2nd International Conference on Experiments/Process/System Modelling/Simulation & Optimization
2nd IC-EpsMsO
Athens, 4-7 July, 2007

AN EFFICIENT CLAHE-BASED, SPOT-ADAPTIVE, IMAGE SEGMENTATION TECHNIQUE FOR IMPROVING MICROARRAY GENES' QUANTIFICATION

Antonis Daskalakis[1,*], Dionisis Cavouras[2], Panagiotis Bougioukos[1], Spiros Kostopoulos[1], Pantelis Georgiadis[1], Ioannis Kalatzis[2], and George Nikiforidis[1]

[1]University of Patras, Department of Medical Physics, School of Medicine, Rio, GR-26503 Greece
[2]Medical Signal and Image Processing Lab, Department of Medical Instrumentation Technology, Technological Educational Institute of Athens, Ag. Spyridonos Street, Aigaleo, 122 10, Athens, Greece
e-mail: cavouras@teiath.gr, web page: http://www.teiath.gr/stef/tio/medisp/

Keywords: microarrays, enhancement, segmentation, cDNA,

Abstract. *An efficient spot-adaptive segmentation technique was developed by suitable combining in a cascade mode the benefits of image enhancement (Contrast Limited Adaptive Histogram Equalization technique (CLAHE)) and image segmentation (Seeded Region Growing technique (SRG)) in order to improve genes' quantification in microarray images. Microarrays utilized for evaluation purposes comprised 7 publicly available images. Initially, an image griding algorithm was employed to divide the image into rectangular image-cells. Subsequently, CLAHE was applied on each individual image-cell, initial SRG-seed was set at the image-cell's center, and SRG-threshold was estimated from the image-cell's background. The spot's boundary was referred to the corresponding cell spot in the original image and the spot's intensity was evaluated. Extracted intensities were comparatively evaluated against a well-established commercial software package (MAGIC TOOL) employing the Jeffrey's divergence-metric. The metric of the spot-adaptive segmentation technique was about double as compared to MAGIC TOOL's metric, with differences ranging between 1.23 and 5.21 in the processed images. Regarding processing time, the proposed method required half the time of MAGIC TOOL's (211 secs against 487 secs) to process the same cDNA image on the same computer.*

1 INTRODUCTION

Complementary DNA (cDNA) microarray technologies are hybridization based methods that enable the simultaneous assessment of the expression levels of thousands of genes [1-3]. In this way, microarrays provide an easy way to compare gene expression profiles between biological samples, by detecting either their expression or differential expression.

Initially, the two messenger RNA (mRNA) samples to be compared are reverse transcripted into cDNA and printed on a glass microscope slide by a robotic arrayer, thus, forming circular spots of known diameter. Subsequently, samples are labeled with red and green fluorescent dye, respectively, are mixed and competitively hybridized to the microarray slide [4, 5]. The end product of the comparative hybridization experiment is scanned, using lasers that excite each dye on the appropriate wavelength. The relative fluorescence between each dye on each spot, representing a gene, is recorded using methods contingent upon the nature of the labeling reaction , i.e. confocal laser scanners and Charged Couple Devices [6].

The output of such systems is two 16-bit TIFF images, one for each fluorescent channel. From the fluorescence intensities of each channel, that are associated to each spot, the relative expression levels of the genes in both samples are estimated [7, 8]. Extraction of genes expression levels is accomplished via image analysis techniques namely griding, spot segmentation, and intensity extraction [9-12]. Griding is the process of identifying and locating the coordinates of each cell containing the spot; the cell is a rectangular region containing the pixels of both the spot and its background. Segmentation refers to the classification of cell-pixels as either signal (spot's foreground) or surrounding area (spot's background). Spots' intensity extraction refers to the calculation of the fluorescence signal's mean intensity for the spot's foreground. Extracted mean intensities correspond to gene expression levels that, in turn, are translated into biological conclusions from molecular biologists, by employing data mining techniques.

However, microarray experiments involve a number of error-prone steps (occurring during fabrication, target labeling, and hybridization), which induce noise on the resulting images [13, 14]. Microarray images are also corrupted by irregularities in the shape, size, and position of the spot [14, 15]. Unless these sources of error are

addressed, they will propagate throughout the stages of the analysis, leading to inaccurate biological inferences.

Although a variety of image pre-processing techniques have been suggested for correcting these error sources on cDNA images [16], existing software tools, utilized for the analysis of microarray images, focus mainly on accurate spot localization and segmentation by various segmentation techniques [17-22]. Only few studies [23-27] have examined the impact of image pre-processing on cDNA image quality, however, without evaluating the effect of image enhancement on spot segmentation.

In the present study, an efficient spot-adaptive segmentation technique was developed by suitable combining in a cascade mode the benefits of image enhancement (Contrast Limited Adaptive Histogram Equalization technique (CLAHE) [16]) and image segmentation (Seeded Region Growing technique (SRG) [28]). Initially, an image griding algorithm was employed to divide the image into rectangular image-cells. Subsequently, CLAHE was applied on each individual image-cell, initial SRG-seed was set at the image-cell's center, and SRG-threshold was estimated from the image-cell's background. The spot's boundary was referred to the corresponding cell spot in the original image and the spot's intensity was evaluated. Extracted intensities were comparatively evaluated against a well-established commercial software package (MAGIC TOOL) employing the Jeffrey's divergence-metric in a publicly available dataset of real cDNA images [29].

2 MATERIAL AND METHODS

Material consisted of 7 microarray images downloaded from a publicly available database of the MicroArray Genome Imaging & Clustering Tool (MAGIC) website [29]. Each image contained 6400 spots investigating the diauxic shift of Saccharomyces cerevisiae. In the particular dataset, the authors [30] have used a common reference messenger RNA pool (green, Cy-3) to control for biological variability [31-33]. Such a design provides an adequate degree of replication, required for the quantitative assessment of image segmentation and subsequent gene quantification.

2.1 Spot Adaptive Segmentation

Prior to spot segmentation, a griding procedure was applied [34] for dividing the cDNA image into rectangular, spot-containing, cell-images. Following griding, individual cell-images, containing spots, were enhanced using the CLAHE method [35-37]. CLAHE is a special case of the histogram equalization technique [16] that functions adaptively on the image to be enhanced. Application of CLAHE individually (in each cell-spot) maximizes the contrast throughout the cell-spots by adaptively enhancing the contrast of each cell-pixel relative to its local neighborhood. The procedure for enhancing individual cell-images by employing the CLAHE technique is described below:

Step 1: Each cell-image was divided into a number of non-overlapping contextual regions of equal sizes, experimentally set to be 2x2, which corresponds to approximately 40 pixels.

Step 2: The histogram of each contextual region was calculated.

Step 3: A clip limit, for clipping histograms, was set (t=0.001). The clip limit was a threshold parameter by which the contrast of the cell-image could be effectively altered; a higher clip limit increased cell-spot contrast.

Step 4: Each histogram was redistributed in such a way that its height did not exceed the clip limit.

Step 5: All histograms were modified by the transformation function

$$T(r_k) = \sum_{j=0}^{k} p_r(r_j) \tag{1}$$

where

$$p_r(r_j) = \frac{n_j}{n} \tag{2}$$

is the probability density function of the input image grayscale value j, n is the total number of pixels in the input image and n_j is the input pixel number of grayscale value j.

Step 6: The neighboring tiles were combined using bilinear interpolation and the cell-image grayscale values were altered according to the modified histograms.

Following individual cell-image enhancement, the corresponding cell-spot was segmented from its background employing the SRG algorithm. SRG segmented the cell-image into pixel regions with respect to a pre-defined seed, set here to be the centre of the rectangular cell-image. Following an iterative procedure, SRG grew pixel regions, by assigning the most homogeneous neighboring pixels, employing a homogeneity criterion: the chosen pixel's intensity should be 1/higher than an estimated noise threshold, which was calculated by the standard deviation of the image-cell's background (using 2-pixels rectangular frame within the cell-image's edges) and 2/ close to the mean intensity of the so far seeded region. This growing procedure was repeated until all pixels in the cell-image were assigned to either the spot or its background. The spot's boundary, thus determined, was referred to the corresponding cell spot on the original image and the spot's intensity was evaluated. This was necessary, since intensities in the processed cell-spots were altered by the enhancement

process.

2.2 Metric for segmentation efficiency

Following segmentation, foreground (spot) and background intensity values for the common reference channel (green, Cy-3) were extracted and, considering all segmented spots, two density distributions were produced, employing a non-parametric kernel density estimation method. The distance between those two distributions was determined employing the Jeffrey's (J) measure of divergence, shown in (3):

$$J(S,B) = \sum_i \left(p_{B,i} - p_{S,i} \right) \log \frac{p_{B,i}}{p_{S,i}} \quad (3)$$

where $p_{S,i}$ and $p_{B,i}$ are the density distributions of the extracted intensities of all image spots and backgrounds respectively [38].

Higher values of J correspond to more distant distributions and, consequently, to more accurate segmentation, considering that intensities are evaluated on the original image alone. Extracted intensities employing the CLAHE-SRG segmentation were comparatively evaluated, in terms of J divergence, against the intensities obtained by recently published commercial software (MAGIC TOOL[39]). For evaluation purposes, the same microarray images were introduced to both methods and the SRG segmentation option of the MAGIC TOOL was chosen.

3 RESULTS

Figure 1 shows the result of the CLAHE-SRG segmentation. Table 1 presents the results of the Jeffrey's divergence between spot and background intensity distributions (of the common reference channel) for the proposed methodology and the MAGIC TOOL respectively.

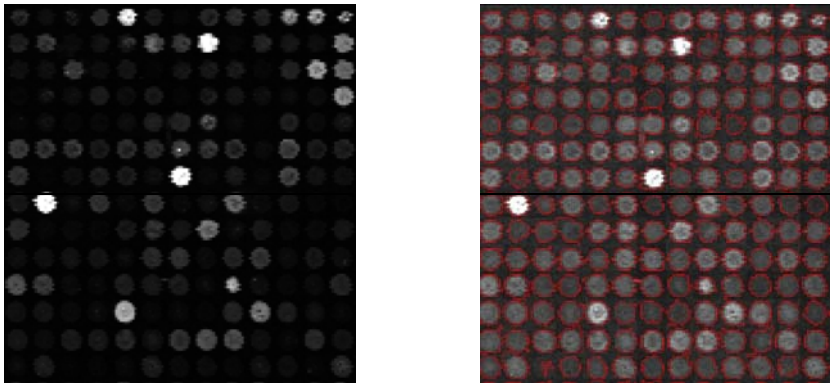

Fig. 1. Original and CLAHE-SRG segmented microarray spots.

Images	CLAHE-SRG	MAGIC TOOL
1302_OD370	8.24	2.71
1303_OD014	6.93	1.33
1309_OD690	8.20	3.76
1310_OD046	4.99	2.19
1311_OD080	2.45	1.99
1312_OD180	4.08	2.43
1313_OD370	3.03	2.44

Table 1 : Jeffrey's divergence metric values (in bits) between spot (signal) and background intensity values for the green (common reference sample) channel, for seven cDNA images.

Figure 2 depicts three randomly selected cell images, the middle row shows the result of the SRG algorithm according to the MAGIC TOOL software, and the bottom row presents the segmentation results based on the CLAHE-SRG segmentation.

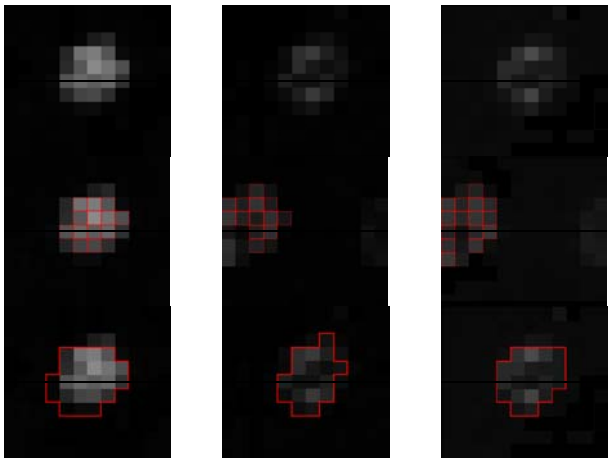

Fig. 2. a/Original cell-images, b/MAGIC TOOL segmentation using SRG, and c/CLAHE-SRG segmentation.

4 DISCUSSION

Microarray technologies have transformed the field of genomic research by allowing the simultaneous profiling of thousands of genes. The microarray process is based entirely on the extraction of quantitative information from images.

In the present study, an efficient spot-adaptive segmentation technique was developed by suitable combining in a cascade mode the benefits of image enhancement (Contrast Limited Adaptive Histogram Equalization technique (CLAHE)) and image segmentation (Seeded Region Growing technique (SRG)) in order to improve the accuracy of microarrays' spot segmentation and consequently genes' quantification. The proposed CLAHE-SRG segmentation comprised 1/a griding algorithm for locating individual cell-images, 2/an enhancement technique (CLAHE) for enhancing individual cell-images and, thus, for facilitating accurate cell-spot detection, and 3/ a segmentation algorithm (SRG) for outlining individual cell-spots.

By visual inspection of the original and the segmented images in Figure 1, it can be observed that the proposed technique improved the display of spots and emphasized the depiction of spot edges. The success of the proposed segmentation scheme is mostly due to its characteristic to perform individually on each spot and not on the whole image and accentuate spots' edges without the requirement for image uniformity, which is prerequisite for most of the common histogram equalization techniques.

The segmentation results of the proposed segmentation scheme were comparatively evaluated with the results obtained using commercial software (MAGIC TOOL) by employing the information theoretic metric of Jeffrey's divergence (Table 1). Results, according to the proposed scheme, were obtained by superimposing spot-outlines on the original cell-spots. In this way, higher divergence achieved by a particular method between the actual spots and surrounding background, would eventual lead to better spot-boundary detection result. Table 1 confirms that the proposed scheme performed as anticipated, by increasing the divergence (J) between signal and background intensity distributions, as compared to corresponding distributions obtained using the MAGIC TOOL software.

Regarding processing time, CLAHE-SRG took 211 seconds against MAGIC TOOL's 487 secs for the same 1024x1024, 16-bit cDNA image, containing 6400 microarray spots, and on the same computer.

5 ACKNOWLEDGMENT

This work was supported by a grant form the General Secretariat for Research and Technology, Ministry of Development of Greece (136/PENED03) to B.A.

REFERENCES

[1] Alizadeh, A., Eisen, M., Botstein, D., Brown, P.O., Staudt, L.M. (1998), "Probing lymphocyte biology by genomic-scale gene expression analysis", J Clin Immunol, Vol. 18, pp. 373-379.

[2] Aston, C., Jiang, L., Sokolov, B.P. (2005), "Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder", Mol Psychiatry, Vol. 10, pp. 309-322.

[3] Taniguchi, M., Miura, K., Iwao, H., Yamanaka, S. (2001), "Quantitative assessment of DNA microarrays--comparison with northern blot analyses", Genomics, Vol. 71, pp. 34-39.

[4] Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E., Davis, R.W. (1998), "Microarrays: Biotechnology's discovery platform for functional genomics", Trends Biotechnol, Vol. 16, pp. 301-306.
[5] Mark Schena. (2000), *Microarray biochip technology*. Eaton Publishing Company,
[6] Jain, K.K. (2004), "Current status of fluorescent in-situ hybridisation", Med Device Technol, Vol. 15, pp. 14-17.
[7] Chen, Y., Dougherty, E., Bittner, M. (1997), "Ratio-based decisions and the quantitative analysis of cdna microarray images", Journal of Biomedical Optics Vol. 2, pp. 364–374.
[8] Schena, M. (2002), *Microarray analysis*. New York.
[9] Angulo, J., Serra, J. (2003), "Automatic analysis of DNA microarray images using mathematical morphology", Bioinformatics, Vol. 19, pp. 553-562.
[10] Barra, V. (2006), "Robust segmentation and analysis of DNA microarray spots using an adaptative split and merge algorithm", Comput Methods Programs Biomed, Vol. 81, pp. 174-180.
[11] Jain, A.N., Tokuyasu, T.A., Snijders, A.M., Segraves, R., Albertson, D.G., Pinkel, D. (2002), "Fully automatic quantification of microarray image data", Genome Res, Vol. 12, pp. 325-332.
[12] Yang, Y.H., Buckley, M.J., Speed, T.P. (2001), "Analysis of cdna microarray images", Brief Bioinform, Vol. 2, pp. 341-349.
[13] Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H., Herzel, H. (2000), "Normalization strategies for cdna microarrays", Nucleic Acids Res, Vol. 28, pp. E47.
[14] Balagurunathan, Y., Wang, N., Dougherty, E.R., Nguyen, D., Chen, Y., Bittner, M.L., Trent, J., Carroll, R. (2004), "Noise factor analysis for cdna microarrays", J Biomed Opt, Vol. 9, pp. 663-678.
[15] Balagurunathan, Y., Dougherty, E.R., Chen, Y., Bittner, M.L., Trent, J.M. (2002), "Simulation of cdna microarrays via a parameterized random signal model", J Biomed Opt, Vol. 7, pp. 507-523.
[16] Gonzalez, R.C., Woods, R.E. (1992), *Digital image processing*
[17] Axon Instruments. (1999), "Genepix4000a user's guide", Vol. pp.
[18] Steinfath, M., Wruck, W., Seidel, H., Lehrach, H., Radelof, U., O'Brien, J. (2001), "Automated image analysis for array hybridization experiments", Bioinformatics, Vol. 17, pp. 634-641.
[19] White, A.M., Daly, D.S., Willse, A.R., Protic, M., Chandler, D.P. (2005), "Automated microarray image analysis toolbox for matlab", Bioinformatics, Vol. 21, pp. 3578-3579.
[20] Zapala, M.A., Lockhart, D.J., Pankratz, D.G., Garcia, A.J., Barlow, C., Lockhart, D.J. (2002), "Software and methods for oligonucleotide and cdna array data analysis", Genome Biol, Vol. 3, pp. SOFTWARE0001,0001-0001,0009.
[21] QuantArray. (1999) "Quantarray analysis software", http://www.bipl.ahc.umn.edu/quantarray.html, 06/12/2006.
[22] Eisen, M.B. (1999) "Scanalyze", http://rana.stanford.edu/software, 06/12/2006.
[23] Wang, X.H., Istepanian, R.S., Song, Y.H. (2003), "Microarray image enhancement by denoising using stationary wavelet transform", IEEE Trans Nanobioscience, Vol. 2, pp. 184-189.
[24] Lukac, R., Plataniotis, K.N., Smolka, B., Venetsanopoulos, A.N. (2005), "Cdna microarray image processing using fuzzy vector filtering framework", Journal of Fuzzy Sets and Systems: Special Issue on Fuzzy Sets and Systems in Bioinformatics, Vol. pp.
[25] Mastriani, M., Giraldez, A.E. (2006), "Microarrays denoising via smoothing of coefficients in wavelet domain", International Journal of Biomedical Sciences, Vol. 1, pp. 1306-1216.
[26] Lukac, R., Smolka, B. (2003), "Application of the adaptive center-weighted vector median framework for the enhancement of cdna microarray", Int. J. Appl. Math. Comput. Sci., , Vol. 13, pp. 369–383.
[27] Daskalakis, A., Cavouras, D., Bougioukos, P., Kostopoulos, S., Argyropoulos, C., Nikiforidis, G.C. (2006), "Improving microarray spots segmentation by k-means driven adaptive image restoration", *Proceedings of ITAB Ioannina, Greece*,
[28] Adams, R., Bischof, L. (1994), "Seeded region growing", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, pp. 641-647.
[29] http://www.bio.davidson.edu/projects/MAGIC/MAGIC.html.
[30] DeRisi, J.L., Iyer, V.R., Brown, P.O. (1997), "Exploring the metabolic and genetic control of gene expression on a genomic scale", Science, Vol. 278, pp. 680-686.
[31] Churchill, G.A. (2002), "Fundamentals of experimental design for cdna microarrays", Nat Genet, Vol. 32 Suppl, pp. 490-495.
[32] Sterrenburg, E., Turk, R., Boer, J.M., van Ommen, G.B., den Dunnen, J.T. (2002), "A common reference for cdna microarray hybridizations", Nucleic Acids Res, Vol. 30, pp. e116.
[33] Yang, Y.H., Speed, T. (2002), "Design issues for cdna microarray experiments", Nat Rev Genet, Vol. 3, pp. 579-588.
[34] Blekas, K., Galatsanos, N., Likas, A., Lagaris, I. (2005), "Mixture model analysis of DNA microarray images." IEEE Trans Med Imaging, Vol. 24, pp. 901-909.

[35] Stefanoyiannis, A.P., Costaridou, L., Skiadopoulos, S., Panayiotakis, G. (2003), "A digital equalisation technique improving visualisation of dense mammary gland and breast periphery in mammography", Eur J Radiol, Vol. 45, pp. 139-149.

[36] Pisano, E.D., Zong, S., Hemminger, B.M., DeLuca, M., Johnston, R.E., Muller, K., Braeuning, M.P., Pizer, S.M. (1998), "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms", J Digit Imaging, Vol. 11, pp. 193-200.

[37] Verellen, D., De Neve, W., Van den Heuvel, F., Coghe, M., Louis, O., Storme, G. (1993), "On-line portal imaging: Image quality defining parameters for pelvic fields--a clinical evaluation", Int J Radiat Oncol Biol Phys, Vol. 27, pp. 945-952.

[38] Kullback, S. (1968), *Information theory and statistics*. Dover Publications,

[39] Heyer, L.J., Moskowitz, D.Z., Abele, J.A., Karnik, P., Choi, D., Campbell, A.M., Oldham, E.E., Akin, B.K. (2005), "Magic tool: Integrated microarray data analysis", Bioinformatics, Vol. 21, pp. 2114-2115.