An Automatic Microarray Image Gridding Technique Based on Continuous Wavelet Transform

Emmanouil Athanasiadis[1], Dionisis Cavouras[2], Panagiota Spyridonos[1], Ioannis Kalatzis[2], and George Nikiforidis[1]

[1] Medical Image Processing and Analysis (*M.I.P.A.*) Group, Laboratory of Medical Physics, School of Medical Science, University of Patras, 26500 Rion - Patras, Greece
mathan@upatras.gr, p_spyridonos@yahoo.com, gnikif@med.upatras.gr
http://mipa.med.upatras.gr
[2] Medical Image and Signal Processing (*MED.I.S.P.*) Laboratory, Department of Medical Instruments Technology, Technological Educational Institute of Athens, Ag. Spyridonos Street, Aigaleo, 122 10, Athens, Greece
{cavouras, ikalatzis}@teiath.gr
http://www.teiath.gr/stef/tio/medisp/

Abstract. In the present study, a new gridding method based on continuous wavelet transform (CWT) was performed. Line profiles of x and y axis were calculated, resulting to 2 different signals. These signals were independently processed by means of CWT at 15 different levels, using daubechies 4 mother wavelet. A summation, point by point, was performed on the processed signals, in order to suppress noise and enhance spot's differences. Additionally, a wavelet based hard thresholding filter was applied to each signal for the task of alleviating the noise of the signals. 14 real microarray images were used in order to visually assess the performance of our gridding method. Each microarray image contained 4 sub-arrays, each sub-array 40x40 spots, thus, 6400 spots totally. Moreover, these images contained contamination areas. According to our results, the accuracy of our algorithm was 98% in all 14 images and in all spots. Additionally, processing time was less than 3 sec on a 1024×1024×16 microarray image, rendering the method a promising technique for an efficient and fully automatic gridding processing.

Keyword: Microarrays, gridding, Continuous Wavelet Transform.

1 Introduction

Microarray imaging is used for the simultaneous identification of thousands of genes in bioinformatics [1]. In a complementary DNA (cDNA) microarray experiment, mean fluorescence intensity values are calculated. These intensities are closely related to the expression level of a specific gene. Hence, the more precise the localization of a spot, the more accurate the intensity measurement is. Consequently, a more precise expression measurement of a gene is obtained.

W.G. Kropatsch, M. Kampel, and A. Hanbury (Eds.): CAIP 2007, LNCS 4673, pp. 864 – 870, 2007.

In order to acquire spot intensity measurements, three major steps are followed [1]-[2]: 1/ the gridding or addressing step, where a precise localization of the spots with their surrounding is determined, 2/ the segmentation step, where a discrimination of the cell's foreground from background is accomplished and 3/ the intensity extraction step, where calculation of the mean fluorescence value of each spot is performed.

Nevertheless, all the above steps are not a trivial issue, due to the fact that microarray images are usually highly contaminated with noise and artifacts during their construction [1]. More precise, rotations, misalignments, and local deformations of the ideally rectangular grids of the image occur, rendering the whole process demanding. Gridding is the first step in the chain of the expression measurement process. It is vital that the gridding should be accurate since any errors during addressing procedure are transferred to the following steps.

A lot of work has already been done for the task of automatic gridding processing. However, initialization or other parameters are needed in order the gridding to be applied. Jain et al. [3] described a gridding method that requires as input the rows and columns of all grids. In Katzer et al [4] technique, gaps between the grids are essential while in Steinfath et al. [5], filter arrays with radioactive label are used. On the other hand, there are many researchers that create manually the grid and then perform an automatic procedure for intensity measurement [1].

In this paper, a novel full automatic gridding approach is presented, based on the properties of the continuous wavelet transform (CWT) [6]. This technique attempts to solve the automatic gridding procedure problem efficiently, providing a tool for an accurate and fast microarray image processing.

2 Material and Methods

In the present study, 9 microarray images, concerning Saccharomyces cerevisiae, were obtained by a publicly available database [7]. In those images, contamination areas, as well as a slight shift of the alignment were visible. Each image consists of 4 subarrays, each subarray consists of 40x40 spots, leading to 6400 spots, totally.

2.1 Continuous Wavelet Transform

CWT is a transform that is used to decompose a signal into wavelets, small oscillations that are highly localized in time and described by eq. 1.

$$C(a,b) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{|a|}} \overline{\psi\left(\frac{t-b}{a}\right)} dt \tag{1}$$

where, *f(t)* is the original signal, *á* represents scale, *b* represents time or space, ϕ represents the mother wavelet and $\overline{z}$ represents the complex conjugate of *z*.

The choice of Daubechies (db) was guided by the fact that this mother wavelet family resulted in better results. Furthermore, we worked with only four coefficients (db 4), since using a greater number of coefficients, have not resulted in any significant change in our results. On the other hand, keeping the number of coefficients, as low as possible it was essential for our methodology since the processing time is an important issue for the gridding procedure.

After applying the CWT, a hard-threshold wavelet based technique [8] was performed according to eq. 2.

$$W_{out} = \begin{cases} W_{in} + T \cdot (G-1) \text{ if } W_{in} > T \\ W_{in} - T \cdot (G-1) \text{ if } W_{in} < -T \\ 0 \qquad \text{otherwise} \end{cases} \tag{2}$$

where *Wout* denotes the output and *Win* the input coefficient values of the details. *T* and *G* are threshold and gain values respectively (Fig 1).

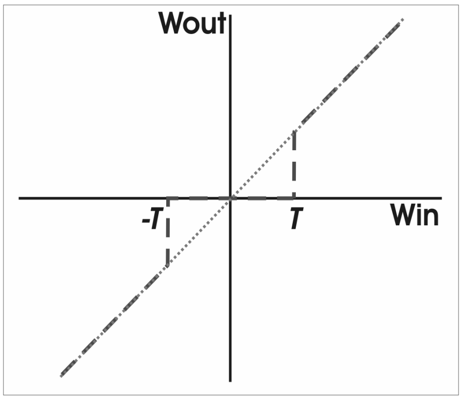


Fig. 1. Wavelet-based hard thresholding mapping layout

2.2 Gridding Procedure

The gridding procedure was accomplished as follow: firstly, line-profiles for x and y axes were calculated and two signals were obtained. Secondly, the CWT was applied to both signals, at 15 scales, using db4 mother wavelet, and 15 detail images for each signal were obtained (Fig.2). Scale was set to 15, due to the fact that in higher scales, no significant information was divulged. During decomposition process, high frequency details are being distinguished more accurately from noise structures, as noise is not present in all decomposed signals, in contradiction to the alternations of spots with their background that can be present in all levels. Thirdly, a summation point by point of the signals of all the 15 scales was performed for the purpose of increasing the actual signal of the spots and decreasing the noise. Fourthly, a hard- thresholding wavelet based technique [6] was applied to each signal for the task of suppressing noise (Fig.3).

Finally, local maxima and local minima were calculated on both signals, corresponding to the centers and the boundaries of the spots respectively. Optimal results were obtained by setting the threshold equal to 10% of the maximum signal value and the gain equal to unity. The described method is briefly summarized in the following steps:

```
Step 1: Load the initial Image
Step 2: Calculate line profiles of x and y axis
Step 3: Apply the CWT on Both x and y axis signals, up to
15 scales using as mother wavelet the db4.
Step 4: For each one of the two signals, make a summation
point by point from 1st to 15th the coefficients of the
CWT.
Step 5: Apply wavelet denoising technique using as
threshold the 10% of the maximum value of each signal.
Step 6: Find the local maxima (centers of spots) and lo-
cal minima (boundaries of spots) in x and y axis.
```

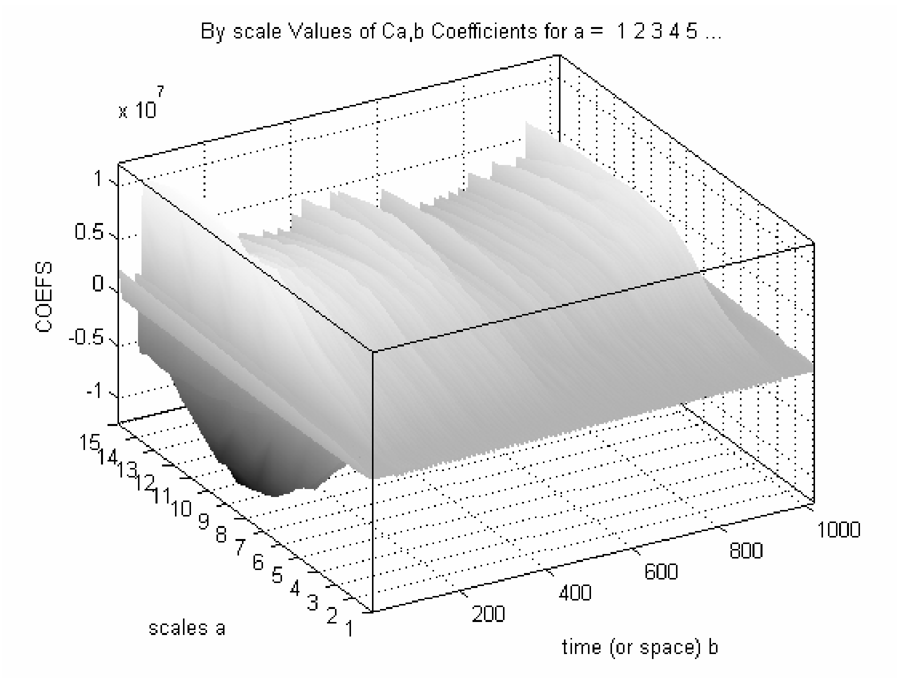
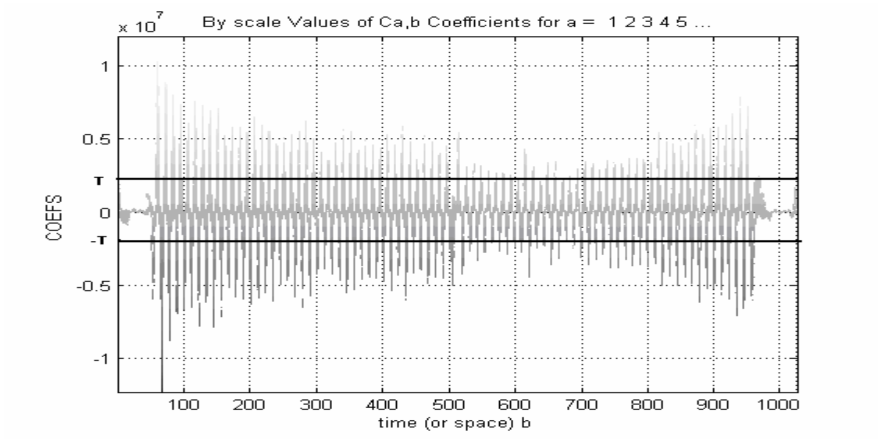
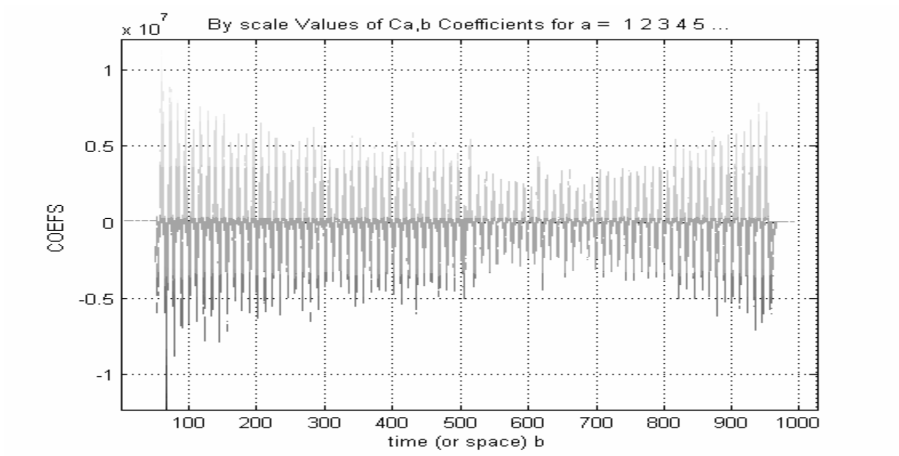


Fig. 2. X-axis signal processed by the CWT in 3 dimensions (coefficient, time or space, scale). In this image, scales a are ranging from 1 to 15, space b is the dimension of the signal (here 1024) and COEFS are the resulted coefficient values of the wavelet transform.



(a)



(b)

Fig. 3. Initial (a) and processed with hard threshold wavelet based filtering technique signal (b). Local maxima corresponded to the centers and local minima corresponded to the boundaries of the cells. In (b) noise has been suppressed by the wavelet threshold filter.

3 Results and Discussion

The proposed method was evaluated by visually inspecting [9] the gridding results and assigning each spot into two categories: either perfectly surrounding the spot (category 1) or not (category 2). Additionally, our method was compared with the method proposed by Blekas et. al. [9]. According to our findings, our algorithm was 98% accurate in all the 14 microarray images in opposition to the other method that scored 95%. It should be noted that microarray images were not pre-processed, thus

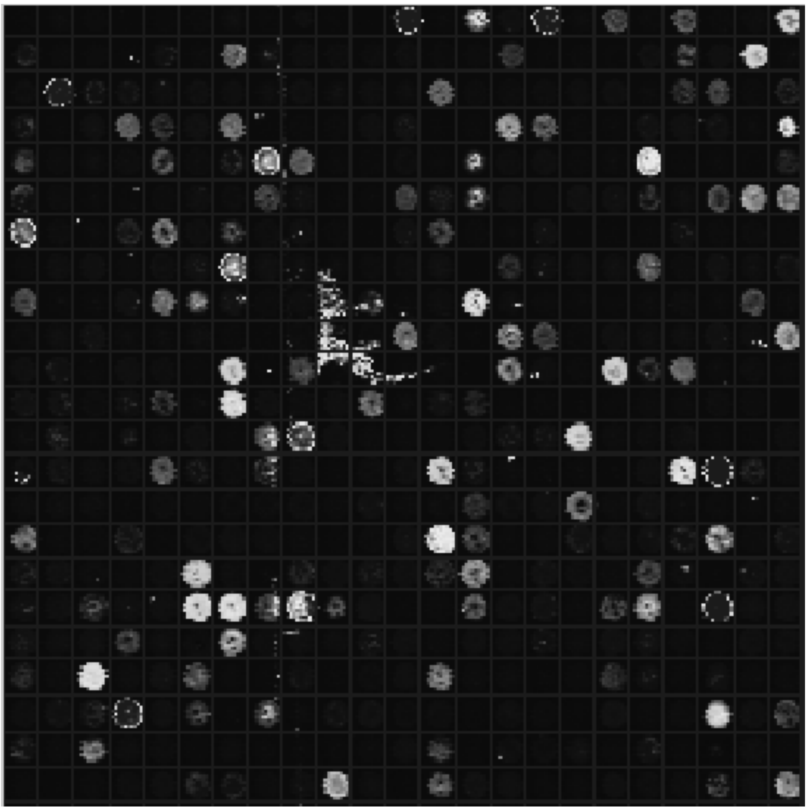


Fig. 4. Gridding results of contaminated arrayer. The contamination is being localized at the middle of the figure.

noise affected the outcome of both gridding algorithms. Nevertheless, in the proposed technique, noise suppression by using the hard-thresholding filter, leaded to higher accuracy results. A typical contaminated area of a microarray image with the grid overlaid is illustrated in Fig. 4. As we can conclude by the results, the contamination at the boundaries of the arrayer is eliminated by the hard thresholding technique.

It should be noted that the processing time for the gridding, on a 1024×1024×16 microarray image, was lower than 3sec (processor: PentiumIV 3.00GHz, 512 MB RAM), rendering the technique a valuable tool for a fully automated microarray image processing application.

4 Conclusion

In the present study we have proposed a new method for automatic addressing of microarray images based on continuous wavelet transform Following this technique, the noise of microarray images was sufficiently suppressed, and the boundaries of the spots were delineated with high accuracy. The processing time was minimal providing the method an effective tool for the demanding task of microarray image processing.

Acknowledgements

We would like to thank the Greek State Scholarships Foundation (I.K.Y.) for funding the above work.

References

1. Yang, Y.H., Buckley, M.J., Duboit, S., Speed, T.P.: Comparison of methods for Image Analysis on cDNA Microarray Data. Journal of Computational and Graphical Statistics 11, 108–136 (2002)
2. Schena, M., Shalon, D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarrray. Science 270, 467–470 (1995)
3. Jain, A., Tokuyasu, T., Snijders, A., Segraves, R., Albertson, D., Pinkel, D.: Fully Automatic Quantification of Microarray Image Data. Genome Res, 325–332 (2002)
4. Katzer, M., Kummert, F., Sagerer, G.: Automatische Auswertung von Mikroarraybildern. In: Workshop Bildverarbeitung f'ur die Medizin, Leipzig (2002)
5. Steinfath, M., Wruck, W., Seidel, H.: Automated image analysis for array hybridization experiments. Bioinformatics 17, T. 7 S., 634–641 (2001)
6. Ingrid Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics), Soc for Industrial & Applied Math (1992)
7. DeRisi, J.L., Iyer, V.R., Brown, P.O.: Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. SCIENCE 278, 680 (1997)
8. Athanasiadis, E., Glotsos, D., Daskalakis, A., Bougioukos, P., Kostopoulos, S., Theocharakis, P., Spyridonos, P., Kalatzis, I., Nikiforidis, G., Cavouras, D.: Microarray Image Enhancement Techniques Using the Discrete Wavelet Transform. In: Second International Conference From Scientific Computing to Computational Engineering (2nd IC-SCCE 2006), Athens (July 5 - 8 2006)
9. Blekas, K., Galatsanos, N., Likas, A., Lagaris, I.E.: Mixture Model Analysis of DNA Microarray Images. IEEE Transactions on Medical Imaging 24, 901–907 (2005)