cDNA Microarray Image Segmentation Using Root Signals

Rastislav Lukac, Konstantinos N. Plataniotis

Multimedia Laboratory, The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada

Received 23 February 2006; accepted 10 August 2006

ABSTRACT: A vector processing based framework suitable for cDNA microarray image segmentation is introduced and analyzed in this paper. By using nonlinear, generalized selection vector filters the framework proposed here classifies the cDNA image data as either microarray spots or image background. The solution converges to a root signal that represents the segmented cDNA microarray image with the regular spots ideally separated from the background and with their coloration uniquely described by dominant color vectors. It will be demonstrated that the framework readily unifies image denoising, enhancement, data normalization, irregular spot rejection, and spot segmentation in one processing step delivering excellent performance at reasonable computational cost. Int J Imaging Syst Technol, 16, 51–64, 2006; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/ima.20067

Key words: cDNA microarray images; microarray analysis; image segmentation; vectorial approach; nonlinear image processing; filter design; root signals; convergence property

I. INTRODUCTION

Recent technological advances have allowed for the combination of various biological, medical, and computational approaches and their application to the field of computational biology, genomic engineering, and bioinformatics. Complementary deoxyribonucleic acid (cDNA) microarray imaging (Fig. 1) is considered one of the most important and powerful technologies used to extract and interpret genomic information (Arena et al., 2002; Whitchurch, 2002; Zhang et al., 2002). Analysis of cDNA microarray data helps in monitoring the expression levels of thousands of genes simultaneously (Eisen and Brown, 1999; Nagarajan and Upreti, 2006) and provides information relevant to cell activity (Damiance et al., 2004). Because of the parallel processing feature and effectiveness of their analysis, cDNA microarrays have found applications in gene and drug discovery, toxicological research, and (cancer, diabetes, and genetic) disease diagnosis (Zhang et al., 2002; Istepanian, 2003).

The cDNA microarray experiment (Fig. 2) requires first to isolate ribonucleic acid (RNA) from both control and experimental sample. In the sequence, the extracted RNAs are converted into cDNAs by the so-called reverse transcription process (Whitchurch, 2002). Using a Cy3/Cy5 system (Nagarajan, 2003), the procedure continues by labeling the cDNAs with fluorescent probes, usually Cy3 for the control and Cy5 for the experimental channel. The fluorescent targets are hybridized to the microarray[1], heated at 65°C, and washed for 16–24 h. Using a specialized laser confocal microscope and a photomultipler tube detector, cDNA microarrays are scanned (Fig. 3) at the ~540 nm (green) for the control (Cy3) and ~630 nm (red) for the experimental channel (Cy5). The scanning procedure produces two monochromatic images, which are further registered into a two-channel, Red–Green image similar to the one depicted in Figure 2. The generated cDNA microarray image is a multichannel vector signal, which can be represented, for storage or visualization purposes, as the RGB color image with a zero blue component (Lukac et al., 2004b).

The image spots' coloration represents the abundance of hybridized RNA in the array (Whitchurch, 2002; Lukac et al., 2004b). The presence of RNA from the experimental (test) or control (reference) population of cells is determined by the red or green spots, respectively. The occurrence of yellow spots suggests that RNAs from both experimental and control population contribute to the abundance while nonfluorescent (black) spots denote no binding of RNA. Based on this simple coloration concept, cDNA microarray based gene expression analysis uses the measurement of the hybridized RNA abundance as a measure of gene expression activity (Katzer et al., 2003). The vast amount of data and calculations needed to obtain the relative expression levels of the genes from the fluorescence intensity at each spot necessitates the development of automated data processing solutions (Filkov et al., 2002; Whitchurch, 2002). Since a number of impairments[2] may affect the cDNA

Correspondence to: Rastislav Lukac; E-mail: lukacr@ieee.org, web: http://www.dsp.utoronto.ca/∼lukacr

[1] cDNAs are spotted on a substrate (mostly glass slide) by an automated robot. Arrays of cDNA spots, usually up to 80 000 probes per 2×4 cm^2 area, are commonly referred to as microarrays (Arena et al., 2002).

[2] Microarray images exhibit variations in intensity due to noise impairments which can be attributed to (Wang et al., 2003; Lukac et al., 2005a) (i) variations in the image background, (ii) variations in the spot sizes and positions, (iii) artifacts caused by laser light reflection and dust on the glass slide, and (iv) photon and electronic noise introduced during scanning.

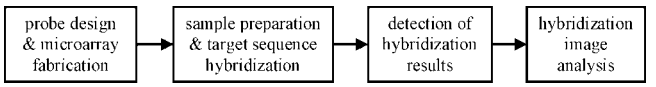

Figure 1. Block diagram representation of the cDNA microarray experiment.

microarray image formation, image processing is necessary in order to eliminate processing errors from propagating further down the processing pipeline to the gene expression analysis tasks (Zhang et al., 2002; Istepanian, 2003; Lukac et al., 2004b).

Most cDNA image processing applications deal with the noise removal and the enhancement of the cDNA microarray images. A number of filtering and enhancement techniques, such as the wavelet-based approach (Wang et al., 2003) and the vector processing approaches (Lukac et al., 2004b, 2005a,b) have been proposed recently. The multichannel nature of the cDNA microarray images led to the introduction of vector processing techniques based on fuzzy logic principles (Lukac et al., 2005a), which remove noise while preserving the structural cDNA image information. The edge-detector schemes (Kim et al., 2001; Lukac and Plataniotis, 2006) can be used to localize the spots in microarray images, while another approach (Lukac et al., 2004b) unifies low-level and high-level image processing operations to form a solution capable of removing noise and localizing spots in a single processing cycle. Other recently proposed microarray imaging solutions focus on data normalization (Kim et al., 2002; Wang et al., 2002; Park et al., 2003), background separation (O'Neill and Magoulas, 2003), and grid adjustment (Bajcsy, 2004).

Image processing techniques such as those mentioned above are usually used before the image segmentation step, which is considered the most important processing operation to be performed prior the determination of the gene expressions. A multitude of cDNA

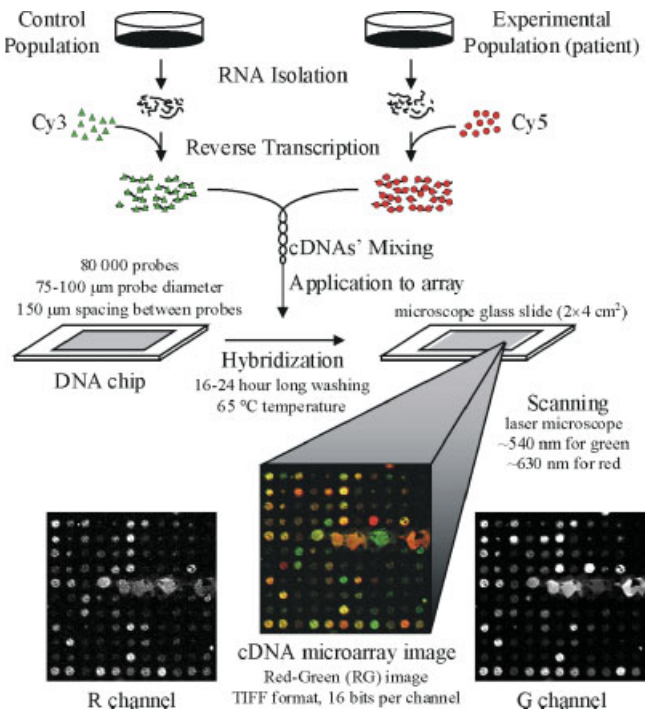

Figure 2. cDNA microarray technology. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

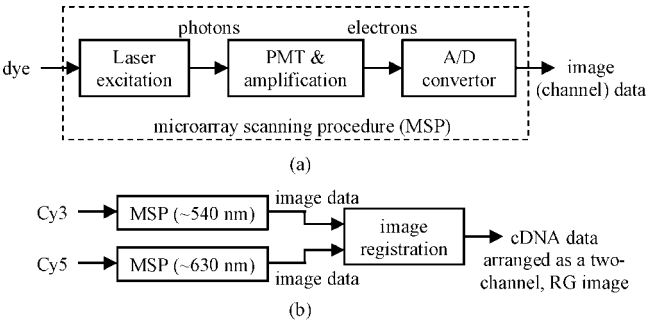

Figure 3. Microarray image scanning: (a) a scanning pipeline is constructed using the laser, photomultipler tube (PMT) detector, and A/D convertor; (b) the scanning procedure is performed for both Cy3 and Cy5 dyes to obtain Red–Green (RG) cDNA microarray image.

image segmentation solutions are available. The cellular neural network scheme (Arena et al., 2002; Zhang et al., 2002) segments the spots by performing a number of operations such as background clean-up, grid analysis, irregular spot elimination, and intensity analysis. The dynamical system modeling based approach (Damiance et al., 2004) performs pixel clustering operations in a parallel manner to speed-up the segmentation process. The combination of Markov random field based grid segmentation and active contour modeling constitutes an approach suitable for spot detection and segmentation (Katzer et al., 2003). The morphology based approach (Hirata et al., 2002) uses a series of optional (correction of rotation) and mandatory (subarray gridding, spot gridding, correction of spot gridding, spot segmentation) steps to segment the microarray image. The two-stage clustering based approach (Nagarajan, 2003) is comprised of spots' boundaries adjusting and intensity-based partitioning operations. The use of adaptive thresholding and statistical intensity modeling is the base for some segmentation schemes (Liew et al., 2003), whereas another approach (Yang et al., 2002) uses a seeded region growing algorithm to identify spots of different shapes and sizes. Histogram and thresholding operations were used to classify microarray image samples into either foreground (spots) or background pixels (Cheng et al., 1997). Finally, correlation of the pixels comprising a microarray spot was used as the segmentation criterion in Nagarajan and Upreti (2006).

In this paper, a unique cDNA image segmentation framework is introduced. The paper extends the preliminary results presented in the conference publication (Lukac and Plataniotis, 2005). The segmentation is performed using root signals obtained through the utilization of a nonlinear filtering operator. Following the multichannel nature of the cDNA image data, the framework uses a new class of generalized selection vector filters (Lukac et al., 2004a) to process the cDNA pixels. For the first time ever in cDNA microarray imaging both the magnitude and directional characteristics of the cDNA vectorial inputs are simultaneously used during processing. Because of the selective nature[3] of the operator the filtering process converges within a few iterations to a root signal, which is not further affected by the processing filter. The repetitive use of an operator capable of normalizing the data population emphasizes the most dominant cDNA vectors in localized neighborhoods. Thus, the

[3] Selection vector filters operate by outputting one of the vectorial inputs inside a supporting window as the result of the filtering operation (Lukac et al., 2004a). In this way, they preserve structural and spectral information in the image and avoid outliers and artifacts in the localized image area.

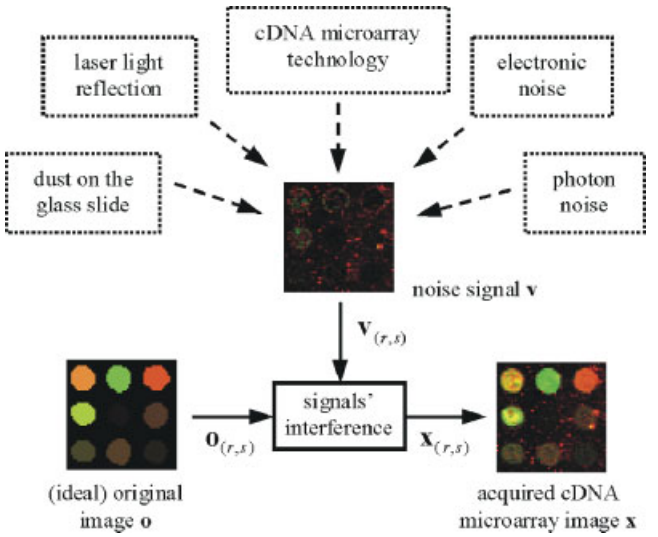

Figure 4. Block diagram representation of the noise process formation in cDNA imaging. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

generated root signal represents the segmented microarray image with the regular spots ideally separated from the background. Moreover, during its convergence phase the proposed solution simultaneously denoises, enhances, normalizes data, rejects irregular spots, and automatically segments spots from the background. Combined with an additional module which thresholds the magnitude of the root signal, the framework can remove residual, irregular foreground information and enhanced perceived and measured differences between foreground and background information in the segmented image. It will be demonstrated that the uniqueness of the root signals, the unification of a number of cDNA microarray image processing operations, and the computational efficiency of the procedure make the proposed framework an attractive proposition for the microarray image processing pipeline.

The rest of the paper is organized as follows. In Section II, the cDNA microarray imaging basics are discussed. Both the image formation and the noise corrupting process are investigated from a microarray image processing point of view. A root signal based vector processing paradigm suitable for cDNA microarray image segmentation is introduced in Section III. Motivation and design characteristics are discussed in detail. Selected solutions designed within the proposed framework are tested using a variety of cDNA microarray images in Section IV. Comparisons, in terms of performance, with the prior art in cDNA image segmentation are also provided. Finally, conclusions are drawn in Section V.

II. cDNA MICROARRAY IMAGING

Let us consider a $K_1 \times K_2$ image $\mathbf{x} : Z^2 \rightarrow Z^2$ representing a two-dimensional matrix of two-component samples $\mathbf{x}_{(r,s)} = [x_{(r,s)1}, x_{(r,s)2}]$, with $r = 1,2,\ldots,K_1$ and $s = 1,2,\ldots,K_2$ denoting the image rows and columns, respectively. Using cDNA microarray technology (Fig. 2), the microarray data form a two-channel, Red–Green (RG) image. The component $x_{(r,s)1}$ indicates the R channel while $x_{(r,s)2}$ indicates the G channel. The two channels are combined to form the cDNA vector $\mathbf{x}_{(r,s)}$ in a two-dimensional vector space (Lukac et al., 2004b, 2005b).

To visualize or store cDNA image data in the familiar RGB color format, the introduction of a zero B component is needed (Lukac et al., 2004b). According to the trichromatic theory of color vision, an arbitrary color is matched by superimposing appropriate amounts of three primary colors (Plataniotis and Venetsanopoulos, 2000) and thus, the cDNA vector $\mathbf{x}_{(r,s)}$ can be considered as point $[x_{(r,s)1}, x_{(r,s)2}, 0]$ in a three-dimensional RGB vector space (Lukac et al., 2004b). Since cDNA microarray technology employs the two (R and G) color base, the two-channel representation of cDNA microarray data should be used in both image processing and gene expression analysis.

Following the modeling scenario shown in Figure 4, the vector $\mathbf{o}_{(r,s)} = [o_{(r,s)1}, o_{(r,s)2}]$ represents the original, noise-free cDNA signal while $\mathbf{v}_{(r,s)} = [v_{(r,s)1}, v_{(r,s)2}]$ is used to denote the various image impairments introduced during processing. The vectorial samples $\mathbf{v}_{(r,s)}$ are considered random in nature and can be modeled through the additive noise model (Nagarajan, 2003). In this paper, the noise signal $\mathbf{v}_{(r,s)}$ is considered either impulsive in nature or it can be modeled as mixed noise—i.e., white additive Gaussian noise followed by impulsive noise (Lukac et al., 2004b). Therefore, the acquired cDNA signal $\mathbf{x}_{(r,s)} = [x_{(r,s)1}, x_{(r,s)2}]$ can be expressed as follows:

$$\mathbf{x}_{(r,s)} = \mathbf{o}_{(r,s)} + \mathbf{v}_{(r,s)} \tag{1}$$

where (r, s) indicates the spatial position of the samples in the image array. In conclusion, based on the microarray imaging basics demonstrated in Figures 2–4 it can be claimed that the microarray image formation is a complicated, nonlinear process influenced by many factors.

As with any vectorial signal, each acquired cDNA sample $\mathbf{x}_{(r,s)}$ is uniquely determined by its magnitude $M_{(r,s)} = \|\mathbf{x}_{(r,s)}\| = \sqrt{(x_{(r,s)1})^2 + (x_{(r,s)2})^2}$ and direction $D_{(r,s)} = \frac{1}{\|\mathbf{x}_{(r,s)}\|}\mathbf{x}_{(r,s)} = \frac{1}{M_{(r,s)}}\mathbf{x}_{(r,s)}$ in the vector space (Lukac et al., 2005b). Because of the various image impairments listed in Figure 4, the cDNA vector fields exhibit considerable variations in intensity. As shown in Figure 5, the data variations affect both the magnitude and directionality of the cDNA vectors. Since the noisy samples deviate from other samples in a given data population, the determination of the outlying cDNA vectors is of a paramount importance in the proposed here cDNA image segmentation framework, which uses uniformity in the characteristics of the cDNA vectors as the base for segmentation. To effectively quantify differences among cDNA vectors the microarray image processing operator should take into consideration both the magnitude and the orientation of the cDNA samples. Although a number of measures can be used to complete the task, the magnitude difference between two vectorial inputs $\mathbf{x}_{(i,j)} = [x_{(i,j)1}, x_{(i,j)2}]$ and $\mathbf{x}_{(g,h)} = [x_{(g,h)1}, x_{(g,h)2}]$, for $\mathbf{x}_{(i,j)} \in \mathbf{x}$ and $\mathbf{x}_{(g,h)} \in \mathbf{x}$, is usually evaluated through the Euclidean metric as follows (Lukac et al., 2005b):

$$d(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)}) = \left(\sum_{k=1}^{2} (x_{(i,j)k} - x_{(g,h)k})^2 \right)^{\frac{1}{2}} \tag{2}$$

The difference in orientation (vector directionality) is usually quantified through the utilization of the angular measure (Lukac et al., 2005b):

$$\mathrm{A}(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)}) = \arccos\left(\frac{\mathbf{x}_{(i,j)} \cdot \mathbf{x}_{(g,h)}}{\|\mathbf{x}_{(i,j)}\| \|\mathbf{x}_{(g,h)}\|} \right) \tag{3}$$

It should be noted that the determination of the similarity (or difference) between two vectorial signals depends critically on the

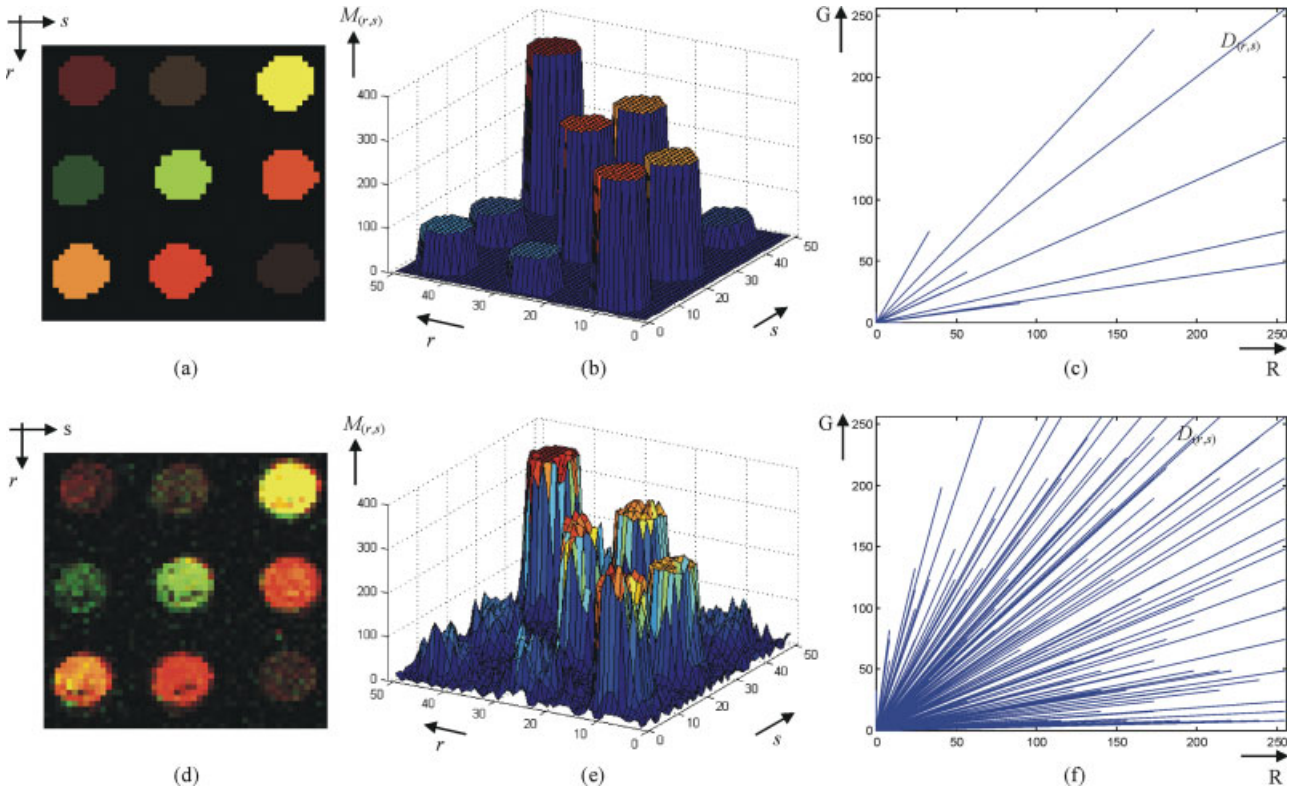

Figure 5. cDNA image characteristics: (a–c) before and (d–f) after the interference of the cDNA microarray image data with noise. (a, d) Image area of interest, (b, e) magnitude characteristics $M_{(r,s)}$, (c, f) directional characteristics $D_{(r,s)}$. The figure follows the scenario shown in Fig. 4. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

measure used to evaluate it. The interested reader should refer to Lukac et al. (2005b), Lukac and Plataniotis (2006), and Plataniotis and Venetsanopoulos (2000) for a more comprehensive treatment of the topic. In the sequel, both Eqs. (2) and (3) will be used to quantify differences between cDNA vectorial inputs and will form the base for the construction of a new cDNA image segmentation framework.

III. PROPOSED cDNA IMAGE SEGMENTATION FRAMEWORK

Image segmentation refers to partitioning an image into different regions that are homogeneous with respect to some image feature (Plataniotis and Venetsanopoulos, 2000; Yang et al., 2002; Nagarajan, 2003). Assuming that the pixels with the same feature characteristics constitute meaningful regions, and certainly this is the case with the spots present in a cDNA microarray image, the problem reduces to pixel classification. Under the ideal conditions depicted in Figures 5a–5c, with the original, noise-free cDNA image samples $\mathbf{o}_{(r,s)}$ of Eq. (1) shown in Figure 5a, each spot has its own unique magnitude (Fig. 5b) and directional (Fig. 5c) characteristics. Therefore, based on the definitions given in Eqs. (2) and (3), we may consider uniformity in vector magnitude and directionality as the criterion for partitioning the cDNA vector field into disjoint regions corresponding to distinguishable spots.

Hence, the cDNA microarray image segmentation process can be defined as follows (Lukac and Plataniotis, 2005):

$$\mathbf{x}_{(r,s)} \in \mathbf{x} \rightarrow \{F, B\} \tag{4}$$

where disjoint sets F and B denote the foreground and background cDNA vectors $\mathbf{x}_{(r,s)}$ in the microarray image $\mathbf{x}$, respectively. As shown in Figure 6, the foreground is constituted by microarray spots. A typical spot in a microarray image has a circular shape and contains ~150–200 cDNA vectors $\mathbf{x}_{(r,s)}$ (Wang et al., 2001). A gap between spots or alternatively the presence of cDNA vectors residing outside spots areas constitute the background. By extracting the

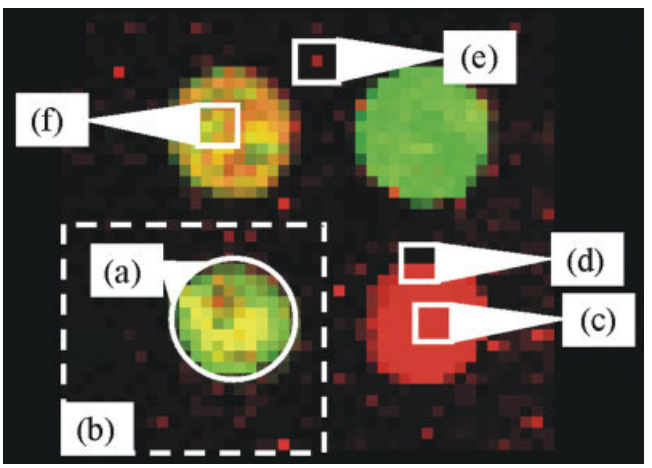

Figure 6. Localized area of the cDNA microarray image formed by (a) pixels inside the circular objects denoting the microarray spots, (b) other nonspotted pixels that belong to the image background. Further inspection reveals basic image signal structures most commonly present in the cDNA microarray images: (c) multichannel constant region, (d) multichannel step edge, (e) multichannel impulse, (f) multichannel oscillation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

spots from the microarray image $\mathbf{x}$, the background can be viewed as a homogeneous region, while the essential foreground should remain heterogonous as a result of the variable spots' coloration.

To develop an effective and automated segmentation solution, we introduce a new framework, which uses root signals obtained via filtering to segment the cDNA image into disjoint areas. The root signal of the cDNA microarray image is an image obtained from the input by repeatedly filtering it until no more changes occur (Burian and Kuosmanen, 2002). Equivalently, root signals can be defined as signals invariant to further processing by the same filtering operator (Astola et al., 1987). The utilization of vector selection filters in the framework eliminates redundant information such as impulses and noise-like variations in the sample population, and converges to a root signal that retains the spatial and spectral characteristics of the input cDNA image. Since root signals consist solely of edges and sequences of identical samples (Astola et al., 1987), the generated filter roots represent a segmented cDNA image.

A. Image Filtering Basics and Root Signals. The filters use a supporting window to process information in a localized area of the microarray image. The window, defined as $\Psi_{(r,s)} = \{\mathbf{x}_{(i,j)}; (i,j) \in \zeta\}$, for $r = 1,2,\ldots,K_1$ and $s = 1,2,\ldots,K_2$, slides over the entire image $\mathbf{x}$, placing, successively, every pixel at the center of a local neighborhood denoted by ζ. The procedure replaces the cDNA vector $\mathbf{x}_{(r,s)}$ located at the window center (r, s) with the output $\mathbf{y}_{(r,s)} = f(\Psi_{(r,s)})$ of a filter function $f(\cdot)$ operating over the samples listed in $\Psi_{(r,s)}$. As shown in Figure 7, the processing window may vary in shape. The type of window determines both the area of support and the overall performance of the procedure. The concept and the properties of the sliding (running) window are discussed in detail in Lukac et al. (2005b). Because of its versatility and demonstrated good performance, the 3×3 rectangular shape window (Fig. 7b) is the most commonly used in image processing and the one to be utilized throughout this paper.

It is well-known that any processing filter can be implemented either in a non-recursive (conventional) or recursive format. For a 3×3 non-recursive operator (Fig. 7b) the localized neighborhood $\Psi_{(r,s)}$ is constituted from the input cDNA vectors $\mathbf{x}_{(i,j)}$ for $(i,j) \in \zeta$ as defined earlier. In the case of a 3×3 recursive operator (Fig. 8b), the set ζ' denotes locations in the output (filtered) image $\mathbf{y}$ whereas other set ζ, with $\zeta' \cap \zeta = \emptyset$, denotes the locations in the input (being processed) image $\mathbf{x}$. Thus, $\Psi_{(r,s)} = \{\mathbf{y}_{(g,h)}, \mathbf{x}_{(i,j)}; (g,h) \in \zeta', (i,j) \in \zeta\}$ consists of the input cDNA vectors $\mathbf{x}_{(i,j)}$ and vectors $\mathbf{y}_{(g,h)}$ outputted by the processing filter in the previous centers of the sliding window. Recursive filters have usually provided better smoothing compared to non-recursive filters at the expense of increased distortion (Burian and Kuosmanen, 2002). To keep the subsequent discussion simple, the conventional, non-recursive filtering formulation is adopted in the paper.

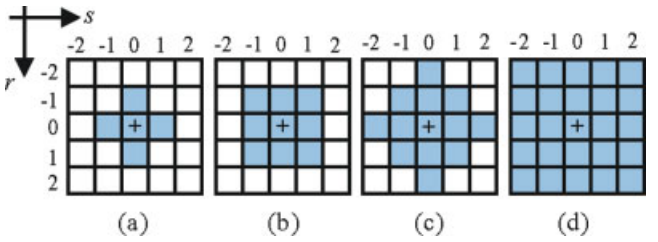

Figure 7. Popular processing windows with the symbol '+' denoting the window center (r, s). The 3×3 square processing window is depicted in (b), with the area of support defined as $\zeta = \{(r-1, s-1), (r-1, s), (r-1, s+1), (r, s-1), (r, s), (r, s+1), (r+1, s-1)\ (r+1, s), (r+1, s+1)\}$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

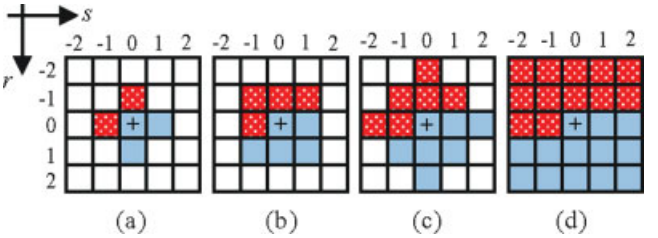

Figure 8. Recursive filtering concept demonstrated for the same processing windows as used in the conventional, non-recursive approach in Fig. 7. The textured locations, such as $\zeta' = \{(r-1, s-1), (r-1, s), (r-1, s+1), (r, s-1)\}$ in (b) denote the samples taken from the filtered image $\mathbf{y}$, while the filled locations such as $\zeta = \{(r, s), (r, s+1), (r+1, s-1)\ (r+1, s), (r+1, s+1)\}$ in (b) denote the samples taken from the input image $\mathbf{x}$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

The proposed framework generates the segmented cDNA microarray images by producing root signals. Such a situation occurs if and only if $\mathbf{y}_{(r,s)} = \mathbf{x}_{(r,s)}$, i.e., the filter output $\mathbf{y}_{(r,s)}$ is identical to a multichannel signal located at the reference window position (r, s). Because of the localized nature of cDNA image features, the analysis of the root signals necessitates the definition of a basic cDNA image structure, which can be observed in the processing window $\Psi_{(r,s)}$. Adopting concepts routinely used in conventional color image processing (Tang et al., 1994), the following basic cDNA multichannel signal structures are defined:

- A multichannel constant region (Fig. 6c) is a neighborhood formed by identical, in terms of both magnitude and direction, cDNA image vectors.
- A multichannel, cDNA step edge (Fig. 6d) is a multichannel constant region followed by another multichannel constant region.
- A multichannel impulse (Fig. 6e) is a cDNA image vector that significantly deviates from a surrounding multichannel constant region.
- Multichannel oscillation[4] (Fig. 6f) is a sequence of cDNA vectors that is not part of a constant region, an edge, or an impulse.

The consideration of the structures defined earlier is essential in the proposed segmentation framework, since the root signals consist solely of constant neighborhoods and edges (Astola et al., 1987). An ideal root signal, defined over a cDNA microarray image, must have the few different forms listed in Figure 9. To obtain such a root, the elimination of impulses and oscillations (data variations) is an essential step. Finally, the presence of the above multichannel structures in $\Psi_{(r,s)}$ determines the speed with which the processing filter converges to a root signal. In this work, we normalize both the magnitude and directional characteristics of the cDNA vectorial inputs and generate root signals by employing a class of nonlinear selection vector operators. Thus, generalized vector filters are described in the next few paragraphs prior to introducing the root signal generation procedure.

[4] The multichannel oscillation consists of diverse data variations and cDNA image vectors which deviate from each other within the local neighborhood.

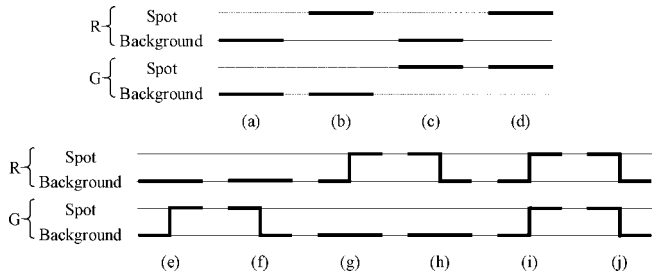

Figure 9. Root signal representation of cDNA vectorial data: (a–d) multichannel constant region, (e–j) multichannel step edge.

B. A Generalized Class of Vector Filters. The structural elements of the cDNA image such as the spots and their edges, as well as the corrupting noise processes are nonlinear in nature. Moreover, the spots vary in the shapes, sizes, and coloration. Therefore, the processing technique should have strong nonlinear characteristics in order to effectively deal with the cDNA image data. The selection weighted vector filters (SWVF)[5] (Lukac et al., 2004a) employed here constitute a generalized and unique class of nonlinear, vector processing operators which are suitable for such a task.

The SWVF structure is characterized by a design parameter p ranging from 0 to 1, and a set of nonnegative real weights $\mathbf{w} = [w_{(i,j)};\ (i,j) \in \zeta]$. For each input sample $\mathbf{x}_{(i,j)}$, $(i,j) \in \zeta$, the weights $w_{(i,j)}$ are used to form a SWVF processing function $f_{\mathrm{SWVF}}(\cdot)$ defined as follows:

$$\mathbf{y}_{(r,s)} = \arg\min_{\mathbf{x}_{(g,h)}} \left[\left(\sum_{(i,j)\in\zeta} w_{(i,j)} d(\mathbf{x}_{(g,h)}, \mathbf{x}_{(i,j)}) \right)^{1-p} \left(\sum_{(i,j)\in\zeta} w_{(i,j)} \mathrm{A}(\mathbf{x}_{(g,h)}, \mathbf{x}_{(i,j)}) \right)^{p} \right] \tag{5}$$

where $\mathbf{y}_{(r,s)} = f_{\mathrm{SWVF}}(\Psi_{(r,s)}, \mathbf{w}, p)$ represents the filter output. The filtering structure considered here outputs the input cDNA vector $\mathbf{x}_{(g,h)} \in \Psi_{(r,s)}$, which minimizes an aggregated distance criterion defined over the cDNA samples inside the processing window $\Psi_{(r,s)}$. The selective nature of the SWVF operator and the use of the minimization concept ensure (i) the outputting of the cDNA vector which is the most similar, under the specific setting of $\mathbf{w}$, to other cDNA samples in $\Psi_{(r,s)}$, (ii) the low-pass nature of the filter, making it capable to denoise and normalize the cDNA data, and (iii) that the SWVF output is restricted to the dynamic range of the input samples and thus, it never introduces new samples.

The weighting coefficient $w_{(i,j)}$ signifies the importance of $\mathbf{x}_{(i,j)}$ in $\Psi_{(r,s)}$. Through the weight vector $\mathbf{w}$ and the design parameter p the SWVF scheme tunes the overall filter's detail-preserving and noise attenuating characteristics and uses both the spatial and spectral characteristics of the cDNA image $\mathbf{x}$ during processing. Depending on the value of parameter $0 \leq p \leq 1$ in Eq. (5), data normalization can be performed in the magnitude ($p = 0$) or directional ($p = 1$) domain. The framework allows for data normalization using equally the magnitude and directional information when $p = 0.5$. Any deviation from $p = 0.5$ to a lower or larger p value places more emphasis on the magnitude or directional characteristics, respectively, and tilts the overall performance of the proposed normalization process towards one or the other extreme.

Visual inspection of the cDNA microarray image (Fig. 5d) and its filtered version (Fig. 10a) suggests that the employed here SWVF operator is robust and excellently preserves edges of the spots. The noise attenuation capability of the proposed framework further increases (Fig. 10d) when the root signal based concept is employed in the processing pipeline. Figures 10a–10c show that the proposed framework, although used in a noniterative manner, minimizes variations in the cDNA image measurements. However, by repeating the SWVF operation until a root signal is obtained, the proposed framework produces normalized image data with characteristics (Figs. 10d–10f) much closer to those of an ideal, noise-free, cDNA microarray image (Figs. 5a–5c).

Since each setting of the filter parameters represents a specific filter that can be used for a specific task, SWVF filters constitute a wide class of vector operators. For example, the use of the unity weight vector $\mathbf{w} = 1$ in Eq. (5) with $p = 0$ denotes a SWVF operator equivalent to the well-known vector median filter (VMF) (Astola et al., 1990) while $\mathbf{w} = 1$ and $p = 1$ denote a SWVF operator with characteristics identical to those of the basic vector directional filter (BVDF) (Trahanias et al., 1996). The complete list of vector filtering operators generalized within the SWVF class is provided in Lukac et al. (2004a).

C. SWVF Weight Vector Adaptation. The use of the SWVF scheme (5) requires the determination of the weight vector $\mathbf{w}$ by the end-user. Since the resolution of a scanned cDNA microarray image is usually determined by processing requirements, the acquisition speed and the calibration of the scanner (Chen et al., 1997) which vary depending on the laboratory procedures and may influence the accuracy of any microarray image processing algorithm, the automated setting of the SWVF parameters is of a great interest. To provide a more comfortable—in terms of the exclusion of the manually justified $\mathbf{w}$—approach, an automated alternative to the user-driven method of Eq. (5) is obtained by employing a multichannel weights' adaptation algorithm of Lukac et al. (2004a). Since the original signal $\mathbf{o}_{(r,s)}$ of Eq. (1) is not available in cDNA microarray applications, the weights $w_{(i,j)}$ in Eq. (5) can be adapted using the approaches depicted in Figure 11. The considered adaptation scheme leads to a number of SWVF filters with different design characteristics. Namely,

- The use of the acquired signal $\mathbf{x}_{(r,s)}$ in Figure 11a is useful, when the corrupting noise power is low and strong detail-preserving characteristics are expected from the SWVF operators.
- The robustness of the SWVF operator and its noise attenuation capability is ensured using a robust, easy to calculate estimate such as the component-wise median $\mathbf{y}^*_{(r,s)} = [\mathbf{y}^*_{(r,s)1}, \mathbf{y}^*_{(r,s)2}]$ used in Figure 11b.
- The approach shown in Figure 11c allows for a filter that balances the noise attenuating and detail-preserving characteristics of the solutions depicted in Figures 11a and 11b.

Since microarray images are usually corrupted by strong noise (Eisen and Brown, 1999; Wang et al., 2003), the approach of Figure 11b is used in the sequel. The weighting coefficients

[5] Although the SWVF operators have been developed to be used primarily for color image filtering (Lukac et al., 2004a), they constitute universal class of processing operators which can process scalar signals such as gray-scale images or multichannel signals such as cDNA microarray images.

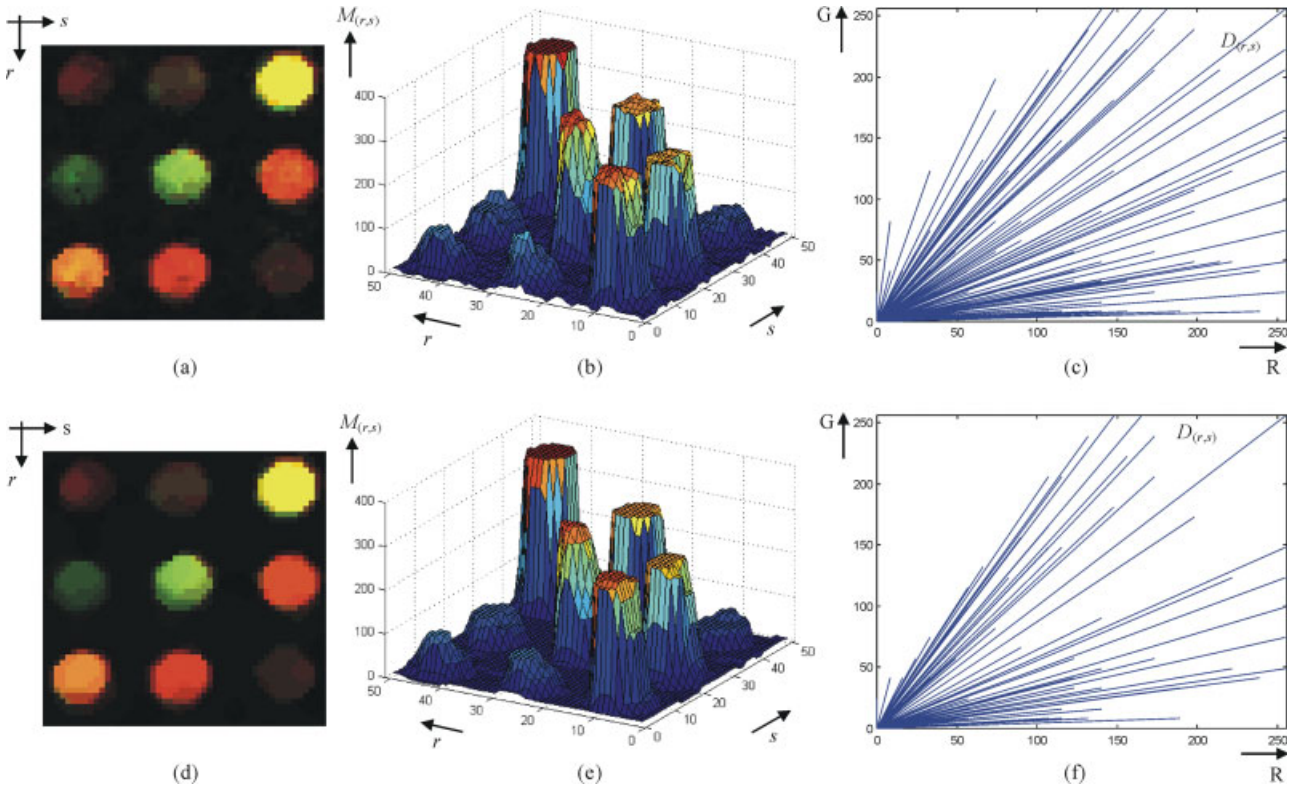

Figure 10. cDNA image characteristics corresponding to (a–c) a single pass through the SWVF operator and (d–f) root signal of the SWVF operator: (a, d) image area of interest, (b, e) magnitude characteristics $M_{(r,s)}$, (c, f) directional characteristics $D_{(r,s)}$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

$w_{(i,j)} \in \mathbf{w}$ are adapted using the cDNA image vectors $\mathbf{x}_{(i,j)}$ as follows:

$$w_{(i,j)} = P\left[w_{(i,j)} + 2\mu R(\mathbf{y}^*_{(r,s)}, \mathbf{y}_{(r,s)})\mathrm{sgn}\left(R(\mathbf{x}_{(i,j)}, \mathbf{y}_{(r,s)})\right)\right] \quad (6)$$

where μ is a regulation factor and $\mathbf{y}^*_{(r,s)} = [y^*_{(r,s)1}, y^*_{(r,s)2}]$ is the component-wise median of $\Psi_{(r,s)}$ defined as

$$y^*_{(r,s)k} = \arg\min_{x_{(g,h)k}} \sum_{(i,j)\in\zeta} |x_{(g,h)k} - x_{(i,j)k}|, \quad \text{for } k = 1,2 \quad (7)$$

Each weight $w_{(i,j)}$ is adjusted by adding the contributions of the corresponding cDNA vector $\mathbf{x}_{(i,j)}$ and the SWVF output $\mathbf{y}_{(r,s)}$. These contributions are measured as the distances to the feature signal $\mathbf{y}^*_{(r,s)}$, which is used to guide the adaptation process. To minimize the influence of the initial setting[6] of the SWVF parameters the adaptation formula should allow for the adjustment of $w_{(i,j)}$ using both positive and negative contributions. Therefore, Eq. (6) is constructed using the sign sigmoidal function $\mathrm{sgn}(a) = 2/(1 + \exp(-a)) - 1$ and the vectorial sign function

$$R(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)}) = S(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)})\left(d(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)})\right)^{1-p}\left(\mathrm{A}(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)})\right)^{p} \quad (8)$$

which considers contributions using both Eqs. (2) and (3) with the polarity $S(\cdot) \in \{-1, 1\}$ defined as follows:

$$S(\mathbf{x}_{(i,j)}, \mathbf{x}_{(g,h)}) = \begin{cases} +1 & \text{for} \|\mathbf{x}_{(i,j)}\| - \|\mathbf{x}_{(g,h)}\| \geq 0 \\ -1 & \text{for} \|\mathbf{x}_{(i,j)}\| - \|\mathbf{x}_{(g,h)}\| < 0 \end{cases} \quad (9)$$

The use of $R(\cdot)$ is essential in $\mathrm{sgn}(\cdot)$ since the positive (or negative) values of $R(\mathbf{x}_{(i,j)}, \mathbf{y}_{(r,s)})$ allow for the corresponding adjustment of $w_{(i,j)}$ in Eq. (6) by adding the negative (or positive) value of $2\mu R(\mathbf{y}^*_{(r,s)}, \mathbf{y}_{(r,s)})\mathrm{sgn}(R(\mathbf{x}_{(i,j)}, \mathbf{y}_{(r,s)}))$. If the sample under consideration $\mathbf{x}_{(i,j)}$ and the actual SWVF output $\mathbf{y}_{(r,s)}$ are identical (i.e. $R(\mathbf{x}_{(i,j)}, \mathbf{y}_{(r,s)}) = 0$), then $\mathrm{sgn}(\cdot) = 0$, which suggests that $w_{(i,j)}$ is kept unchanged at the moment. In all other cases, $w_{(i,j)}$ is adjusted based on the difference between the SWVF output $\mathbf{y}_{(r,s)}$ and the feature signal $\mathbf{y}^*_{(r,s)}$ because the value of $\mathrm{sgn}(\cdot)$ is usually very close to –1 or 1. To keep the aggregated distances in Eq. (5) positive, and thus to ensure the unbiased low-pass characteristics of the SWVF filters, a projection function $P(\cdot)$, defined as $P(w_{(i,j)}) = 0$ for $w_{(i,j)} < 0$ and $P(w_{(i,j)}) = w_{(i,j)}$ for $w_{(i,j)} \geq 0$, is used to project the updated weight $w_{(i,j)}$ onto the constraint space of $\mathbf{w}$ during the adaptation process in Eq. (6).

The algorithmic steps performed during the optimization of $\mathbf{w}$ are summarized, in pseudo-code format, in Figure 12. The weight adaptation in Eq. (6) is performed in all spatial locations (r, s) of the microarray image $\mathbf{x}$, i.e., for $r = 1,2,\ldots,K_1$ and $s = 1,2,\ldots,K_2$. When the optimization of $\mathbf{w}$ is completed, the constructed SWVF filter can be used to generate root signals by repeating the procedure summarized in Figure 13, and thus segment the cDNA microarray

[6] The initial weight vector can be set to any arbitrary positive value, but equally aligned weighting coefficients such as $w_{(i,j)} = 1$, for $(i,j) \in \zeta$, corresponding to the robust smoothing functions and $\mu \ll 0.5$ are the values recommended in (Lukac et al., 2004a) for conventional color image processing applications.

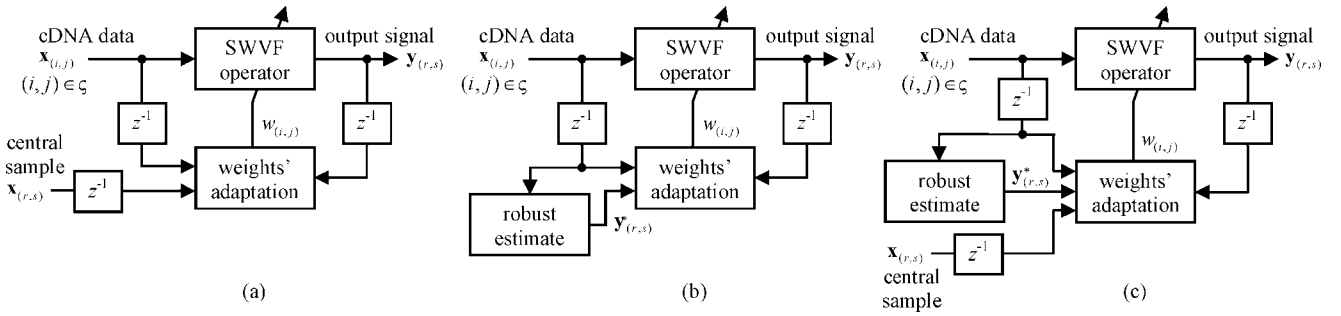

Figure 11. SWVF adaptation schemes adjusting the weights using: (a) the central sample $\mathbf{x}_{(r,s)}$, (b) the robust estimate $\mathbf{y}^{\star}_{(r,s)}$, (c) both the central sample $\mathbf{x}_{(r,s)}$ and the robust estimate $\mathbf{y}^{\star}_{(r,s)}$.

image. To demonstrate the suitability of the proposed framework, the robust straightforward solutions obtained through the end-user's choice of the parameters in Eq. (5) and the adaptive solutions in Eq. (6) are used for comparison purposes in the sequel. Studies on determining the optimal SWVF structure and the enumeration of the various adaptation solutions have been addressed in Lukac et al. (2004a) and are beyond the scope of this paper.

D. SWVF Root Signals. The proposed framework generates the segmented cDNA microarray images by producing SWVF root signals. The SWVF response to any input signal is uniquely defined in Eq. (5) with the SWVF output $\mathbf{y}_{(r,s)} \in \Psi_{(r,s)}$. Therefore, the root signal

$$\mathbf{x}_{(r,s)} = f_{\mathrm{SWVF}}(\Psi_{(r,s)}, \mathbf{w}, p) \tag{10}$$

can be obtained by filtering repeatedly with a SWVF operator any finite-length cDNA signal.

It should be emphasized that not all SWVF operators are suitable for cDNA image segmentation. For example, the SWVF filtering class includes two special subclasses of the so-called identity and idempotent operators, respectively, which have an extremely fast convergence to a root signal. Any SWVF operator with $w_{(r,s)} \geq \sum w_{(i,j)}$, for $(i,j) \in \zeta$ and $(i,j) \neq (r,s)$, reduces to the identity filter, which leaves the central sample $\mathbf{x}_{(r,s)}$ unchanged (Lukac et al., 2004a). Using an identity operator any input signal is always invariant to filtering, with the output $\mathbf{y}_{(r,s)} = \mathbf{x}_{(r,s)}$. The second subclass of operators, the so-called idempotent filters, produce a signal, which is invariant to a second pass from the same filtering operator (Zeng, 1994). Both special cases as well as other SWVF operators with inefficient smoothing capability preserve structural information and noise in cDNA microarray image and thus, are neither capable of normalizing the variations in the cDNA vector field nor separating spots from the background. It is commonly accepted that both the robustness and the smoothing capability of the selection type filters increase with the degree of uniformity in the entries of $\mathbf{w}$, and that the most robust SWVF operators are those with the unity weight vector[7]. The use of the SWVF operators with robust smoothing characteristics makes the spots uniform and removes noisy foreground information. Thus, these operators repeatedly used to process the microarray image emphasize both the measured and perceptual differences between foreground and background, and meet the objective in Eq. (4).

```
Inputs:  NumberOfRows × NumberOfColumns input image x
         Initial weight vector w = [w_(i,j); (i,j) ∈ ς]
         Design parameter p
         Regularization factor μ
Output:  Suboptimal weight vector w

For r = 1 to NumberOfRows
  For s = 1 to NumberOfColumns
    Let the input set Ψ_(r,s) = {x_(i,j); (i,j) ∈ ς}
    Determine y_(r,s) = f_SWVF(Ψ_(r,s), w, p) using (5)
    Determine y*_(r,s) = [y*_(r,s)1, y*_(r,s)2] using (7)
    Adapt each w_(i,j) ∈ w using (6)
  End
End
Output, suboptimal weight vector w
```

Figure 12. Pseudo-code format of the SWVF weight vector adaptation algorithm operating within the area of support ζ.

Assuming that $\mathbf{y}^{n}_{(r,s)}$ is a vector in the image $\mathbf{y}^{n}$ obtained after filtering n times the input cDNA image $\mathbf{x}$, the convergence rate expressed as a function of the difference between two successive filtering results can be defined as follow (Lukac and Plataniotis, 2005):

$$\gamma = \left(R(\mathbf{y}^{n}, \mathbf{y}^{n-1})\right)^2 \tag{11}$$

where $\mathbf{y}^{0} = \mathbf{x}$ denotes the (input) cDNA microarray image that undergoes segmentation. The proposed segmentation procedure is completed when $\gamma = 0$, indicating that there are no changes in the filtered signal and that the root signal has been reached. If real-time processing aspects are of paramount interest, the segmentation process can be stopped for γ smaller than a user defined threshold. When the stopping condition is satisfied, the value of $n - 1$ denotes the iteration for which the root signal has been reached.

The root signal convergence process performs morphological operations such as various compositions of erosion and dilatation operations defined over $\Psi_{(r,s)}$, which is considered as the structuring element. The impairments present in background have magnitude larger than the desired background samples and thus, they can be considered "positive" noise, which can be efficiently removed by either multichannel morphological erosion or opening (Magaros and Schafer, 1987). In addition to positive noise, a number of spots

[7] The multiplication of the weights in $\mathbf{w}$ by a positive constant does not affect the performance of the SWVF operator defined in Eq. (5).

```
Inputs:   NumberOfRows × NumberOfColumns input image x
          Weight vector w = [w_(i,j); (i,j) ∈ ς]
          Design parameter p
Output:   NumberOfRows × NumberOfColumns image y

For r = 1 to NumberOfRows
  For s = 1 to NumberOfColumns
    Let the input set Ψ_(r,s) = {x_(i,j); (i,j) ∈ ς}
    Determine y_(r,s) = f_SWVF(Ψ_(r,s), w, p) using (5)
  End
End
Output, filtered image y
```

Figure 13. Pseudo-code format of the SWVF image processing algorithm (one pass only) operating within the area of support ζ.

contain (i) cDNA vectors that are, in terms of amplitude of their components, smaller than neighboring vectors, and (ii) holes that have been created by cDNA vectors of zero-like magnitude. These impairments can be viewed as "negative" noise which can be efficiently handled using dilatation or closing morphological operators. Figure 14 shows that any high-frequency impairments such as outliers in the cDNA image data population (Fig. 14a) are smoothed in a single pass using the robust SWVF operator (Fig. 14b). As shown in Figures 14c and 14d, the framework completes morphological processing like operations by repeatedly filtering the acquired image. In this case, any small signal structures contained in $\Psi_{(r,s)}$, such as irregular spots or holes present in the microarray spots are removed from the segmented image using root signals, which can be seen as performing erosion/opening or dilatation/closing operations.

The replacement of the window center $\mathbf{y}_{(r,s)}^{n-1}$ with the statistically, under the setting of SWVF parameters, most similar, to the cDNA samples within $\Psi_{(r,s)}$, vector $\mathbf{y}_{(r,s)}^{n}$ produces a spot which is uniquely described by dominant color vectors. The generated segmented images have normalized intensity in both the background

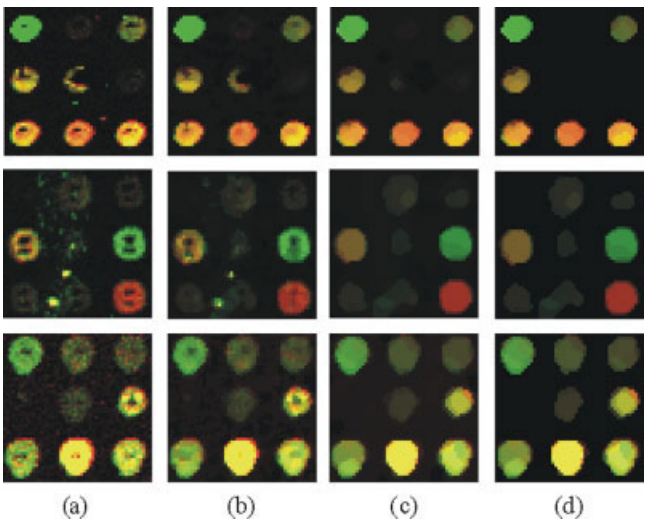

Figure 14. Comparison of the images produced at the different stages in the proposed segmentation framework: (a) input cDNA microarray image, (b) SWVF output, (c) SWVF root signal, (d) SWVF root signal after postprocessing. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

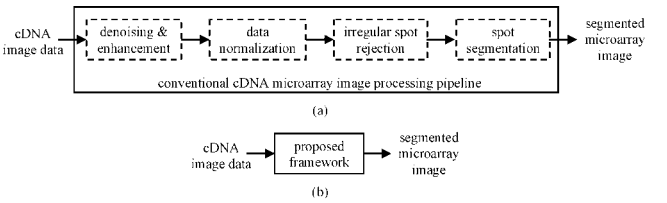

Figure 15. cDNA microarray image segmentation pipeline architectures: (a) the conventional pipeline, (b) the proposed framework. Employing the root signal concept, the proposed vectorial processing based segmentation approach replaces the set of four processing operations used in the conventional processing pipeline.

and the spot locations. Furthermore, the difference, both visually perceived and objectively measured between foreground and background information, has been enhanced. This difference can be further increased by root signal post-processing operations. By performing thresholding operations over the magnitude of the root signal, the framework removes residual irregular spots and idealizes the background in the segmented microarray image (Fig. 14d).

A pipeline architecture of the proposed cDNA microarray image segmentation framework is depicted in Figure 15b. As it can be seen, the proposed segmentation approach should be considered as a replacement of the conventional microarray image processing pipeline (Fig. 15a) operating in a cascaded processing mode. Furthermore, visual inspection of the block diagram representation given in Figure 16 reveals that by (i) setting the filter parameters, (ii) choosing the window shape, (iii) determining the SWVF processing mode, and (iv) employing the proper weights' adaptation mechanism, the framework can offer solutions which differ in their design philosophy, characteristics, computational complexity, and performance. The enumeration of all available options or the determination of the best configuration of construction elements, according to specified criteria, is beyond the scope of the paper. To

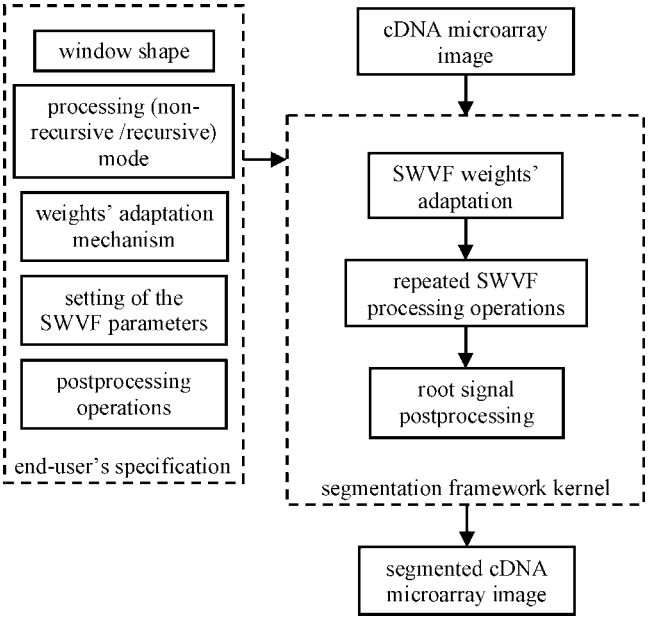

Figure 16. Block diagram representation of the proposed cDNA microarray image segmentation framework.

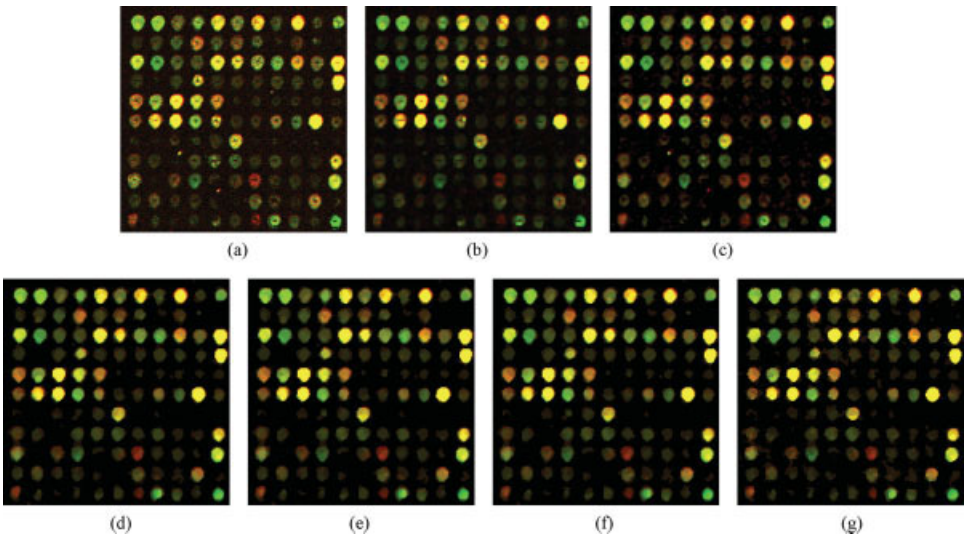

Figure 17. cDNA microarray images corresponding to (a) input cDNA microarray image and (b–g) its segmented versions obtained using (b) morphological approach, (c) clustering technique, (d) proposed nonadaptive approach with $p = 0$, (e) proposed nonadaptive approach with $p = 1$, (f) proposed nonadaptive approach with $p = 0.5$, (g) proposed adaptive approach with $p = 0.5$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

demonstrate the suitability of the proposed framework, the most typical solutions obtained in Eqs. (5) and (6) are used in the sequel for comparison purposes.

IV. EXPERIMENTAL RESULTS

A number of microarray images has been used to test and evaluate the performance of the proposed framework. Examples of 200 $\times$ 200 patterns (eight bits per component) cropped from the images captured using laser microscope scanners are shown in Figures 17a and 18a. The test images vary in both complexity and noise appearance. Therefore, they can be used to test the robustness of the proposed framework. In addition to the real cDNA microarray images depicted in Figures 17a and 18a, phantom images (Figs. 19a and 20a) with a spatial resolution of 200 $\times$ 200 pixels are used so that an objective evaluation of the proposed framework via the comparative evaluation of the produced output images and the original phantom images is possible.

For the comparisons three robust SWVF operators, defined in Eq. (5) using the unity weight vector $\mathbf{w}$ are considered. Namely, we choose those which operate in the magnitude domain ($p = 0$), directional domain ($p = 1$), and the SWVF operator which uses both the magnitude and the directional characteristics of the cDNA vectors simultaneously ($p = 0.5$). An SWVF operator ($p = 0.5$) using a weight vector $\mathbf{w}$ determined adaptively from the microarray image via Eq. (6) is considered, as well. In all proposed solutions, a 3 $\times$ 3 square window, and a non-recursive processing mode are employed. The proposed solutions are compared, in terms of performance, against other segmentation techniques, such as the morphological approach and clustering based segmentation technique which were shown to be quite appropriate for cDNA image segmentation.

A. Examination of the Performance Using Real cDNA Microarray Images.

Figures 17 and 18 allow for the visual comparison of the input cDNA microarray images and the corresponding segmented images. The results are evaluated subjectively since for realistic applications ground-truth (original) images are usually not available. Therefore, image quality is evaluated in terms of spot

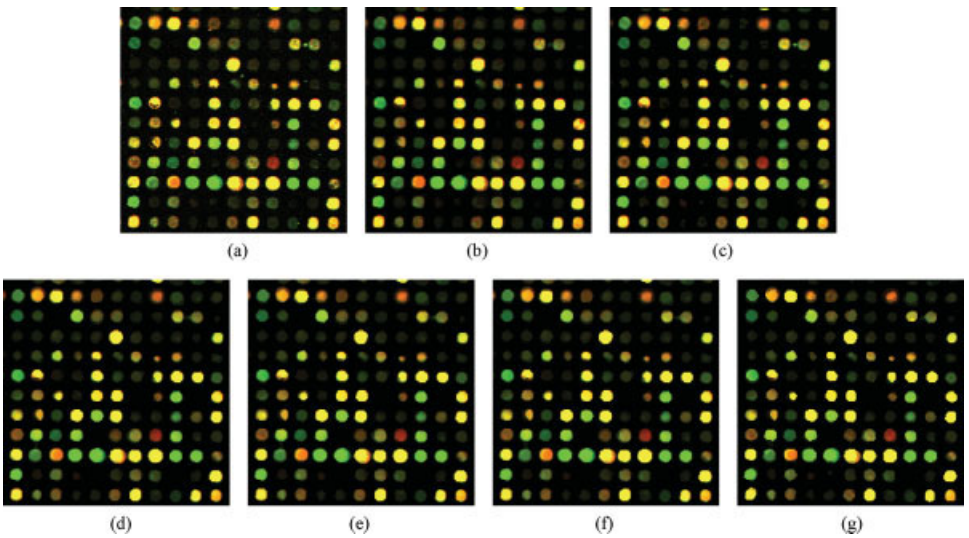

Figure 18. Another example of cDNA microarray images corresponding to (a) input cDNA microarray image and (b–g) its segmented versions obtained using (b) morphological approach, (c) clustering technique, (d) proposed nonadaptive approach with $p = 0$, (e) proposed nonadaptive approach with $p = 1$, (f) proposed nonadaptive approach with $p = 0.5$, (g) proposed adaptive approach with $p = 0.5$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

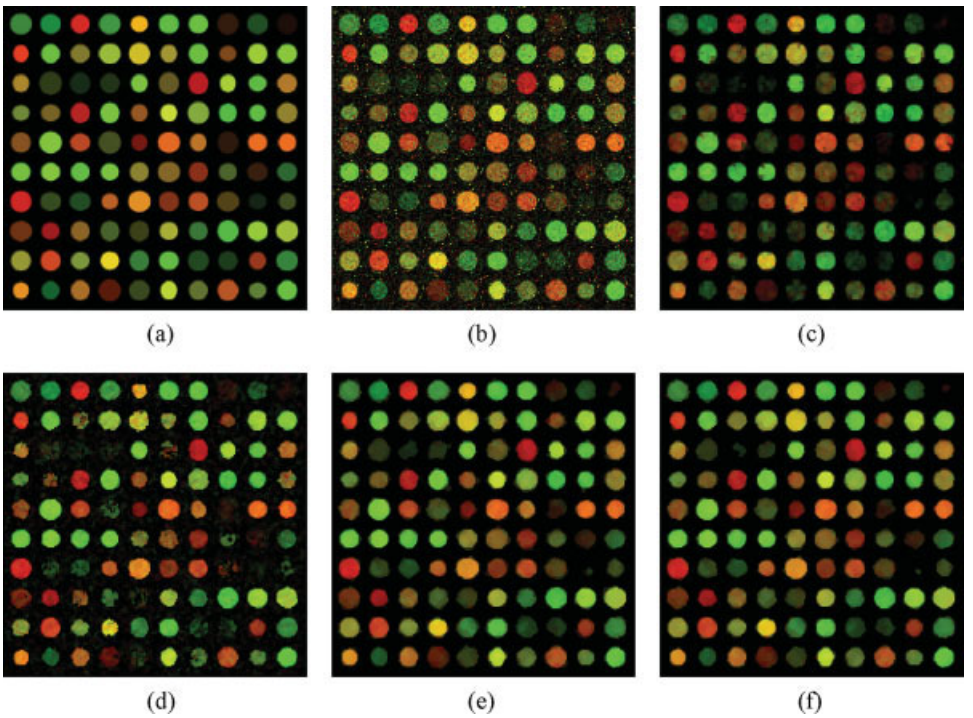

Figure 19. Results obtained using the artificially corrupted phantom images: (a) original image, (b) noisy image, (c) morphological approach, (d) clustering technique, (e) proposed nonadaptive approach with $p = 0$, (f) proposed adaptive approach with $p = 0$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

preservation and the presence of residual noise which can be viewed as the result of faulty processing.

Visual inspection of Figure 17 indicates that the input cDNA image (Fig. 17a) contains various foreground and background impairments. Moreover, a number of spots have holes and varying colorations. Morphological processing (Fig. 17b) removes the background noise present in the input cDNA image and normalizes the variations of the spots' intensities. However, it does not completely eliminate the holes presents in the spots while at the same time enhances irregular spots in various image locations, thus introducing additional errors that will affect the accuracy of the subsequent processing tasks such as gene expression analysis. Figure 17c depicts the image produced by a clustering based segmentation technique subsequently followed by component-wise median filtering. Although the clustering approach is usually sensitive to noise, the cascade of clustering and median filtering operations overcomes

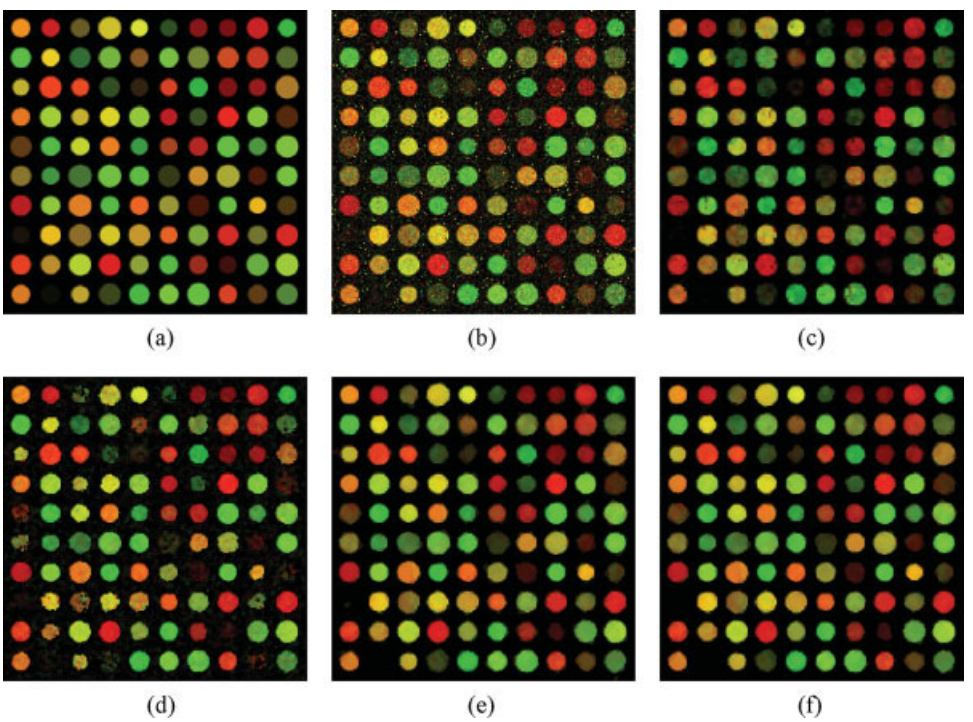

Figure 20. Another set of results obtained using the artificially corrupted phantom images: (a) original image, (b) noisy image, (c) morphological approach, (d) clustering technique, (e) proposed nonadaptive approach with $p = 0$, (f) proposed adaptive approach with $p = 0$. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com]

Table I. Objective comparison of the considered solutions.

	Fig. 19		Fig. 20	
Solution/Criterion	MAE	PSNR	MAE	PSNR
Identity (noisy image)	17.2	18.6	17.4	18.6
Morphological	10.8	21.6	11.6	21.1
Clustering	11.5	20.5	12.1	19.8
Proposed non-adaptive	5.4	26.4	5.7	26.2
Proposed adaptive	5.2	26.3	5.5	26.1

this limitation and often outperforms the morphological approach in terms of spot intensity/color normalization. However, as it can be seen, the segmentation performance is still rather insufficient, especially in the heavily noise-corrupted areas. This is not the case when the proposed segmentation framework is used. By employing a solution that operates in the magnitude domain, the proposed framework produces the output image shown in Figure 17d. Visual inspection of the result reveals that noise pixels have been eliminated, the coloration of the spots is normalized, and the foreground information is readily separated from the background. Figure 17e displays the result obtained when the input cDNA image was processed with a SWVF operator operating solely in the directional domain. The result indicates that directional processing based operators have worse smoothing characteristics compared to the magnitude processing based approaches (Lukac et al., 2004a) and may not be suitable for processing two-component vectorial signals such as cDNA microarray images. Better results are obtained when the proposed solutions simultaneously utilize both magnitude and orientation of cDNA vectors to segment the image (Figs. 17f and 17g). However, the performance of such SWVF solutions seems to be worse compared to their variants operating mainly in the magnitude domain (i.e. $0.5 > p \geq 0$). Note that the above discussion on the suitability of directional processing operators in cDNA microarray imaging coincides with the discussion in Lukac et al. (2004b).

The robust behavior of the proposed cDNA microarray image segmentation solution is confirmed by the results depicted in Figure 18. Although in this case the acquired cDNA image (Fig. 18a) contains a smaller amount of impairments compared to the input image of Figure 17a, neither the morphological (Fig. 18b) nor the clustering (Fig. 18c) methods are capable of producing a segmented image of higher quality compared to the images generated using the proposed framework (Figs. 18d–18g).

B. Examination of the Performance Using Phantom Images. To allow for the objective comparisons of all solutions considered here, phantom images are used in the sequel. The ground-truth (original) images shown in Figures 19a and 20a have been corrupted using mixed noise, comprised of additive Gaussian noise with the standard deviation equal to 30 and impulsive noise with the impulse probability equal to 0.02 in order to generate the noisy images shown in Figures 19b and 20b. These noisy images have been subsequently processed and the segmented outputs are shown in Figures 19c–19f and 20c–20f. Differences between the segmented results and the original ground-truth images have been calculated using the mean absolute error (MAE) and the peak signal-to-noise ratio (PSNR), which are commonly used in the image processing community. The interested reader may find additional information on the employed noise model and definitions of the objective criteria in Lukac and Plataniotis (2006).

Results reported in Table I demonstrate that the proposed framework clearly outperforms the competition in terms of both MAE and PSNR. Note that the images obtained using clustering were subsequently enhanced via component-wise median smoothing. A parameter value of $p = 0$ was used in the proposed framework. As it can be seen, the best results (lowest MAE values and highest PSNR values) were obtained using the proposed framework. Inspection of the MAE values corresponding to the proposed framework reveals a performance improvement due to weight adaptation in the adaptive SWVF solution, which allows for better spot preservation compared to its nonadaptive variant. On the other hand, robust nonadaptive solutions such as one employed in the experiment usually produce images with higher PSNR values due to their extensive smoothing characteristics.

Detailed inspection of the images shown in Figures 19 and 20 confirm the performance evaluation summarized in Table I. Comparing the segmented images, in terms of the spot preservation, the presence of residual noise, and the similarity to the ground-truth images, reveals that the proposed framework produced better results than the morphological and clustering segmentation approaches. Namely, Figures 19e and 19f and Figures 20e and 20f depict images with properly extracted and smoothed spots, whereas Figures 19c and 19d and Figures 20c and 20d list images which suffer from residual noise and spot intensity variations. Moreover, among the considered solutions, images segmented by the proposed framework exhibit the highest similarity to the original images. This suggests that the proposed framework allows for simultaneous high-quality image enhancement and spot segmentation.

C. Computational Complexity Analysis. Apart from the numerical behavior (actual performance) of any algorithm, its computational complexity is a realistic measure of its practicality and usefulness. Therefore, solutions designed within the proposed cDNA microarray segmentation framework are analyzed here in terms of normalized operations, such as additions (ADDs), subtractions (SUBs), multiplications (MULTs), divisions (DIVs), square roots (SQRTs), comparisons (COMPs), and arc cosines (ARCCOSs).

Table II summarizes the total number of operations for the most important solutions designed within the SWVF framework. Note that the use of non-recursive or recursive processing mode keeps the computational complexity of the operators unchanged (Burian and Kuosmanen, 2002). The cost of the implementation of the

Table II. Cost of the SWVF solutions in Eq. (5) for a 3×3 supporting window.

Filter/Operation	ADDs	SUBs	MULTs	DIVs	SQRTs	COMPs	ARCCOSs
$\mathbf{w} = 1; p = 0$	108	72	72	–	36	8	–
$\mathbf{w} = 1; p = 1$	180	–	162	36	36	8	36
$\mathbf{w} = 1; p = 0.5$	288	72	243	36	90	8	36
Arbitrary $\mathbf{w}$; $p = 0$	108	72	144	–	36	8	–
Arbitrary $\mathbf{w}$; $p = 1$	180	–	234	36	36	8	36
Arbitrary $\mathbf{w}$; $p = 0.5$	288	72	387	36	90	8	36

Table III. Obtained weight vectors normalized according to their corresponding maximum values.

Image	Weight vector $\mathbf{w} = [w_{(i,j)}; (i,j) \in \zeta]$								
	$w_{(r-1,\,s-1)}$	$w_{(r-1,\,s)}$	$w_{(r-1,\,s+1)}$	$w_{(r,\,s-1)}$	$w_{(r,\,s)}$	$w_{(r,\,s+1)}$	$w_{(r+1,\,s-1)}$	$w_{(r+1,\,s)}$	$w_{(r+1,\,s+1)}$
Fig. 17a	0.9157	0.9338	0.9120	0.9185	1.0000	0.9278	0.9122	0.9374	0.9244
Fig. 18a	0.7440	0.7929	0.7410	0.7922	1.0000	0.7578	0.7451	0.7917	0.7574
Fig. 19b	0.9783	0.9862	0.9848	0.9859	1.0000	0.9879	0.9535	0.9996	0.9799
Fig. 20b	0.9231	0.9420	0.9711	0.9816	1.0000	0.9768	0.9989	0.9883	0.9849

SWVF operators increases with the level of the generalization. The most attractive solution, from a computational point of view, is obtained when a SWVF operator with $p = 0$ and $\mathbf{w} = 1$ is considered.

Since the cDNA image processing is commonly performed using a PC, the efficiency of the SWVF operators can also be measured, in terms of the execution time in a typical computing platform. When implemented in software, on a PC equipped with an Intel Pentium IV 2.40 GHz CPU, 512 MB RAM, Windows XP operating system, and MS Visual C++ 5.0 programming environment, the proposed SWVF operator with $p = 0$ and $\mathbf{w} = 1$ processing a 200 $\times$ 200 cDNA microarray image requires (on average) 0.43 s per single pass. Using the SWVF weights' adaptation tool requires an additional processing time of 0.92 s.

Note that the objective of this analysis is to provide qualitative benchmark information regarding implementation issues and not to exhaustively cover all possible implementations. The development of software-optimized realizations of the algorithms under consideration is beyond the scope of this paper.

D. Discussion. Summarizing the findings obtained from the experimental analysis, the following claims can be made:

- The root-signal based solutions constitute a powerful cDNA microarray segmentation framework. The proposed framework excellently extracts spots while preserving both spatial and spectral characteristics of the cDNA vectors.
- The repetitive use of the robust SWVF operators eliminates noise, rejects irregularities present in the cDNA microarray images, and extracts perfectly the desired spots.
- The proposed framework is robust and outperforms other commonly used cDNA image segmentation solutions.

It should be also noted that the use of the proposed framework necessitates the division of large cDNA microarray image into smaller patterns, such as those shown in Figures 17a and 18a, which are comprised of $\sim 10 \times 10$ microarray spots, so that faster convergence to root signals and better spot segmentation are obtained. Therefore, the use of the proposed framework in practice will most probably require high-power processor arrays and parallel processing of scanned high-resolution cDNA images. Note that our framework can be applied to images of any spatial resolution. However, the use of cDNA microarrays with a resolution of at least 80 pixels per spot (in average) is recommended for high-quality results to be obtained.

In addition, as the recorded processing time suggests, repetitive filtering of input images may be time consuming. Our experimentation showed that sufficiently good results are produced by the introduced framework within 10–15 iterations. Since this number of iterations requires ~6 s per 200 $\times$ 200 pixel pattern, the development of fast SWVF algorithms will be of interest. One possible way of speeding up the process is to search for root signals such as multichannel constant regions and multichannel step edges inside a filter window. Knowing the root signals of a given filter allows for the elimination of extensive calculations due to the fact that local root signals are invariant to filtering, and thus their presence in the filter window directly determines the filter output. Alternatively, as the numbers listed in Table II indicate, magnitude processing based SWVF operators (i.e. $p = 0$) should be used. This choice provides the best trade-off between performance and computational efficiency. Finally, visual inspection of the SWVF weights listed in Table III suggests that although the suboptimal weight vector obtained in Eq. (6) varies depending on the image complexity and the noise characteristics, the proposed SWVF weight adaptation usually results in the center-weighted structures (i.e., structures with the highest value of the central weight $w_{(r,s)}$ and lower, approximately equal values of the remaining weights $w_{(i,j)}$, for $(i,j) \neq (r,s)$). The contribution of the neighboring cDNA vectors located in $(i,j) \in \zeta$, for $(i,j) \neq (r,s)$, increases with the noise corruption to achieve the balance between spot preservation and noise attenuation. Tuning only the center weight $w_{(r,s)}$ instead of the complete weight vector $\mathbf{w}$ will allow for the additional boosting of the computational efficiency.

V. CONCLUSION

This paper introduced a unique, root-signal based cDNA microarray image segmentation framework. The proposed framework employs generalized vector filtering operators to utilize essential spatial (edges, structural information) and spectral (magnitude and direction of cDNA vectorial inputs) image characteristic to track the changes in the structural content of the microarray image. By repeating the robust filtering process over the input image the framework produces a root signal which is invariant to further processing with the same type of the filtering operator. The achieved root signal represents a segmented microarray image with normalized data and enhanced spot information, thus suggesting that the roots are perfectly suitable for subsequent analysis and gene expression determination related tasks.

REFERENCES

P. Arena, M. Bucolo, L. Fortuna, and L. Occhipinty, Cellular neural networks for real-time DNA microarray analysis, IEEE Eng Med Biol 21 (2002), 17–25.

J. Astola, P. Haavisto, and Y. Neuvo, Vector median filters, Proc IEEE 78 (1990), 678–689.

J. Astola, P. Heinonen, and Y. Neuvo, On root structures of median and median type filters, IEEE Trans Acoust Speech Signal Process 35 (1987), 1199–1202.

P. Bajcsy, Gridline: Automatic grid alignment in DNA microarray scans, IEEE Trans Image Process 13 (2004), 15–25.

A. Burian and P. Kuosmanen, Tuning the smoothness of the recursive median filters, IEEE Trans Signal Process 50 (2002), 1631–1639.

Y. Chen, E. Dougherty, and M. Bittner, Ratio-based decisions and the quantitative analysis of cDNA microarray images, J Biomed Opt 2 (1997), 364–374.

A.P.G. Damiance, L. Zhao, and A.C.P.L.F. Carvalho, A dynamic model with adaptive pixel moving for microarray images segmentation, Real-Time Imaging 10 (2004), 189–195.

M.B. Eisen and P.O. Brown, DNA arrays for analysis of gene expression, Methods Enzymol 303 (1999), 179–205.

V. Filkov, S. Skiena, and J. Zhi, Analysis techniques for microarray time-series data, J Comput Biol 9 (2002), 317–330.

R. Hirata Jr, J. Barrera, R.F. Hashimoto, D.O. Dantas, and G.H. Esteves, Segmentation of microarray images by mathematical morphology, Real-Time Imaging 8 (2002), 491–505.

R.S.H. Istepanian, Microarray image processing: Current status and future directions, IEEE Trans Nanobiosci 2 (2003), 173–175.

M. Katzer, F. Kummert, and G. Sagerer, Methods for automatic microarray image segmentation, IEEE Trans Nanobiosci 2 (2003), 202–213.

J.H. Kim, H.Y. Kim, and Y.S. Lee, A novel method using edge detection for signal extraction from cDNA microarray image analysis, Exp Mol Med 33 (2001), 83–88.

J.H. Kim, D.M. Shin, and Y.S. Lee, Effect of local background intensities in the normalization of cDNA microarray data with a skewed expression profiles, Exp Mol Med 34 (2002), 224–232.

A.W.C. Liew, H. Yana, and M. Yang, Robust adaptive spot segmentation of DNA microarray images, Pattern Recognit 36 (2003), 1251–1254.

R. Lukac and K.N. Plataniotis, Vector median root signals determination for cDNA microarray image segmentation, Proc Int Conf Image Anal Recognit, 2005, pp. 879–885.

R. Lukac and K.N. Plataniotis, "A taxonomy of color image filtering and enhancement solutions," InAdvances in imaging and electron physics, Vol. 140, P.W. Hawkes (Editor), Academic Press, Elsevier, 2006, pp. 187–264.

R. Lukac, K.N. Plataniotis, B. Smolka, and A.N. Venetsanopoulos, Generalized selection weighted vector filters, EURASIP J Appl Signal Process (2004a), 1870–1885. Special Issue on Nonlinear Signal and Image Processing 2004.

R. Lukac, K.N. Plataniotis, B. Smolka, and A.N. Venetsanopoulos, A multichannel order-statistic technique for cDNA microarray image processing, IEEE Trans Nanobiosci 3 (2004b), 272–285.

R. Lukac, K.N. Plataniotis, B. Smolka, and A.N. Venetsanopoulos, cDNA microarray image processing using fuzzy vector filtering framework, J Fuzzy Sets Syst152 (2005a), 17–35. Special Issue on Fuzzy Sets and Syst in Bioinformatics.

R. Lukac, B. Smolka, K. Martin, K.N. Plataniotis, and A.N. Venetsanopulos, Vector filtering for color imaging, IEEE Signal Process Mag 22 (2005b), 74–86. Special Issue on Color Image Processing.

P. Maragos and R.W. Schafer, Morphological filters, Part 2: Their relations to median, order-statistic, and stack filters, IEEE Trans Acoust Speech Signal Process 35 (1987), 1170–1184.

R. Nagarajan, Intensity-based segmentation of microarrays images, IEEE Trans Med Imaging 22 (2003), 882–889.

R. Nagarajan and M. Upreti, Correlation statistics for cDNA microarray image analysis (accepted for publication in IEEE/ACM Trans Comput Biol Bioinformatics).

P. O'Neill and G.D. Magoulas, Improved processing of microarray data using image reconstruction techniques, IEEE Trans Nanobiosci 2 (2003), 176–183.

T. Park, S.G. Yi, S.H. Kang, S.Y. Lee, Y.S. Lee, and R. Simon, Evaluation of normalization methods for microarray data, BMC Bioinformatics 4 (2003), 13.

K.N. Plataniotis and A.N. Venetsanopoulos, Color image processing and applications, Springer Verlag, Berlin, Germany, 2000.

K. Tang, J. Astola, and Y. Neuvo, Multichannel edge enhancement in color image processing, IEEE Trans on Circuits Syst Videotechnol 4 (1994), 468–479.

P.E. Trahanias, D. Karakos, and A.N. Venetsanopoulos, Directional processing of color images: Theory and experimental results, IEEE Trans Image Process 5 (1996), 868–881.

X. Wang, S. Ghosh, and S.W. Guo, Quantitative quality control in microarray processing and data acquisition, Nucleic Acids Res 29 (2001), 1–8.

X.H. Wang, R.S.H. Istepian, and Y.H. Song, Microarray image enhancement using stationary wavelet transform, IEEE Trans Nanobiosci 2 (2003), 184–189.

Y. Wang, J. Lu, R. Lee, Z. Gu, and R. Clarke, Iterative normalization of CDNA microarray data, IEEE Trans Inf Technol Biomed 6 (2002), 29–37.

A.K. Whitchurch, Gene expression microarrays, IEEE Potentials 21 (2002), 30–34.

Y.H. Yang, M.J. Buckley, S. Dudoit, and T.P. Speed, Comparison of methods for image analysis on cDNA microarray data, J Comput Graph Stat 11 (2002), 108–136.

B. Zeng, Convergence properties of median and weighted median filters, IEEE Trans Signal Process 42 (1994), 3515–3518.

X.Y. Zhang, F. Chen, Y.T. Zhang, S.G. Agner, M. Akay, Z.H. Lu, M.M.Y. Waye, and S.K.W. Tsui, Signal processing techniques in genomic engineering, Proc IEEE 90 (2002), 1822–1833.