SIViP
DOI 10.1007/s11760-015-0800-6

ORIGINAL PAPER

A generalized methodology for the gridding of microarray images with rectangular or hexagonal grid

Nikolaos Giannakeas[1,2] · **Fanis Kalatzis**[2] · **Markos G. Tsipouras**[2] · **Dimitrios I. Fotiadis**[2]

Received: 21 January 2014 / Revised: 7 July 2015 / Accepted: 8 July 2015

Abstract Microarrays provide a simple way to measure the level of hybridization of known probes of interest with one or more samples under different conditions. The rapid development of microarray technology requires the implementation of smart and flexible algorithms to deal either with the great amount of data or with the variations of the used hardware. In this paper, a generalized methodology for spot addressing and gridding of microarray images is presented. The methodology can cope with both rectangular and hexagonal grids, which are used for the probes placement onto the substrate. Initially, the methodology identifies the structure of the image, and an efficient spot-by-spot approach has been developed for the detection of all spots in the image. The evaluation of the methodology was performed using both rectangular and hexagonal structured images, merged in a single dataset. The methodology results in high accuracy in the spots detection, ranging from 92.8 to 99.8 % depending on the dataset used.

Keywords Microarray image processing · Rectangular and hexagonal grids · Microarray gridding

1 Introduction

Several types of microarrays have been developed concerning the probe sequences or the manufacture of a microarray slide. Earliest types of microarrays (cDNA microarrays) [1,2] aim to investigate the expression of genes under several conditions, using complementary DNA (cDNA). Nowadays, microarrays for detection of single nucleotide polymorphisms (SNPs) [3], protein microarrays [4], and CGH microarrays [5] have been developed. Fortunately, most of microarray types generate images with similar characteristics.

The scanned microarray is a dual-channel image, where each channel corresponds to a scanning wavelength [1]. Usually, one channel is pseudocolored red, while the other is pseudocolored green. The intensity of a spot is related to the absolute hybridization amount between the sample and the corresponding probe. The spots placement into blocks usually follows the structure of a rectangular grid (i.e., the spots are located into rows or columns). Although rectangular structure is the most common one, some manufacturers of microarray slides prefer to place spots in a hexagonal structure. The hexagonal structure of spots allows more probes to be packed in the substrate than the rectangular structure. However, hexagonal structured images require more sophisticated methods for their analysis.

Dutoit et al. [6] introduced three stages of microarray image processing. The first stage, called spot addressing and gridding, provides spot detection and isolation of each detected spot into a single cell. In the second stage, which is known as segmentation stage, each pixel in the spot cell is characterized as signal or background. Finally, in the third stage, the intensity of each spot is extracted and several quantities are calculated. Many studies investigate the whole procedure of microarray imaging, including all three stages [7,8]. However, most of the studies focus either on spot addressing [9,10] or on segmentation [11,12], considering each two stages individually. Hardware improvements reconsider both gridding and segmentation stages. Spot addressing

✉ Dimitrios I. Fotiadis
fotiadis@cc.uoi.gr

[1] Laboratory of Biological Chemistry, Medical School, University of Ioannina, 45110 Ioannina, Greece

[2] Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, PO Box 1186, 45110 Ioannina, Greece

Published online: 28 July 2015

stage, which is mainly affected by the hardware improvements, is still a field of interest for several research groups [9,10,13–15].

In this work, a spot addressing methodology that can operate in both rectangular and hexagonal structured microarray images is introduced. It provides a generalized approach to process both types of images, which can be extracted from different sources. To manage the generalization issues, a four-step methodology has been proposed. During the first step, the image blocks are separated and identification of the grid type is performed, by extracting the projection profiles of the image. A new algorithm based on watershed transform is employed, to deal with the incongruity between the projection profiles extracted from the two different structures (i.e., rectangular and hexagonal). Also, identification of the grid type (rectangular or hexagonal) is performed using the projection profiles in different angles. In the second step, high-intensity spots are detected using simple segmentation techniques. Then (third step), the growing concentric polygon (GCP) algorithm is applied, which is a generalization of the Growing Concentric Hexagon algorithm (presented in [9]), able to process both grid types. The GCP algorithm detects a number of non-hybridized spots in each iteration, with respect to different grid type images. Finally, the fourth step employs calculation of the Voronoi diagram, to isolate each spot in a single cell.

2 Related work

Proposed methods for spot addressing and gridding can be grouped into two different approaches: (a) the holistic approach and (b) the spot-by-spot approach. Holistic approaches detect the optimum linear paths of the image, in order to build the vertical and horizontal lines of the grid [14]. These methods do not detect the location of each spot in the image, but assign an area to each spot using the generated areas between the intersections of the lines. Alternatively, spot-by-spot approaches detect the position of all spots in an image and directly assign a cell for each detected spot [9]. Holistic approaches are less computationally expensive and time-consuming than spot-by-spot approaches; however, they can deal with ideal microarray images that have fundamental properties [16].

Holistic approaches can be categorized as manual, semi-automated and automated, according to the user interference. Earliest manual and semi-automated methods [17,18] require a number of parameters provided by the user, such as the number of blocks, the number of spots in each block, and the radius of each spot. Most of the automated methods extract the vertical and the horizontal projections [19–21] of an image, summing the rows and columns of pixel intensities. The projections are processed using either morphological operators [22,23] or smoothing filters [24], to provide two enhanced signals. Apart from the methods which are based on image projections, several holistic methods are based on more sophisticated techniques, such as support vector machines (SVMs) [24,25] or genetic algorithms (GA) [27]. SVMs approach extracts support vectors for each spot of two consecutive rows and then generate an optimal line between these rows, minimizing the vectors. Zacharia et al. [27] proposed the use of GAs for the optimization of the line position between two consecutive spot rows.

Spot-by-spot approaches attempt to detect the position of each spot in the image. Empty spots are generated due to the low degree of hybridization between the corresponding probe and the samples, affecting the accuracy results. Methods of this category initially detect all high-intensity spots, using simple segmentation techniques, such as histogram [28] and template matching [29–32]. Then, the positions of the empty spots are estimated, using geometrical characteristics from the already detected high-intensity spots [29,30], or applying graph models [28]. The main disadvantage of these methods is that they detect one spot in each iteration, increasing processing time of large images. Sometimes the estimated positions of non-hybridized spot are improved employing Affine [31] or Radon [33] transforms. Galinsky et al. modified their method [31], to operate in hexagonal structured image [34]. Nonetheless, these works process each grid type separately and not as a single approach. Finally, commercial applications, such as Scanalyze [17] and Genepix [18], which are embedded in microarray stations, are based on manual or semi-automated methods for gridding.

3 Materials and methods

Both real and simulated images are employed for the evaluation of the proposed methodology. Real images are collected from two different sources, in order to have images structured in both rectangular and hexagonal grid. More specifically, real images from the Stanford Microarray Database (SMD) [35] and from the CNV370 Beadchip of Illumina [36], which use rectangular and hexagonal grid structure, respectively, have been used. Real datasets are annotated, providing the spots' positions. Apart from the real images, a number of simulated images are generated using the Nykter [37] simulator. The simulator has been modified, in order to generate images structured in hexagonal grids. Apart from the differences in grid type, there are several different properties among the employed datasets such as the number of spots per block, the spot size and shape, the distances between spots and the level of noise (Fig. 1).

The evaluation dataset included 4 images from SMD, 1 image from CNV370, 6 simulated images with rectangular grids (Nykter_rec) and 6 simulated images with hexagonal

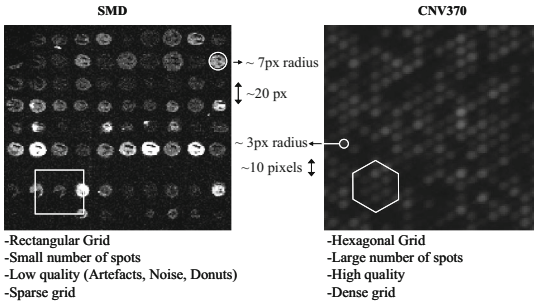

Fig. 1 Differences between SMD and CNV370 beadchip of Illumina

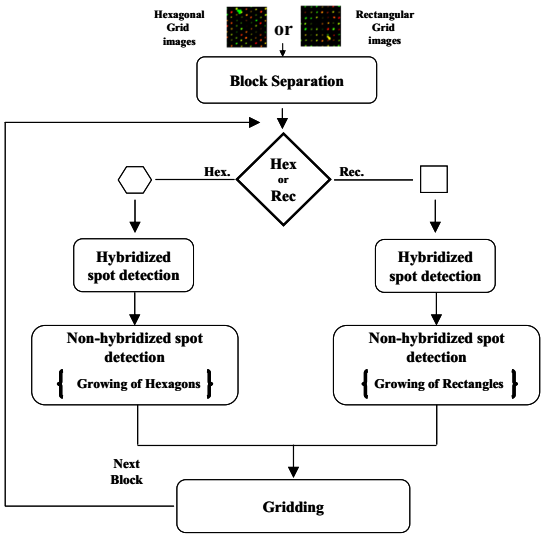

Fig. 2 Flowchart of the proposed methodology

grids (Nykter_hex), resulting to a total of 17 images. The total number of spots for the evaluation of the proposed method is approximately 720,000 spots, coming from different microarray sources.

The proposed methodology begins with the separation of the blocks, which also includes the identification of the grid type (rectangular of hexagonal). In the second step, all high-intensity objects of the image are detected, using an adaptive thresholding technique. Objects are either high-intensity spots that correspond to hybridized probes, or artifacts generated by dust or other contamination on the slide. The term "object" is used because it is not clear during this step if they are spots or artifacts. The third step of the methodology is the non-hybridized spot detection, where the positions of empty spots are estimated by the GCP algorithm. According to the identified grid type, this step grows an appropriate polygon and generates the non-hybridized spots on the polygons' contour. Finally, to isolate each spot into a cell, the gridding step employs the Voronoi diagram to the centers of the detected spots. Figure 2 presents the flowchart of the proposed methodology.

3.1 Block separation

In the first step of the proposed methodology, the image is split into individual blocks, which are individual rectangular areas in the images, and they are processed sequentially. Block separation uses the projection profiles (vertical or horizontal) of the images, which are produced by summing the intensity of pixels in rows or columns, respectively. As a result, two one-dimensional signals are extracted given by:

$$P_{\text{hor}_i} = \sum_i I_{ij} \quad \text{and} \quad P_{\text{ver}_j} = \sum_j I_{ij}, \tag{1}$$

where I_{ij} is the intensity of the ith row and the jth column pixel in the image. The processing of the projection profiles aims to: (a) detect the valleys between blocks and (b) calculate a measure that corresponds to the distance between neighboring spots. This measure will be employed by the GCP algorithm, during the non-hybridized spot detection.

One of the projections of hexagonal structured images has different characteristics than the projections of rectangular structured images. It presents shallow valleys due to the existence of spots in the intermediate positions. Due to this diversity of the projection profiles, thresholding technique fails to detect the valleys in both grid simultaneously. To avoid the thresholding, an approach which is inspired from the watershed transform [38] is proposed. Suppose that the valleys of the projections are filling with water. While the moving catchment of the water is going up, it works like a sliding threshold. At any point of the catchment, the number of valleys is known. The methodology stops to fill the valleys with water, when the water of a valley is brimmed to a neighbor valley.

Since the valleys are now full of water, the volume of the water in each valley is computed. As a result, two types of valleys are presented, the low-volume valleys and the high-volume valleys. The high-volume valleys correspond to the points between the blocks of the image, while the low-volume valleys correspond to the points between the rows and the columns of spots (Fig. 3). The wide dark path between the block in the images generates a high-volume valley in projection profiles. For the block separation task, the high-volume valleys are used. The K-means algorithm [39] is employed to group valleys into two clusters. For this reason, the volume of water is used as feature, and the number of cluster is set to 2.

After the clustering, the high-volume valleys have been identified and the image is split using their coordinates. The clustering of the valleys also identifies all low-volume valleys, which are useful to estimate a representative distance between two neighboring spots. This distance will be used for the non-hybridized spot detection step. To estimate this distance, all the distances between consecutive low-volume

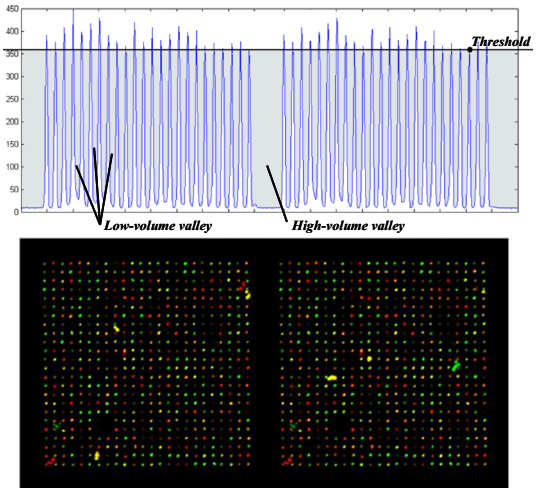

Fig. 3 Valleys in the projection profile full of water after the threshold estimation

valleys are computed, and the mode distance (*mode_dist*) is selected.

Using the projection of the image under different orientation, the grid type of the image can be recognized: The image projections for the two different grids have different characteristics in the orientations of 45° and 60°. If the image uses hexagonal grid, the projection in 60° should be a high standard deviation signal, while its 45° projection should be a smoothed signal. Both these projections of a hexagonal structured image are shown in Fig. 4. Conversely, the standard deviation of the 45° projection of a rectangular structured image is higher than the projection in 60° orientation. According to the above assumption, the grid of the image is indentified under two simple rules:

$$\begin{aligned} &\text{IF std}\left(60^\circ\right) > \text{std}\left(45^\circ\right) \text{THEN “hex}''\text{,} \\ &\text{IF std}\left(45^\circ\right) > \text{std}\left(60^\circ\right) \text{THEN “rec}''\text{.} \end{aligned} \tag{2}$$

3.2 Hybridized spot detection

The hybridized spot detection step focuses on the revealing of the high-intensity objects in the image. The detected objects are probably spots with high degree of hybridization, but it could also be artifacts. The hybridized spot detection procedure is independent from the grid type. The Otsu method [40] is used to convert the image into binary, and then, all 8-connected pixel areas are marked as objects. Thus, if two 8-neighbor pixels have been set to “1,” then they will be included in the same object. The center of the mass (X_C, Y_C) for each object is calculated as:

$$X_C = \frac{\sum_{i\in D} x_i I_i}{\sum_{i\in D} I_i}, \quad Y_C = \frac{\sum_{i\in D} y_i I_i}{\sum_{i\in D} I_i}, \tag{3}$$

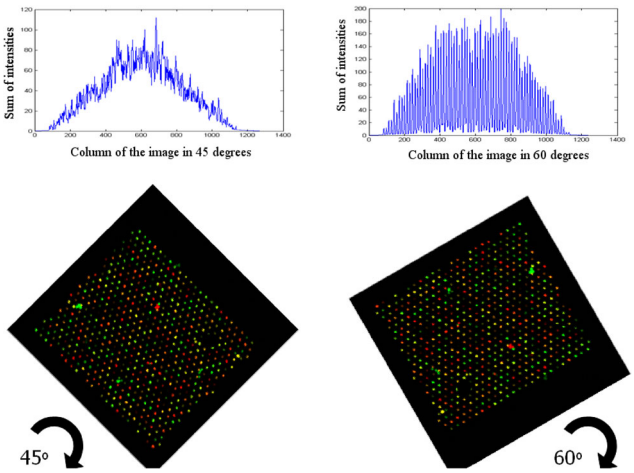

Fig. 4 Projection of a hexagonal structured microarray image in the orientations 45° and 60°

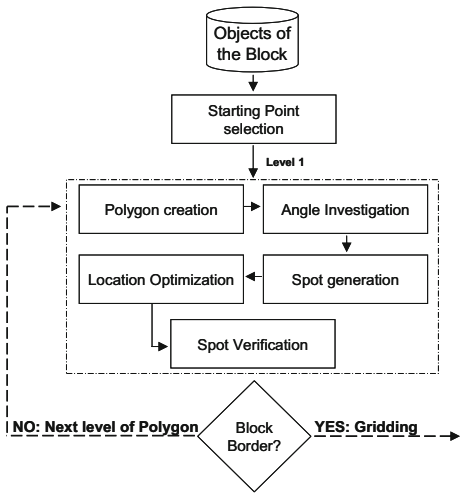

Fig. 5 Flowchart of the non-hybridized spot detection step

where x_i and y_i are the coordinates of the pixel, I_i is the intensity of the pixel in the domain of spot, and D is the domain defining the area of the object.

3.3 Non-hybridized spot detection

After the hybridized spot detection step, all high-intensity objects of the image have been found. However, a spot-by-spot approach requires the detection of all spots in the image, including the dark ones (low degree of hybridiztion). Additionally, the detected high-intensity objects must be clarified if they are spots or artifacts. The non-hybridized spot detection employs the GCP algorithm (Fig. 5). The concept of the algorithm is to identify all spots of the image (hybridized or not) on the contour of concentric polygons.

Initially, the algorithm randomly selects starting points from the centers of the already detected objects, and then appropriate polygons (rectangle or hexagon) are generated

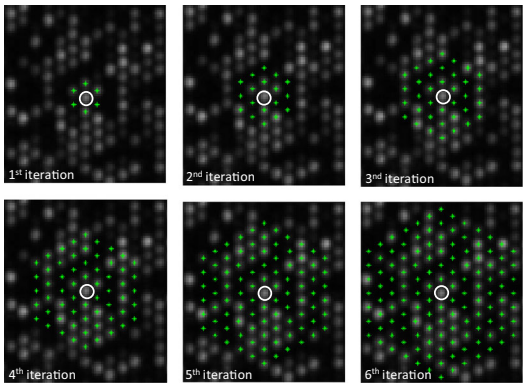

Fig. 6 Six iterations of the GCP algorithm around a starting point. High-intensity objects are verified as spots, while the positions of non-hybridized spots are estimated

around these points. Each iteration increases by one the level of the polygon, recalculating its radius and the number of spots on its contour. Most of the already detected objects (from the second step of the methodology) are verified to be on the generated polygons. All other are rejected as artifacts. An illustration of the algorithm application is shown in Fig. 6, where six iterations of the algorithm are presented.

3.3.1 Starting points selection

One or more spots are randomly selected as starting points of the algorithm. The appropriate polygon (according to the grid type identification) is grown around the selected starting point during the iterative procedure. If the number of spots in a block is extremely large, high-level polygons are required. However, high-level polygons fail to detect accurately the spots on their contour due to small distortion effect, which is frequently occurred. To address this issue, multiple starting points, which are growing simultaneously, are used. Each starting point is responsible to cover a part of the block, so that the maximum level of the polygons is drastically reduced.

Since the starting points are selected from the high-intensity objects, there is a possibility to select an artifact as starting point. If this happens, the growing concentric polygons fail to detect any other spot, due to the wrong location of the polygon's center. For this reason, the algorithm optimizes the location of all generated polygons in each iteration. More specifically, during the ith iteration of the algorithm, the already generated spots are assumed as a rigid grid of points (i.e., no independent movement for each point is allowed), and the sum of the initial intensities of these points is computed. The position of the rigid grid, which maximizes the sum of the intensities, is chosen to be the correct position.

3.3.2 Polygon creation

Each iteration of the algorithm begins with the polygon creation. According to the identification of the grid type (i.e., "rec" or "hex" in Fig. 2), the algorithm creates either a rectangle or a hexagon, respectively. Given the coordinates of the center of the polygon (which are the coordinates of the selected starting point) and the radius of the circumcircle of the polygon, the coordinates of its corners can be computed. The radius of the circumcircle is a multiple of the mode distance (*mode_dist*) between neighboring spots (computed using the low-volume valleys during the block separation step) given as:

$$R = level \times mode\_dist, \tag{4}$$

where *level* is the level of the polygon equal to the iteration of the algorithm. By generating concentric polygons around the starting point, all spots of the block are detected. The number of the spots on the polygon contour also depends only on the level of the polygon and is given as:

$$n_r = 8 \times level, \tag{5}$$

$$n_h = 6 \times level, \tag{6}$$

where n_r is the number of spots for the rectangular grid, and n_h is the number of spots for the hexagonal grid. If the generated spots of a concentric polygon lay out of the block boundaries (the boundaries of the block are known from the first step of the methodology), then they are eliminated. The iteration stops when all generated centers of a concentric polygon are out of the block boundaries.

Due to rounding errors, the computed radius for each level of the created polygon requires an extra validation. After few iterations of the algorithm, the contour of the polygon is deviated from the correct locations because the product R is few pixels smaller or larger than the real radius. This phenomenon is caused by the difference of the digitalized distance (in pixels) between two spots and the real distance (in μm) between two probes on the microarray slide. If the real distance between two neighboring spots is not a rounded but a decimal value, the accumulation of the rounding errors will affect the correctness of the radius. To deal with this issue, five values for the radius are investigated in each iteration: (i) the product R of Eq. (4), (ii) $R-1$, (iii) $R-2$, (iv) $R+1$, and (v) $R+2$, creating five different polygons. For each of these cases, the number of identified detected object (detected from the 2nd step) is countered. An object is identified, when its center and the center of a generated spot on the contour of the current polygon are both in the same 3×3 neighborhood. The radius with the maximum count is selected.

3.3.3 Verification of high-intensity objects

Finally, verification between the generated spots on contour and the detected objects from the hybridized spot detection step is utilized. The 3×3 neighborhood of each generated spot is investigated for the existence of already detected objects. If there is an object in this neighborhood, the generated spot is eliminated and the object is verified remaining on its position. In contrast, the entire set of nonverified object is eliminated as artifact.

3.4 Gridding

Since the centers of the entire set of spots have been detected, each of the spot must be isolated into a cell. The isolated spots are the input of the segmentation stage of microarray image processing. Most of the already developed methods [17,21,28,31] assign a rectangular region around each spot to define the cell. However, assigning rectangular regions around the spots in hexagonal structured images could be ineffective. In this case, a number of signal pixels, which belong to neighboring spots, probably exist into a single cell. The proposed methodology employs the Voronoi diagram [41], using as nodes the centers of the detected spots. The Voronoi diagram generates the mediator of each pair of spots, assigning the optimal area in each spot. The intersection points of these mediators are equidistance with three neighboring spots. In this way, a grid of cells is generated in the block, where each cell contains the pixels of a single spot.

4 Results

For the evaluation of the methodology, the images from the four datasets are merged in one dataset. In this way, all steps of the generalized methodology are evaluated providing results independently of the structure of the input images. Figure 7 presents the results of the methodology.

Figure 7a presents a randomly chosen block of a hexagonal structured simulated image. Figure 7b–d presents the progress of the proposed methodology, after the hybridized spot detection, the non-hybridized spot detection, and the gridding procedure, respectively.

A secure way to evaluate the spot addressing of the proposed methodology is the comparison between the positions of the spot centers, which are provided by the annotation of the image, and the positions of the centers of the spot, which are detected by the methodology. In extremely large images, such as microarray images, the accurate detection (in pixels) of an object is a very difficult task. On the other hand, the spots are too small (having diameter of some pixels), so that an error equal to 5 or 10 pixels probably results in definitely wrong detection. In addition, the center of the mass of a spot

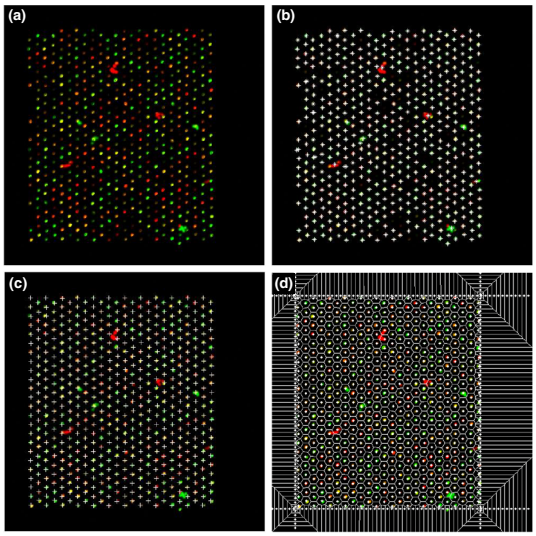

Fig. 7 Results of the proposed methodology for each step, **a** initial block after the block separation step, **b** results of the hybridized spot finding step, **c** results of the non-hybridized spot finding step, **d** results after the gridding procedure

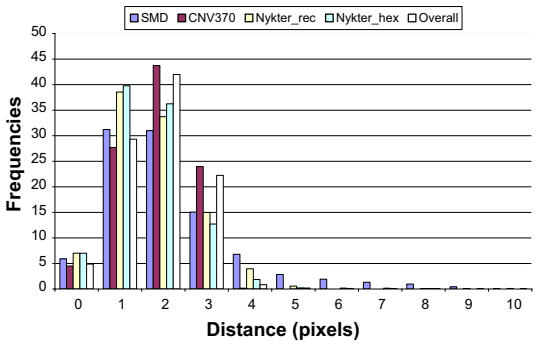

Fig. 8 Histogram of the distances between the annotated centers and the detected centers by the proposed method (for each set of images and in overall)

could be everywhere in the area of the signal pixels, according to the intensity distribution of the spot. For all these reasons, it is meaningful to present the distances between the centers of the annotation and the centers of the detected spots.

The histograms of the distances, for the merged dataset and for each dataset separately, are presented in Fig. 8, while the boxplots for the merged dataset and for each set of images are also presented in Fig. 9. The boxplot diagrams present that the estimated mean distance for all dataset is about 2 pixels. The added value of these diagrams is the presentation of the outliers. As it is shown in Fig. 9, there are several outliers of the computed distances, for the two simulated set of images, compared to the CNV370 dataset, where there is only one outlier value.

To measure the performance of the proposed methodology, the mean and the standard deviation of the distances between the center of the annotation and the detected spots are computed. The accuracy of the methodology depends

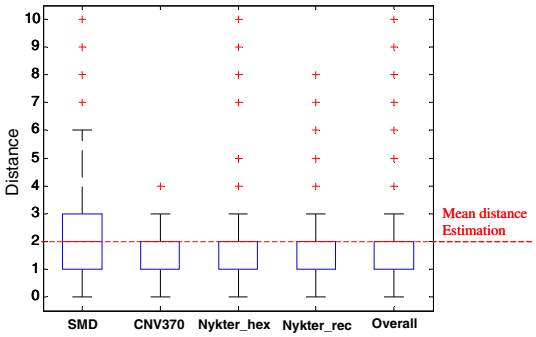

Fig. 9 Boxplots of the distances for each set of images, and in overall

on the theoretical radius of the current set of images. For real images, the theoretical radius is provided by the dataset according to the digitalization of the scanned image ($r = 3$ pixels for CNV370 and $r = 7$ pixels for SMD). It yields by the physical radius of the printed probe and the physical size that represent each pixel. For the simulated dataset, the theoretical radius is the radius that is used for the corresponding parameter of the Nykter simulator ($r = 3$ to $r = 7$ pixels). Thus, it is meaningful to consider a correct detected center, when it is detected within the area of the theoretical radius. The accuracy is given by:

$$\text{Acc} = \frac{n_{\text{correct}}(r)}{n_{\text{all}}} \times 100\,\%, \tag{7}$$

where n_{correct} is the number of the centers, which are detected in the circle area around the annotated center with radius r and n_{all} is the number of all image spots. Table 1 presents the results of the proposed methodology for all sets of images.

As it is shown in Table 1, the accuracy of the methodology for the entire dataset is 99.3 %, while the mean distance and the standard deviation of the distances between the detected and the annotated centers are 1.91 and 1.03 pixels, respectively. Processing time for each set of images is presented.

5 Discussion

A generalized methodology for the spot addressing and the gridding of microarray images is presented. For spot-by-spot methods, the distance between the center of the annotation and the detected center is the most valuable metric to measure performance; most methods presented the literature employ distance-based metrics [9,26,27,30,32]. For the evaluation of the current work, the above distance, the mean and standard deviation for all spots, as well as the accuracy metric, are computed. The result for the overall dataset is 99.3 % accuracy for the spot addressing procedure, while the mean distance is computed less than two pixels error. However, the errors are not uniformly spread over the different image datasets.

As it is shown in Fig. 9, the frequency rapidly decays for the CNV370 image after 2 pixel distance, while it slowly decays for the SMD set; thus, hexagonal structured real images are more accurately processed than rectangular ones. Furthermore, better accuracy results are obtained for the CNV370 image (99.8 %) than the SMD set of images (92.8 %). All the above are partially due to the better quality of the CNV370 image compared to the SMD set of images, as well as the smaller spot theoretical radius in CNV370 image (3 pixels) than the theoretical spot radius of the SMD set of images (5 pixels). In SMD images, both large number of artifacts and high background noise affect the quality of images. The artifacts are generated from contaminations or scratches, inner holes in spots (donuts spots), as well as from dyes or samples which have been spilled on the substrate (bleeding effect). The later (low-quality, large theoretical radius) interpreters the detection of several centers 4 or 5 pixels away from the annotated centers, affecting the accuracy of the method in SMD set of images. The accuracy results for the simulated sets of images are better for the rectangular structured images (98.9 %) than the hexagonal ones (97.5 %). Since the simu-

Table 1 Results of the proposed methodology: mean and standard deviation of the distances between the annotated and the detected centers, and accuracy of the proposed method

Images	Description	Image size (px)	# of spots per image	Processing time per image (s)	Radius	Mean (px)	STD (px)	Acc (%)
Merged dataset	Both the grids	–	–	–	–	1.91	1.03	99.3
SMD	Real, rectangular grid	1980 × 1917	9196	389	5	2.52	2.59	92.8
Nykter_rec	Simulated, rectangular grid	3188 × 9552	576,756	20,802	5 or 7	1.77	1.16	98.9
CNV370	Real, hexagonal grid	1832 × 936−2800 × 2800	4608–9216	198–821	3	1.88	0.82	99.8
Nycter_hex	Simulated, hexagonal grid	1832 × 936−2800 × 2800	4608–9216	186–795	5 or 7	1.94	2.32	97.5

Table 2 Review of existing methods, employed datasets, and evaluation metrics

	Work	Description	Dataset	Metric
1	Jung et al. [28]	Gridding based on K-nearest neighbor	10 artificial, 3 from NIH, 3 from SMD	[a]
2	Galinsky [34]	Incomplete Voronoi diagram	Several images	[a]
3	Bajcsy [19]	Gridline: automatic grid alignment	–	[a]
4	Ceccarally et al. [33]	Deformable grid-matching approach	Artificial images, 6 randomly chosen images from SMD	MSE = 1.53−5.03 pixels (MSE of center distances)
5	Zacharia et al. [27]	Gridding using generic algorithms	25 randomly chosen images from SMD	Acc = 94.6 % (correct detected spot within the cell area)
6	Bariamis et al. [26]	Gridding based on support vector machines	54 randomly chosen images from SMD	Acc = 96.4 % (correct detected spot within the cell area)
7	Proposed method	A generalized approach for both grid types	SMD, Beadchip CNV370, simulated images.	Acc = 99.3 % (correct detected spot within the circle area of radius r)

[a] The authors present only qualitative results

lated sets of images are generated using the same parameters (concerning the quality of the image), the variation in the performance could reveal a difficulty of processing hexagonal than rectangular structure.

Table 2 summarizes several methods for microarray image analysis. Although a direct comparison is not feasible, since different datasets and evaluation metrics have been employed in each study, the proposed methodology compares well with the other methods presented in the literature. The proposed method presents higher accuracy results than the other methods, about 3–5 percentage units.

A direct comparison between the proposed methodology and two previous works has been performed, employing the same set of images. The first work [30] presents a spot-by-spot approach for the spot addressing of rectangular structured images (requires the placement of the spot in the image to be in rows and columns). The second work [9,42] presents a preliminary version of the proposed one, which has been developed to deal with hexagonal structured images. The concept of this comparison is to investigate the performance of a generalized methodology versus two other methods developed for specific grid type.

As it is shown in Table 3, the obtained results of the proposed methodology are similar to the results of each of the specialized method, for all sets of images, except of SMD. The results of the generalized methodology for the SMD set of images present lower accuracy than the specialized method for the rectangular grid type image. Furthermore, in CNV370 dataset, the accuracy results are slightly improved, while the mean distance and the standard deviation are decreased. This could be explained due to the increased number of outliers, which are detected 4 pixels away from the annotated center, as it is shown also in Fig. 9.

Table 3 Comparative results between the proposed methodology and two previous works

Images	Metric	[30]	[9]	Proposed methodology
SMD	Mean	1.35	–	2.52
	Std	1.30		2.59
	Acc	98.7		92.8
Nykter_rec	Mean	1.74	–	1.77
	Std	1.03		1.17
	Acc	99.3		99.0
CNV370	Mean	–	1.10	1.88
	Std		0.59	0.82
	Acc		99.5	99.8
Nykter_hex	Mean	–	2.00	1.94
	Std		2.30	2.32
	Acc		97.2	97.5

6 Conclusions

Two significant requirements must be addressed for a microarray image processing software. The first is related to the variety of the microarray experiments, while the second is associated with the continuously increasing amount of the extracted data. The variety of approaches in microarray technology requires generalized methods for image processing and data analysis. In this direction, the proposed methodology recognizes two different types of printing and has the ability to perform spot addressing for both of them. The proposed methodology faces also the increasing number of spots, since it is based on an algorithm for the detection of the non-hybridized spots, which can be easily parallelized.

Acknowledgments This work is part funded by the European Commission (POCEMON Project, FP7-ICT-2007-216088).

References

1. Schena, M., Shalon, D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270**, 467–470 (1995)
2. Eisen, M.B., Brown, P.O.: DNA arrays for analysis of gene expression. Methods Enzymol. **303**, 179–205 (1999)
3. Shen, R., Fan, J.B., Campbell, D., Chang, W., Chen, J., Doucet, D., Yeakley, J., Bibikova, M., Wickham-Garcia, E., McBride, C., Steemers, F., Garcia, F., Kermani, B.G., Gunderson, K., Oliphant, A.: High-throughput SNP genotyping on universal bead arrays. Mutat. Res. **573**, 70–82 (2005)
4. MacBeath, G., Stuart, L.: Schreiber printing proteins as microarrays for high-throughput function determination. Science **289**, 1760–1763 (2000)
5. Shinawi, M., Cheung, S.W.: The array CGH and its clinical applications. Drug Discov. Today **13**, 760–770 (2008)
6. Dudoit, S., Yang, Y.H., Callow, M.J., Speed, T.P.: Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat. Sin. **12**, 111–139 (2002)
7. Beleana, B., Bordaa, M., Galc, B.L., Terebesa, R.: FPGA based system for automatic cDNA microarray image processing. Comput. Med. Imaging Graph. **36**(5), 419–429 (2012)
8. Athanasiadis, E.I., Cavouras, D.A., Spyridonos, P.P., Glotsos, D.T., Kalatzis, I.K., Nikiforidis, G.C.: Complementary DNA microarray image processing based on the fuzzy Gaussian mixture model. IEEE Trans. Inf. Technol. Biomed. **13**(4), 419–425 (2009)
9. Giannakeas, N., Kalatzis, T., Fotiadis, D.I.: Spot addressing for microarray images structured in hexagonal grids. Comput. Methods Programs Biomed. **106**(1), 1–13 (2012)
10. Shao, G., Yang, F., Zhang, Q., Zhou, Q., Luo, L.: Using the maximum between-class variance for automatic gridding of cDNA microarray images. IEEE/ACM Trans. Comput. Biol. Bioinform. **PP**(99), 1–10 (2012)
11. Harikiran, J., Rama-Krishna, D., Phanendra, M.L., Lakshmi, P.V., Kiran -Kumar, R.: Fuzzy C-means with Bi-dimensional empirical mode decomposition for segmentation of microarray image. IJCSI **9**(5), 316–321 (2012)
12. Weng, G., Hu, Y., Li, Z.: cDNA microarray image segmentation using shape-adaptive DCT and K-means clustering. In: International Conference in Electrics, Communication and Automatic Control, pp. 317–324 (2012)
13. Liu, J., Feng, Y., Liu, W., Wang, T.: A microarray image gridding method based on image projection difference sequences analysis and local extrema searching. In: 10th World Congress on Intelligent Control and Automation, pp. 4961–4964 (2012)
14. Yao, Z., Shunxiang, W.: Statistics-adaptive method for cDNA microarray images gridding. In: 4th International Conference on Digital Home, pp. 380–383 (2012)
15. Labib, F.E.-Z., Fouad, I., Mabrouk, M., Sharawy, A.: An efficient fully automated method for gridding microarray images. Am. J. Biomed. Eng. **2**(3), 115–119 (2012)
16. Schena, M.: Microarray Biochip Technology. Eaton Publishing, Natick (2000)
17. Eisen, M.B.: ScanAlyse. http://rana.Stanford.EDU/software/ (1999)
18. Fielden, M.R., Halgren, R.G., Dere, E., Zacharewski, T.R.: GP3: GenePix post-processing program for automated analysis of raw microarray data. Bioinformatics **18**, 771–773 (2002)
19. Bajcsy, P.: Gridline: automatic grid alignment in dna microarray scans. IEEE Trans. Image Process. **13**, 15–25 (2004)
20. Blekas, K., Galatsanos, N., Likas, A., Lagaris, I.E.: Mixture model analysis of DNA microarray images. IEEE Trans. Med. Imaging **24**(7), 901–909 (2005)
21. Jain, A.N., Tokuyasu, T.A., Snijders, A.M., Segraves, R., Albertson, D.G., Pinkel, D.: Fully automated quantification of microarray image data. Genome Res. **12**, 325–332 (2002)
22. Hirata, R., Barrera, J., Hashimoto, R.F., Dantas, D.: Microarray gridding by mathematical morphology. In: Proceedings of the Ijth Brazilian Symposium on Computer Graphics and Image Processing, pp. 112–119 (2001)
23. Bengtsson, A., Bengtsson, H.: Microarray image analysis: background estimation using quantile and morphological filters. BMC Bioinformat. **7**, 96–105 (2006)
24. Lonardi, S., Yu, L.: Gridding and compression of microarray images. In: Proceedings of IEEE Computational Systems Bioinformatics Conference-Workshops (CSBW'05), pp. 122–130 (2004)
25. Bariamis, D., Maroulis, D., Iakovidis, D.K.: Automatic DNA microarray gridding based on support vector machines. In: The Proceedings of 8th IEEE International Conference on BioInformatics and BioEngineering, pp. 1–5 (2008)
26. Bariamis, D., Maroulis, D., Iakovidis, D.K.: Unsupervised SVM-based gridding for DNA microarray images. Comput. Med. Imaging Graph. **34**, 418–425 (2010)
27. Zacharia, E., Maroulis, D.: An original genetic approach to the fully-automatic gridding of microarray images. IEEE Trans. Med. Imaging **27**, 805–813 (2008)
28. Jung, H.-Y., Cho, H.-G.: An automatic block and spot indexing with k-nearest neighbors graph for microarray image analysis. Bioinformatics **18**, S141–S151 (2002)
29. Giannakeas, N., Fotiadis, D.I., Politou, A.S.: An automated method for gridding in microarray images. In: 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5876–5879 (2006)
30. Giannakeas, N., Fotiadis, D.I.: An automated method for gridding and segmentation Of cDNA microarray images. Comput. Med. Imaging Graph. **33**, 40–49 (2009)
31. Galinsky, V.L.: Automatic registration of microarray images. I. Rectangular grid. Bioinformatics **19**, 1824–1831 (2003)
32. Steinfath, M., Wruck, W., Seidel, H., Lehrach, H., Radelof, U., O'Brien, J.: Automated image analysis for array hybridization experiments. Bioinformatics **17**, 634–641 (2001)
33. Ceccarelli, M., Antoniol, G.: A deformable grid-matching approach for microarray images. IEEE Trans. Image Process. **15**, 3178–3188 (2006)
34. Galinsky, V.L.: Automatic registration of microarray images. I. Hexagonal grid. Bioinformatics **19**, 1832–1836 (2003)
35. Gollub, J., Ball, C.A., Binkley, G., Demeter, K., Finkelstein, D.B., Hebert, J.M., Hernandez-Boussard, T., Jin, H., Kaplper, M., Matese, J.C., Schroeder, M., Brown, P.O., Botstein, D., Sherlock, G.: The Stanford Microarray Database: data access and quality assessment tools. Nucleic Acids Res. **31**(1), 94–96 (2003)
36. Ionita-Laza, I., Rogers, A.J., Lange, C., Raby, B.A., Lee, C.: Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics **93**(1), 22–26 (2009)
37. Nykter, M., Aho, T., Ahdesmäki, M., Ruusuvuori, P., Lehmussola, A., Yli-Harja, O.: Simulation of microarray data with realistic characteristics. BMC Bioinformat. **7**, 349–365 (2006)
38. Roerdink, J.B.T.M., Meijster, A.: The watershed transform: definitions, algorithms and parallelization strategies. Fundamenta Informaticae **41**, 187–228 (2000)
39. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley. University of California Press, vol. 1, pp. 281–297 (1967)
40. Otsu, N.: A threshold selection method for gray-levels histograms. IEEE Trans. Pattern Anal. Mach. Intell. **13**, 583–598 (1979)

41. Aurenhammer, F.: Voronoi diagrams—a survey of a fundamental geometric data structure. In: Proceedings of ACM Computing Surveys, vol. 23, pp. 345–405 (1991)
42. Kalatzis, F.G., Giannakeas, N., Exarchos, T.P., Lorenzelli, L., Adami, A., Decarli, M., Lupoli, S., Macciardi, F., Markoula, S., Georgiou, I., Fotiadis, D.I.: Developing a genomic-based point-of-care diagnostic system for rheumatoid arthritis and multiple sclerosis. In: 31st Annual International Conference of IEEE Engineering in Medicine and Biology Society, pp. 827–830 (2009)