

# FUZZY VECTOR FILTERS FOR MICROARRAY IMAGE ENHANCEMENT

*R. Lukac\* K.N. Plataniotis\* B. Smolka\*\* and A.N. Venetsanopoulos\**

\* Bell Canada Multimedia Laboratory, The Edward S. Rogers Sr. Department of ECE,  
University of Toronto, 10 King's College Road, Toronto, Canada  
*lukacr@ieee.org, {kostas, anv}@dsp.utoronto.ca*

\*\* Laboratory of Multimedia Communication, Silesian University of Technology,  
Rudzka Str., 44-200, Rybnik, Poland, *bsmolka@ia.polsl.gliwice.pl*

## ABSTRACT

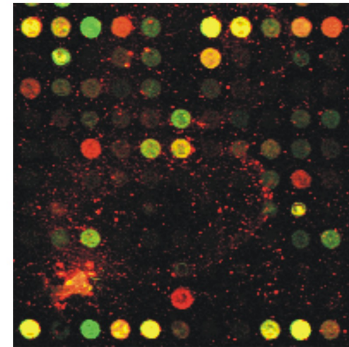
In this paper, the images of gene chips (microarray images) are enhanced applying the vector fuzzy filtering framework. Using a generalized fuzzy concept, we adaptively determine weights in the filtering structure and provide different filter designs. The strong potential of fuzzy adaptive filters for microarray image enhancement, is illustrated with several examples. We demonstrate that the fuzzy vector filters are capable of reducing outliers present in microarray images and simultaneously preserve the spot edges. This can significantly help in spot detection and estimation of the gene expression level.

## 1. INTRODUCTION

The cDNA microarray is a popular and effective method for simultaneous assaying the expression of large numbers of genes and is perfectly suited for the comparison of gene expression in different populations of cells [2],[8]. A microarray (*Fig.1*) is a collection of green, red and yellow spots (of different hue, saturation and intensity) containing DNA, deposited on the surface of a glass slide. Each of the spots 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.

This paper deals with the noise removal in microarray images. The problem is important for following reasons:

- The fluorescent intensities for each of the two dyes are measured separately, producing a two channel image. Therefore, vector processing of the acquired image data is necessary [4],[5].
- The image is false colored using red and green for each image components, which represent the light intensity emitted by the two fluorescent dyes [1],[2].
- The major sources of uncertainty in spot finding and measuring the gene expression are variable spot sizes



**Fig. 1.** A typical DNA microarray.

and positions, variation of the image background and various image artifacts [2].

- The natural fluorescence of the glass slide and non-specifically bounded DNA or dye molecules add a substantial noise floor to the microarray image along with discrete image artifacts such as highly fluorescent dust particles, unattached dye, salt deposits from evaporated solvents, fibers and various airborne debris [3].
- Vector filtering techniques [5] help in image denoising and digital interpretation of microarrays and can also enable correct spot segmentation.

Vector filtering methods used as a preprocessing tool for subsequent processing tasks such as spot segmentation and gene expression analysis are required to eliminate the noise present in corresponding digital data and simultaneously preserve color image edges making the spots easier to detect. Furthermore, filtering methods designed to process vector fields such as microarray images should take into consideration the vector nature of the data, the nonlinear characteristics of the image formation and the possible nonlinear nature of the noise corruption.

## 2. PROBLEM FORMULATION

As described in the previous section, noise introduced into the microarray images is present in forms of artifacts significantly deviating from neighboring pixels. This results in color distortions and outliers affecting the original data. In order to model such a corruption in image processing applications, the additive noise model is commonly used [5],[6]:

$$\mathbf{x}_i = \mathbf{o}_i + \mathbf{v}_i \quad (1)$$

where  $\mathbf{x}_i = (x_{i1}, x_{i2}) \in Z^2$ , represents observation (noisy) sample,  $\mathbf{o}_i = (o_{i1}, o_{i2}) \in Z^2$ , is desired (noise free) sample,  $\mathbf{v}_i = (v_{i1}, v_{i2}) \in Z^2$ , is the vector describing the noise process, and  $i$ , characterizes the spatial position of the samples in the image.

## 3. FUZZY VECTOR FILTERS

Let us consider a  $K_1 \times K_2$  two-channel image  $\mathbf{x}(i) : Z^2 \rightarrow Z^2$  representing a two-dimensional matrix of two-component samples (pixels)  $\mathbf{x}_i = (x_{i1}, x_{i2}) \in Z^2$  [5]. Let us consider sliding (moving, running) window  $W = \{\mathbf{x}_i \in Z^2; i = 1, 2, \dots, N\}$  of finite odd size  $N$ , which usually affects one image sample (mostly the sample  $\mathbf{x}_{(N+1)/2}$  placed in the center of the window) at a time, changing its value by some function of a local neighborhood area  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ . This window operator slides over the image to affect individually all the image pixels. The rationale of this approach is to minimize the local distortion and especially ensure the stationarity of the processes (including noise and blurring) generating the image. Based on the theoretical model (1) and the relationship between  $\mathbf{x}_i$  and  $\mathbf{o}_i$ , the filter should produce the output close as much as possible to the original signal  $\mathbf{o}_i$ . Since the original signal is not available in microarray imaging, the filter should be designed in such a way that it will be capable of removing the samples deviating from the local neighborhood area.

One of the most frequently used adaptive techniques is based on fuzzy logic principles. Data-dependent fuzzy vector filters operating on supporting filter window  $W$  are constructed by fuzzy rules that allow filter adaptation to local data as follows [6],[7]:

$$\mathbf{y} = f \left( \frac{\sum_{i=1}^N w_i \mathbf{x}_i}{\sum_{i=1}^N w_i} \right) = f \left( \sum_{i=1}^N w'_i \mathbf{x}_i \right) \quad (2)$$

where  $f(\cdot)$  is a nonlinear function that operates over the weighted average of the input set and  $w_i$  is the filter weight equivalent to the fuzzy membership function associated with the input color vector  $\mathbf{x}_i$ . Note that two constraints are necessary to ensure that the filter output is an unbiased estimator:

- Each weight is a positive number,  $w'_i \geq 0$ .

- The summation of all the weights is equal to unity  $\sum_{i=1}^N w'_i = 1$ .

In this adaptive design the weights provide the degree to which the input vector contributes to the output of the filter. Utilizing the sigmoidal membership function, the weights adaptation in (2) is performed [6] by

$$w_i = \beta (1 + \exp \{d_i\})^{-r} \quad (3)$$

where  $r$  is a parameter adjusting the weighting effect of the membership function and  $\beta$  is a normalizing constant.

The quantity

$$d_i = \sum_{j=1}^N A(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

denotes the aggregated distance measure defined via the angles between the two-dimensional vectors  $\mathbf{x}_i = (x_{i1}, x_{i2})$  and  $\mathbf{x}_j = (x_{j1}, x_{j2})$ :

$$A(\mathbf{x}_i, \mathbf{x}_j) = \cos^{-1} \left( \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{|\mathbf{x}_i| |\mathbf{x}_j|} \right) \quad (5)$$

$$= \cos^{-1} \left( \frac{x_{i1}x_{j1} + x_{i2}x_{j2}}{\sqrt{x_{i1}^2 + x_{i2}^2} \sqrt{x_{j1}^2 + x_{j2}^2}} \right) \quad (6)$$

Within the general fuzzy adaptive filter framework of (2), numerous filters may be constructed by changing the form of the nonlinear function  $f(\cdot)$ , as well as the way the fuzzy weights are calculated [6]. The choice of these two parameters determines the filter characteristics.

In this paper, except the angular measure of (3), we make use of the aggregated distance measure calculated via the magnitude differences between the vector components:

$$d_i = \sum_{j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|_L \quad (7)$$

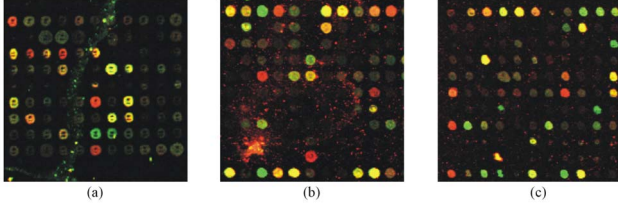
where

$$\|\mathbf{x}_i - \mathbf{x}_j\|_L = \left( \sum_{k=1}^m |x_{ik} - x_{jk}|^L \right)^{\frac{1}{L}} \quad (8)$$

denotes the generalized Minkowski metric [7] determining the distance between two multichannel samples  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Note that (7) is regulated by the norm  $L$  corresponding to the city-block distance ( $L = 1$ ) or Euclidean distance ( $L = 2$ ).

Based on the weighting coefficients of (3), the output of the fuzzy weighted vector filter (FWVF) [6] is given as follows:

$$\mathbf{y} = \sum_{i=1}^N w'_i \mathbf{x}_i \quad (9)$$



**Fig. 2.** Real microarray images used for testing purposes.

where  $w'_i$ , for  $i = 1, 2, \dots, N$ , are the normalized weight coefficients.

This filter provides a vector-valued signal which is not included in the original set of inputs. The weighted average form of the filter provides a compromise between a non-linear order statistics filter and an adaptive filter with data dependent coefficients.

Another possible choice of nonlinear function  $f(\cdot)$  is the maximum selector. In this case, the output of the maximum fuzzy vector filter (MFVF) [6] is the input vector that corresponds to the maximum fuzzy weight. The form of this filter is

$$\mathbf{y} = \mathbf{x}_i \text{ with } i = \arg \max w_i; \text{ for } i = 1, 2, \dots, N \quad (10)$$

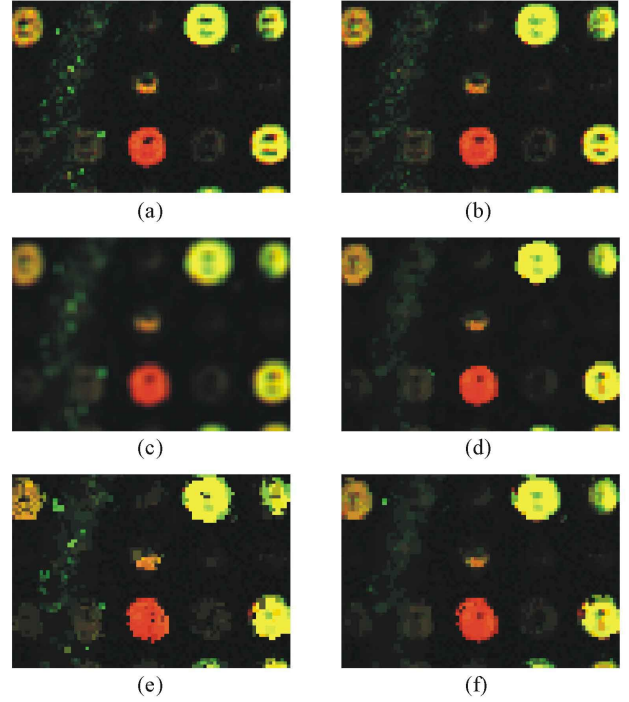
Using the maximum selector concept, as an output the input vector associated with the maximum fuzzy weight is selected. In other words, the output of the filter is a part of the original input set. This property is useful in the image areas with noise in the form of impulsive sequences. Note that due to the average nature of (9) and the selection capability of (10), the described FWVF and MFVF filters represent the robust filtering operators.

#### 4. APPLICATION TO MICROARRAY IMAGES

The proposed zooming method is tested using the color images shown in Fig.2. These images have been captured using specialized laser scanners.

The purpose of image filtering is to estimate the desired (original) signal  $\mathbf{o}$  as much as possible. Since the original image is unavailable in real applications such as microarray image enhancement, the achieved results can be evaluated using the subjective evaluation approach, which can be summarized into the three main points [7]: i) Is the noise removed?, ii) Is the structural content (edges, textures and fine details) of the image preserved?, and iii) Are there some color artifacts as a result of faulty processing?

Note that noise introduced into the microarray images inhibits the correct understanding of image information and avoids the correct spot localization and segmentation. Simultaneously, utilized image filters have to be capable of preserving the edges, which provide an indication of the

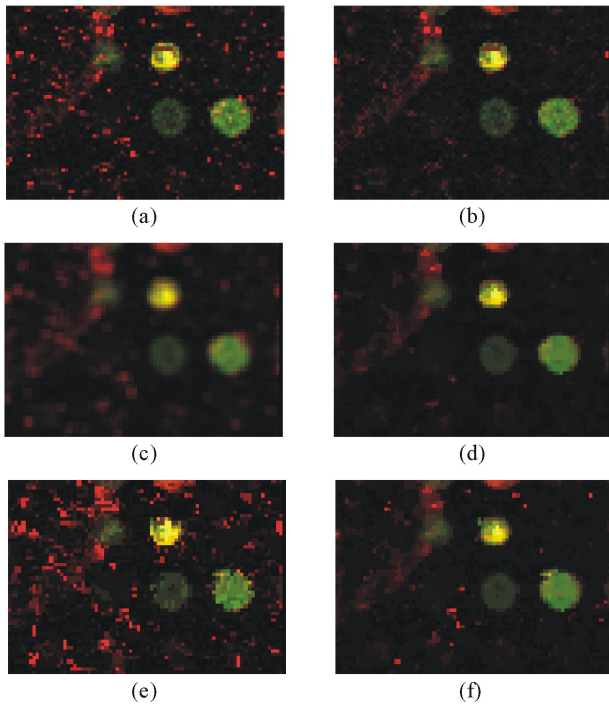


**Fig. 3.** Zoomed results corresponding to the image Fig.2a: (a) original image, (b) AVMF output, (c) FWVF output based on (4), (d) FWVF output based on (7), (e) MFVF output based on (4), (f) MFVF output based on (7).

shape of the spots in the microarray. Therefore, it is important to distinguish the fine structures from the noise, so that they can be preserved during the filtering process. The last requirement follows the classification of any imperfection such as blocking artifacts or new color pixels that were not present in the input image. It is necessary to evaluate color appearance which refers to color sharpness, the distinctness of boundaries among colors, and color uniformity which refers to the consistency of the color in uniform areas.

The proposed vector fuzzy filters (FWVF and MFVF) are compared, in terms of subjective evaluation, with the adaptive vector median filter (AVMF) of [4].

Figs.3-5 illustrate the performance of the methods using enlarged parts of the test microarrays. These results clearly show that fuzzy vector filters based on the Minkowski metric remove noise from the microarray images in a robust way and outperform the AVMF technique of [4]. It can be observed that the FWVF filter based on the angular distance measure and the MFVF filter based on the Minkowski metric preserve spot edges and color information while being robust against noise and outliers present in input microarray images.



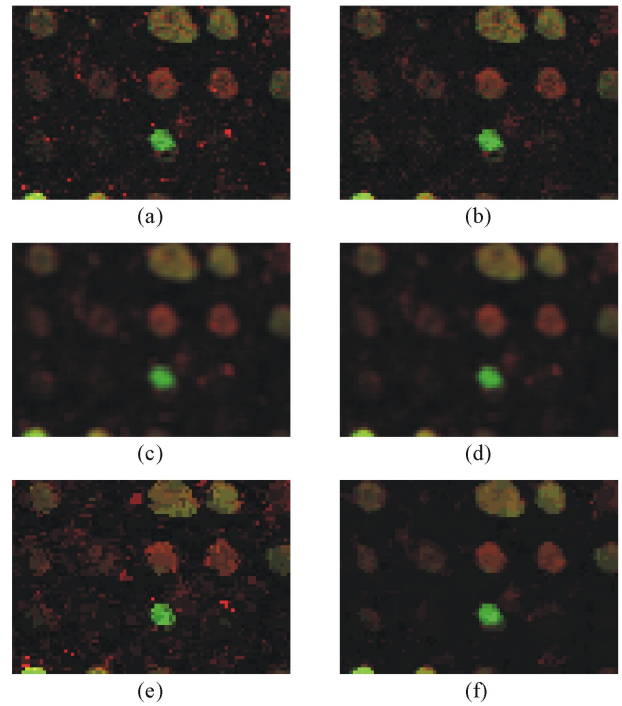
**Fig. 4.** Zoomed results corresponding to the image Fig.2b: (a) original image, (b) AVMF output, (c) FWVF output based on (4), (d) FWVF output based on (7), (e) MFVF output based on (4), (f) MFVF output based on (7).

## 5. CONCLUSIONS

In this paper, fuzzy vector algorithms of the noise reduction in microarray chip images has been presented. It was observed that presented fuzzy vector filters based on the Minkowski metric remove the outliers affected the spots while preserving the spot edges. Therefore, the presented fuzzy filtering framework can serve as an efficient low-level processing tool for microarray image enhancement, which can enable better spots localization and the estimation of their intensity.

## 6. REFERENCES

- [1] N. Ajay, T. Tokuyasu, A. Snijders, R. Segraves, D. Albertson, and D. Pinkel, "Fully automatic quantification of microarray image data," *Genome Research*, vol.12, pp.325-332, 2002.
- [2] J. Dopazo, "Microarray data processing and analysis," in *Microarray Data Analysis II*, ed. SM Lin and KF Johnson, Kluwer Academic, pp.43-63, 2002.
- [3] L. Hsiao, R. Jensen, T. Yoshida, K. Clark, J. Blumenstock, and S. Gullans, "Correcting for signal saturation errors in the analysis of microarray data," *Biotechniques*, vol.32, pp.330-336, 2002.



**Fig. 5.** Zoomed results corresponding to the image Fig.2c: (a) original image, (b) AVMF output, (c) FWVF output based on (4), (d) FWVF output based on (7), (e) MFVF output based on (4), (f) MFVF output based on (7).

- [4] R. Lukac and B. Smolka, "Application of the adaptive center-weighted vector median framework for the enhancement of cDNA microarray images," *Int. J. of Applied Mathematics and Computer Science*, vol.13, pp. 101-115, October 2003.
- [5] R. Lukac, B. Smolka, K. Martin, K.N. Plataniotis, and A.N. Venetsanopoulos, "Vector filtering for color imaging," *IEEE Signal Processing Magazine, Special Issue on Color Image Processing*, to appear, April 2004.
- [6] K.N. Plataniotis, D. Androutsos, and A.N. Venetsanopoulos, "Adaptive fuzzy systems for multichannel signal processing," *Proceedings of the IEEE*, vol. 87, pp. 1601-1622, September 1999.
- [7] K.N. Plataniotis, A.N. Venetsanopoulos, *Color Image Processing and Applications*. Springer Verlag, 2000.
- [8] M. Schena, D. Shalon, R.W. Davis, and P.O. Brown, "Quantitative monitoring of gene expression patterns with a complimentary DNA microarray," *Science*, vol.270, pp.467-470, 1995.