

Vector Median Root Signals Determination for cDNA Microarray Image Segmentation

Rastislav Lukac and Konstantinos N. Plataniotis

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

Abstract. This paper presents a new cDNA microarray image segmentation framework. The framework uses robust vector median filtering to generate a root signal which is an image obtained from the input by repeatedly filtering it until no more changes occur. During the convergence to the root signal, the framework classifies the cDNA image data as either microarray spots or image background, and ideally separates the regular spots from the background. Thus, the obtained root signal represents the segmented microarray image. In addition, the framework excellently removes noise present in the cDNA microarray images and normalizes spots' intensities.

1 Introduction

Complementary Deoxyribonucleic Acid (cDNA) microarray technology is a powerful tool used to extract and interpret genomic information [1]-[5]. It has found applications in gene and drug discovery, toxicological research, and diagnosis of cancer, diabetes and genetic diseases [1],[2]. During the cDNA microarray experiment, Ribonucleic Acid (RNA) from both control and experimental sample are isolated and converted into cDNAs by the so-called reverse transcription process [6], and the obtained cDNAs are labelled with fluorescent probes, usually Cy3 for the control and Cy5 for the experimental channel [7]. After hybridization, heating and washing, a specialized scanner is used to acquire cDNA microarrays at the ~ 540 nm (Green) for the control (Cy3) and ~ 630 nm (Red) for the experimental channel (Cy5) respectively [2],[7]. The scanning procedure produces a Red-Green image (*Fig.1*) which is a multichannel vector signal which can be represented, for storing or visualization purposes, as the RGB color image with a zero blue component [2],[8],[9].

The objective of the microarray experiment is to measure the abundance of hybridized RNA and analyze the gene expression activity in the recorded samples [6],[8]. Based on a simple coloration concept, the observance of non-fluorescent black spots in the microarray image denotes no binding of RNA while the occurrence of red, green, or yellow spots suggests that RNA sequences from the experimental, control, or combined population of cells contribute to the abundance. Since strong noise and various impairments are present in cDNA microarray images, image processing is used prior the analysis step.

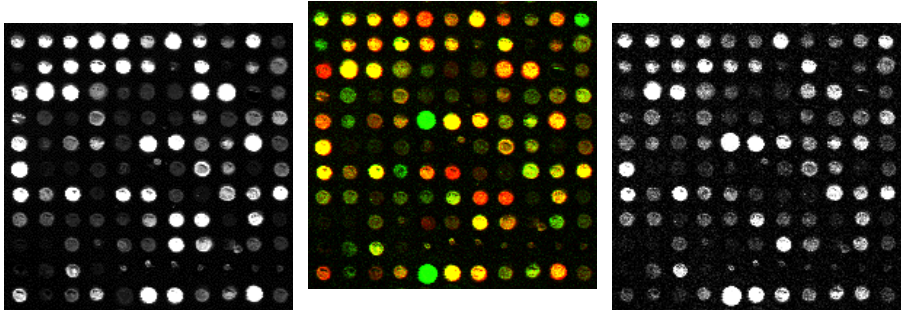


Fig. 1. Typical cDNA microarray image (middle) with the decomposed Red (left) and Green (right) channels

2 cDNA Microarray Imaging Basics

Let us consider, a $K_1 \times K_2$ Red-Green (RG) 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 [8],[9].

cDNA microarray images (*Fig.1*) suffer from noise introduced during image formation which is a complicated, nonlinear process influenced by many factors [2]. Therefore, image processing is necessary in order to eliminate processing errors from propagating further down the processing pipeline to the gene expression analysis tasks [1],[8]. The acquired cDNA signal $\mathbf{x}_{(r,s)} = [x_{(r,s)1}, x_{(r,s)2}]$ can be expressed as follows [9]:

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

where $\mathbf{o}_{(r,s)} = [o_{(r,s)1}, o_{(r,s)2}]$ represents the original, noise-free cDNA signal while $\mathbf{v}_{(r,s)} = [v_{(r,s)1}, v_{(r,s)2}]$ is used to denote the various image impairments introduced during image formation. The vectorial samples $\mathbf{v}_{(r,s)}$ are considered random in nature and can be modelled through the additive noise model [7]. The noise signal $\mathbf{v}_{(r,s)}$ is considered either impulsive in nature or it can be modelled as mixed noise (white additive Gaussian noise followed by impulsive noise) [2],[8].

Each acquired cDNA sample $\mathbf{x}_{(r,s)}$ is uniquely determined by its magnitude $M_{(r,s)} = \|\mathbf{x}_{(r,s)}\| = \sqrt{(x_{(r,s)1})^2 + (x_{(r,s)2})^2}$ and direction $D_{(r,s)} = \frac{1}{\|\mathbf{x}_{(r,s)}\|} \mathbf{x}_{(r,s)} = \frac{1}{M_{(r,s)}} \mathbf{x}_{(r,s)}$ in the vector space [2]. Due to the various image impairments, such as those attributed to [10],[11]: i) variations in the image background, ii) variations in the spot sizes and positions, iii) artifacts caused by laser light reflection and dust on the glass slide, and iv) photon and electronic noise introduced during scanning, the cDNA vector fields exhibit considerable variations in intensity (*Fig.1*) [2],[8].

3 Root Signals Based Segmentation Framework

The proposed cDNA image segmentation framework uses uniformity in the characteristics of the cDNA vectors as the base for segmentation [2]. Since the noisy samples deviate from other samples in a given data population [8], the determination of the outlying cDNA vectors is of a paramount importance. Due to the vectorial nature of the cDNA microarray image, the microarray image processing operator should process the cDNA image data as the set of vectors. The most popular vector processing techniques, such as those listed in [9], operate over magnitude or direction of cDNA vectors, or combine both magnitude and directional characteristics to ensure proper processing of vectors within the localized image area. Such solutions have been used to denoise and enhance cDNA microarray image, and detect spots' edges [8],[11],[12].

In this work, we use the vector processing concepts to build a unique segmentation framework. The framework employs the vector processing operator to produce a root signal which represents a segmented microarray image. As it is demonstrated in [2], by employing various vector processing operators the framework offers solutions which differ in their design philosophy, characteristics, and performance. To demonstrate the suitability of the proposed framework, the well-known vector median filter (VMF) [13] will be used in the sequence.

3.1 Vector Median Filter

The most popular vector operators such as the VMF are based on the theory of order-statistics [9],[14]. Since outliers are associated with the maximum extremes of aggregated distances to other input vectors in the sliding window, the output of the ordering-based vector filters is the lowest ranked vector in a predefined sliding window. Such an approach makes the processing operators robust to noise and various impairments present in cDNA microarray images.

The VMF operates over samples $\mathbf{x}_{(i,j)}$ localized within a sliding window $W_{(r,s)} = \{\mathbf{x}_{(i,j)}; (i,j) \in \zeta = \{(r-1, s-1), (r-1, s), \dots, (r+1, s+1)\}\}$. The filtering procedure replaces the sample $\mathbf{x}_{(r,s)}$ placed in the window center through a function $\mathbf{y}_{(r,s)} = f(W_{(r,s)})$ applied to a local neighborhood area $W_{(r,s)}$:

$$\mathbf{y}_{(r,s)} = \min_{\mathbf{x}_{(i,j)} \in W_{(r,s)}} \arg \sum_{(g,h) \in \zeta} \|\mathbf{x}_{(i,j)} - \mathbf{x}_{(g,h)}\|_L \quad (2)$$

This window operator slides over the entire image, for $r = 1, 2, \dots, K_1$ and $s = 1, 2, \dots, K_2$, to cover all the pixels in the microarray image [9]. In the expression above the term $\|\mathbf{x}_{(i,j)} - \mathbf{x}_{(g,h)}\|_L$ represents the L -metric (Minkowski metric [14],[15]) used to quantify the distance between two cDNA vectors $\mathbf{x}_{(i,j)}$ and $\mathbf{x}_{(g,h)}$:

$$\|\mathbf{x}_i - \mathbf{x}_j\|_L = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^L \right)^{\frac{1}{L}} \quad (3)$$

where L denotes the norm parameter, e.g. the city-block distance ($L = 1$) or considered here Euclidean distance ($L = 2$).

3.2 Root Signals

The VMF produces root signals if $\mathbf{y}_{(r,s)} = \mathbf{x}_{(r,s)}$, i.e. the filter output $\mathbf{y}_{(r,s)}$ is identical to a multichannel signal located at the reference window position (r, s) . Due to the localized nature of cDNA image features, the analysis of the root signals necessitates the definition of the following basic cDNA image structures which can be observed in the processing window $W_{(r,s)}$, [2],[16]: i) a multichannel constant region is a neighborhood formed by identical, in terms of both magnitude and direction, cDNA image vectors, ii) a multichannel cDNA step edge is a multichannel constant region followed by another multichannel constant region, iii) a multichannel impulse is a cDNA image vector which significantly deviates from a surrounding multichannel constant region, and iv) multichannel oscillation is a sequence of cDNA vectors which is not part of a constant region, an edge, or an impulse. The consideration of the structures defined above is essential in the proposed segmentation framework since the root signals consist solely of constant neighborhoods (background and spots) and edges (step transitions between spots and background) [2]. In order to obtain a root, defined over a cDNA microarray image, the elimination of impulses and oscillations (data variations) is an essential step.

Due to the low-pass nature of the smoothing operator, the utilization of the VMF (as well as some other robust vector filters) in a unique cascade cycle [2]: i) eliminates redundant information such as impulses and noise-like variations in the sample population, ii) converges to VMF roots (signals invariant to further processing by the VMF operator) that retain the spatial and spectral characteristics of the input cDNA image, and iii) produces a root signal which represents a segmented cDNA image. Since the VMF response to any input signal is uniquely defined in (2) with the VMF output $\mathbf{y}_{(r,s)} \in W_{(r,s)}$, the root signal [2]:

$$\mathbf{x}_{(r,s)} = f(W_{(r,s)}) \quad (4)$$

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

Assuming that $\mathbf{y}_{(r,s)}^n$ is a vector in the image \mathbf{y}^n obtained after filtering n times the input cDNA image \mathbf{x} , the convergence to a root signal can be expressed as a function of the difference between two successive filtering results [2],[17]:

$$\gamma = \|\mathbf{y}^n, \mathbf{y}^{n-1}\|_L \quad (5)$$

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

The VMF root signal based procedure divides an image into different regions that are homogeneous with respect to its magnitude characteristics. Since pixels with the same magnitude characteristics constitute meaningful regions, such as the spots present in a cDNA microarray image, the problem reduces to pixel classification defined as follows [2]:

$$\mathbf{x}_{(r,s)} \in \mathbf{x} \rightarrow \{F, B\} \quad (6)$$

where disjoint sets F and B denote the foreground and background cDNA vectors $\mathbf{x}_{(r,s)}$ in the microarray image \mathbf{x} , respectively. The foreground is constituted by microarray spots, typically of a circular shape of 150-200 cDNA vectors $\mathbf{x}_{(r,s)}$. A gap between spots or alternatively the presence of cDNA vectors residing outside spots areas constitute the background. By extracting the spots from the microarray image \mathbf{x} , the background can be viewed as a homogeneous region, while the essential foreground should remain heterogonous as a result of the variable spots' coloration [2]. The use of the VMF operator makes the vectors within individual spots uniform and removes noisy foreground information.

It is argued in [2] that the root signal convergence process performs morphological operations such as various compositions of erosion and dilatation operations defined over $W_{(r,s)}$, which is considered as the structuring element. By employing the robust filter, the proposed segmentation procedure removes "positive" noise (impairments present in background have magnitude larger than the desired background samples) similarly to morphological erosion or opening. At the same time, the procedure removes negative noise (cDNA vectors which are, in terms of amplitude of their components, smaller than neighboring vectors, and holes which have been created by cDNA vectors of zero-like magnitude) similarly to morphological dilatation or closing. In addition, the procedure removes any high-frequency impairments such as outliers in the cDNA image data population, and any small signal structures such as irregular spots or holes present in the microarray spots, performing thus erosion/opening or dilatation/closing operations.

The replacement of the window center $\mathbf{y}_{(r,s)}^{n-1}$ with the statistically most similar, to the cDNA samples within $W_{(r,s)}$, vector $\mathbf{y}_{(r,s)}^n$: i) produces an image which has normalized intensity in both the background and the spot locations, ii) enhances the difference between foreground and background information [2],[8]. As it is suggested in [2], the root-signal based segmentation procedure can be completed by performing thresholding operations over the magnitude of the root signal in order to remove residual irregular (small) spots and idealize the background in the segmented microarray image.

4 Experimental Results

In order to demonstrate the performance of the proposed segmentation framework, we used a number of cDNA microarray images, such as those shown in *Fig.1* and *Fig.2a*. The images have been captured using laser microscope scanners and vary in both complexity and noise appearance. The proposed method is compared, in terms of performance, against other segmentation techniques such as the clustering based segmentation technique and the morphological approach.

Fig.2 shows the input cDNA microarray image and the corresponding segmented images. As it can be seen in *Fig.2a*, the input cDNA image contains various foreground and background impairments, which usually affect the performance of the processing and analysis techniques. Visual inspection of the

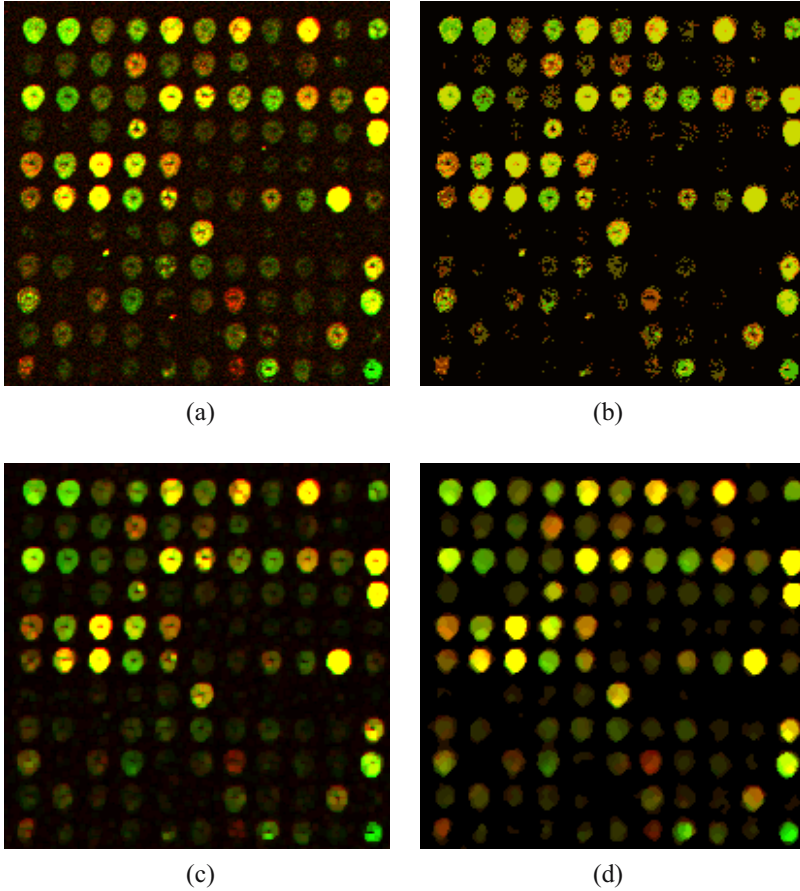


Fig. 2. Experimental results: (a) input cDNA microarray image, (b)-(d) segmented images obtained using (b) clustering technique, (c) morphological approach, (d) VMF root-signals based framework

image shown in *Fig.2b* reveals that the clustering approach is sensitive to noise. *Fig.2c* shows that the morphological approach delivers better performance, although it does not completely eliminate the holes present in the spots while at the same time enhances irregular spots in various image locations. However, the best results are obtained using the proposed solution. *Fig.2d* clearly shows that the proposed method removes noise, preserves the coloration of the spots and readily separates the foreground information from the background.

5 Conclusion

This paper presented a root-signal based concept for cDNA microarray image segmentation. A popular vector median filter was employed here to demonstrate

the concept of the proposed framework by producing a root signal which is invariant to further processing with the same type of the filtering operator. The achieved root signal represents a segmented microarray image which exhibits enhanced spot information and data normalization.

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. *Pattern Recognition, Special Issue on Bioinformatics*, submitted (2005)
3. 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
4. Dopazo, J.: Microarray data processing and analysis. In *Microarray Data Analysis II*, ed. SM Lin and KF Johnson, Kluwer Academic (2002) 43–63
5. Eisen, M.B., Brown, P.O.: DNA arrays for analysis of gene expression. *Methods in Enzymology* **303** (1999) 179–205
6. Whitchurch, A.K.: Gene expression microarrays. *IEEE Potentials* **21** (2002) 30–34
7. Nagarajan, R.: Intensity-based segmentation of microarrays images. *IEEE Transactions on Medical Imaging* **22** (2003) 882–889.
8. 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
9. 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
10. Wang, X.H., Istepian, R.S.H., Song, Y.H.: Microarray image enhancement using stationary wavelet transform. *IEEE Trans. Nanobioscience* **2** (2003) 184–189
11. 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: Spec. Iss. Fuzzy Sets and Systems in Bioinformatics* **152** (2005) 17–35
12. Lukac, R., Plataniotis, K.N., Smolka, B., Venetsanopoulos, A.N.: An automated multichannel procedure for cDNA microarray image processing. *Lecture Notes in Computer Science* **3212** (2004) 1–8
13. Astola, J., Haavisto, P., Neuvo, Y.: Vector median filters. *Proceedings of the IEEE* **78** (1990) 678–689
14. Lukac, R.: Adaptive vector median filtering. *Pattern Recognition Letters* **24** (2003) 1889–1899
15. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification and Scene Analysis*. John Wiley, 2nd Edition, (2000)
16. Tang, K., Astola, J., Neuvo, Y.: Multichannel edge enhancement in color image processing. *IEEE Trans. Circuits & Systems for Videotechnology* **4** (1994) 468–479
17. Burian, A., Kuosmanen, P.: Tuning the smoothness of the recursive median filters. *IEEE Transactions on Signal Processing* **50** (2002) 1631–1639