

A Data-Adaptive Approach to cDNA Microarray Image Enhancement

Rastislav Lukac¹, Konstantinos N. Plataniotis¹,
Bogdan Smolka^{2*}, and Anastasios N. Venetsanopoulos¹

¹ The Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering,
University of Toronto, 10 King's College Road, Toronto, M5S 3G4, Canada
{lukacr, kostas, anv}@dsp.utoronto.ca

<http://www.dsp.utoronto.ca/~lukacr>

² Polish-Japanese Institute of Information Technology,
Koszykowa 86 Str, 02-008 Warsaw, Poland
bsmolka@ia.polsl.gliwice.pl

Abstract. A data-adaptive approach for cDNA microarray image enhancement is presented. Through the weighting coefficients adaptively determined from local microarray image statistics, the proposed technique tunes the overall filter's detail-preserving and noise-attenuating characteristics and uses both the spatial and spectral correlation of the cDNA image during processing. Noise removal is performed by tuning a membership function which utilizes the aggregated absolute differences between the cDNA microarray inputs localized within a processing window sliding over the image.

1 Introduction

cDNA microarray technology [1, 2, 3] is an advanced tool used to analyze changes in genome-wide patterns of gene expression in different populations of cells. Based on the widely used two-color Cy3/Cy5 system [4, 5], the control DNA sequence is usually fluor-tagged with green (Cy3), whereas the experimental sequence of the patient is tagged with red (Cy5). The composed Red-Green (RG) image (*Fig.1*) 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 [5]. 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 [1, 2, 6].

Although recognition of spots in either control or experimental channels seems to be straightforward, the task is complicated and challenging. Variation in image background and image artifacts, as well as spot sizes and positions

* This research has been supported by a grant No. **PJ/B/01/2004** from the Polish - Japanese Institute of Information Technology.

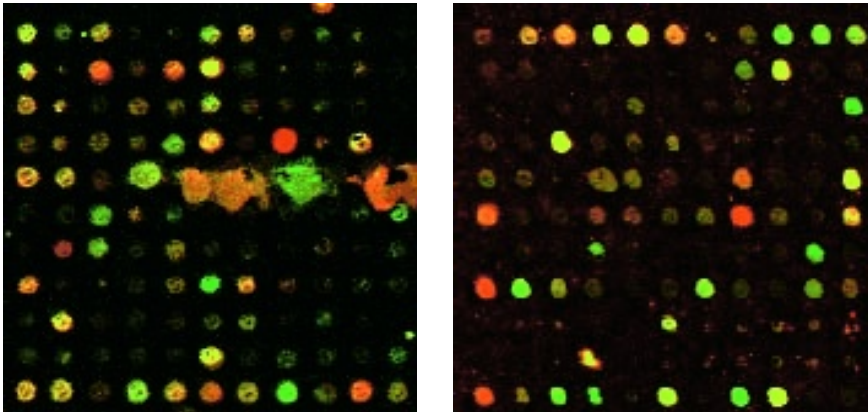


Fig. 1. cDNA microarray images

represent the major sources of uncertainty in spot finding and gene expression determination [5, 7]. In particular, non-specifically bounded DNA or dye molecules and the natural fluorescence of the glass slide result to a substantial noise floor in the microarray image. Since any errors or noise introduced usually limit the performance of the essential analysis procedure, if not defeat it altogether, image processing is necessary in order to eliminate processing errors from propagating further down the processing pipeline to the gene expression analysis tasks [1, 2, 6, 8]. Due to thousands of spots, the microarray image enhancement procedure, such as the one presented in this paper, should be fully automated.

2 Problem Formulation

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 $r = 1, 2, \dots, K_1$ and $s = 1, 2, \dots, K_2$ denoting the image rows and columns, respectively. 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 [2, 5]. Thus, the microarray data form a two-channel, Red-Green (RG) image [9, 10].

The cDNA microarray image suffers from high level of noise and edge uncertainty [7]. 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 [2]. These imperfections along with the background fluorescence introduce into the image considerable variability in intensity both within and between the individual spots, as shown in *Fig. 1*. 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 [4]. Isolated discrete artifacts and

outliers present in the cDNA microarray image can be attributed to impulsive noise [5]. In conventional image processing applications, noise corruption of this nature is most often modelled through a mixture of additive Gaussian noise and impulsive noise [11]. Therefore, the acquired cDNA signal $\mathbf{x}_{(r,s)} = [x_{(r,s)1}, x_{(r,s)2}]$ can be expressed through an additive noise model as follows [5]:

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

where (r, s) indicates the spatial position of the samples in the image array. The vector $\mathbf{o}_{(r,s)} = [o_{(r,s)1}, o_{(r,s)2}]$ represents the original, noise-free cDNA signal, while the vectorial samples $\mathbf{v}_{(r,s)} = [v_{(r,s)1}, v_{(r,s)2}]$ are considered random in nature and are used to denote the various image impairments introduced during processing.

3 Proposed Method

Since the noisy samples deviate from other samples in a given data population [5, 11], we may consider uniformity in cDNA pixels inside a localized image area as the objective for enhancing the cDNA microarray image. Due to the various image impairments introduced during microarray image formation, there is no method to objectively determine whether or not a cDNA sample is noisy. Therefore, the proposed microarray image enhancement solution utilizes local image statistics of the cDNA samples localized inside a processing window. The window, defined as $\Psi_{(r,s)} = \{\mathbf{x}_{(i,j)}; (i, j) \in \zeta\}$, for $p = 1, 2, \dots, K_1$ and $q = 1, 2, \dots, K_2$, 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 sample $\mathbf{x}_{(r,s)}$ 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 $\Psi_{(r,s)}$.

The filter function $f(\cdot)$ and the area of support ζ can have numerous forms, such as those listed in [11]. Both $f(\cdot)$ and ζ determine the complexity and the overall performance of the procedure. A 3×3 rectangular shape window with $\zeta = \{(r-1, s-1), (r-1, s), \dots, (r+1, s+1)\}$ is commonly used in image processing. Due to its versatility and demonstrated good performance, the same window shape is utilized throughout this paper. It has been proven that the conventional approaches such as component-wise median filter (MF) [12] and vector median filter (VMF) [13] produce robust estimates [5, 11]. However, noise and other impairments present in the acquired cDNA microarrays usually prohibit to accurately recognize between noise and edge pixels. Therefore, we will enhance the CDNA images using a fuzzy logic based smoothing approach (*Fig. 2*) which is capable of tuning the overall filtering characteristics by employing a membership function defined over local image statistics [9].

To preserve the spectral characteristics of the cDNA microarray image, the proposed method constitutes the Red/Green (R/G) ratios from the cDNA vectorial inputs. The color-ratio approach takes advantage of the expected relative uniformity of the local R/G ratios. Due to the decreased high-frequency portion

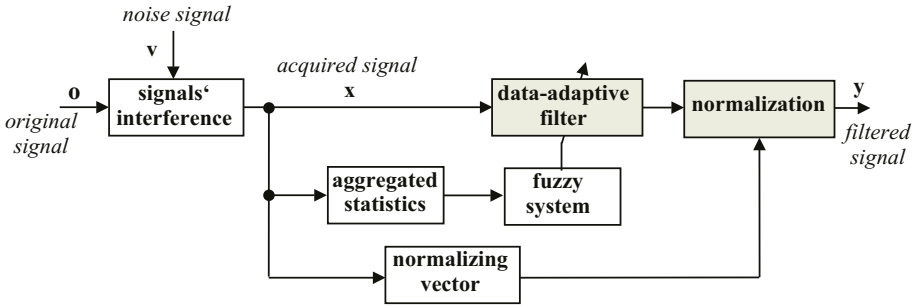


Fig. 2. Data-adaptive cDNA microarray image enhancer

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. Each sample $\mathbf{x}_{(i,j)}$ is used to produce the Red/Green color ratio $x_{(i,j)1}/x_{(i,j)2}$. The produced ratio-values $x_{(i,j)1}/x_{(i,j)2}$, for $(i,j) \in \zeta$, are used to determine local image statistics, expressed here in the form of aggregated absolute differences between color ratios [9]:

$$D_{(i,j)} = \sum_{(g,h) \in \zeta} |x_{(i,j)1}/x_{(i,j)2} - x_{(g,h)1}/x_{(g,h)2}| \tag{2}$$

Following the processing pipeline depicted in *Fig.2*, the scalar values $D_{(i,j)}$, for $(i,j) \in \zeta$, are used to determine the fuzzy weights $\mu_{(i,j)}$ using the popular sigmoidal membership function [14] as follows:

$$\mu_{(i,j)} = \beta (1 + \exp \{D_{(i,j)}\})^{-r} \tag{3}$$

where r is used to tune the weighting effect of the membership function and β is a normalizing constant. Note that each fuzzy weight $\mu_{(i,j)}$, for $(i,j) \in \zeta$, is associated with the input cDNA sample $\mathbf{x}_{(i,j)}$.

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 $\mathbf{x}_{(r,s)}$ at the window center with a suitable vector representative for the local image area $\Psi_{(r,s)}$. Using the color ratios $x_{(i,j)1}/x_{(i,j)2}$ and the defuzzified weighting coefficients $w_{(i,j)} = \mu_{(i,j)} / \sum_{(g,h) \in \zeta} \mu_{(g,h)}$, the output vector $\mathbf{y}_{(r,s)} = [y_{(r,s)1}, y_{(r,s)2}]$ is obtained as follows [9]:

$$y_{(r,s)1} = x_{(r,s)2}^* \sum_{(i,j) \in \zeta} \{w_{(i,j)} x_{(i,j)1}/x_{(i,j)2}\} \tag{4}$$

$$y_{(r,s)2} = x_{(r,s)1}^* \left[\sum_{(i,j) \in \zeta} \{w_{(i,j)} x_{(i,j)1}/x_{(i,j)2}\} \right]^{-1} \tag{5}$$

where $\mathbf{x}_{(r,s)}^* = [x_{(r,s)1}^*, x_{(r,s)2}^*]$ is a vector, whose components are used to normalize the output (smoothed) color-ratio value $\sum_{(i,j) \in \zeta} \{w_{(i,j)} x_{(i,j)1} / x_{(i,j)2}\}$ to the individual intensities $y_{(i,j)1}$ and $y_{(r,s)2}$ corresponding to the recovered R and G color channels, respectively. The normalization vector $\mathbf{x}_{(r,s)}^*$ can be considered as equivalent to a robust estimate which statistically represents the input set $\Psi_{(r,s)}$. Using robust order-statistics principle, a normalization factor $\mathbf{x}_{(r,s)}^*$ is defined here as the component-wise MF [12]. Thus, the data-adaptive filter combines the fuzzy weighted average operations and the nonlinear median operations to normalize the output of the proposed microarray image enhancer.

4 Experimental Results

A number of cDNA microarray images have been used to evaluate the proposed filtering framework. Examples are shown in *Fig. 3a* and *Fig. 4a*. These images have

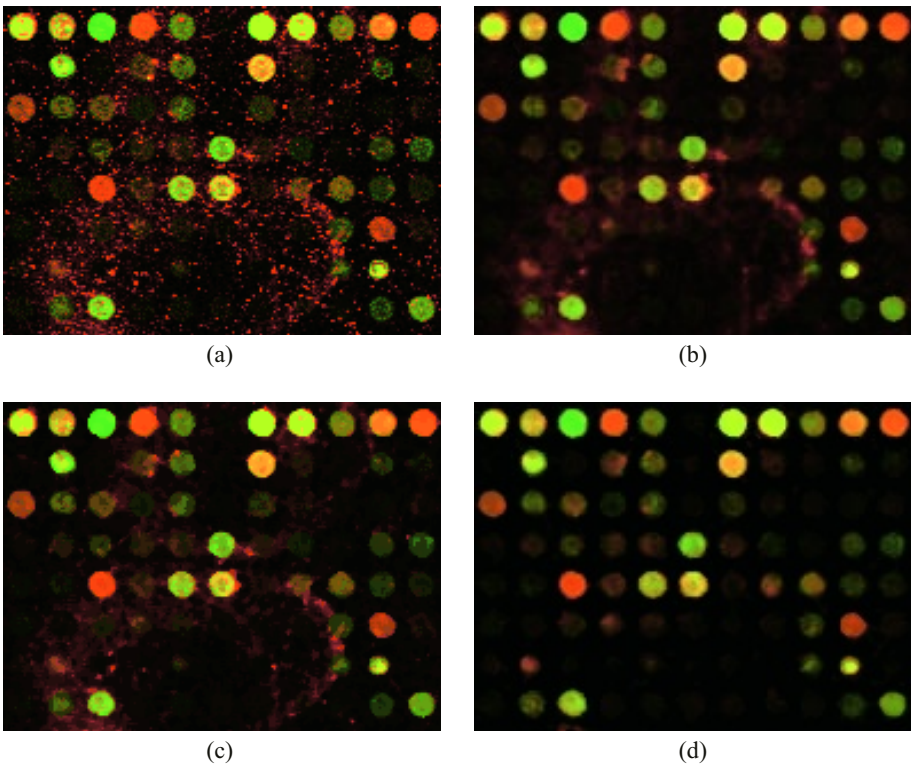


Fig. 3. Experimental results: (a) acquired image, (b) component-wise MF output, (c) VMF output, (d) the proposed method output

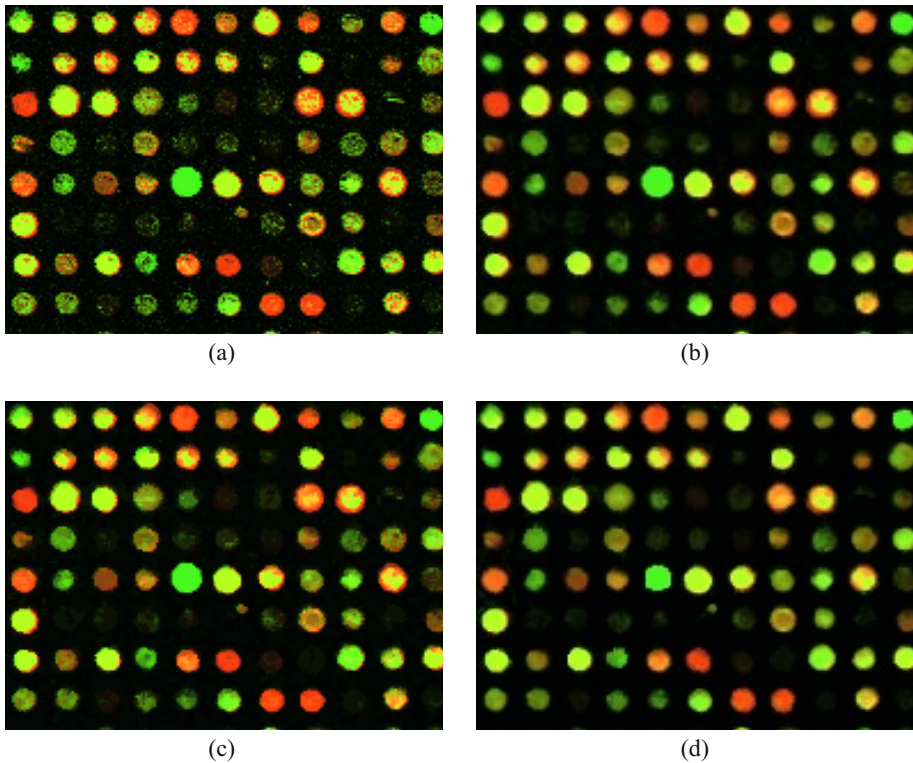


Fig. 4. Experimental results: (a) acquired image, (b) component-wise MF output, (c) VMF output, (d) the proposed method output

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.

The noise attenuation properties and detail-preserving capability of the proposed method are compared, in terms of performance, with the most popular and robust image filters such as the component-wise median filter (MF) [12] and the VMF technique [13].

Figs. 3-4 show the acquired cDNA microarray images (a) and the denoised images (b-d) obtained using the filtering techniques under consideration. The result indicate that the proposed scheme achieves excellent balance between signal-detail preservation and noise attenuation. It can be seen that the well-known MF and VMF schemes as well as the proposed method remove well foreground noise such as the various impulses and other high-frequency image impairments. However, only the proposed method is capable of removing strong varied background intensities introduced in course of cDNA microarray technology. Similar

conclusions can be drawn by visually inspecting the 3-D plots of cDNA microarray images shown in *Fig.5*. This suggests that the proposed method should be used as a preprocessing tool for cDNA microarray spot localization.

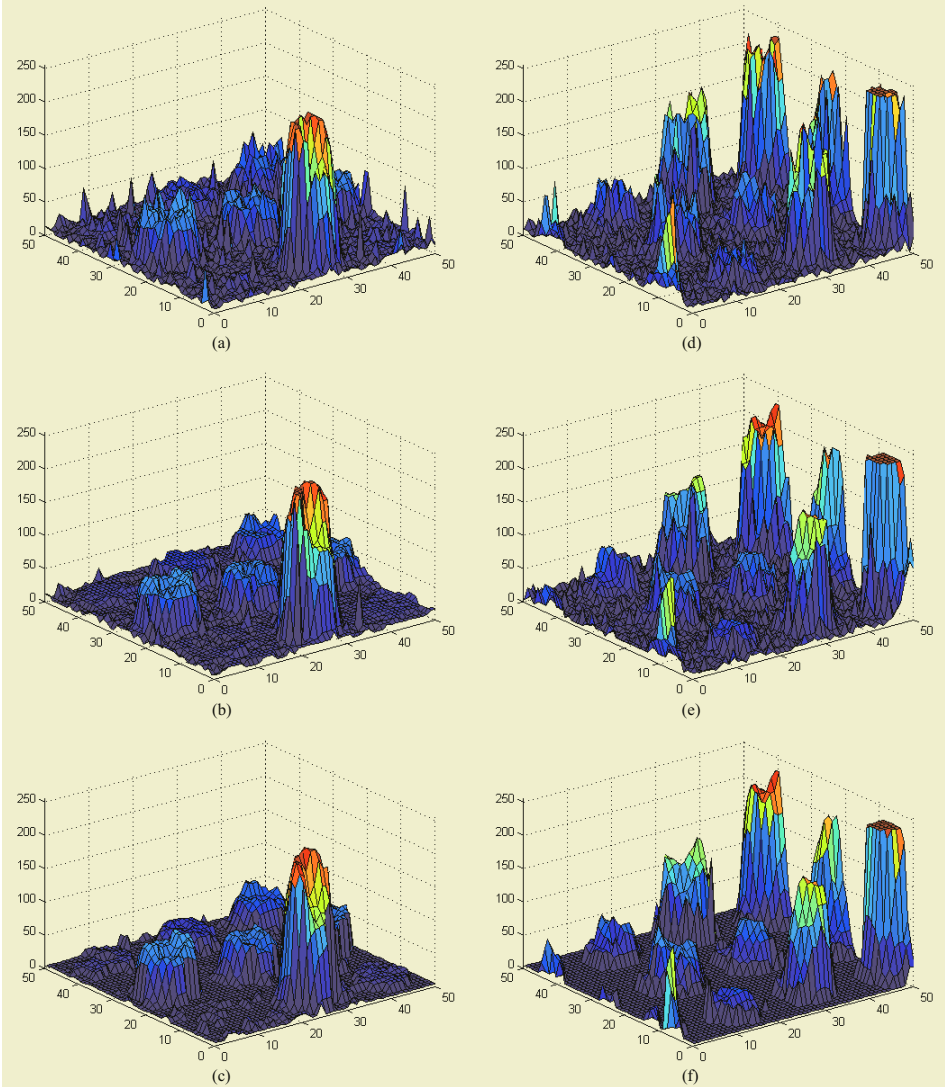


Fig. 5. 3-D visualization of cDNA spots: (a,d) acquired image, (b,e) VMF output, (c,f) the proposed method output

5 Conclusion

A new cDNA microarray processing tool has been introduced. The proposed method uses principles of fuzzy logic in conjunction with data-adaptive filtering concept defined over cDNA image data. Experimental results indicate that the proposed method is capable of removing noise in cDNA microarray images while preserving required data features for subsequent processing tasks.

References

1. Zhang, X.Y., Chen, F., Zhang, Y.T., Agner, S.G., Akay, M., Lu, Z.H., Waye, M.M.Y., Tsui, S.K.W.: Signal processing techniques in genomic engineering. *Proceedings of the IEEE* **90** (2002) 1822–1833
2. Lukac, R., Plataniotis, K.N.: cDNA microarray image segmentation using root signals. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, submitted (2004)
3. Eisen, M.B., Brown, P.O.: DNA arrays for analysis of gene expression. *Methods in Enzymology* **303** (1999) 179–205
4. Nagarajan, R.: Intensity-based segmentation of microarrays images. *IEEE Transactions on Medical Imaging* **22** (2003) 882–889.
5. Lukac, R., Plataniotis, K.N., Smolka, B., Venetsanopoulos, A.N.: A multichannel order-statistic technique for cDNA microarray image processing. *IEEE Transactions on Nanobioscience* **3** (2004) 272–285
6. Arena, P., Bucolo, M., Fortuna, L., Occhipinty, L.: Cellular neural networks for real-time DNA microarray analysis. *IEEE Engineering in Medicine and Biology* **21** (2002) 17–25
7. Wang, X.H., Istepian, R.S.H., Song, Y.H.: Microarray image enhancement using stationary wavelet transform. *IEEE Transactions on Nanobioscience* **2** (2003) 184–189
8. Dopazo, J.: Microarray data processing and analysis. In *Microarray Data Analysis II*, ed. SM Lin and KF Johnson, Kluwer Academic (2002) 43–63
9. Lukac, R., Plataniotis, K.N., Smolka, B., Venetsanopoulos, A.N.: cDNA microarray image processing using fuzzy vector filtering framework. *Journal of Fuzzy Sets and Systems: Special Issue on Fuzzy Sets and Systems in Bioinformatics*, to appear (2005)
10. Whitchurch, A.K.: Gene expression microarrays. *IEEE Potentials* **21** (2002) 30–34
11. Lukac, R., Smolka, B., Martin, K., Plataniotis, K.N., Venetsanopoulos, A.N.: Vector filtering for color imaging. *IEEE Signal Processing Magazine; Special Issue on Color Image Processing* **22** (2005) 74–86
12. Pitas, I., Venetsanopoulos, A.N.: Order statistics in digital image processing. *Proceedings of the IEEE* **80** (1992) 1892–1919
13. Astola, J., Haavisto, P., Neuvo, Y.: Vector median filters. *Proceedings of the IEEE* **78** (1990) 678–689
14. Plataniotis, K.N., Androutsos, D., Venetsanopoulos, A.N.: Adaptive fuzzy systems for multichannel signal processing. *Proceedings of the IEEE* **87** (1999) 1601–1622