33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Automatic segmentation of time-lapse microscopy images depicting a live Dharma embryo

Eleni Zacharia, Maria Bondesson, Anne Riu, Nicole A. Ducharme, Jan-Åke Gustafsson and Ioannis A. Kakadiaris, *Senior Member, IEEE*

***Abstract*—Biological inferences about the toxicity of chemicals reached during experiments on the zebrafish Dharma embryo can be greatly affected by the analysis of the time-lapse microscopy images depicting the embryo. Among the stages of image analysis, automatic and accurate segmentation of the Dharma embryo is the most crucial and challenging. In this paper, an accurate and automatic segmentation approach for the segmentation of the Dharma embryo data obtained by fluorescent time-lapse microscopy is proposed. Experiments performed in four stacks of 3D images over time have shown promising results.**

I. Introduction

Chemicals, which can be defined as artificially created forms of matter that have constant chemical compositions and characteristic properties, are found practically everywhere in our environment and have a profound impact on our contemporary lifestyles. According to the latest statistical studies, the typical consumer is exposed to thousands of different chemicals every day, while 30,000 industrial chemicals at quantities of over 4,000 million tons are produced per year on average [1]. This ground truth could not have passed unnoticed and has effectively forced a number of international environmental organizations - most notably the WWF [2], EPA [3], and OECD [4] – to take action and invest considerable sums on determining the chemicals' toxicity and in particular on understanding the relationship between the overwhelming and pervasive exposure to chemicals and its possible environmental and health implications.

Under certain conditions, the use of zebrafish embryos as a model system can serve as an excellent way forward in determining the toxicity of a given chemical (Fig. 1). This fact is attributed to the fish's genetic features that additionally make it cost-effective [5]: Zebrafish are small in size and can produce hundreds of embryos per pair on a weekly basis. Furthermore, its embryo is completely transparent, an indispensable characteristic that allows scientists to inspect *in vivo* the development of every individual cell - even those that are located deep in tissues - using confocal time-lapse microscopy. Last but not least, the developmental expression of specific genes can be followed by green fluorescent protein (GFP) markers under the promoter control of the gene in question in transgenic zebrafish. One such transgenic fish is the Dharma fish, in which GFP is expressed after the *dharma* promoter. The developmental expression of *dharma* can be visually recorded by GFP signaling during the first 24 hours of development in live Dharma embryos (Fig. 1). This attribute is highly significant because the time of GFP appearance/disappearance during development as well as the spatial distribution and intensity of the GFP expression can vary according to the presence of normal developmental conditions or to the lack thereof after exposure to chemicals. Crucially, these variations could reveal the degree of toxicity of the chemicals.

Thus, there is a critical need to carefully and correctly analyze the 3D stack of images depicting Dharma embryos which are acquired during a laboratory-controlled experiment from the middle to the end of the blastula phase. The first stage of the analysis is the segmentation of embryo. This is in fact a rather challenging task because of the nature of the 3D stack of images. The quality of these images is often degraded due to the existence of noise and/or artifacts as well as due to inhomogeneous background. Therefore, the intensity values fluctuate noticeably even between two consecutive pixels.

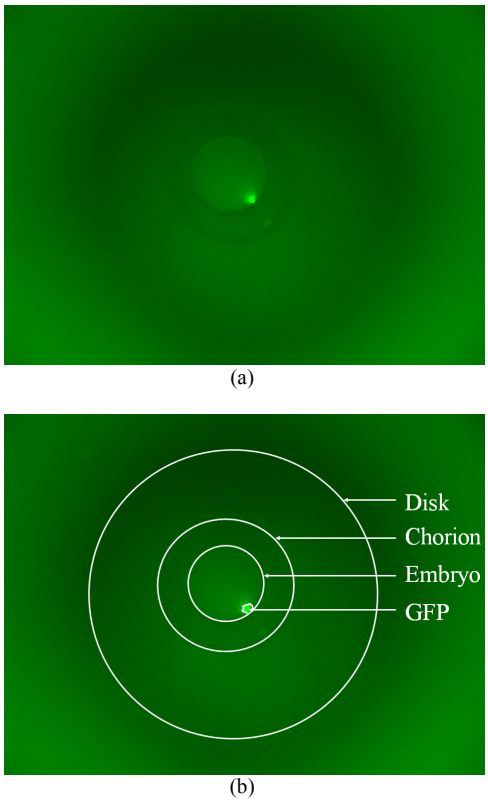
To the best of our knowledge, there is no previous work on the segmentation of time-lapse microscopy images depicting Dharma embryos from the middle to the end of the blastula phase. Furthermore, due to the aforementioned attributes and the additional absence of any distinctly visible edges, the exact region of the embryo cannot be easily and definitively determined when using fluorescence microscopy, at this stage of embryo development. Therefore, image segmentation techniques based either on thresholding or on determining the regional edges of images alone are not sufficient to segment the embryo.

In this paper, an automatic approach to the segmentation of time-lapse microscopy images depicting a live Dharma embryo is presented. The proposed approach relies on a genetic algorithm, which determines very effectively the optimal circle inside which the Dharma embryo is located.

Manuscript received April 15, 2011. This work was supported in part by a training fellowship from the Keck Center Computational Cancer Biology Training Program of the Gulf Coast Consortia (CPRIT Grant No. RP101489).

E. Zacharia and I.A. Kakadiaris are with the Computational Biomedicine Lab, Department of Computer Science, University of Houston, Houston TX 77204-3010, USA. (eezacharia@gmail.com, IKakadia@central.uh.edu).

M. Bondesson, Anne Riu, Nicole A. Ducharme and Jan-Åke Gustafsson are with the Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5056, USA. (ariu@uh.edu, nducharme-ross@uh.edu, mbondessonbolin@uh.edu, jgustafsson@uh.edu).



(a)

(b)

Fig. 1. (a) Image obtained by fluorescent time-lapse microscopy and depicting a Dharma embryo and (b) annotation of the different regions in the image (Color image).

The proposed method works in the presence of inhomogeneous background, noise and artifacts, as the chosen genetic algorithm is not based on the edges of the image.

The remainder of this paper is structured as follows: the proposed segmentation approach is described in Section II, the results of the proposed approach are presented in Section III. Finally, our conclusions are presented in Section IV.

II. Proposed Approach

The segmentation of time-lapse microscopy images depicting the Dharma embryo is equivalent to the determination of a circle, which encloses the embryo. Therefore, it can be viewed as an optimization problem, which is tackled by using the proposed genetic algorithm that determines the exact values of the variables of the optimal circle surrounding the embryo.

A. Chromosome

Let chromosome m represent, on a two-dimensional Cartesian system, a circle $C(x,y)$ which can be described algebraically by the following equation:

$$(x - x_o)^2 + (y - y_o)^2 = r^2 \quad (1)$$

where (x_o,y_o) are the coordinates (x,y) of the circle's center and r is its radius.

The chromosome m encodes the values of the three parameters of the circle, that is the x_o, y_o, and r.

B. Fitness Function

Each chromosome m in every population is evaluated using a fitness function, $F(m)$, which assigns to the chromosome a degree of how appropriate a solution to the optimization problem is. The higher the value of the fitness function, the more appropriate the chromosome becomes.

Regarding the specific optimization problem, the genetic algorithm should progressively assign – from (a) to (d) – a higher fitness value to the chromosomes representing the red circles in Fig. 2. Therefore, there are two primary objectives when evaluating the chromosome, which are: (i) to maximize the total intensity value inside the circle against the total intensity outside the circle, and (ii) to maximize the total intensity value of the pixels which are located inside the circle and near its circumference against the total intensity value of the pixels which are located outside the circle and near its circumference.

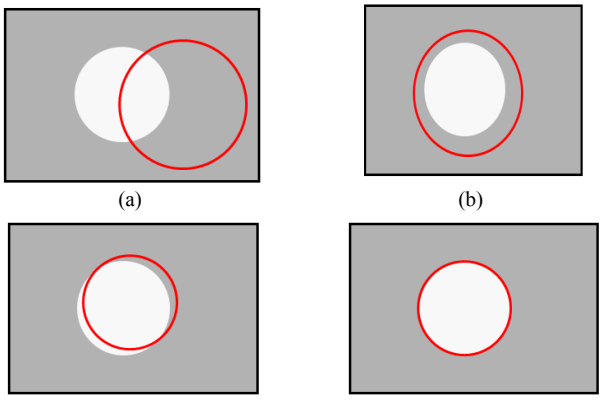
Subsequently, each chromosome m – representing the circle C_m – defines four regions in the image: (i) C_m^I containing the pixels of the image located inside the circle C_m (Fig. 3a), (ii) C_m^O containing the pixels of the image located outside the circle C_m and whose distance from C_m center is less than $\sqrt{2}r_m$, where r_m is the radius of the circle C_m (Fig. 3b), (iii) R_m^I containing the pixels of the image in a ring inside the circle C_m (Fig. 3c), and (iv) R_m^O containing the pixels of the image in a ring outside the circle C_m (Fig. 3d).

The fitness function is defined as:

$$F(m) = \begin{cases} P_1(m) & , \text{if } P_1(m) < T_p \\ P_1(m) + P_2(m) + P_3(m), & \text{otherwise} \end{cases} \quad (2)$$

where

$$P_1(m) = \frac{\#\{(x,y) \mid (x,y) \in C_m^I \wedge I(x,y) > \bar{\mathrm{I}}_{C_m^O}\}}{\#\{(x,y) \mid (x,y) \in C_m^I\}}, \quad (3)$$



(a) (b) (c) (d)

Fig. 2. An outline of an image containing Dharma embryo (white area). The red circles depict three possible circles encoded by three chromosomes.

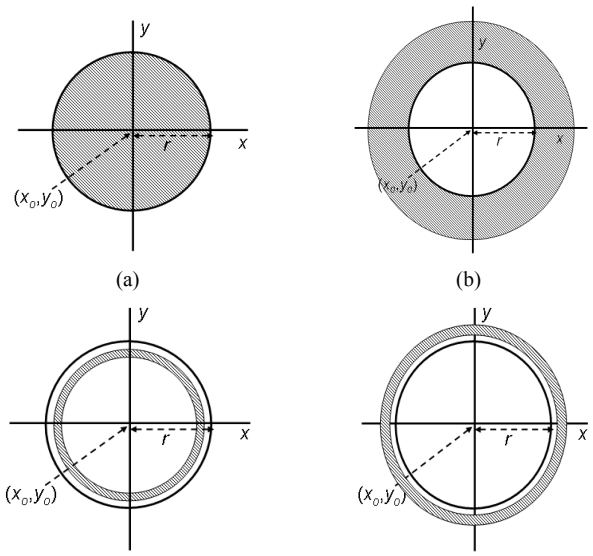


Fig. 3. The four regions defined by each chromosome m. (a) C_m^I, (b) C_m^O, (c) R_m^I, and (d) R_m^O.

$$P_2(m) = \frac{\#\{(x,y) \mid (x,y) \in R_m^I \wedge I(x,y) > \bar{\mathrm{I}}_{R_m^O}\}}{\#\{(x,y) \mid (x,y) \in R_m^I\}}, \quad (4)$$

and

$$P_3(m) = \frac{\#\{(x,y) \mid (x,y) \in R_m^O \wedge I(x,y) \leq \bar{\mathrm{I}}_{R_m^I}\}}{\#\{(x,y) \mid (x,y) \in R_m^O\}}. \quad (5)$$

The symbol # denotes the number of elements of the set that is defined by the brackets {}. The intensity value of the pixel with coordinates (x,y) is denoted as $I(x,y)$. The mean intensities values of the C_m^O, R_m^O, and R_m^I regions are denoted as $\bar{\mathrm{I}}_{C_m^O}$, $\bar{\mathrm{I}}_{R_m^O}$ and $\bar{\mathrm{I}}_{R_m^I}$ respectively. The minimum acceptable percentage $P_1(m)$ for a circle to be acceptable as the solution is denoted as T_P.

The fitness function of a chromosome m equals to $P_1(m)$ or to $P_1(m)+P_2(m)+P_3(m)$ according to the percentage of $P_1(m)$. If the percentage $P_1(m)$ is less than a threshold T_P, it indicates that the circle represented by the chromosome m is ill-defined (i.e. it is located mostly in the background area instead of the area of the embryo (Figs. 2a,b)). On the other hand, if the percentage $P_1(m)$ of a chromosome m is higher than the threshold T_P, it implies that the chromosome is well defined (i.e. it is located mostly in the area of the embryo (Fig. 2c,d)). Using the fitness function $F(m)$, the genetic algorithm can assign to the chromosome m of the first case a lower fitness value than to the one of the second case.

It is worth pointing out that due to the inhomogeneous background of the image as well as due to the presence of noise, it is possible for some regions located inside the embryo to have smaller intensity values than regions located outside the embryo. To overcome this problem, the Eqs. (3-5) are computed not based on the total area of the circle or ring but on sectors of them (Fig. 4). Specifically, each quadrant q of each circle or ring is split into four sectors $S_{q,k}$ of equal size. In other words, the circle is divided in sixteen equal sectors. q denotes the corresponding quadrant each sector belongs to and k the specific placement of a sector within the quadrant (q, k=1,...,4). Each of the Eqs. (3-5) is applied in turn to a pair of sectors that are positioned anti-diametrically (Fig. 4, gray sectors):

$$(S_{q,k}, S_{q+2,k}) \quad (6)$$

This process is repeated for a total of eight times so as to include all eight pairs of anti-diametrically positioned sectors. Therefore, each of the Eqs. (3-5) yields a set of eight results. The average value of each one of these three sets of eight results is then consider as the final result of each equation.

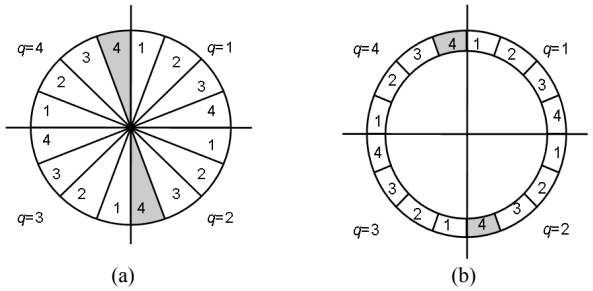


Fig. 4. Illustration of the sectors into which each quadrant q of each circle or ring is split.

C. Evolutionary circle and termination criterion

The genetic algorithm searches for the optimal circle as follows: First, it creates an initial population (P_1) of chromosomes. Subsequently, the chromosomes constituting the P_1 are evaluated using the fitness function. Then, the genetic algorithm makes the population P_1 evolve into a new population P_2 using progressively: (i) the elitist reproduction, (ii) the combination of the BLX-a crossover as well as the Dynamic Heuristic one [6] and (iii) the Wavelet mutation [7]. New populations are produced until the genetic algorithm is executed up to a maximum number of populations G_F for which the best fitness value remains unchanged.

III. Results

Several experiments were performed to evaluate the accuracy of the proposed approach. The dataset of images used for validation of our segmentation method were obtained by fluorescent time-lapse microscopy depicting a Dharma embryo from the middle to the end of the blastula phase. The images were acquired in parallel to the focal plane xy, at layers regularly spaced at z depth. The size of each image is 1040 x 1392 pixels with a resolution of 1.61 $\mu m/pixel$. The distance between the layers is 40 $\mu m/pixel$. Repeating this procedure over time, four 3D stacks of 19 images were acquired, each containing an untreated Dharma embryo from the middle to the end of the blastula phase.

TABLE I
PERFORMANCE OF THE PROPOSED SEGMENTATION METHOD

	Accuracy	Precision	True Positive rate	True Negative rate
Average	0.93	0.98	0.94	0.05
Standard deviation	0.04	0.01	0.04	0.04

During the experiments the T_P of 0.9 was adopted as the most suitable in order to distinguish an acceptable solution against a non-acceptable solution. Furthermore, the G_F was set equal to 100.

The accuracy of the proposed method was analyzed by means of statistical analysis. The statistical analysis was based on the ground truth of expert empirical observation. In other words, the resulting embryonic depictions of the proposed method were compared to the depictions drawn by experts: a circle in each image indicating the boundaries of the embryo is manually drawn twice to reduce the error risk factor. Therefore, there are two ground truths (two manually drawn circles) for each embryonic depiction.

The statistical analysis can be described in terms of TP, TN, FP, and FN. The pixels that belong to the region of embryo according to both the experts and the proposed method are denoted as TP. The pixels that belong to the region of embryo according to the experts but not according to the proposed method are denoted as TN. The pixels that belong to the region of embryo according to the proposed method but not according to the experts are denoted as FP. The pixels that do not belong to the region of embryo according to the experts and the proposed method are denoted as FN.

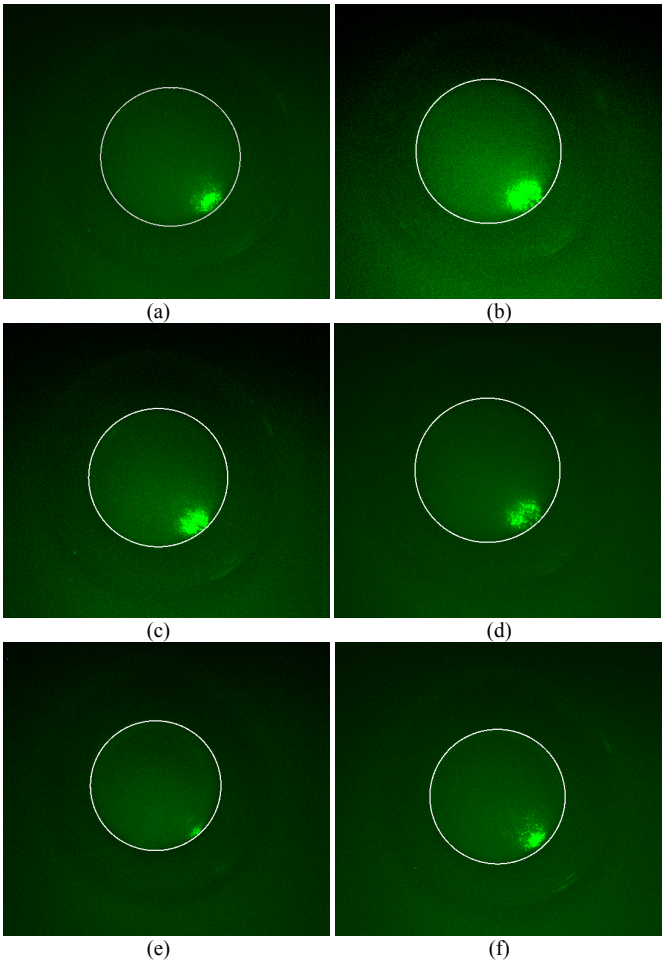


Fig.5. Segmentation results of the proposed method (color images).

The results were quantitatively evaluated using four metrics: the accuracy, the precision, the true positive rate and the true negative rate where calculated. Table I presents the average and the standard deviation of these four metrics. It is evident that the proposed method segments the Dharma embryo almost perfectly. The accuracy is (93%±0.04), the precision is (98%±0.01), and the true positive rate is (94%%±0.04). It is worth pointing out that the performance of our proposed method is very high, irrespective of which of the two ground truths was used.

Six segmentation examples of images - obtained from the 3D stacks depicting the Dharma embryo - are depicted in Fig. 5. Note the high segmentation quality of the proposed method is high even though the background of the image is inhomogeneous and the image is contaminated with noise and artifacts, (that is high intensity pixels outside the embryo).

IV. CONCLUSION

In this paper, a method for the segmentation of 3D time-lapse microscopy images depicting a live Dharma embryo has been presented. The proposed method is based on a genetic algorithm, which determines the boundaries of the dharma embryo very effectively. The experimental results demonstrate that the proposed method achieves (93%±0.04) accuracy and (98%±0.01) precision. It is also noise-resistant and can be applied to images, which do not contain edges.

REFERENCES

[1] G. Watson, *The Generations X report*, WWF DetoX Campaign, WWF, 2005.
[2] World Wildlife Fund. Available: http://www.worldwildlife.org
[3] Environmental Protection Agency. Available: http://www.epa.gov
[4] Organization for Economic Co-operation and Development. Available: http://www.oecd.org
[5] W. Driever, D. Stemple, A. Shier and L.S. Krezel, "Zebrafish: genetic tools for studying vertebrate development," *Trends in Genetics*, vol. 10, no. 5, 1994.
[6] F. Herrera, M. Lozano and A.M. Sanchez, "Hybrid crossover operators for real-coded genetic algorithms: An experimental study", *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, Spinger, vol. 9, no.4, April 2005, pp. 280-298.
[7] S.H. Ling, F.H.F. Leung, "An improved genetic algorithm with average-bound crossover and wavelet mutation operations", *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, Spinger, vol. 11, no.1, Aug. 2006, pp. 7-31.