

Adaptive Noise Reduction in Microarray Images Based on the Center-Weighted Vector Medians

Rastislav Lukac¹, Bogdan Smolka^{2*}, Andrzej Swierniak²,
Konstantinos N. Plataniotis³, and Anastasios N. Venetsanopoulos³

¹ Slovak Image Processing Center,
Jarkova 343, 049 25 Dobsina, Slovak Republic
lukacr@ieee.org

² Department of Automatic Control,
Silesian University of Technology, Akademicka 16 Str., 44-101 Gliwice, Poland
bsmolka@ia.polsl.gliwice.pl

³ The Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering,
University of Toronto, 10 King's College Road, Toronto, Canada
{kostas, anv}@dsp.utoronto.ca

Abstract. In this paper, we provide a new noise reduction method for the enhancement of the images of gene chips. We demonstrate that the new technique is capable of reducing outliers present in microarray images while preserving the spot edges. This paper contains the comparisons of the proposed technique with the standard schemes used in color image filtering, as well as examples of their efficiency when these methods are applied to typical microarray images.

1 Introduction

This paper focuses on a new nonlinear vector filtering scheme taking the advantage of center-weighted vector median filters and optimal filtering situation. The proposed method improves the signal detail preservation capability of the standard vector filtering schemes used for multichannel image filtering and provides the adaptive alternation between the identity operation and the robust vector median filtering. In addition, the computational complexity of the proposed method is still acceptable, which allows its application for the enhancement of the cDNA microarray images [1],[3]. Using the new filtering scheme, it is possible to achieve excellent balance between the signal-detail preservation and the noise attenuation. These properties of the proposed method were tested for a wide range of multichannel image signals such as phantom (artificial) images and real microarray chips. Using the phantom images, we can test the efficiency of the methods depending on the intensity of impulsive noise corruption, which is frequently introduced into the microarray images. We also provide a short description of the microarray cDNA images.

* B. Smolka is supported by the KBN grant 4T11F01824.

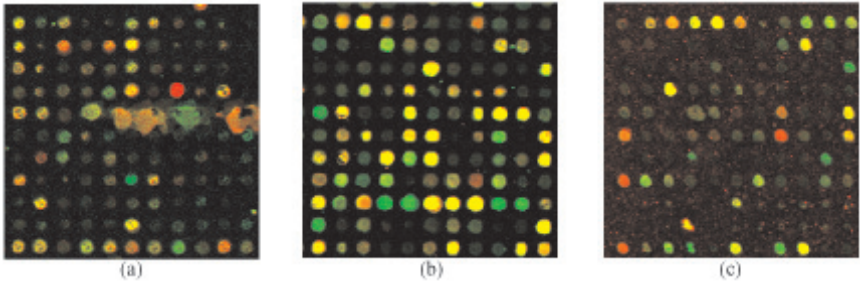


Fig. 1. Real cDNA microarray images.

2 Vector Median Filter

Concerning the multichannel image filtering [10], standard color images represent the vector-valued image signals, in which each image point can be considered as a vector of three components associated with intensities of color channels consisting of red (R), green (G) and blue (B). Thus, it is necessary to consider the correlation that exists between color channels and to apply the vector processing. If the existing correlation is not taken into account and color channels are processed independently, then the filtering operation is applied componentwise. In general, componentwise (marginal) approaches produce new vector samples, i.e. color artifacts, caused by composition of reordered channel samples.

Vector filters represent a natural approach to the noise removal in multichannel images, since these filters utilize the correlation between color channels. Therefore, the vector methods represent optimal and attractive approaches for studying and processing of noisy color images [2],[8],[10],[12].

The most popular nonlinear, multichannel (vector) filters are based on the ordering of vectors in a predefined sliding window. Designed to perform low-pass filtering, vector filtering schemes output the lowest ranked vector according to a specific ordering technique [10].

Let $y(x) : Z^l \rightarrow Z^m$ represent a multichannel image, where l is an image dimension and m denotes a number of channels. Let $W = \{\mathbf{x}_i \in Z^l; i = 1, 2, \dots, N\}$ represent a filter window of a finite size N , where $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ is a set of noisy samples centered around the central sample $\mathbf{x}_{(N+1)/2}$.

Let us consider input sample \mathbf{x}_i , for $i = 1, 2, \dots, N$, associated with the distance measure L_i given by

$$L_i = \sum_{j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|_\gamma, \text{ for } i = 1, 2, \dots, N \tag{1}$$

where

$$\|\mathbf{x}_i - \mathbf{x}_j\|_\gamma = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^\gamma \right)^{\frac{1}{\gamma}} \tag{2}$$

characterizes the generalized Minkowski distance [10] between two multichannel samples \mathbf{x}_i and \mathbf{x}_j . Notation γ characterizes the used norm, m is the dimension of vectors and x_{ik} is the k -th element of the sample \mathbf{x}_i . Note that the well-known Euclidean distance is associated with $\gamma = 2$.

Assuming that the ordering of $L_{(i)}$ implies the same ordering of the input set $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ results in the ordered set $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(N)}$, where $\mathbf{x}_{(i)}$ is associated with $L_{(i)}$. The vector median output [2] is defined as the lowest order-statistics $\mathbf{x}_{(1)}$. It is evident that the VMF output is restricted to the dynamic range of the input samples and thus, it can never cause new outliers.

Because the VMF minimizes the aggregated distance to other multichannel samples inside the filter window, it can be equivalently defined by

$$\sum_{j=1}^N \|\mathbf{y}_{VMF} - \mathbf{x}_j\|_{\gamma} \leq \sum_{j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|_{\gamma}, \text{ for } i = 1, 2, \dots, N \quad (3)$$

where $\mathbf{y}_{VMF} = \mathbf{x}_{(1)} \in \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ denotes the VMF output.

3 Proposed Method

Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ be an input set determined by a filter window and N represent a window size. Let us assume that w_1, w_2, \dots, w_N represent a set of nonnegative integer weights so that each weight w_i , for $i = 1, 2, \dots, N$, is associated with the input sample \mathbf{x}_i . Thus, the weighted vector distance is defined as follows [13]:

$$J_i = \sum_{j=1}^N w_j \|\mathbf{x}_i - \mathbf{x}_j\|_{\gamma}, \text{ for } i = 1, 2, \dots, N \quad (4)$$

The sample $\mathbf{x}_{(1)} \in \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ associated with minimal aggregated weighted distance $J_{(1)} \in \{J_1, J_2, \dots, J_N\}$, is the sample that minimizes the sum of weighted vector distances.

The sample $\mathbf{x}_{(1)}$ associated with the minimum aggregated weighted distance $J_{(1)}$ represents the output of the weighted vector median filter (WVMF) [13]. WVMF is also defined by

$$\sum_{j=1}^N w_j \|\mathbf{y}_{WVMF} - \mathbf{x}_j\|_{\gamma} \leq \sum_{j=1}^N w_j \|\mathbf{x}_i - \mathbf{x}_j\|_{\gamma}, \text{ for } i = 1, 2, \dots, N \quad (5)$$

It is clear that in the dependence on the weight coefficients w_1, w_2, \dots, w_N the WVMFs can perform a wide range of smoothing operations so that the optimal weight vector may be practically found for each filtering problem.

Because both VMF and WVMF will perform the smoothing operation with the fixed amount of smoothing, these filters can fail in situation with different statistical properties especially caused by a low degree of the observed noise, when they introduce too much smoothing into the image. Note the blurring introduced by a filter may be more objectionable than original noise.

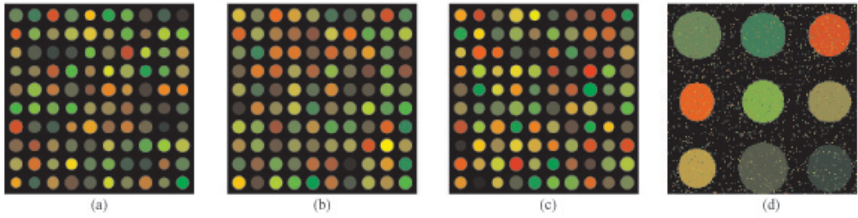


Fig. 2. cDNA phantom (artificial) test images: (a) cDNA1, (b) cDNA2, (c) cDNA3, (d) detailed view on the test image cDNA1 corrupted by 5% impulsive noise.

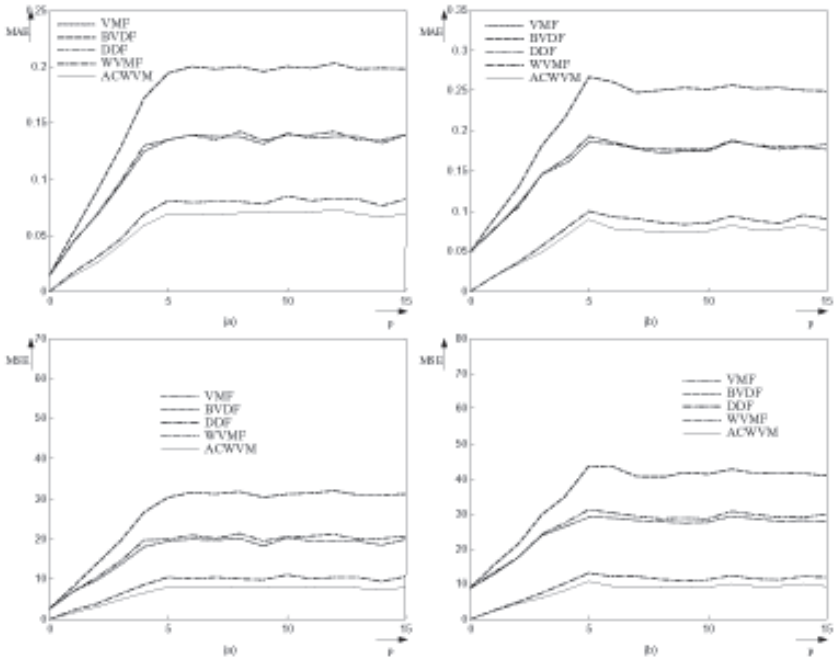


Fig. 3. Dependence of the MAE and MSE criteria on the impulsive noise probability p_v : (a) cDNA1 test image, (b) cDNA3 test image.

Now, we provide the adaptive center weighted vector median filter (ACWVM). Let us assume the weight vector given by

$$w_j = \begin{cases} N - 2k + 2 & \text{for } j = (N + 1)/2 \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

i.e. the weight vector of nonnegative integer weights, where only the central weight $w_{(N+1)/2}$ associated with the central sample $\mathbf{x}_{(N+1)/2}$ can be alternated, whereas other weights associated with the neighboring samples remain equal to one. Note that $k = 1, 2, \dots, (N + 1)/2$ is a smoothing parameter. If the smoothing

parameter k is equal to one, then the center-weighted vector median (CWVM) [8] is equivalent to the identity operation and no smoothing will be provided. In the case of $k = (N + 1)/2$, the maximum amount of the smoothing will be performed and the CWVM filter is equivalent to WVMF. By varying the smoothing parameter k between one and $(N + 1)/2$, it is possible to achieve the best balance between the noise suppression and the signal-detail preservation.

In the case of outliers, bit errors and color artifacts (all above-mentioned events can affect the cDNA microarray images), it is desirable to filter only corrupted samples, whereas the desired image features are invariant to the filtering operation [8]. For that reason, the ACWVM is based on dividing the samples into two classes, namely corrupted samples and noise-free samples. The central sample $\mathbf{x}_{(N+1)/2}$ is examined according to the following rule:

$$\begin{aligned} \text{IF } Val \geq Tol \quad \text{THEN } \mathbf{x}_{(N+1)/2} \text{ is impulse} \\ \text{ELSE } \mathbf{x}_{(N+1)/2} \text{ is noise - free} \end{aligned} \quad (7)$$

where Tol is the threshold parameter and Val is the operation value given by

$$Val = \sum_{k=\lambda}^{\lambda+2} \|\mathbf{y}_k - \mathbf{x}_{(N+1)/2}\|_{\gamma} \quad (8)$$

where $\mathbf{x}_{(N+1)/2}$ is the central sample of the input set W and \mathbf{y}_k is the output of center-weighted median filter with the smoothing parameter k . Note that the optimal values of Tol and λ was found as 80 and 2, respectively. If the operation value Val is greater than or equal to the threshold value Tol , then the central input sample $\mathbf{x}_{(N+1)/2}$ is probably corrupted and it will be processed by VMF with a robust noise attenuation capability. Otherwise, the central sample $\mathbf{x}_{(N+1)/2}$ is noise-free and it should be invariant against a filtering operation.

Table 1. Results achieved using the phantom images corrupted by 5% impulsive noise.

Image	cDNA1		cDNA2		cDNA3	
Method	MAE	MSE	MAE	MSE	MAE	MSE
Noisy	3.325	525.2	3.231	502.5	3.673	583.2
VMF	0.135	19.4	0.152	21.9	0.186	29.0
BVDF	0.194	30.4	0.224	35.0	0.267	43.7
DDF	0.135	19.9	0.160	23.7	0.193	31.2
WVMF	0.081	10.4	0.085	10.6	0.100	13.1
ACWVM	0.069	8.0	0.070	7.7	0.089	10.9

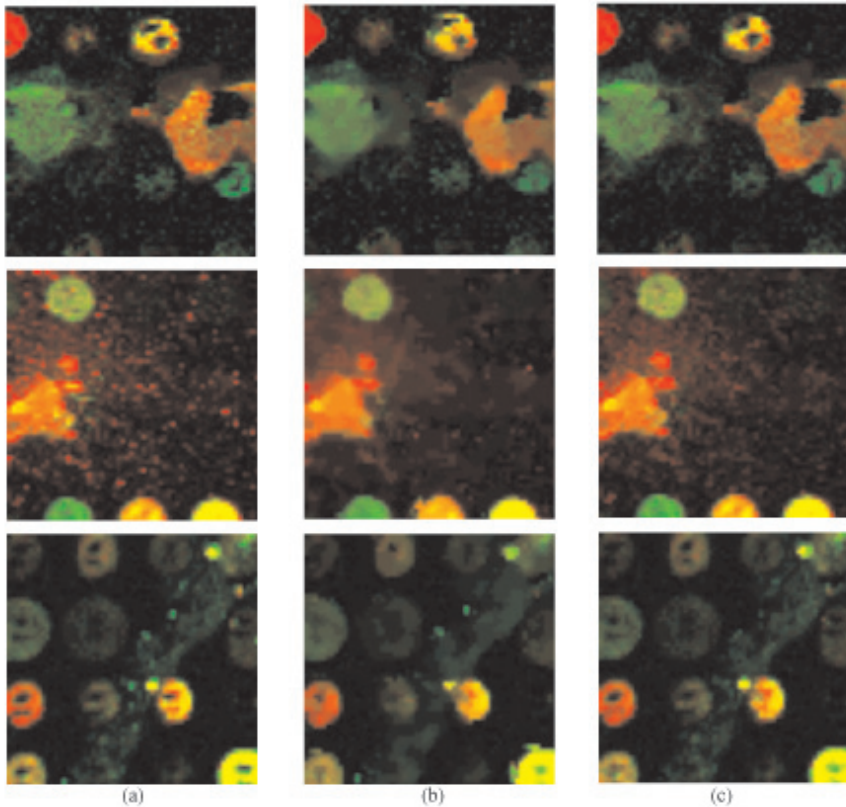


Fig. 4. Enhanced images obtained using real cDNA images: (a) observed image, (b) VMF output, (c) ACWVM output.

4 Experimental Results

The cDNA microarray [3],[6],[14] 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. A cDNA microarray is a collection of spots 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. The fluorescent intensities for each of the two dyes are measured separately, producing a two channel image. The image is false colored using red and green for each image components, which represent the light intensity emitted by the two fluorescent dyes. So the microarray images look like a collection of green, red and yellow spots of different hue, saturation and intensity. The intensities provided by the array image can be quantified by measuring the

average or integrated intensities of the spots. The major sources of uncertainty in spot finding and measuring the gene expression are variable spot sizes and positions, variation of the image background and various image artifacts. Spots vary significantly in size and position within their vignettes despite the use of precise robotic tools to lay them out onto the slide. Additionally 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. So, the task of microarray image enhancement and is of great importance, as it enables correct block and spot segmentation.

In order to compare the performance of used filtering schemes, we used some artificial images (Fig.2a-c) degraded by impulsive noise (Fig.2d) and also the natural microarray images shown in Fig.1. By using the artificial images, the undistorted, original, images are available and the obtained results can be evaluated objectively (Tab.1 and Fig.3) via the mean absolute error (MAE) and the mean square error. In the case of the natural chip images depicted in Fig.1, we can compare only the subjective results, in the form of visual assessment of the filter outputs depicted in Fig.4.

As can be seen the proposed method significantly outperforms the commonly used multichannel image processing noise reduction techniques such as VMF [2], basic vector directional filter (BVDF) [12], directional distance filter (DDF) [7] and non-adaptive WVMF [13] with weight coefficients [1, 2, 1, 4, 5, 4, 1, 2, 1]. This is also confirmed by its performance on real images shown in Fig.4, in which the noise component was successfully removed, while preserving the texture and sharpness of the spot edges.

5 Conclusion

In this work a novel algorithm of the noise reduction in microarray chip images has been presented. During the filtering process the outliers affected the spots are being detected and removed while the edges remain well preserved. The proposed technique can serve as an efficient low-processing tool for microarray image enhancement, which can enable better spots localization and the estimation of their intensity.

Future research will focus on the segmentation and the spot localization, where contour smoothing algorithms of [4],[5] can play a key role. Combined with the enhancement method introduced in this paper it is expected that image segmentation techniques will bring another improvement on the cDNA microarray image analysis.

References

- [1] Ajay, N., Tokuyasu, T., Snijders, A., Segraves, R., Albertson, D., Pinkel, D.: Fully automatic quantification of microarray image Data. *Genome Research* **12** (2002) 325–332

- [2] Astola, J., Haavisto, P., Neuvo, Y.: Vector median filters. *Proceedings of the IEEE* **78** (1990) 678–689
- [3] Dopazo, J.: Microarray data processing and analysis. In *Microarray Data Analysis II*, ed. SM Lin and KF Johnson, Kluwer Academic, 2002, 43–63
- [4] Glasa, J.: On derivatives estimation of smoothed digital curves. *Computer and Artificial Intelligence* **19** (2000) 235–249
- [5] Glasa, J., Podhajecky, G.: On feasible digital contour smoothing for biomedical image analysis. *Proc. Embec'02*, (2002) 840–841
- [6] Hsiao, L., Jensen, R., Yoshida, T., Clark, K., Blumenstock, J., Gullans, S.: Correcting for signal saturation errors in the analysis of microarray data. *Biotechniques* **32** (2002) 330–336
- [7] Karakos, D.G., Trahanias, P.E.: Generalized multichannel image-filtering structure. *IEEE Trans. Image Processing* **6** (1997) 1038–1045
- [8] Lukac, R.: Vector LUM smoothers as impulse detector for color images. *Proc. ECCTD '01* **3** (2001) 137–140
- [9] Pitas, I., Venetsanopoulos, A.N.: Order statistics in digital image processing. *Proceedings of the IEEE* **80** (1992) 1892–1919
- [10] Plataniotis, K.N., Venetsanopoulos, A.N.: *Color image processing and applications*. Springer Verlag (2000)
- [11] Schena, M., Shalon D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complimentary DNA microarray. *Science* **270** (1995) 467–470
- [12] Trahanias, P.E., Karakos, D., Venetsanopoulos, A.N.: Directional processing of color images: theory and experimental results. *IEEE Trans. Image Processing* **5** (1996) 868–881
- [13] Viero, T., Oistamo, K., Neuvo, Y.: Three-dimensional median related filters for color image sequence filtering, *IEEE Trans. Circ. Syst. Video Tech* **4** (1994) 129–142.
- [14] Yang, Y., Buckley, M., Dudoit, S., Speed, T.: Comparison of methods for image analysis on cDNA microarray data. *Jour. Computational and Graphic Statistics* **11** (2002) 108–136