Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2021) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Edge-based Active Contours for Microarray Spot Segmentation

Carmen Costea[1], Bogdan Gavrea[1], Mihaela Streza[2], Bogdan Belean[2]*

[1]*Faculty of Automation and Computer Science, Department of Mathematics, Technical University of Cluj-Napoca, Romania*
[2]*National Institute for Research and Development of Isotopic and Molecular Technologies, Center for Research and Advanced Technologies for Alternative Energies, Cluj-Napoca, Romania*

Abstract The paper proposes a novel approach for image segmentation in case of bio-medical applications. The medical images considered for our analysis are drawn from microarray image databases and they are generally used for the estimation of gene expression levels, which correspond to the average intensities of the microarray spots represented as circular shapes within the image under analysis. The novelty of the proposed image analysis approach consists of a segmentation procedure which uses an active contour model (ACM) based on edge information. Basically, a predefined curve is evolved towards the object edges (i.e. microarray spots). For a precise determination of image object, a cellular neural network approach (CNN) is employed for edge detection of each microarray spot. This is used further on in the curve evolution process for a more accurate segmentation of microarray spots. Specific measures for the characterization of microarray spots features are used to compare the obtained results with the ones delivered by GenePix microarray software platform.

Click here and insert your abstract text.

Keywords: image segmentation, edge detection, active contours, celular neural networks, microarray, microfluidics devices, gene expression, cell clusters

1. Introduction

In recent years, image segmentation has gained a lot of research attention as an important area of computer vision. The aim of image segmentation is to divide an image into multiple parts with similar features such as texture, strength, or color. The resulted image parts are described as meaningful related to the image structure and elements and make it easier to examine objects within the image. Image segmentation has numerous applications in industry and research, including image or object recognition, image processing and comprehension, medical diagnosis of diseases, computer-

* Corresponding author. Tel.: 40 264 58 40 37; Fax: +40 264 42 00 42
E-mail address: bogdan.belean@itim-cj.ro

guided surgery, treatment planning and anatomical structure. Many conventional image segmentation techniques have been suggested, including threshold-based segmentation [1], edge-based segmentation [2], region-based segmentation [3], hybrid segmentation based on region and edge information [4,5], clustering-based segmentation [6], segmentation based on Cellular Neural Networks (CNNs) and active contour models – ACM [7, 8]. Considering the last category, they are defined as an evolving contour towards the object edges. Edge-based models use the image gradient information to define object boundaries and stop the contour from evolving. Additionally, atlas-based segmentations are also available, whereas the use of a probabilistic atlas together level-set - based segmentation methods leads to increased segmentation accuracy [9].

Further on, we proposed an edge based active contour, where edge features determine by the CNN are embedded within the curve evolution process. Thus, the reminder of the document is structured as follows: the basic principle of edge- based ACM together with the CNN approach for edge feature computation are presented within the introduction; in the second section, the proposed CNN-driven active contour approach is employed in case of microarray images; the third section presents the conclusions.

1.1. Active contours for image segmentation

The segmentation of microarray spots by mean of active contours starts with the selection of the rectangular area for each spot using the method proposed in [10]. Consequently, we obtained one rectangular image I_s on the Ω domain, confining each microarray spot. The rectangular area includes both the background and the foreground pixels associated to the microarray spot. The foreground pixels correspond to pixels that are part of the microarray spot, whereas the background pixels to the local spot background. Consequently, the second step, which is inspired by Agilent's "cookie cutter" approach, is to divide pixels within each rectangular area into foreground and background. An edge-based level set approach is used to represent the contour of the microarray spot C_S for each rectangular area determined by grid alignment, as the zero-level set of the level set function (LSF) denoted by $\varphi(x, y, t)$, as expressed by equation (1).

$$C_S(t) = \{(x, y)/\varphi(x, y, t) = 0\} \quad (1)$$

The determination of the contour C_S is converted into finding solution of the partial differential equation (PDE) from eq. (2), which is referred to as the level set evolution equation [11], where F is the speed function that controls the motion of the curve on its normal direction.

$$\frac{\partial \varphi}{\partial t} = F\|\nabla\varphi\| \quad (2)$$

The LSF must be smooth and accurate in image segmentation applications, particularly near its zero-level set, where it describes the contour of the object to be determined. In order to apply the level set approach for object boundaries determination, we consider the motion of the curve by mean curvature expressed by the following equation (3), where $F = div\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right)$ is the curvature of the level-set curve [1].

$$\frac{\partial \varphi}{\partial t} = div\left(\frac{|\nabla\varphi|-1}{|\nabla\varphi|}|\nabla\varphi|\right) - \frac{\partial e_{ext}}{\partial \varphi} \quad (3)$$

The e_{ext} is chosen to describe the edge information. According to [11], the classic approach makes use of the following edge indicator functions, $e_{ext} = \lambda L(\varphi) + \alpha A(\varphi)$, with $L(\varphi) = \int_\Omega g\delta(\varphi)\,|\nabla\varphi|dx$ and $A(\varphi) = \int_\Omega gH(-\varphi)\,dx, \lambda > 0$, $\alpha \in R$ and

$$g = \frac{1}{1+|\nabla G_\sigma * I_S|^2}. \quad (4)$$

It is to be mentioned that g represents the edge indicator function for a given microarray spot image Is define on the Ω domain, with G_σ a Gaussian kernel with standard deviation σ. The results of such segmentation using edge-based active contours in case of some circular shapes representing microarray spots are given in figure 1. It can be seen that, in spite of using a 3 by 3 size window with reduced standard deviation σ =1.1 for the gaussian kernel, the obtained edges are missing some foreground pixels in case of the 2nd and 4th images from Fig.1, whereas background is included in the foreground areas in case of the first spot. Further on, a CNN based approach is used to determine

the edge information function. The aforementioned shortcomings demanded a curve evolution process which does not depend only on the image gradient, but also on the intensity distribution pattern within the object to be detected. Complex mathematical apparatus is available for modeling fluid dynamics in porous media [12]. The evolution of an initial curve towards image objects which would account for intensity inhomogeneities within objects is envisaged. Thus, shortcomings similar to the ones underlined in Fig. spots 2 and 4, can be eliminated by the curve evolution process which is meant to include at the iterations end bright pixels as the one excluded in fig 1 from the segmented object within the red curve.

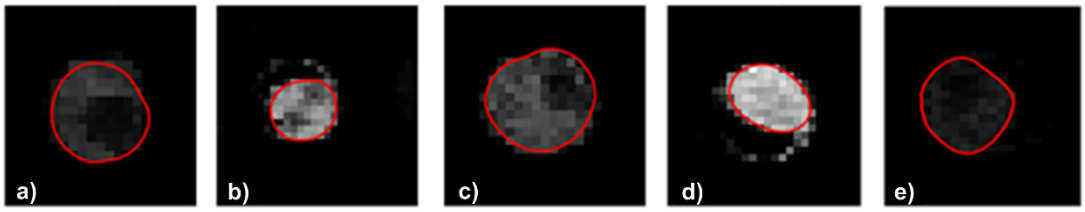

Fig. 1. Results of the classic edge based active contour procedure model for a set of 5 microarray spots corresponding to the following categories, donut shape spot (a,c), spots with outliers (b, d), weakly expressed spot (e)..

1.2. Cellular Neural Networks - CNN

Artificial and natural vision systems differ in that the latter is characterized by continuous time and signal values, while the former is characterized by discrete time and signal values. The CNN universal machine (CNN–UM) and cellular neural networks (CNN) were introduced in 1988 and 1992 [13, 14]. Such networks are described as arrays of identical dynamical structures, known as cells, that are linked locally [14]. A CNN's basic unit, the cell, is a one-dimensional dynamic structure connected only to its neighbors, i.e. adjacent cells interact directly with each other Fig. 2. Because of the propagation effects of the network's dynamics, cells in near proximity have an indirect impact. x_{ij} stands for the cell in place *(i,j)* of a two-dimensional *M x N* sequence, and N^r_{ij} stands for its neighbors (1).

$$N^r{}_{ij} = \{x_{kl} | \, max\{|k-i|, |l-j|\} \le r, 1 \le k \le M, 1 \le l \le N\} \qquad (5)$$

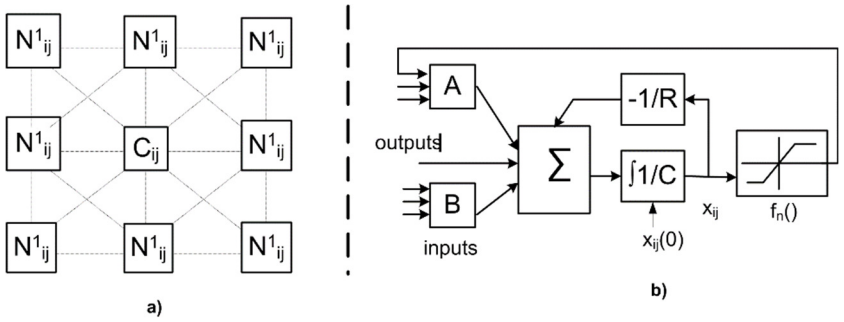

Fig. 2. (a) Cells interconnection within the CNN approach; (b) recurrent computation of the resulted matrix , based on the inputs B and the CNN formalism

A cell is the basic circuit unit of a CNN. It has linear and nonlinear elements that function as linear capacitors, linear resistors, linear and nonlinear controlled sources, and independent sources, among other things. Each neural network cell has two key characteristics: a constant external input u and an output y. Each neural network cell's corresponding block diagram is shown in Fig. 2.a. The dynamics of a CNN are defined by the first-order non-linear differential equation:

$$C\frac{\partial x_{ij}(t)}{\partial t} = -\frac{1}{R}x_{ij}(t) + \sum_{C_{kl}\in N^r{}_{ij}} A(i,j;k,l)y_{kl}(t) + \sum_{C_{kl}\in N^r{}_{ij}} B(i,j;k,l)u_{kl}(t) + I \qquad (6)$$

where

- x_{ij} is the state of cell C_{ij};
- C and R represent system integration time constants;
- I represents an independent bias constant;
- $y_{ij}(t) = f(x_{ij}(t)),\ f(x) = \frac{1}{2}(|x+1| - |x-1|)$

The cloning templates are the matrices A(.) and B(.), where A(.) affects the output control of neighbouring cells and is known as the feedback operator whereas B(.) affects the input control and is known as the control operator. A(.) and B(.) are selected based on the application the CNN are used for. When it comes to image processing, an image is a rectangular array of N and M rows and columns, respectively. Each element of the array corresponds to a cell in a CNN. A two-dimensional CNN can be thought of as a parallel non-linear two-dimensional filter that can be used for noise removal [15], shape extraction, edge detection, and inpainting [16] by calibrating the A, B, and I parameters.

Thus, if we consider the time-dependent equation 3, we get the following discrete model for the CNN, which is similar to the continuous one and can be used in image processing.

$$x_{ij}[n] = \sum_{C_{kl} \in N^r{}_{ij}} A(i,j;k,l) y_{kl}[n-1] + \sum_{C_{kl} \in N^r{}_{ij}} B(i,j;k,l) u_{kl}[n-1] + I \quad (7)$$

with

$$y_{ij}[n] = (1/2)(|x_{ij}[n] + 1| - |x_{ij}[n] - 1|) \quad (8)$$

Edge features detection

Edge detection is a common task in image processing generally used to extract object borders within the image under analysis. The previous task is achieved using gradient information. For edge features detection in image processing, state-of-the-art research proposes several hypotheses, including cellular neural networks, genetic algorithms, and wavelet transforms [17, 18]. The pixel intensity values of bio-medical images significantly vary in the areas corresponding to image object edges. Moreover, in case of images obtain by means of light microscopy, cases of inhomogeneity correction and noisy edges are present. These phenomena lead to unexpected results when we process noisy images by classical edge detection operators, such as Roberts, Sobel, Prewitt. Because of inhomogeneous foreground and background and noisy edges, it is unrealistic to find a uniform threshold suitable for every image object. Thus, the proposed CNN based edge feature detection is sensitive to all edges. It actually delivers a mapping of all pixel intensities according to the magnitude and sign and direction of the image gradient. According to the concept of the CNN, the image evolves in time and converges to an image with visible edges, described by the differential equations, eq. (3). (4). A block diagram is shown in the next figure 3, for a more detailed look at how CNNs function. The CNN approach can be seen as an iterative spatial convolution process on bi-dimensional matrices corresponding to image pixel intensity values. In [18] X-ray images are processed for edge detection. Nevertheless, the proposed approach is not limited to bi-dimensional filtering. It also can be employed in processing uni-dimensional image profiles for the determination of the magnitude and location of profile peaks. As a result, an increased accuracy for the determination of structural ordering parameters is obtained [19, 20].

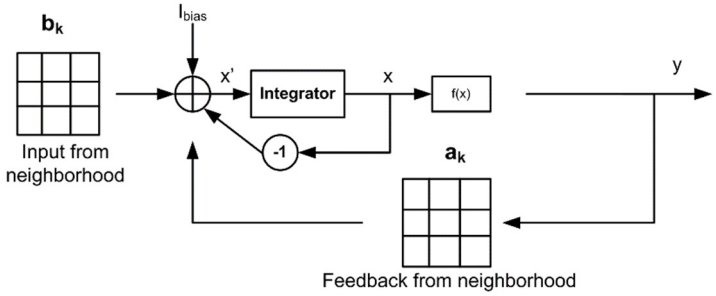

Fig. 3. (a) CNN computation of the edge feature matrix a_k.

Up to this point, the determination of the edge information function g_{CNN} is described. Thus, after an iterative convolutional process applied to the x_{ij} inputs corresponding to the initial image, the resulted image becomes $g_{CNN}(ij)$, and replaces the g function from equation (4).

2. Experimental results in bio-medical imaging

2.1. Microarray Data-sets

Nowadays, the interrogation of genomic functionality relies on microarray technology to assess the gene expression levels simultaneously for all cellular transcripts (mRNAs) in a single experiment. By measuring the mRNA levels for the whole genome, the microarray experiments are capable to study functionality, pathological phenotype, and response of cells to a pharmaceutical treatment [21]. The workflow of a microarray experiment includes, besides the procedure of measurement, a step of extensive data analysis. Standardized protocols and design methods exist for measurements but the processing of the extensive number of non-homogeneous data is often still a challenge. In microarray experiments, RNA extracted from biological sample is synthesized to microarray targets. The targets are either single-stranded DNAs or RNAs labeled with fluorescent markers. One or two labels (e.g. the dyes Cy3, and Cy5) can be utilized in the same hybridization measurement. The microarray targets hybridize on a microarray slide with spots of sub-sequences (probes) of the genes of the whole genome. A laser scanning with appropriate wavelengths produces a TIFF image I for each fluorescent label (a lower index may denote the dye, e.g., ICy3 denotes a microarray image recorded of Cy3 dye). The microarray images thus obtained represents a collection of microarray spots, each spot corresponding to a specific gene. The expression levels of genes are calculated based on intensities of the fluorescent light [22]. The calculation of expression levels is a crucial step to extract valuable information from a microarray measurement and involves three major tasks: (1) a grid alignment (addressing), to determine the spatial coordinates of each spot; (2) a segmentation to classify pixels either as foreground, representing the DNA spots, or as background; (3) an extraction of intensity of each spot and its individual background. Results of the image analysis are the layout of the spot array, the spot sizes and shapes, the spot intensities (i.e., gene expression levels), and the background intensities values.

The data set consists of 4 pair of images corresponding to the microarray samples having the following ID GSM333336, GSM333353, GSM333337 and GSM333341 (from Gene Expression Omnibus). The 4 pairs represent biological replicates of the same experiment. Each pair consists of two images corresponding to the I_{cy3} and I_{cy5} fluorescent dyes. Each microarray image has the size of 4000x1944 pixels and contains 32 spot groups with 380 spots per group. Visual results are presented in case of our proposed active contour-based segmentation using CNN edge information in case of the series of microarray spots from Fig. 4.

2.2. Microarray Image Processing using CNN and Active Contour Models

For the entire population of spots included in the selected microarray image samples, we use our proposed approach to extract spot features. In terms of reproducibility and reliability, the obtained results are compared to those provided by GenePix Software.

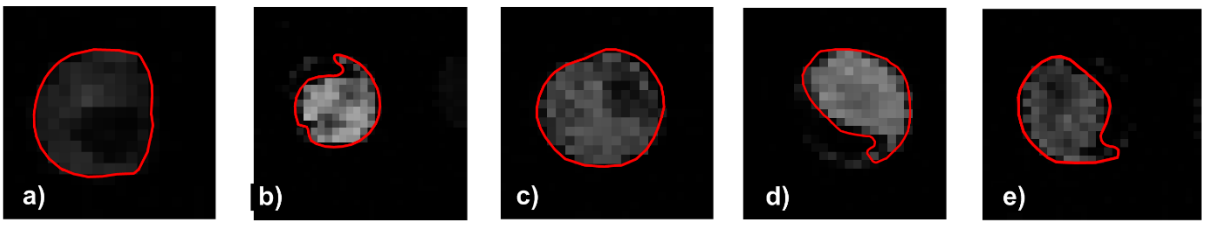

Fig. 4. Resulted edges obtained using the CNN based active contour model for a set of 5 microarray spots corresponding to the following categories, donut shape spot (a,c), spots with outliers (b, d), weakly expressed spot (e).

The reproducibility and biological significance of the proposed image processing techniques are assessed. The coefficient of variation (CV), which express the variation of spot intensities are computed to quantify the reproducibility of the proposed segmentation techniques.

The CV is given by:

$$CV_{spot} = \frac{\sigma}{\mu} \tag{9}$$

where σ and μ represent the standard deviation of spot intensity with subtracted background and the mean spot intensity, respectively.

The proposed method's efficiency increases as the CV values decrease. Both our proposed segmentation method and the GenePix approach had CV values computed. As shown in the next table, the following parameter values were obtained for the E1 experiment: In terms of CV coefficient, the level-set segmentation protocol produces comparable results to the GenePix method. In the case of spot pixel intensity values, the CV represents a standardized metric of dispersion that is independent of the unit in which the measurement was taken. A small CV shows that the pixel intensity values for a given microarray spot slightly differ. As a consequence, CV table shows a more accurate spot description delivered by our proposed approach.

Experiment ID (Image channel)	***Mean CV our results***	***Mean CV (GenePix)***
GSM333336 (I_{Cy3})	0. 592	0.663
GSM333336 (I_{Cy5})	0.498	0.534
GSM333353 (I_{Cy3})	0.452	0.635
GSM333353 (I_{Cy5})	0.712	0.706
GSM333337 (I_{Cy3})	0.597	0.684
GSM333337 (I_{Cy5})	0.549	0.766
GSM333341 (I_{Cy3})	0.486	0.552
GSM333341 (I_{Cy5})	0.658	0.795

Fig. 5. Table with the coefficient of variation obtained using our segmentation approach compared to the ones resulted using GenePix software,

3. Conclusions

The paper proposes a novel segmentation approach for bio-medical images which makes use of active contour models (ACM) and cellular neural networks (CNN). The novelty of the proposed method consists of an edge features matrix computed based on the CNN approach, which is embedded in the curved evolution process, acting as an edge enhancement technique. Results of the proposed method applied on a set of microarray images are delivered. The CV coefficients were used to underline the benefits of our proposed approach in terms of a more accurate microarray spot segmentation.

Acknowledgements

This work was supported by a grant of Romanian Ministry of Education and Research, CNCS – UEFISCDI, project number PN-III-P4-ID-PCE-2020-0368, within PNCDI III.

References

[1] L.A. Vese, C.L. Guyader (2015) "Variational Methods in Image Processing", Chapman & Hall/CRC *Mathematical and Computational Imaging Sciences Series*, Taylor & Francis

[2] P.S. Jadhav (2015) "A review on: "image segmentation based on level set method", *J. Eng.*, (**3**): 168-172.

[3] K. Suresh, P. Srinivasa Rao (2018) "Various image segmentation algorithms: A survey", *Proceedings of the 2nd International Conference on Smart Computing and Informatics*, pp. 233-239.

[4] Behiye Kaya, Evgin Goceri, Aline Becker et al. (2017) "Automated fluorescent miscroscopic image analysis of PTBP1 expression in glioma", *Plos One*, **12 (3)**: e0170991 .

[5] Evgin Goceri, Mehmet Zubeyir, Oguz Dicle (2015) "A comparative performance evaluation of various approaches for liver segmentation from SPIR images", *Turk J Elec Eng & Comp Sci*, (**23**): 741 – 768.

[6] Goceri E. (2016) "Automatic labeling of portal and hepatic veins from MR images prior to liver transplantation", *International Journal of Computer Assisted Radiology and Surgery,* (**11**) : 2153–2161.

[7] J. Ma, X. Yang (2019) "Automatic dental root CBCT image segmentation based on CNN and level set method", *In: Proceedings of Medical Imaging,* SPIE, San Diego, CA, United states, 10949.

[8] Le Zou, Liang-Tu Song, Thomas Weise, Xiao-Feng Wang, Qian-Jing Huang, Rui Deng, Zhi-Ze Wu (2020) "A survey on regional level set image segmentation models based on the energy functional similarity measure", *Neurocomputing*, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2020.07.141.

[9] Dura, E., Domingo, J., Göçeri, E. et al. (2018) "A method for liver segmentation in perfusion MR images using probabilistic atlases and viscous reconstruction", *Pattern Anal Applic,* (**21**) : 1083–1095.

[10] Bogdan Belean, Robert Gutt, Carmen Costea, and Ovidiu Balacescu (2020) "Microarray image analysis: from image processing methods to gene expression levels estimation", *IEEE Access*, pp. 159196 – 159205, Electronic ISSN: 2169-3536.

[11] C. Li, C. Xu, C. Gui, M.D. Fox (2010) "Distance Regularized Level Set Evolution and Its Application to Image Segmentation", *IEEE Trans. on Image, Processing*, **19** (**12**): 3243 - 3254.

[12] Gutt, Robert (2020) "Boundary Element Method For The Lid-Driven Problem Related To The Darcy-Forchheimer-Brinkman System", *Journal Of Porous Media*, **23** (**8**): 821-835.

[13] Chua, L.O. and Yang, L. (1988) "Cellular Neural Networks: Theory and Applications", *IEEE Trans. on Circuits and Systems*, Vol.**35** : 1257-1290.

[14] Roska, T. and Vandewalle, J. (1993) "Cellular Neural Networks". (John Wiley&Sons),

[15] Te-Jen Su, Cian-Pin Wei, Shih-Chun Huang, Chia-Ling Hou (2008) "Image noise cancellation using linear matrix inequality and cellular neural network", *Optics Communications*, **281**(**23**): 5706-5712.

[16] P.Elango a and K.Murugesan (2009) "Digital Image Inpainting Using Cellular Neural Network," *Int. J. Open Problems Compt. Math.*, **2**(**3):** 440-450.

[17] Gema Bello-Orgaz, Sancho Salcedo-Sanz, David Camacho (2018) "A Multi-Objective Genetic Algorithm for overlapping community detection based on edge encoding, Information Sciences", **462**: 290-314.

[18] P.M.K. Prasad, D.Y.V. Prasad, G. Sasibhushana Rao (2016) "Performance Analysis of Orthogonal and Biorthogonal Wavelets for Edge Detection of X-ray Images", *Procedia Computer Science*, **87**: 116-121.

[19] D. Benea, R. Gavrea, M. Coldea, O. Isnard, V. Pop (2019) "Half-metallic compensated ferrimagnetism in the Mn-Co-V-Al Heusler alloys", *Journal of Magnetism and Magnetic Materials*, **475** : 229-233.

[20] Radu Gavrea, Alfred Bolinger, Viorel Pop ert al. (2017) "Influence of Cu Doping on the Electronic Structure and Magnetic Properties of the Mn2VAl Heusler Compound", *Physica Status Solidi B-Basic Solid State Physics*, **254**(**11**): 1700160.

[21] Schena M (2003) Microarray Analysis. New York, John Wiley & Sons

[22] Campbell AM, Hatfield WT, Heyer LJ (2007) Make microarray data with known ratios. *CBE, Life Sciences Education*, **6**: 196-197