

SPOT IDENTIFICATION IN MICROARRAY IMAGES USING GAUSS-LAGUERRE WAVELETS

S. Pallavaram^{1}, J.A. Berger¹, S.K. Mitra¹, M. Carli² and A. Neri²*

¹ Dept. of Electrical & Computer Engineering, University of California
Santa Barbara, CA 93106-9560, USA

² Dept. of Electrical Engineering, University of Roma TRE
via della vasca, Navale, 84, 00146 Rome, Italy

ABSTRACT

In this paper, we introduce a new technique for the spot identification and quantification problem for DNA microarray images. Microarray printing, hybridization, and staining may all create substantial variability in the quality of data. Robust image processing algorithms are required to extract spot signal information from the slides. To this end we describe a strategy based on the representation of the image components by a family of circular harmonic functions (CHF) defining the Gauss Laguerre transform (GLT). We utilize the outputs of the CHF operators to identify low intensity spots. Our method has been found to be comparable to the DeArray microarray image processing package. The proposed algorithm clearly separates the signal or foreground pixels from the background in the enhanced output. This aids in the process of spot segmentation and subsequent spot quantification, and we demonstrate our technique on a publicly available set of microarray gene expression image data for a breast cancer cell line.

1. INTRODUCTION

Microarray hybridization experiments are used to profile transcription in parallel biological samples. A microarray slide contains thousands of spots and each spot corresponds to a unique gene. Separate fluorescent images of the test and reference dyes are taken using a confocal laser scanning microscope. A common problem in information extraction from these images is the identification of the intensity values for the spots [1]. Each spot provides quantitative information about the expression of a distinct DNA sequence. It is thus imperative that spots are found and quantified accurately.

* This work was supported in part by a University of California MICRO grant with matching support from Philips Research Laboratories and in part by Microsoft Corp.

One of the key problems is the detection and measurement of low intensity spots, some of which are nearly invisible. This phenomenon is especially frustrating because dim spots often correspond to interesting gene expressions that are unfortunately expressed at low levels [2]. We propose a method to process the original microarray image and identify spot foreground pixels, even for spots of poor quality. Our proposed technique is wavelet-based and robust for handling this kind of data.

We demonstrate our proposed technique by examining the gene expression image data for the BT474 cell line [3]. Moreover, we hand select genes that have been separately verified by reverse transcription-polymerase chain reaction (RT-PCR) experiments and compare the results of our quantification algorithm to those of DeArray [4]. Concluding remarks are made to address how our algorithm performs.

2. METHOD

The quality of microarray data varies across the slide. Small measurement errors may have a drastic impact on the analysis [5]. Therefore, it is crucial to discard low quality observations in the early phases of a microarray experiment to prevent erroneous biological conclusions. To identify spots after quality filtering, the Gauss-Laguerre transform (GLT) has been chosen because its formulation fits closely with the human visual system. In addition, this method has been shown to successfully enhance low intensity images for further analysis [6].

2.1. The Gauss- Laguerre Transform

Circular harmonic functions (CHF)s are complex polar-separable filters defined by a point spread function of the form

$$h^{(n)}(r, \theta) = v^{(n)}(r) e^{-jn\theta}, \quad n = 0, 1, 2, \dots \quad (1)$$

where r and θ are in polar coordinates, n is the order of the CHF and $v^{(n)}(\bullet)$ is the radial profile [7]. We employ

the specific family of CHF's with Gauss-Laguerre (GL) radial profiles

$$\mathcal{L}_k^{(n)}(r, \theta) = (-1)^k 2^{(n+1)/2} \pi^{n/2} \left[\frac{k!}{(|n|+k)!} \right]^{1/2} r^{|n|} L_k^{(|n|)}(2\pi r^2) \times e^{-\pi r^2} e^{jn\theta}, \quad (2)$$

where $\mathcal{L}_k^{(n)}$ are generalized Laguerre polynomials

$$L_k^{(n)}(t) = \sum_{h=0}^k (-1)^h \binom{n+k}{k-h} \frac{t^h}{h!}. \quad (3)$$

2.2. Algorithm

Let $f(x, y)$ be the input image and $f_P(r, \theta)$ be its polar form. The foreground of the input image $f_F(r, \theta)$ is extracted using a thresholding algorithm. We employed the triangle thresholding algorithm due to Zach et al [8]. The reason for choosing this technique is that the histogram of a microarray image usually has a smooth gradient decay with only a little random variation.

Next, we project the original image onto the edge domain, where it is easy to differentiate the spot signal from the background. This edge extraction is based on the Gauss Laguerre wavelet (GLW) decomposition of the first angular harmonic. In fact, first-order GL functions $\mathcal{L}_0^{(1)}(r, \theta)$ act as differential operators whose outputs are complex images with magnitudes proportional to their edge strengths, and phases equal to their edge orientations. The subscript k represents the degree and the superscript (l) indicates the first order GL function. Since our objective is to determine the edge strength we are concerned only with the magnitude. For example, the complex operator $\mathcal{L}_0^{(1)}(r, \theta)$ produces a complex output with real and imaginary components proportional to the respective horizontal and vertical derivatives of the low pass filtered input. This low pass filter has a Gaussian transfer function. A filter scale of unity is chosen to obtain the GLW filter. The response of GLW filter $\mathcal{L}_0^{(1)}(r, \theta)$ of first order and degree 0 for the input image is computed as follows

$$f_0(r, \theta) = \left| f_P(r, \theta) * \mathcal{L}_0^{(1)}(r, \theta) \right|, \quad (4)$$

where $*$ represents convolution. The foreground of this response is then extracted using the triangle thresholding algorithm to arrive at $f_{F0}(r, \theta)$. We then combine the foreground of the original image with the foreground of the response of the GLW filter to obtain the enhanced image $f_E(r, \theta)$, namely,

$$f_E(r, \theta) = f_F(r, \theta) + f_{F0}(r, \theta). \quad (5)$$

The flowchart of the proposed spot boundary detection algorithm is shown in Figure 1.

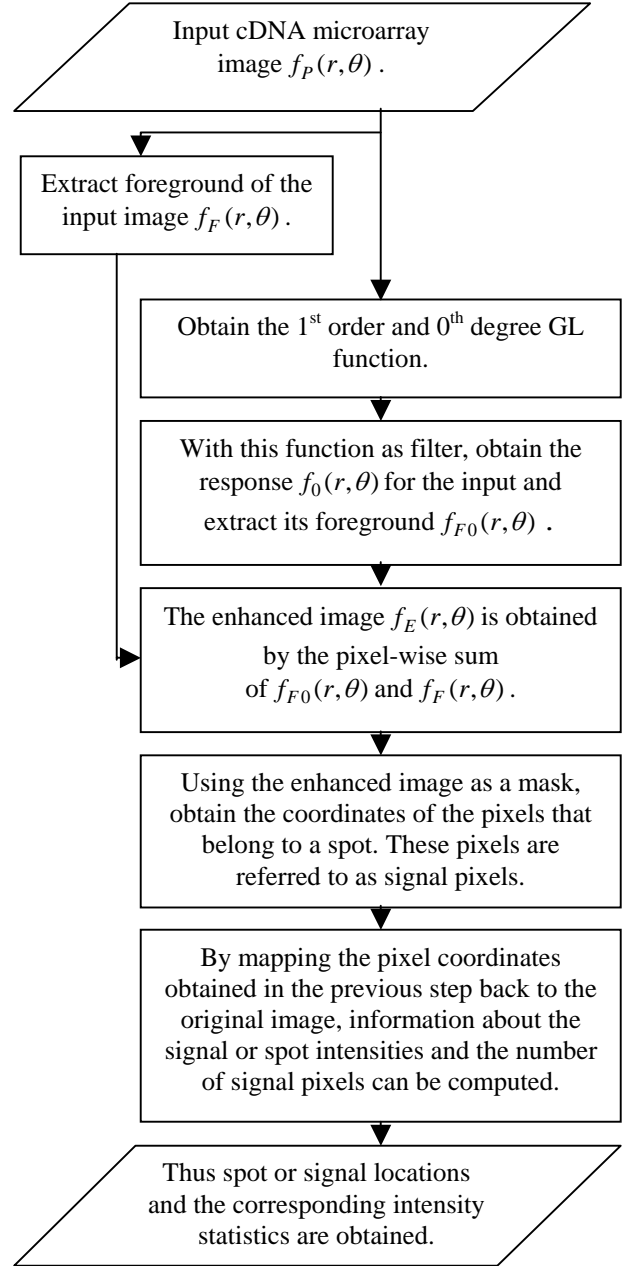


Figure 1. GLW decomposition flowchart for spot identification and quantification.

Since the contribution of the background of the input image to the process of obtaining the enhanced image is minimal, the effect of background noise in the input image is greatly suppressed. Further research is required with respect to handling bleeding between spots.

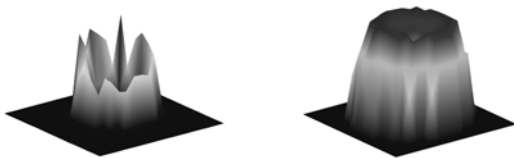


Figure 2. (a) Original spot, and (b) Enhanced mask.

2.3. Results and Discussion

The surface plots shown in Figures 2(a) and 2(b) provide good visualization of the degree of enhancement achieved. The height at a given point on the flat rectangular surface is an indication of the intensity of the pixel at that location. The contour at the top of the enhanced mask is the contour of the detected spot. The annular area between the larger contour, which is at the bottom of the mask, and the detected spot contour denotes edge enhancement. This edge enhancement aids in the segmentation of spots. Only those pixels that belong to the contour at the top of the enhanced mask are identified as pixels belonging to the spot.

Next, the coordinates of these signal pixels are obtained. This process of identifying signal pixels can be done either by a thresholding process of the intensities in the enhanced image or simply by choosing those pixels that have the maximum intensity value in the enhanced image. The latter method of identifying signal pixels from the enhanced mask guarantees, in general, that none of the background pixels are identified as signal. The trade-off is that it is possible to miss some signal as background.

It has been confirmed by repeated experimentation of our algorithm that the enhanced image is able to locate foreground pixels accurately. The intensities for the pixels identified as signal pixels using the enhanced mask are then extracted from the original image for analysis. Segmentation of spots in the original image shown in Figure 3 is performed using the enhanced image as a mask. Canny's edge detection method [10] is used. The segmentation result is shown in Figure 4.

Identifying signal pixels from the enhanced image, our algorithm performs comparably to DeArray in the analysis of most of the spots. DeArray is the central processing tool controlling most of the image processing tasks in the MicroArray suite software developed by Scanalytics [4]. There are roughly 10,000 different genes and comparing spot by spot is difficult to do. In this contribution, for comparative analysis, we randomly selected 100 spots from the BT474 microarray image data. Spot identification and quantification was performed and analyzed for both methods. In 66 of these spots our method detected more signal pixels than DeArray, while



Figure 3. Original image (18 spots shown).



Figure 4. Segmentation of original image.

in 17 spots DeArray detected 15% or more signal pixels than our method. This consequently affects how the foreground and background are computed. We note that even a small number of pixels that exhibit differences in assignments of foreground or background result in large effects on spot quantification. Overall, for these 100 spots that we looked at, our algorithm performed comparably to DeArray, however a few spots were assigned different intensity values.

Our algorithm's performance is further compared to DeArray in detail for a small set of over-expressed genes. We illustrate this using select genes from the BT474 cell line presented in Table 1. The BT474.tif image has been acquired from the website [11]. It has been verified by Hyman et al [9] that the cDNA clones listed in Table 1 are over-expressed. It can be seen from Table 1 that our method reports over-expression of all the genes considered while DeArray is able to detect over-expression only for the second gene in the table. Further experimentation on these genes is needed for conclusive analysis, but our results correspond to other experimental results using RT-PCR.

3. CONCLUDING REMARKS

A new method for the spot identification problem in cDNA microarray images has been presented. The method is based on image decomposition using CHF's for spot enhancement. We selected genes highly influential to breast cancer detection in the BT474 cell line and our experimental results show the effectiveness of the proposed scheme for quantifying these spots. This algorithm can be used for accurately locating spot boundaries and spot signal pixels. Using the enhanced image as a mask, the intensity values of the spot pixels in the original image can now be determined in a robust manner. The procedure can be used to address the challenging problem of identifying and eventually

Table 1. Comparison between DeArray and the proposed GLW algorithm for selected clones from the BT474 cell line (over-expression is denoted by O/E, under-expression is denoted by U/E and not available is denoted by N/A).

Clone Name	Cytoband	Detected signal pixels		Log(R/G)		Classification	
		DeArray	GLW	DeArray	GLW	DeArray	GLW
v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (replicate 1)	17q11.2-q12	0	96	0	0.13	N/A	O/E
v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (replicate 2)	17q11.2-q12	74	75	2.94	0.24	O/E	O/E
MLN51 protein	17q11-q21.3	0	36	0	0.73	N/A	O/E
homeo box B2	17q21-q22	88	115	-3.22	0.41	U/E	O/E
homeo box B7	17q21-q22	32	74	-0.63	1.45	U/E	O/E

assessing the quality of spots that have very low intensity. This idea is currently under study. Low intensity spots may represent genes that are vital to the experiment outcome, but have been expressed at a very low level in one of the two channels. In this work, we used the spot quality assessment procedure in [3] to reject spots of poor quality. Since the contribution of the background of the input image to the process of obtaining the enhanced image, and thus in the identification of signal pixels is minimal, the effect of background noise in the input image is greatly suppressed. Spot segmentation from the enhanced output obtained using our algorithm has also been demonstrated to show the location of spots.

4. REFERENCES

[1] M. Katzer, F. Kummert and G. Sagerer, "Robust automatic microarray image analysis," *Proc. of the International Conference on Bioinformatics: North-South Networking*, Bangkok, 2002.

[2] J. Buhler, T. Ideker and D. Haynor, "Dapple: Improved techniques for finding spots on DNA microarrays," *University of Washington Technical Report*, Aug. 2000.

[3] A.-K. Järvinen, S. Hautaniemi, H. Edgren, P. Auvinen, J. Saarela, O.-P. Kallioniemi and O. Monni, "Are data from different gene expression microarray platforms comparable?" *Genomics*, in press.

[4] Y. Chen, V. Kamat, E.R. Dougherty, M.L. Bittner, P.S. Meltzer and J.M. Trent, "Ratio statistics of gene expression levels and applications to microarray data analysis," *Bioinformatics*, pp. 1207-1215, Vol. 18, 2002.

[5] S. Hautaniemi, H. Edgren, P. Vesänen, M. Wolf, A.-K. Järvinen, O. Yli-Harja, J. Astola, O. Kallioniemi and O. Monni, "A novel strategy for microarray quality control

using Bayesian networks," *Bioinformatics*, pp. 2031-2038, Vol. 19, 2003.

[6] M. Carli, A. Neri, S. Hatipoglu and S.K. Mitra, "Texture based classification and segmentation using the Laguerre-Gauss transform," *Proc. 2nd International Symposium on Image and Signal Processing and Analysis*, Pula, Croatia, pp. 19-21, June 2001.

[7] G. Jacovitti and A. Neri, "Multiresolution circular harmonic decomposition," *IEEE Transaction on Image Processing*, pp. 3242-3247, Vol. 48, 2000.

[8] G.W. Zack, W.E. Rogers, and S.A. Latt, "Automatic measurement of sister chromatid exchange frequency," *Journal of Histochemistry and Cytochemistry*, pp. 741-753, Vol. 25, 1977.

[9] E. Hyman, P. Kauraniemi, S. Hautaniemi, M. Wolf, S. Mousses, E. Rozenblum, M. Ringnér, G. Sauter, O. Monni, A. Elkahoun, O.-P. Kallioniemi and A. Kallioniemi, "Impact of DNA amplification on gene expression patterns in breast cancer," *Cancer Research*, pp. 6240-6245, Vol. 62, 2002.

[10] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 679-698, Vol. PAMI-8, 1986.

[11] http://sigwww.cs.tut.fi/TICSP/Jarvinen_et_al_2003/