Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fuzzy Sets and Systems 152 (2005) 17–35
FUZZY sets and systems
www.elsevier.com/locate/fss

cDNA microarray image processing using fuzzy vector filtering framework

Rastislav Lukac[a,*], Konstantinos N. Plataniotis[a], Bogdan Smolka[b], Anastasios N. Venetsanopoulos[a]

[a]*The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto ON, Canada M5S 3G4*
[b]*Department of Automatic Control, Silesian University of Technology, Akademicka 16 Str., 44-101 Gliwice, Poland*
Available online 11 November 2004

Abstract

This paper presents a novel filtering framework capable of processing cDNA microarray images. The proposed two-component adaptive vector filters integrate well-known concepts from the areas of fuzzy set theory, nonlinear filtering, multidimensional scaling and robust order-statistics. By appropriately setting the weighting coefficients in a generalized framework, the method is capable of removing noise impairments while preserving structural information in cDNA microarray images. Noise removal is performed by tuning a membership function which utilizes distance criteria applied to cDNA vectorial inputs at each image location. The classical vector representation, adopted here for a two-channel processing task, as well as a new color-ratio model representation are used. Simulation studies reported in this paper indicate that the proposed adaptive fuzzy vector filters are computationally attractive, yield excellent performance and are able to preserve structural information while efficiently suppressing noise in cDNA microarray data.

Keywords: Microarray; cDNA microarray technology; Image processing; Smoothing; Noise removal; Enhancement; Vector processing; Fuzzy filters; Data-adaptive filtering concept; Aggregated distances; Euclidean metric

1. Introduction

Complementary deoxyribonucleic acid (cDNA) microarray technology [1,7] is an advanced tool used in the investigation of toxicological problems determined via cellular response to low-dose ionizing

* Corresponding author.
E-mail address: lukacr@ieee.org (R. Lukac)
URL: http://www.dsp.utoronto.ca/∼lukacr.

0165-0114/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2004.10.012

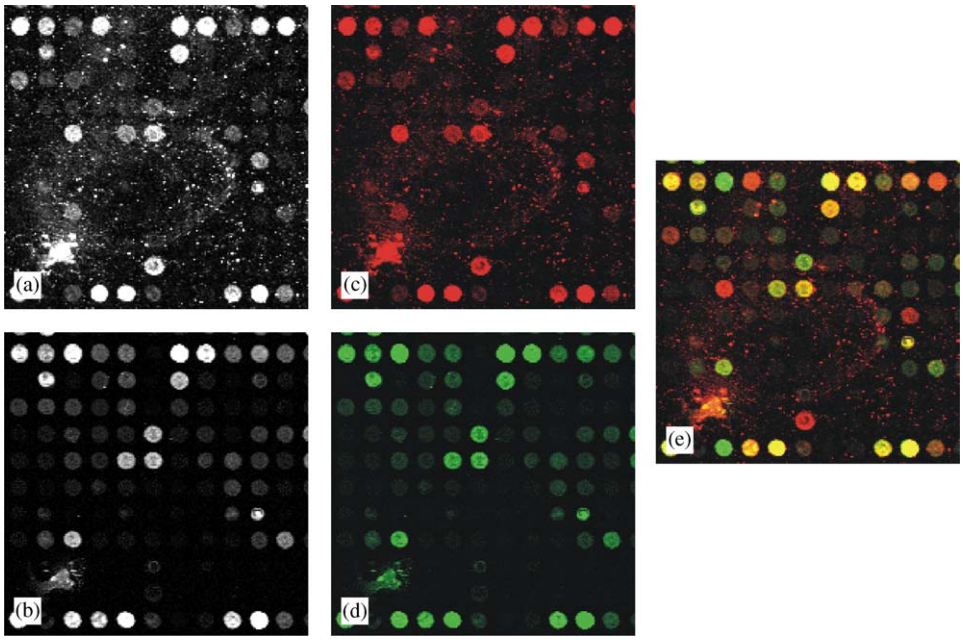

Fig. 1. Red–Green image channels of the cDNA microarray: (a) Red channel as a gray-scale image, (b) Green channel as a gray-scale image, (c,d) colored channels, (e) cDNA microarray visualized as the RGB image with zero B components.

radiation. This new methodological advancement is used to analyze changes in genome-wide patterns of gene expression in different populations of cells. It is therefore not surprising that cDNA technology is used to identify potentially hazardous substances, such as carcinogens and reproductive toxins.

A microarray (Fig. 1) is a collection of green, red and yellow discrete spots containing DNA, deposited on the surface of a microscope glass slide [6]. Each spot contains multiple copies of a single DNA sequence. The spots occupy a small fraction of the image area and they have to be individually located and isolated from the image background prior to the estimation of its mean intensity.

The cDNA microarray image represents a two-channel Red–Green (RG) image (Figs. 1 and 2) and thus should be considered a vector-valued image signal or vector field [17]. In this set-up, variation in image background and image artifacts, as well as spot sizes and positions represent the major sources of uncertainty in spot finding and gene expression determination. In particular, non-specifically bounded DNA or dye molecules and the natural fluorescence of the glass slide result in a substantial noise floor in the microarray image. These noise impairments along with discrete image artifacts necessitate the use of image filtering prior to subsequent analysis [6,12]. The purpose of subsequent processing tasks is to analyze the spots, normalize the arrays and to identify which genes each type of cells are expressing. Due to thousands of spots, the procedure should be fully automated. Note that any errors or noise introduced is being propagated through subsequent analysis. Therefore, removing noise in cDNA microarray images makes the spots easier to detect and analyze, and in the end of the process obtained gene expression measurements are more accurate to interpret [30].

It has been observed that changes of the pixel intensities from the foreground to the background can be attributed to the Gaussian nature of noise corrupting cDNA chips [19]. Isolated discrete artifacts and

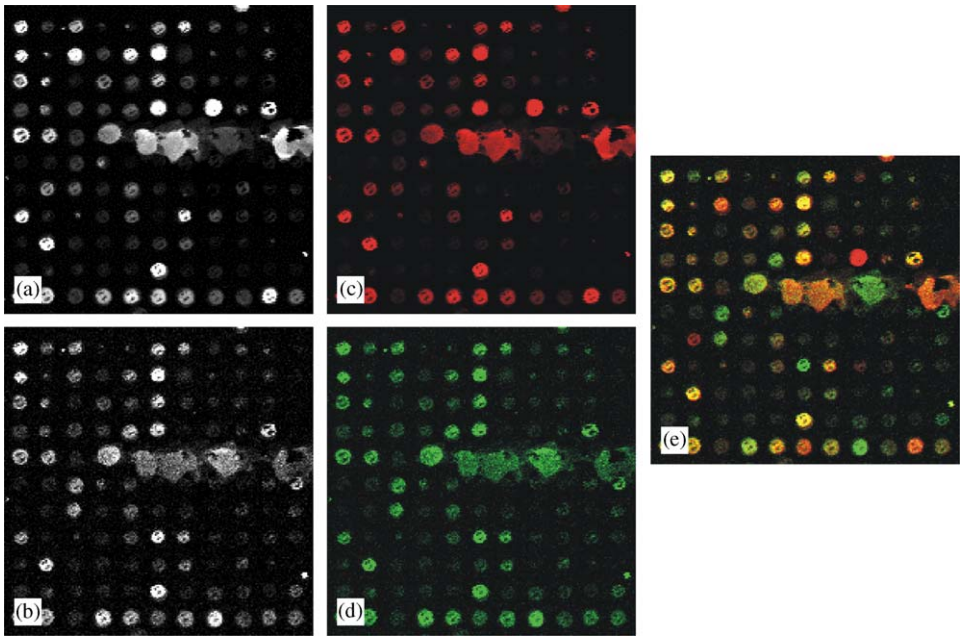

Fig. 2. Another example of the cDNA microarray and real noise affecting both Red and Green image channels: (a) Red channel as a gray-scale image, (b) Green channel as a gray-scale image, (c,d) colored channels, (e) cDNA microarray visualized as the RGB image with zero B components.

outliers present in the cDNA microarray image can be attributed to impairments which are impulsive in nature [14]. In conventional image processing applications, noise corruption of such nature is most often modelled through a mixture of additive Gaussian noise and impulsive noise [26]. Due to the vectorial nature of the microrray image data and taking into consideration the noise characteristics in cDNA microarray images, fuzzy logic-based techniques are adopted here to form a two-channel image processing framework and used as a solution to the estimation problem.

Through the utilization of linguistic terms, a fuzzy rule-based approach to signal processing allows for the incorporation of human knowledge and intuition into the design, which cannot be achieved via traditional mathematical modelling techniques. However, there is no optimal way to determine the number and type of fuzzy rules required for the fuzzy image operation. Usually, a large number of rules are necessary and the designer has to compromise between quality and number of rules used, since even for a moderate processing window, a large number of linguistic rules are required. Therefore, data-dependent filters adopting fuzzy reasoning have been proposed to overcome these difficulties. These designs combine fuzzy concepts, such as membership functions, and fuzzy aggregators with nonlinear filters [23].

This paper builds on the developments in the area of data-dependent fuzzy systems suitable for vector signal processing. It has been observed that fuzzy filters can be successfully applied to many noise removal problems [22] and thus, there is a supposition that this signal processing framework may also be suitable in cDNA microarrays. To the best of the authors knowledge it is the first time ever that vector processing filters based on fuzzy logic concepts are designed and used to attenuate noise in two-channel images such as cDNA microarrays. Since in this application environment the original noise-free data are not available to the designers, the problem of determining the optimal filtering structure becomes more challenging

compared to the case of natural color image filtering design. To minimize distortion introduced during processing and to increase the precision in estimating true cDNA values, the filtering scheme considered here utilizes weighting coefficients which are adaptively determined on the basis of local signal context expressed via aggregated distances between the vectorial inputs. The adaptive approaches adopted here for the two-channel, cDNA microarray image processing integrate well-known concepts from the areas of fuzzy set theory, nonlinear filtering, multidimensional scaling and robust order-statistics. In addition to the classical vector distance based solutions, a novel design based on the color-ratio model is also introduced, studied and applied. Utilizing the existing correlation between the R and G channels in the cDNA image, the color-ratios serve as the input to filter's membership function. Since the color-ratio quantities have significantly decreased the high-frequency portion of the signal, the filtering procedure is able to faithfully preserve edges and structural content of the cDNA image.

The rest of this paper is organized as follows. In Section 2, the fundamentals of cDNA imaging are introduced. The formulation of the problem is given and the state-of-the-art in vector filtering techniques is discussed. A generalized framework for filtering noise in cDN A microarrays using fuzzy logic principles is introduced in Section 3. Motivation and design characteristics are discussed in detail. Variations of the proposed structure are recommended and analyzed with respect to their properties and parameters used. In Section 4, the proposed methods are tested in a variety of cDNA microarray images. Conclusions are offered in Section 5.

2. cDNA imaging

2.1. Biological background

Deoxyribonucleic acid (DNA) is a chemical structure that forms chromosomes. A piece of a chromosome that dictates a particular feature is called a gene. Note that many genes are used to specify features unique to each type of cell and DNA repair is one of the most critical cellular functions. Therefore, deficiencies in repair commonly relate to susceptibility to genetic diseases, chromosome abnormalities, and cancers. Recent developments in the microarray technology allow for looking at many genes at once and determine a particular cell type with specific gene expression [7,8].

2.2. Microarray basics

The cDNA microarray technology is based on arraying of known cDNA sequences, the so-called cDNA probes, onto a glass slide with labelled cDNA target sequences [3]. Expression arrays containing up to 80,000 probes are printed onto a $2 \times 4\,\mathrm{cm}^2$ area on a microscope glass slide. The arrays parameters are 75–100 μm probe diameter and 150 μm spacing between probes. Based on the widely used two-color Cy3/Cy5 system [18], the control sequence is usually fluor-tagged with green (Cy3), where as the experimental (target) sequence is tagged with red (Cy5). The hybridization procedure refers to the pairing of the fluorescent cDNA to the spotted DNA. Following a 16–24 h long washing at 65 °C temperature, the slides are scanned at the corresponding wavelengths (i.e. $\sim$ 540 nm for green and $\sim$ 630 nm for red) generating two 16-bit images [10]. The composed RG color image contains red, green and yellow spots which indicate that a particular gene is expressed in the control channel, experimental channel or both of them, respectively.

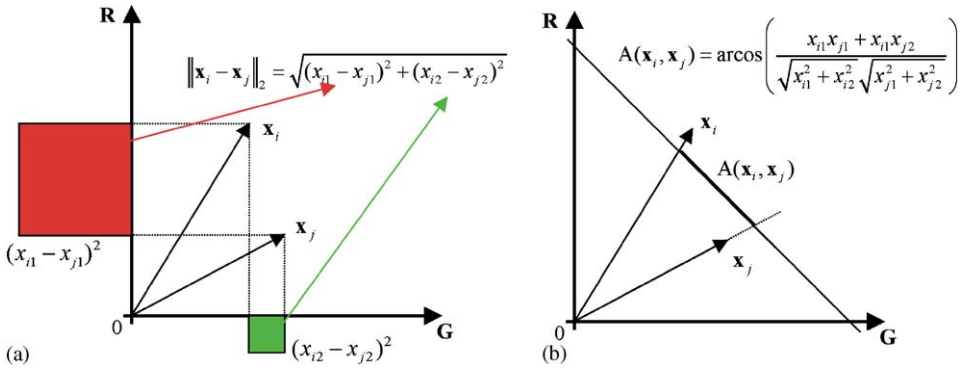

Fig. 3. Differences between the two-dimensional vectors $\mathbf{x}_i = [x_{i1}, x_{i2}]^{\mathrm{T}}$ and $\mathbf{x}_j = [x_{j1}, x_{j2}]^{\mathrm{T}}$ expressed in the RG color space: (a) expressed through the Euclidean metrics $\|\mathbf{x}_i - \mathbf{x}_j\|_2$ in the magnitude domain, (b) expressed through the angle $\mathrm{A}(\mathbf{x}_i, \mathbf{x}_j)$ in the directional domain.

2.3. Image representation

Let us consider, a $K_1 \times K_2$ two-channel image $\mathbf{x} : Z^2 \to Z^2$ representing a two-dimensional matrix of 2-component samples $\mathbf{x}_i = [x_{i1}, x_{i2}]^{\mathrm{T}}$. Note that cDNA microarrays represent RG images [14,25], as shown in Fig. 1 and 2. Components x_{ik}, for $k = 1, 2$ and $i = 1, 2, \ldots, Q$; $Q = K_1K_2$, represent the kth elements of the vectorial input $\mathbf{x}_i$.

Each two-channel sample $\mathbf{x}_i$ can be considered as a two-dimensional vector in the vector space (Fig. 3). As such, each vector is uniquely defined by its length (magnitude) and orientation (direction) in the vector space [17]. For a color vector $\mathbf{x}_i$, its magnitude $M_{\mathbf{x}} : Z^2 \to R^+$ is defined as follows:

$$M_{\mathbf{x}_i} = \|\mathbf{x}_i\| = \sqrt{(x_{i1})^2 + (x_{i2})^2}. \tag{1}$$

The directionality of $\mathbf{x}_i$ is expressed via $D_{\mathbf{x}} : Z^2 \to S$ defined as

$$D_{\mathbf{x}_i} = \frac{1}{\|\mathbf{x}_i\|}\mathbf{x}_i = \frac{1}{M_{\mathbf{x}_i}}\mathbf{x}_i, \tag{2}$$

$$d_{x_{ik}} = \frac{x_{ik}}{\|\mathbf{x}_i\|} = \frac{x_{ik}}{M_{\mathbf{x}_i}}, \quad \text{for } k = 1, 2, \tag{3}$$

where S is a unit ball in R^2 and $\|D_{\mathbf{x}_i}\| = 1$.

Assuming zero B components, a cDNA microarray can be easily visualized in the RGB color space or stored as a conventional RGB color image [17]. Following this interpretation, as shown in Fig.4, the magnitude $M_{\mathbf{x}_i}$ obtained in (1) and direction $D_{\mathbf{x}_i} = [d_{x_{i1}}, d_{x_{i2}}, 0]^{\mathrm{T}}$, for $D_{\mathbf{x}} : Z^2 \to S^2$, of the RGB vectors $\mathbf{x}_i = [x_{i1}, x_{i2}, 0]^{\mathrm{T}}$ of the three-channel image $\mathbf{x} : Z^2 \to Z^3$ constitute a measure of their brightness and chromaticity, respectively.

2.4. Problem formulation

Although recognition of spots in either control or experimental channels seems to be straightforward, the task is complicated and challenging. The cDNA microarray image suffers from high-level noise and

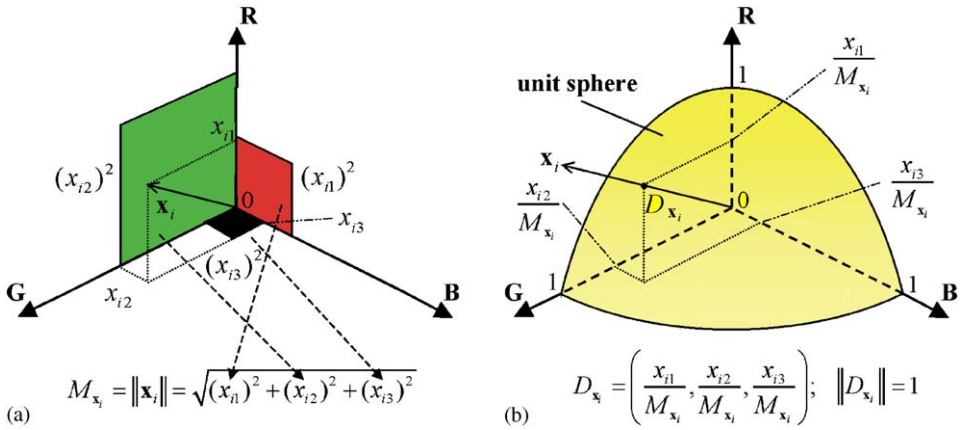

Fig. 4. Basic quantities of the color vector $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$ used in RGB color image processing: (a) brightness $M_{\mathbf{x}_i}$, (b) chrominance defined as the point $D_{\mathbf{x}_i}$ on unit sphere.

edge uncertainty [29]. A number of noise sources mostly in the form of photon noise, electronic noise, laser light reflection and dust on the slide contribute to impairments and defects. These imperfections along with the background fluorescence introduce into the image considerable variability in intensity both within and between the individual spots [30]. It has been observed that changes of the pixel intensities from the foreground to the background can be attributed to the Gaussian nature of noise [19]. Isolated discrete artifacts and outliers present in the cDNA microarray image can be attributed noise impulsive in nature [14]. Heterogeneous brightness of outliers as well as variability of multipixel artifacts in shape and size make spots hard to detect and remove automatically. Automated spot finding tools can mistakenly detect bright artifacts as spots. It has to be mentioned that spots are themselves of variable size and brightness and this significantly complicates the task.

It is therefore reasonable to assume that unprocessed cDNA microarray data can be represented via the commonly used additive noise model [4,22]:

$$\mathbf{x}_i = \mathbf{o}_i + \mathbf{v}_i, \tag{4}$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}]^{\mathrm{T}}$ represents the observed (noisy), two-channel cDNA sample, $\mathbf{o}_i = [o_{i1}, o_{i2}]^{\mathrm{T}}$ is the desired (noise free) sample, $\mathbf{v}_i = [v_{i1}, v_{i2}]^{\mathrm{T}}$ is the vector describing the impairment with i denoting the spatial position of the samples in the image array.

2.5. Background on vector signal processing

Based on these noise characteristics, local filtering operators are designed to replace the corrupted cDNA RG vectorial input with the vectors which are statistically close to its neighbors. Such a filter is operating on some type of sliding window $W = \{\mathbf{x}_i \in Z^2; i = 1, 2, \ldots, N\}$ of finite size N (Fig. 5). The filtering procedure usually affects one image sample as a time (mostly the sample $\mathbf{x}_{(N+1)/2}$ placed in the center of the window), changing its value through a function applied to a local neighborhood area $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ [21,26]. This window operator slides over the image to individually affect all the image pixels.

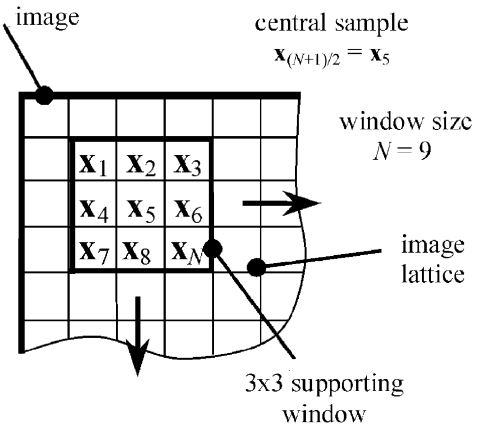

Fig. 5. Arrangements of the vectorial inputs in the sliding supporting window W.

2.5.1. Vector median filter

Probably the most well-known color image filter is the vector median filter (VMF) [4]. The VMF can be derived as a maximum likelihood estimate (MLE), when the underlying probability densities of $\mathbf{v}_i$ are double exponential. The output of the VMF scheme is the input vector $\mathbf{x}_{(1)} \in W$ minimizing the distance to other samples inside the input set W:

$$\min \underset{\mathbf{x}_{(1)} \in W}{\arg} \sum_{i=1}^{N} \|\mathbf{x}_{(1)} - \mathbf{x}_i\|_L, \tag{5}$$

where $\|\mathbf{x}_i - \mathbf{x}_j\|_L$ is the generalized Minkowski metric [23] determining here the distance between two cDNA image vectors which represent two-channel samples $\mathbf{x}_i = [x_{i1}, x_{i2}]^{\mathrm{T}}$ and $\mathbf{x}_j = [x_{j1}, x_{j2}]^{\mathrm{T}}$:

$$\|\mathbf{x}_i - \mathbf{x}_j\|_L = \left(\sum_{k=1}^{2} |x_{ik} - x_{jk}|^L \right)^{1/L}, \tag{6}$$

where L denotes the norm parameter, e.g. the city-block distance ($L = 1$) or Euclidean distance ($L = 2$) and x_{ik} is the kth element of $\mathbf{x}_i$.

The output of the VMF filter can be equivalently determined using the vector order-statistics. Let us denote

$$D_i = \sum_{j=1}^{N} \|\mathbf{x}_i - \mathbf{x}_j\|_L \quad \text{for } i = 1, 2, \ldots, N \tag{7}$$

as the aggregated measure associated with $\mathbf{x}_i$, then the ordered sequence of $D_1, D_2, \ldots, D_N$ is given by $D_{(1)} \leqslant D_{(2)} \leqslant \cdots \leqslant D_{(N)}$. Assuming that the ordering of $D_{(i)}$'s implies the same ordering of the input set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, the procedure results in ordered set $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(N)}$, where $\mathbf{x}_{(i)}$ is associated with $D_{(i)}$. In this case, the VMF output is defined as the lowest order-statistics $\mathbf{x}_{(1)}$, which is equivalent to the sample minimizing earlier definition (5).

Based on the VMF-framework operating on the magnitude of the vectorial inputs, the adaptive VMF scheme [14] and the digital path approach [25] were successfully used to remove noise in cDNA microarrays, which are the special case of a color image. Whereas the conventional VMF as well as the adaptive VMF scheme are designed to remove impulses or isolated outliers, the digital path filtering technique is capable of removing noise which is statistically close to the additive Gaussian noise model.

2.5.2. Vector directional filters

Within the vector processing framework, the class of vector directional filters (VDF) [28], is operating on the directional domain of the color image. It has been observed that the output of the basic vector directional filter (BVDF) defined within the VDF class is the color vector $\mathbf{x}_{(1)} \in W$ whose direction is the MLE of directions of the input vectors [20]. Thus, the BVDF output $\mathbf{x}_{(1)}$ minimizes the angular ordering criteria to other samples inside the sliding filtering window W:

$$\min \underset{\mathbf{x}_{(1)} \in W}{\arg} \sum_{i=1}^{N} \mathrm{A}(\mathbf{x}_{(1)}, \mathbf{x}_i), \tag{8}$$

where

$$\mathrm{A}(\mathbf{x}_i, \mathbf{x}_j) = \arccos\left(\frac{\mathbf{x}_i \cdot \mathbf{x}_j}{|\mathbf{x}_i||\mathbf{x}_j|}\right) = \arccos\left(\frac{x_{i1}x_{j1} + x_{i2}x_{j2}}{\sqrt{x_{i1}^2 + x_{i2}^2}\sqrt{x_{j1}^2 + x_{j2}^2}}\right) \tag{9}$$

represents the angle between two cDNA image vectors $\mathbf{x}_i$ and $\mathbf{x}_j$.

Similarly as in the VMF scheme, the BVDF filter can be equivalently expressed through the order-statistic approach. Assuming that each vectorial input is associated with the aggregated angular measure

$$\alpha_i = \sum_{j=1}^{N} \mathrm{A}(\mathbf{x}_i, \mathbf{x}_j) \quad \text{for } i = 1, 2, \ldots, N \tag{10}$$

and the ordering of α_i's implies the same ordering of the input vectors $\mathbf{x}_i$, the sample $\mathbf{x}_{(1)}$, i.e. the lowest ranked vector or the lowest order-statistics, associated with the minimum aggregated angular measure $\alpha_{(1)} \in \{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ is the BVDF output.

The angular minimization approach is useful for color data as it preserves chromaticity. Since automated gene expression techniques utilize the color information to indicate a channel (control and/or experimental) and spots as well as noise vary in color, directional processing may be also useful in cDNA microarrays. Except pure directional processing, the filtering techniques can be designed to combine both directional and magnitude processing. Such a filtering class includes the generalized vector directional filter (GVDF) and the double window GVDF [28]. These filters eliminate the color vectors with atypical directions in the vector space and the vectors with the most similar orientation are then processed according to their magnitude. Thus, the GVDF splits the color image processing into the directional processing and the magnitude processing. Selection weighted vector directional filters [15,16] minimize the aggregated weighted angles between the color vectors and improve the detail-preserving filtering characteristics of the conventional VDF schemes.

3. Fuzzy vector filtering framework

Many fuzzy techniques developed for low-level image processing tasks have been proposed for monochromatic images [2,5,24]. Most of these image processing tasks relate to gray-scale edge detection, image sharpening, video coding and noise removal. Recent works incorporate fuzzy logic, fuzzy set theory and fuzzy rules to develop efficient and cost-effective color image processing systems [13,22,23,27]. Before the introduction of adaptive fuzzy vector filters designed to remove noise and simultaneously preserve the color/structural content of the cDNA microarrays, let us briefly describe fuzzy logic basics.

3.1. Fuzzy logic basics

For understanding of fuzzy logic it is important to discuss fuzzy sets [33]. Fuzzy sets are commonly considered as sets with unsharp boundaries and thus, they are better suited to deal with tolerance for some inexactness and imprecision compared to a conventional set theory approach. The characteristic function of a fuzzy set is called the membership function and depending on definition it can take a variety of different shapes [32,33].

The fuzzy rule-based system shown in Fig. 6 utilizes fuzzy logic to convert the linguistic term into the fuzzy quantities. In order to use fuzzy logic in a particular processing task, the fuzzification procedure transforms the input data into fuzzy values. These are processed in the inference engine using the set of IF-THEN-ELSE fuzzy-rules usually constituted in the if-then format. Defuzzification procedure converts the fuzzy output into the original (crisp) application format.

3.2. Adaptive filter design based on vectorial inputs

Since the images are highly non-stationary in edges and due to the difficulty in distinguishing between noise and edge pixels, fuzzy sets are highly appropriate for image filtering tasks [23]. A number of fuzzy filters adopt a window-based, rule-driven approach leading to data-dependent fuzzy solution. Using a bank of fuzzy rules the fuzzy filter directly yields the filtered output taking into account selected patterns in the neighborhood of the element to be processed. Since the antecedents of fuzzy rules can be composed of several local characteristics, it is possible for the fuzzy filter to adapt to local data. Local correlation in the data is utilized by applying the fuzzy rules directly on the signal elements which lie within the operational window. Thus the output of the fuzzy filter depends on the fuzzy rule and the defuzzification process, which combines the effects of the different rules into an output value.

To design the fuzzy-system, as the one shown in Fig. 6, for a particular processing task (e.g. filtering), the fuzzy rules must be optimally set using the optimization procedure [11]. Such an approach is appropriate in applications, where the original signal and the time enough for learning are available. Most often the training is performed via genetic algorithm optimization [9] which can significantly increase the overall computational cost. Such an approach for color images has been introduced in [27]. With respect to the application considered here, where the original data are not available, the fuzzy vector filters are designed to remove noise in cDNA microarrays operating directly on (noisy) cDNA vectorial inputs. In this way, no fuzzy rules are required in designing the proposed microarray image processing tools.

Since the most commonly used method to decrease the level of random noise present in the signal is smoothing, an averaging operation is required in order to replace the noisy vector $\mathbf{x}_{(N+1)/2}$ at the window center with a suitable vector representative for the local image area $W = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. The general

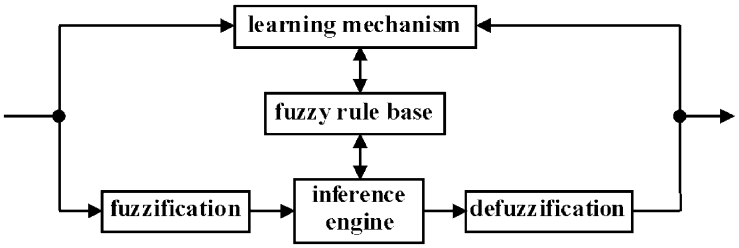

Fig. 6. Block scheme of the conventional fuzzy system with the learning mechanism.

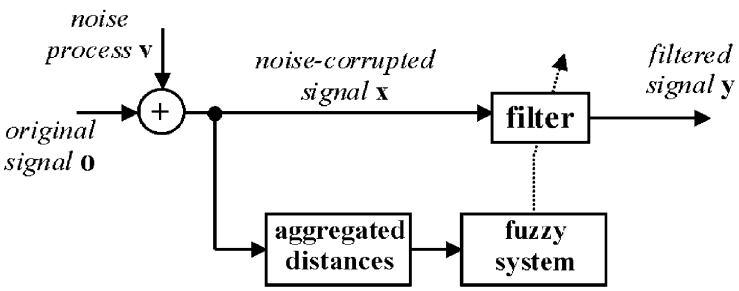

Fig. 7. Block scheme of the adaptive fuzzy vector filter considered here.

form of the system presented here is given as a fuzzy weighted average [22] of the input vectors inside the supporting window W:

$$\mathbf{y} = f\left(\sum_{i=1}^{N} w_i \mathbf{x}_i\right) = f\left(\sum_{i=1}^{N} \mu_i \mathbf{x}_i / \sum_{i=1}^{N} \mu_i\right), \tag{11}$$

where $f(\cdot)$ is a nonlinear function that operates over the weighted average of the input set and w_i is the filter weight and μ_i is the fuzzy membership function, both associated with the input color vector $\mathbf{x}_i$. The weights w_i, for $i = 1, 2, \ldots, N$, provide the degree to which an input vector contributes to the output of the filter. The relationship between the central sample $\mathbf{x}_{(N+1)/2}$ and its neighbors $\mathbf{x}_i \in W$ should be reflected in the decision for the weights of the filter. Since the solution is unbiased, the fuzzy estimator (11) must satisfy two conditions [22]: i) $w_i \geqslant 0$, and ii) $\sum_{i=1}^{N} w_i = 1$.

Operating on the cDNA vectorial inputs $\mathbf{x}_i$, the weights w_i of (11) are determined adaptively here using functions of a distance criterion between the input vectors (Fig. 7). Such a distance criterion can be based on Minkowski metrics (6), angular measure (9) or other distance measures which can be found in [23]. The weights can be considered to be a membership function based on a given vectorial input set W. As it can be seen from (11), the proposed fuzzy vector filtering scheme uses the membership values μ_i to calculate the final filtered output. Using a distance between input vectors, a membership function value can be used to quantify the degree of similarity of the sample $\mathbf{x}_i$ to the other samples in W. Such a membership function can be defined as

$$\mu_i = \frac{1}{1 + f(d(\mathbf{x}_i, \mathbf{x}_j))}, \tag{12}$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ denotes the distance between input vectors $\mathbf{x}_i$ and $\mathbf{x}_j$. Based on the above definition, $\mu_i \to 0$ for $d(\mathbf{x}_i, \mathbf{x}_j) \to \infty$ and $\mu_i = 1$ for $d(\mathbf{x}_i, \mathbf{x}_j) = 0$. Depending on the specific distance measure that is applied to the input data, a different fuzzy membership function can be devised.

Since the relationship between distances measured in physical units and perception is generally exponential [22], an exponential type of function maybe suitable to be used in the weighting formulation. Utilizing the sigmoidal membership function and the aggregated chrominance (angular) criteria α_i of (10), the weight adaptation in (11) is performed as follows:

$$\mu_i = \beta \left(1 + \exp\{\alpha_i\}\right)^{-r}, \tag{13}$$

where β is a normalizing constant and r is a parameter adjusting the weighting effect of the membership function. Based on the aggregated Minkowski metrics D_i of (7), the adaptation formula (13) is redefined as follows:

$$\mu_i = \beta \left(1 + \exp\{D_i\}\right)^{-r}. \tag{14}$$

Thus, the proposed fuzzy vector filters can operate either on magnitude or direction of the vectorial inputs. Since α_i and D_i significantly differ in the values, (13) and (14), respectively, produce different degrees of membership and the filtering scheme (11) will result in two different outputs. This increases the degree of freedom in design of the proposed method.

To avoid optimization of fuzzy rules, the proposed adaptive fuzzy system utilizes the inference engine in the form of transformed distance metrics between the vectorial inputs of W. Since the output of the adaptive fuzzy system considered here depends on local neighborhood W, the system is capable of tracking the varying image and noise statistics. The training or learning of the weighting coefficients is only based on local image features without the use of linguistic fuzzy rules or local statistics estimation.

The cost-effective defuzzification step is realized via the filtering procedure, which replaces the noise-corrupted cDNA image vector $\mathbf{x}_{(N+1)/2}$ located at the window center by a prototype vector $\mathbf{y}$. Such an output has the minimization property, since it determines the most appropriate vectorial value to represent a collection of cDNA inputs $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ whose membership functions have been constructed over a universe of discourse. It has been proved that a widely used centroid defuzzification approach, the so-called center of gravidity, generates the defuzzified value which is at the center of the values of a fuzzy set [22]. Therefore, the generalized filtering scheme of (11) based on membership functions defined via the distance concept also satisfies minimization property required in noise removal applications.

The defuzzified vector $\mathbf{y}$ obtained through the centroid defuzzification approach is not part of the original input set W. In some image processing application [17], constrained solutions such as the VMF of (5) and the BVDF of (8) that can provide higher preservation of image details compared to the unconstrained solutions, are required. Therefore, the different defuzzification strategy should be used and the adaptive weights in (11) can be re-defined as follows:

$$w_i = \frac{\mu_i^{\lambda}}{\sum_{i=1}^{N} \mu_i^{\lambda}} = \frac{\left(\mu_i / \mu_{(\max)}\right)^{\lambda}}{\sum_{i=1}^{N} \left(\mu_i / \mu_{(\max)}\right)^{\lambda}}, \tag{15}$$

where $\mu_{(\max)} \in \{\mu_1, \mu_2, \ldots, \mu_N\}$ is the largest membership value.

Given that $\mu_i < \mu_{(\max)}$ and $\lambda \to \infty$, the weighting coefficients of (15) are obtained via the maximum defuzzifier strategy as follows [23]:

$$w_i = \begin{cases} 1 & \text{if } \mu_i = \mu_{(\max)}, \\ 0 & \text{if } \mu_i \neq \mu_{(\max)}. \end{cases} \tag{16}$$

If the maximum value occurs at a single point only, the output of an adaptive fuzzy system is defined as

$$\hat{\mathbf{y}} = \mathbf{x}_i, \text{ for} \mu_i = \mu_{(\max)}. \tag{17}$$

In this case, the fuzzy adaptive filter is designed to perform a selection filtering operation, which identifies the one of the samples inside the processing window W as the filter output. This property is essential for preserving the structural image content. However, unlike selection vector filters which are primary designed to remove impulsive noise [16], the proposed fuzzy filters can be optimized for any noise model by appropriately tuning their membership function. Therefore, they can be used to remove the additive noise corrupting cDNA chips.

3.3. Adaptive filter design based on color-ratios

The proposed here filtering class can exploit the correlation between the two channels in the cDNA image using not only differences between the vectorial inputs but Red/Green (R/G) ratios as well. The color-ratio approach takes advantage of the expected relative uniformity of the local R/G ratios. Due to the decreased high-frequency portion of the signal, color-ratio based processing is able to preserve edges and structural content of the cDNA image better than the conventional approaches operating in the intensity domain.

Let us consider the cDNA vectorial inputs $\mathbf{x}_i$, for $i = 1, 2, \ldots, N$. Each sample $\mathbf{x}_i$ can be used to produce the R/G color ratio x_{i1}/x_{i2}. Using the concept of aggregated distances between the vectorial inputs determined inside the processing window W, procedure (7) or (10) reduces to calculation of absolute differences between color ratios:

$$c_i = \sum_{j=1}^{N} |x_{i1}/x_{i2} - x_{j1}/x_{j2}| \quad \text{for } i = 1, 2, \ldots, N. \tag{18}$$

Based on aggregated scalar measure (18), the fuzzy weights are calculated as follows:

$$\mu_i = \beta(1 + \exp\{c_i\})^{-r}, \tag{19}$$

where r is used to tune the weighting effect of the membership function and β is a normalizing constant.

Using the color ratios x_{i1}/x_{i2}, for $i = 1, 2, \ldots, N$, and the defuzzified weighting coefficients $w_i = \mu_i / \sum_{i=1}^{N} \mu_i$, the output vector $y = [y_1, y_2]^{\mathrm{T}}$ is obtained as follows:

$$y_1 = x_2^* \sum_{i=1}^{N} w_i x_{i1}/x_{i2}, \tag{20}$$

$$y_2 = x_1^* \left[\sum_{i=1}^{N} w_i x_{i1}/x_{i2} \right]^{-1}, \tag{21}$$

where $\mathbf{x}^* = [x_1^*, x_2^*]^T$ is a vector, whose components are used to normalize the output color ratio $\sum_{i=1}^{N} w_i x_{i1}/x_{i2}$ to the individual intensities y_1 and y_2 corresponding to the recover Red and Green color channels, respectively.

The normalization vector $\mathbf{x}^*$ can be considered as equivalent to a robust estimate which statistically represents the input set W. Using robust order-statistics [21] principle, $\mathbf{x}^*$ is defined as the component-wise median filter (MF) [31]. Note that in most cases $\mathbf{x}^*$ is considered to be the output of a filtering procedure. In this scheme $\mathbf{x}^*$ serves only as a normalization factor. Thus, the adaptive fuzzy filter based on color-ratios combines the fuzzy weighted averages and the nonlinear median operation to normalize the output of the color-ratio based filter.

4. Application to cDNA microarray images

A number of cDNA microarray images have been used to evaluate the proposed filtering framework. Examples are shown in Fig. 8a and 9a. These images have been captured using laser microscope scanners. The images vary in complexity and noise appearance. Note that all filtering results presented in this paper were obtained with a 3×3 square window, i.e. for $N = 9$.

The adaptive fuzzy filters (AFF) considered here for comparison purposes are the magnitude-based AFF_1 defined by (11) and (14), the directional AFF_2 of (11) and (15), and the color-ratio-based AFF_3 defined in (18)–(21). The noise attenuation properties and detail-preserving capability of the adaptive fuzzy filters are compared, in terms of performance with a set of color image filters widely used in color image processing applications. The comparisons include component-wise MF [31], VMF [4], BVDF [28] and adaptive VMF (AVMF) [14].

Since the original signal is not available, we make use of the subjective evaluation approach (Table 1). In this approach, the image quality is evaluated with respect to the structural content (edges, textures and fine details) preservation and the presence of unremoved impulses or introduced artifacts as a result of faulty processing.

Table 2 allows for the comparison of a variety of filtering techniques applied to cDNA microarray images. The results indicate that the proposed AFF_3 provides best results among the tested filters and its performance is satisfactory robust for a wide range of cDNA microarray images.

Figs. 8,9 show that all the proposed AFF schemes achieve excellent balance between signal-detail preservation and noise attenuation. However, only AFF_3 is capable of removing the complete set of impairments (including strong varied background intensities) introduced in course of the cDNA microarray technology.

By the visual inspection of the corresponding enlarged parts of the obtained outputs (Figs.10 and 11) it can be easily observed that the filtering techniques such as MF, VMF and AVMF excellently suppress impulses present in the image. However, these schemes are not capable of removing variations in background. Moreover, the BVDF scheme enhances this noise and introduces additional impairments into the image. The proposed AFF_1 and AFF_2 attenuate noise including background imperfections in a better way compared to the previous schemes. It should be emphasized that only the color-ratio-based AFF_3 technique among all the tested schemes removes fluorescence artifacts and background noise. This is also confirmed by the results depicted in Figs.12 and 13 corresponding to the 3-D plots of cDNA microarray image intensity corresponding to the selected areas of cDNA image inputs. Note that the areas plotted in these figures have been found difficult, in terms of spot preservation, for the examined

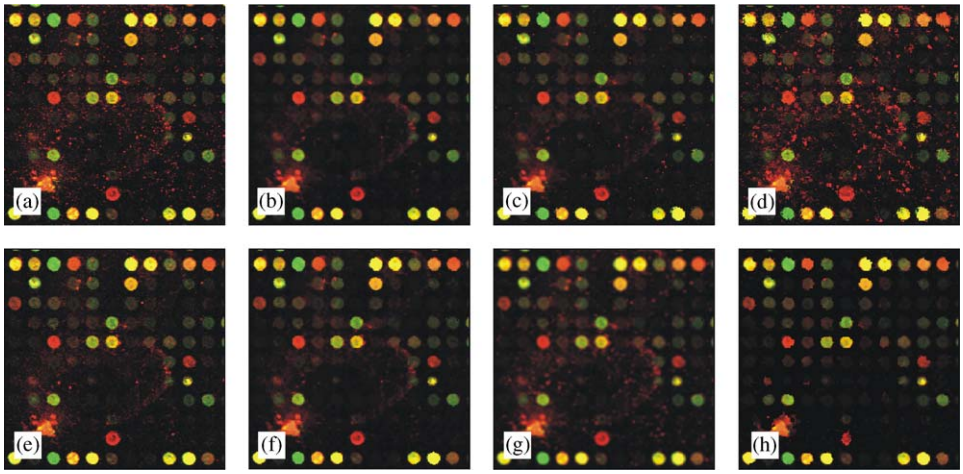

Fig. 8. Results obtained using the cDNA image with noise impairments affecting mainly the Red channel: (a) original image, (b) MF output, (c) VMF output, (d) BVDF output, (e) AVMF output, (f) AFF_1 output, (g) AFF_2 output, (h) AFF_3 output.

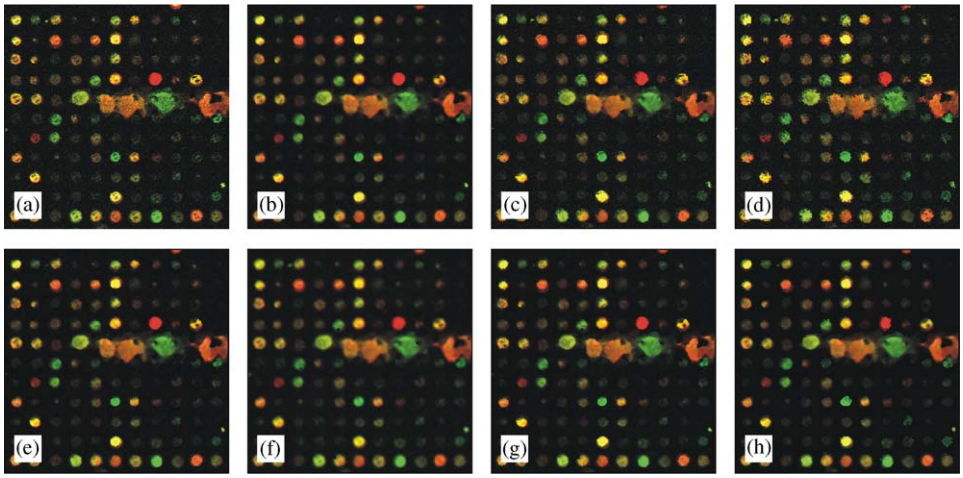

Fig. 9. Results obtained using the cDNA image with noise present mainly in the Green channel: (a) original image, (b) MF output, (c) VMF output, (d) BVDF output, (e) AVMF output, (f) AFF_1 output, (g) AFF_2 output, (h) AFF_3 output.

filtering schemes. It can be seen that the VMF does not remove variations from the backgrounds, whereas the proposed AFF_3 scheme produces output intensities close to the ideal case.

Apart from the numerical behavior (actual performance) of any algorithm, its computational complexity is a realistic measure of its practicality and usefulness. Therefore, the selected filtering classes are analyzed here in terms of normalized operations, such as additions (ADDs), subtractions (SUBs), multiplications (MULTs), divisions (DIVs), square roots (SQRTs), comparisons (COMPs), exponents (EXPs), absolute values (ABVs) and arc cosines (ARCCOSs). Table 3 summarizes the total number of operations for VMF, BVDF and AFF schemes. The computational complexity analysis of the adaptive designs requires knowledge of the membership function used to calculate the adaptive weights and the exact form of the

Table 1
Subjective image evaluation guidelines

Score	Overall evaluation of the distortion	Noise removal evaluation
1	Very disruptive	Poor
2	Disruptive	Fair
3	Destructive but not disruptive	Good
4	Perceivable but not destructive	Very good
5	Imperceivable	Excellent

Table 2
Filters' practicability expressed via subjective evaluation

Filter	Reference	Score
MF	[31]	4.05
VMF	[4]	4.00
BVDF	[28]	2.80
AVMF	[14]	3.90
AFF_1	(11), (14)	4.15
AFF_2	(11), (15)	3.95
AFF_3	(18)–(21)	4.80

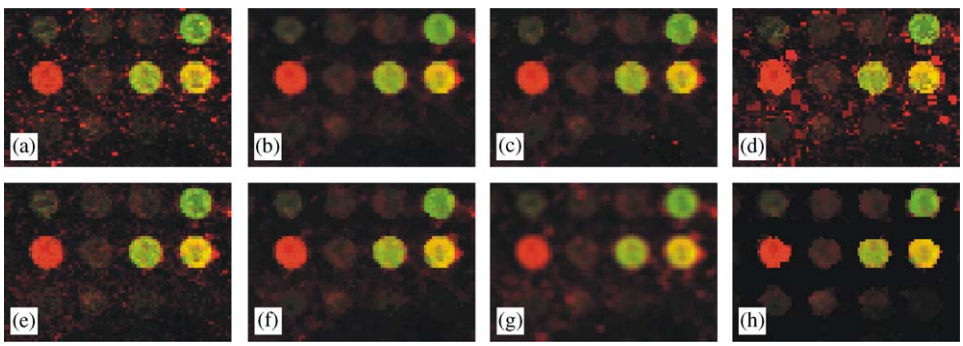

Fig. 10. Enlarged parts of the images shown in Fig. 8: (a) original image, (b) MF output, (c) VMF output, (d) qBVDF output, (e) AVMF output, (f) AFF_1 output, (g) AFF_2 output, (h) AFF_3 output.

selected distance measure used. The computationally intensive part of the AFF_1 and AFF_2 schemes rests in calculating the distance. This part, however, is common to both VMF and BVDF techniques. On the other hand, due to absolute differences between the color-ratios, the AFF_3 is a computationally most attractive case among the considered filtering schemes.

In conclusion, the following can be stated:

- In the proposed framework, there is no requirement for fuzzy rules or explicit determination of local statistics. Features extracted from local data, in the form of aggregated vector distances between the

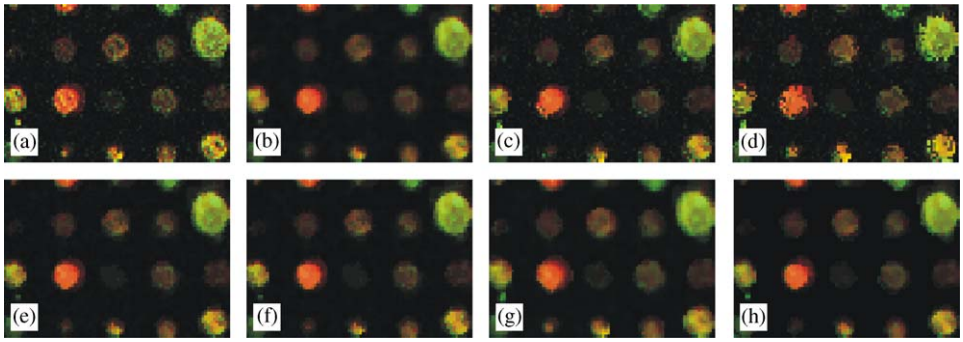

Fig. 11. Enlarged parts of the images shown in Fig. 9: (a) original image, (b) MF output, (c) VMF output, (d) BVDF output, (e) AVMF output, (f) AFF_1 output, (g) AFF_2 output, (h) AFF_3 output.

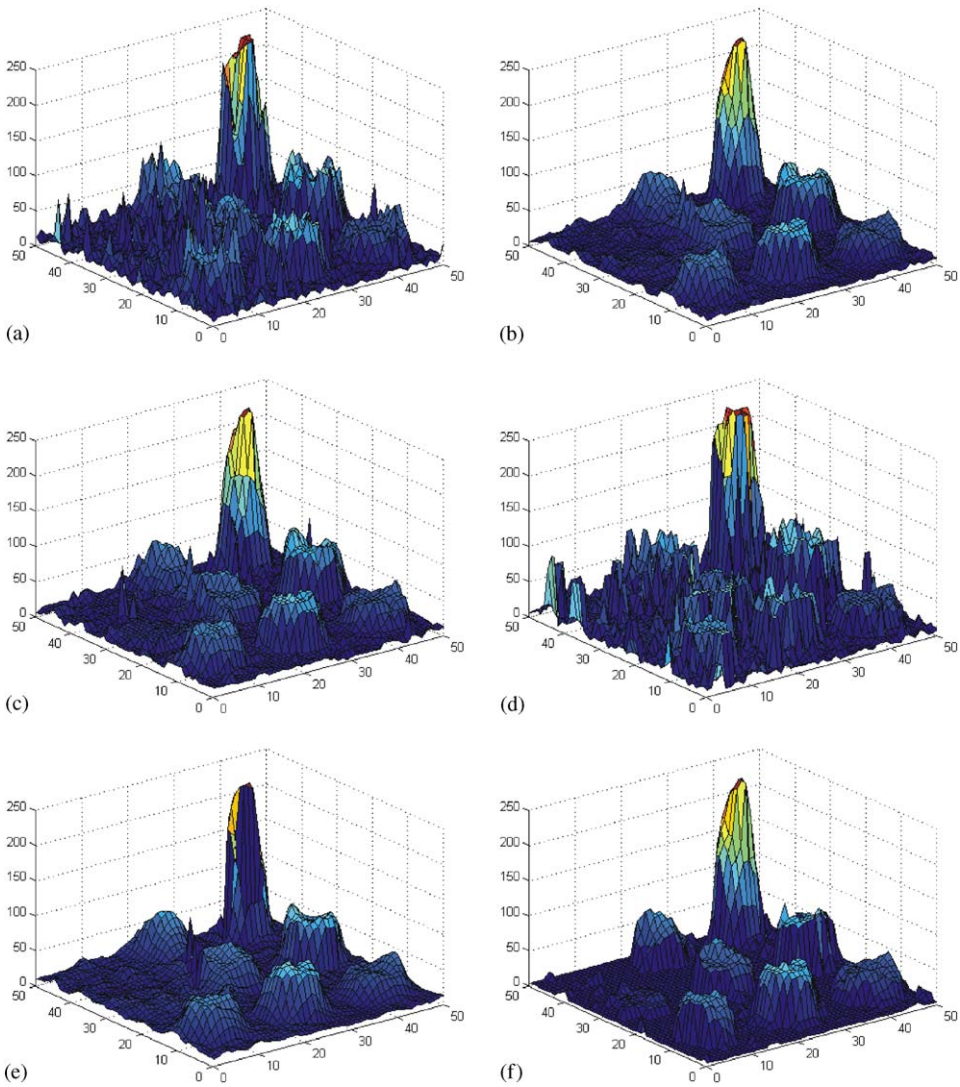

Fig. 12. Microarray image intensity plotted for selected parts of the achieved results: (a) original image, (b) MF output, (c) VMF output, (d) BVDF output, (e) AFF_1 output, (f) AFF_3 output.

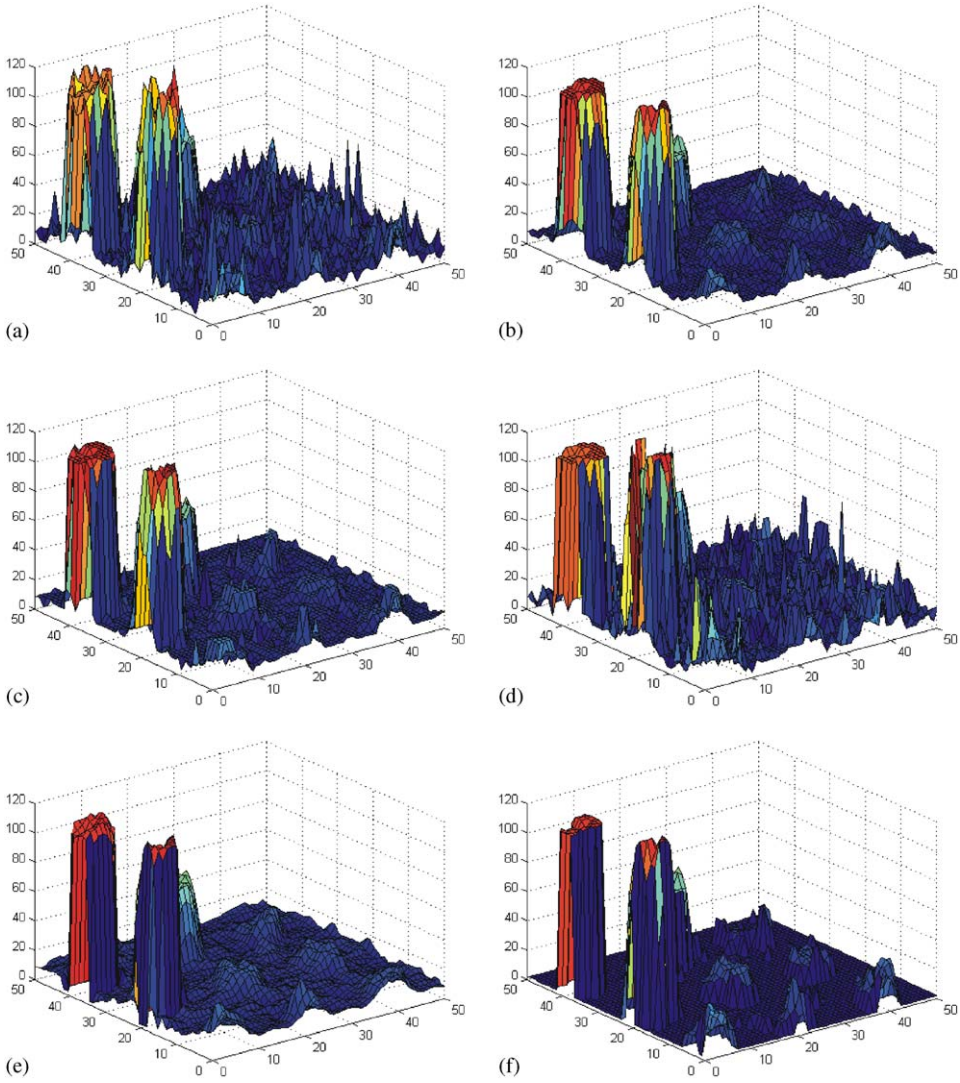

Fig. 13. 3-D plots of microarray image intensity selected in the areas with strong variations in the image background: (a) original image, (b) MF output, (c) VMF output, (d) BVDF output, (e) AFF_1 output, (f) AFF_3 output.

cDNA vectors or aggregated absolute differences between the color-ratios, are used as inputs to the membership function.

- The proposed fuzzy filters remove noise present in the cDNA microarray images in a robust way.
- The signal-detail preserving capability of the filters designed within the fuzzy filtering framework is sufficient for a particular task (cDNA image processing).
- The proposed fuzzy filtering framework produces better results than well-known vector filtering schemes (VMF, BVDF) based on a sliding supporting window.

Table 3
Cost-effectiveness of 3×3 vector filtering schemes expressed via the number of elementary operations for a complete processing cycle

Operation	ADDs	SUBs	MULTs	DIVs	SQRTs	EXPs	ARCCOSs	ABVs	COMPs
VMF	108	72	72	—	36	—	—	—	8
BVDF	180	—	162	36	9	—	36	—	8
AVMF	252	72	288	—	36	—	—	—	25
AFF_1	133	72	89	18	36	9	—	—	—
AFF_2	205	—	179	54	9	9	36	—	—
AFF_3	97	36	19	27	—	9	—	36	58

- The filter complexity of AFF_1 and AFF_2 techniques is comparable with widely used VMF and BVDF schemes. The AFF_3 approach based on the color-ratios is the most cost-effective solution among the considered vectorial schemes.

5. Conclusion

The paper introduced a new generalized cDNA processing tool based on principles of fuzzy logic. The proposed adaptive fuzzy filters operate either on cDNA vectorial inputs or color-ratios defined over cDNA data. The behavior of the introduced filtering class was analyzed in details. Simulation results and comparisons reported here indicate that the proposed framework is sufficiently robust and capable of removing noise in cDNA microarray images while preserving required data features for subsequent analysis.

Acknowledgements

The work of the first author is supported by a NATO/NSERC Science ward.

References

[1] N. Ajay, T. Tokuyasu, A. Snijders, R. Segraves, D. Albertson, D. Pinkel, Fully automatic quantification of microarray image data, Genome Res. 12 (2002) 325–332.
[2] K. Arakawa, Median filters based on fuzzy rules and its application to image processing, Fuzzy Sets and Systems 77 (1996) 3–13.
[3] P. Arena, M. Bucolo, L. Fortuna, L. Occhipinty, Cellular neural networks for real-time DNA microarray analysis, IEEE Eng. Med. Biol. 21 (2002) 17–25.
[4] J. Astola, P. Haavisto, Y. Neuvo, Vector median filters, Proc. IEEE 78 (1990) 678–689.
[5] Y.S. Choi, R. Krishnapuram, A robust approach to image enhancement based on fuzzy logic, IEEE Trans. Image Process. 6 (1997) 808–825.
[6] J. Dopazo, Microarray Data Processing and Analysis, in: S.M. Lin, K.F. Johnson (Eds.), Microarray Data Analysis II Kluwer Academic Publihers, Dordrecht, 2002, pp. 43–63.
[7] M.B. Eisen, P.O. Brown, DNA arrays for analysis of gene expression, Methods Enzymol. 303 (1999) 179–205.
[8] V. Filkov, S. Skiena, J. Zhi, Analysis techniques for microarray time-series data, J. Comput. Biol. 9 (2002) 317–330.

[9] D. Goldberg, Genetic Algorithms in Search, Optimisation, and Machine Learning, AddisonWesley, Reading, MA, 1989.
[10] J. Goncalves, W.L. Marks, Roles and requirements for a research microarray database, IEEE Eng. Med. Biol. 21 (2002) 154–154.
[11] L.R. Hou, Z. Yi, Fuzzy logic controller based on genetic algorithms, Fuzzy Sets and Systems 83 (1996) 1–10.
[12] L. Hsiao, R. Jensen, T. Yoshida, K. Clark, J. Blumenstock, S. Gullans, Correcting for signal saturation errors in the analysis of microarray data, Biotechniques 32 (2002) 330–336.
[13] L. Khriji, M. Gabbouj, Adaptive fuzzy order statistics-rational hybrid filters for color image processing, Fuzzy Sets and Systems 128 (2002) 35–46.
[14] R. Lukac, B. Smolka, Application of the adaptive center-weighted vector median framework for the enhancement of cDNA microarray images, Special Issue on Cancer Growth and Progression, Mathematical Problems and Computer Simulations, Internat. J. Appl. Math. Comput. Sci. 13 (2003) 369–383.
[15] R. Lukac, Adaptive color image filtering based on center-weighted vector directional filters, Multidimensional Systems Signal Process. 15 (2004) 169–196.
[16] R. Lukac, B. Smolka, K.N. Plataniotis, A.N. Venetsanopoulos, Selection weighted vector directional filters, Special Issue on Colour for Image Indexing and Retrieval, Comput. Vision Image Understanding, 94 (2004) 140–167.
[17] R. Lukac, B. Smolka, K.N. Plataniotis, A.N. Venetsanopulos, Vector filtering for color imaging, Special Issue on Color Image Processing, IEEE Signal Process. Mag. 22 (2005).
[18] R. Nagarajan, C.A. Peterson, Identifying spots in microarray images, IEEE Trans. Nanobiosci. 1 (2002) 78–84.
[19] R. Nagarajan, Intensity-based segmentation of microarrays images, IEEE Trans. Med. Imaging 22 (2003) 882–889.
[20] N. Nikolaidis, I. Pitas, Nonlinear processing and analysis of angular signals, IEEE Trans. Signal Process. 46 (1998) 3181–3194.
[21] I. Pitas, A.N. Venetsanopoulos, Order statistics in digital image processing, Proc. IEEE 80 (1992) 1892–1919.
[22] K.N. Plataniotis, D. Androutsos, A.N. Venetsanopoulos, Adaptive fuzzy systems for multichannel signal processing, Proc. IEEE 87 (1999) 1601–1622.
[23] K.N. Plataniotis, A.N. Venetsanopoulos, Color Image Processing and Applications, Springer, Berlin, 2000.
[24] F. Russo, G. Ramponi, A fuzzy filter for images corrupted by impulse noise, IEEE Signal Process. Lett. 3 (1996) 168–170.
[25] B. Smolka, K.N. Plataniotis, R. Lukac, A.N. Venetsanopoulos, Noise reduction in DNA microarray images based on digital path approach, Proc. 5th IFAC Symp. 2003 on Modelling and Control in Biomedical Systems, Melbourne, Australia 2003, pp. 17–22.
[26] K. Tang, J. Astola, Y. Neuvo, Nonlinear multivariate image filtering techniques, IEEE Trans. Image Process. 4 (1995) 788–798.
[27] H.H. Tsai, P.T. Yu, Genetic-based fuzzy hybrid multichannel filters for color image restoration, Fuzzy Sets and Systems 114 (2000) 203–224.
[28] P.E. Trahanias, D. Karakos, A.N. Venetsanopoulos, Directional processing of color images: theory and experimental results, IEEE Trans. Image Process. 5 (1996) 868–881.
[29] X.H. Wang, R.S.H. Istepian, Y.H. Song, Microarray image de-noising using stationary wavelet transform, Proc. 4th IEEE Conf. Information Technology Applications in Biomedicine, 2003, pp. 15–18.
[30] X.Y. Zhang, F. Chen, Y.T. Zhang, S.G. Agner, M. Akay, Z.H. Lu, M.M.Y. Waye, S.K.W. Tsui, Signal processing techniques in genomic engineering, Proc. IEEE 90 (2002) 1822–1833.
[31] J. Zheng, K.P. Valavanis, J.M. Gauch, Noise removal from color images, J. Intell. Robotic Systems 7 (1993) 257–285.
[32] A. Zilouchian, M. Jamshidi, Intelligent Control Systems Using Soft Computing Methodology, CRC Press, Boca Raton, 2001.
[33] H. Zimmermann, Fuzzy Set Theory and Its Applications, Second ed., Kluwer Academic Publishers, Dordrecht, 1991.