, we understand the mean intensity of all pixels that fall within the microarray spot i with subtracted mean background intensity for the spot in question. For each microarray image, the resulted mean spots intensities $\bar{I}_{\text{spot}}^{\text{PDE}}$ were correlated with the $\bar{I}_{\text{spot}}^{\text{GEO}}$ intensities drawn from GEO data repository. An overall $r = 0.977$ Pearson correlation coefficient was obtained for the whole data set. Considering the aforementioned mean spots intensities were mapped into [0, 1] interval, the mean difference Av. Diff. and the standard deviation SD were also computed for each microarray image (see Table 3).

2.1 Computational complexity

We estimate the computational complexity both for the state-of-the-art approaches and for our proposed approach for automatic grid alignment and segmentation, considering an M × N pixels size image.

The autocorrelation, commonly used for microarray grid alignment [2, 5, 35], has reduced complexity, but, on the other hand, has major disadvantages such as failure to correctly align the microarray grid in case of irregular image profiles and spots with different radii. Morphological operators, used in [4, 37] for automatic microarray image addressing, have a computational complexity of $O(2S_eMN)$ where S_e is the size in pixels (approx. 10^3) of the structural element for dilation and erosion. As far as the SVM-based approaches [7, 8] are concerned, the computational complexity is $O(MN(M + k))$, one order of magnitude lower than the one associated with the genetic algorithm [38], as reported in [32]. The parameter k represents the number of selected microarray spots to train the SVM, in the case of real microarray images this order being 10^3. For the fully automatic microarray grid alignment performed using an

optimal multilevel threshold approach [32], the reported computational complexity is $O(t_sN^2)$, where t_s denotes the threshold set size.

For our method, the computational cost for the grid alignment procedure, including the refinement procedure, is given by the upper bound function $f(M, N) = 2MN \cdot s + 8p(M + N)s$, with s representing one computational step and p denoting the number of iteration necessary for the profiles evolution. The order of growth for the computational cost is $O(f(M, N)) = 2MN + p(M + N)$ and represents the computational complexity of the proposed method. As denoted by Table 4, reduced computational complexity is achieved in spite of the iterative nature of the shock filters, taking into account that shock filters are applied only on 1D image profiles.

The computational complexity of our PDE-based segmentation procedure was estimated as follows. Let α and β represent the number of microarray spots on each line and columns, respectively, and d the average microarray spot diameter. The average width for a line or a column of spots is 2d. We computed for each spot line and spot column, the horizontal and vertical image profiles, respectively, with the total complexity of $2\alpha dM + 2\beta dN = 4MN$. Shock filters were further on applied on each of the determined profiles having a complexity of $p\alpha M + p\beta N$, where $p\alpha M$ represents p iterations performed on a number of α profiles (i.e., one profile for each line of spots), each profile having the size M. The previous estimations, together with the computational cost for grid alignment, led to the order of growth for the total computational cost for segmentation of $6MN + p(\alpha M + \beta N)$.

As reported in [11], the pixel clustering approach achieves lowest computational complexity by using a k-means clustering algorithm which has a time complexity of $O(rkMN)$. Spot segmentation using mathematical morphology [3] has a computational cost $\gg S_eMN$, due to the morphological filtering by area opening with a structural element of the size of spot S_e used to detect the initial markers for the watershed transform. Taking into account Markov random field modeling for image segmentation

[14], the computational cost is mMNlog(M^2N^2/m), with m being the number of arcs within the minimum cut problem on a graph with MNU nodes and mU arcs, as reported in [22]. The computational complexity of image segmentation using active contours can be reduced to n^2MN, as reported in [41]. The n factor represents the size of a Gaussian kernel $\ll$ MN.

In Table 4, an overall view on the resulted computational complexity of our method with regard to state-of-the-art approaches for microarray image processing was presented. As illustrated by these results, our proposed approach has reduced computational complexity for both microarray image addressing and segmentation, being a strong candidate to be integrated in future software packages. Moreover, the reduced computational complexity of the proposed approaches for automatic grid alignment and segmentation is of high interest in case of application specific future devices for microarray image processing, as the ones presented in [9, 25]. By adding robust processing methods for gene expression microarray analysis and interpretation [19, 24, 28], future devices for medical applications which integrate the complete gene expression analysis chain can be developed.

The proposed methods are also evaluated with regard to their computational time, in order to have an in-depth view on the computational complexity. The workstation used for evaluation is built around an Intel i5, 3.3 MHz processor with 4 GB of RAM. In case of the AT-20385 5,550 × 1,910 pixel size microarray image, with 48 spot groups, each group including 324 spots, grid alignment is performed in 14 s, while the segmentation procedure lasts for 35 s. Considering the same microarray image and the same processing platform, snake fisher model and k-means clustering were used for the segmentation of each microarray spot. The total computational time, including the grid alignment procedure, was 132 s for the k-means pixel clustering procedure and 340 s for the active contours-based approach. Using a similar workstation as processing platform, in [7] grid alignment is achieved in 10 s, in case of a 450 × 450 pixel size image block containing 870 spots.

the first step for including the image processing part of the microarray experiment within the scanner level, which would significantly reduce the costs of a microarray experiment. Reduced complexity of the image processing algorithms will ease their integration into the microarray scanner. Our proposed PDE-based approach for automatic microarray image processing showed significantly reduced computational complexity compared with the state-of-the-art methods. Timing considerations detailed on the Sect. 2 were consistent with the estimated computational complexity.

Our proposed PDE-based grid alignment procedure perfectly determined grid cells for each microarray spot within the first data set. In case of the second and third data sets, our results showed improvement of grid alignment compared with the OMTG methods. The percentage of perfectly identified spots characterizes grid alignment for each microarray image. For the second data set, the averaged percentage of perfectly identified spots is 1.3 % higher than the one resulted by applying the OMTG method. In case of the third data set, the results do not show significant grid alignment improvement compared with the OMTG method (the averaged percentage of perfectly identified spots is 0.3 % higher). Nevertheless, the images within the third data set have a different layout than the two other data sets. Thus, by successfully applying the proposed method on different data sets which include microarray image having different layouts, the generality of our PDE-based approach for microarray image processing is proven.

In order to provide information on the results relevance and reliability, the results delivered by our proposed approach for spots segmentation were compared with the ones delivered by Agilent Feature Extraction and GenePix Pro software platforms, drawn from GEO and SMD data repositories. The low average difference and standard deviations between the coefficients of determination, mass centers and mean spots intensities, in addition to the increased Pearson correlation coefficients revealed by the performed comparisons, showed that the proposed approach delivers accurate and reliable results.

3 Discussion

Automation and reliability are open subjects in microarray image processing. For speeding up the analysis process, it is desirable that grid alignment and spot intensity extraction to be performed in an automated manner, without user intervention. Moreover, user intervention in microarray image processing brings up the need of a workstation, a processing platform together with a bioinformatician, increasing the cost of a microarray experiment. Fully automated image processing techniques are

4 Conclusions

In this paper, a novel partial differential equation-based approach was proposed for fully automatic microarray grid alignment and spot segmentation. The luminance function profiles of the microarray image are evolved using a hyperbolic PDE, which marks the inflexion points over the profiles. Further on, using an autocorrelation-based refinement procedure, the correct grid lines are determined based on the detected inflexion points. Moreover, by applying the PDE formalism on the horizontal and

vertical profiles of each line and column of spots, respectively, segmentation was also achieved. The proposed approach was tested on real-life microarray images with various spots layouts, drawn from SMD and GEO public data repositories. Reduced complexity and a high degree of accuracy in the presence of noise and artifacts were obtained compared with state-of-the-art methods for grid alignment and segmentation in case of microarray image processing.

References

1. Agilent Technologies (2012) Feature extraction reference guide. Agilent Technologies Inc, Santa Clara
2. Alhadidi B, Fakhouri HN, AlMousa OS (2006) cDNA microarray genome image processing using fixed spot position. Am J Appl Sci 3:1730–1734
3. Angulo J, Serra J (2002) Automatic analysis of DNA microarray images using mathematical morphology. Bioinformatics 19(5):553–562
4. Angulo J, Serra J (2003) Automatic analysis of DNA microarray images using mathematical morphology. Oxf Bioinform 19(5):553–562
5. Bajcsy P (2004) Gridline, automatic grid alignment in DNA microarray scans. IEEE Trans Image Process 13:15–25
6. Bajcsy P (2006) An overview of DNA microarray grid alignment and foreground separation approaches. EURASIP J Appl Sig Process 2:1–13
7. Bariamis D, Iakovidis DK, Maroulis D (2010) M^{3}G: maximum margin microarray gridding. BMC Bioinformatics 11:49
8. Bariamis D, Maroulis D, Iakovidis DK (2010) Unsupervised SVM-based gridding for DNA microarray images. Comput Med Imaging Graph 34:418–425
9. Belean B, Borda M, Le Gal B, Terebes R (2012) FPGA based system for automatic cDNA microarray image processing. Comput Med Imaging Graph 36(5):419–429
10. Blekas K, Galatsanos NP, Likas A, Lagaris IE (2005) A Mixture model analysis of DNA microarray images. IEEE Trans Med Imaging 24(7):901–909
11. Bozinov D, Rahnenfuhrer J (2002) Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering. Bioinformatics 18(5):747–756
12. Brandle N, Bischof H, Lapp H (2003) Robust DNA microarray image analysis. Mach Vis Appl 15:11–28
13. Campbell AM, Hatfield WT, Heyer LJ (2007) Make microarray data with known ratios. CBE Life Sci Educ 6:196–197
14. Demirkaya O, Asyali MH, Shoukri MM (2005) Segmentation of cDNA microarray spots using markov random field modelling. Bioinformatics 21(13):2994–3000
15. Emerich S, Lupu E, Arsinte R (2011) A new approach to iris recognition. In: Proceedings of the IEEE international symposium on signals, circuits and systems, ISSCS, pp 1–4
16. Florea L, Florea C, Vertan C, Sultana A (2011) Automatic tools for diagnosis support of total hip replacement follow-up. Adv Electr Comput Eng 11(4):55–62
17. Giannakeas N, Fotiadis D (2009) An automated method for gridding and clustering-based segmentation of cDNA microarray images. Comput Med Imaging Graph 33(1):40–49
18. Giannakeas N, Kalatzis F, Tsipouras M, Fotiadis D (2012) Spot addressing for microarray images structured in hexagonal grids. Comput Methods Programs Biomed 106(1):1–13
19. Gómez P, Díaz F, Martínez R, Malutan R, Rodellar V, Puntonet CG (2006) Robust preprocessing of gene expression microarrays for independent component analysis. Lect Notes Comput Sci ICA 3889:714–721
20. Handran S, Zhai YZ (2003) Biological relevance of GenePix results. Molecular Devices - Application Notes, pp 1–9
21. Ho J, Hwang W (2008) Automatic microarray spot segmentation using a snake-fisher model. IEEE Trans Med Imaging 27(6):847–857
22. Hochbaum D (2001) An efficient algorithm for image segmentation, markov random fields and related problems. J ACM 48(4):686–701
23. Katzer M, Kummert F, Sagerer G (2003) Methods for automatic microarray image segmentation. IEEE Trans Nanobiosci 2(4):202–214
24. Kim KY, Kim J, Kim HJ, Nam W, Cha IH (2010) A method for detecting significant genomic regions associated with oral squamous cell carcinoma using aCGH. Med Biol Eng Compu 48(5):459–468
25. Kornaros G, Blionas S (2008) Microarchitecture of a multicore SoC for data analysis of a lab-on-a-chip microarray. EURASIP J Adv Signal Process 520641:1–11
26. Li Q, Fraley C, Bumgarner R, Yeung K, Raftery A (2005) Donuts, scratches and blanks: robust model-based segmentation of microarray images. Bioinformatics 21(12):2875–2882
27. Ludusan C, Lavialle O (2012) Multifocus image fusion and denoising: a variational approach. Pattern Recogn Lett 33(10):1388–1396
28. Malutan R, Gómez P, Borda M (2010) Independent component analysis algorithms for microarray data analysis. Intell Data Anal 14(2):193–206
29. Osher S, Rudin L (1990) Feature-oriented image enhancement using shock filters. SIAM J 27:919–940
30. Rahnenfuhrer J, Bozinov D (2004) Hybrid clustering for microarray image analysis combining intensity and shape features. BMC Bioinformatics 5:47
31. Roy K, Bhattacharya P (2009) Iris recognition in non-ideal situations. Lect Notes Comput Sci Inf Secur 5735:143–150
32. Rueda L, Rezaeian I (2011) A fully automatic gridding method for cDNA microarray images. BMC Bioinformatics 12(113):1–17
33. Rueda L, Vidyadharan V (2006) A Hill-climbing approach for automatic gridding of cDNA microarray images. IEEE/ACM Trans Comput Biol Bioinf 3(1):72–83
34. Schena M (2003) Microarray analysis. Wiley, New York
35. Steinfath M, Wruck W (2001) Automated image analysis for array hybridization experiments. Oxf J Bioinform 17(7):634–641
36. Verdnik D (2004) Guide to microarray analyses. GenePix Pro, MDS Analytical Technologies, Sunnyvale, CA
37. Wang Y, Marc QM, Zhang K, Shih YF (2007) A hierarchical refinement algorithm for fully automatic gridding in spotted DNA microarray image processing. Inf Sci Int J 177(4):1123–1135
38. Zacharia E, Maroulis D (2008) An original genetic approach to the fully automatic gridding of microarray images. IEEE Trans Med Imaging 27(6):805–813
39. Zacharia E, Maroulis D (2010) 3D spot-modeling for automatic segmentation of microarray images. IEEE Trans Nanobiosci 9(3):181–192
40. Zhang M, Mao K, Tao W, Tarn T (2006) A computational method to geometric measure of biological particles and application to DNA microarray spot size estimation. Med Biol Eng Compu 44:275–279
41. Zhang K, Song H, Zhang L (2010) Active contours driven by local image fitting energy. Pattern Recogn 43:1199–1206