34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Segmentation of Zebrafish Embryonic Images using a Geometric Atlas Deformation*

Eleni Zacharia, Maria Bondesson, Jan-Åke Gustafsson and Ioannis A. Kakadiaris

Abstract— **Transgenic zebrafish expressing fluorescent proteins in specific tissues or organs are promising models for studies of normal developmental processes, or perturbations of these. However, for widespread use, reliable quantification of the observed effects is necessary. Therefore, accurate and automatic analysis of images obtained by fluorescent microscopy and depicting zebrafish embryos becomes crucial. In this paper, a segmentation approach for such images is proposed. The segmentation is achieved by fitting a species-specific 2D atlas to the zebrafish data depicted in the images. Experiments performed in a set of 50 images have provided promising results.**

I. Introduction

Zebrafish (Danio rerio), has been a valuable vertebrate model organism in scientific research for decades. The most significant attributes that make zebrafish an ideal model system include [1][2]: (i) its small size, which facilitates simple staining/imaging techniques; (ii) its transparent embryos, that allow scientists to inspect the development of organs and cells in vivo – even those that are located deep in tissues – using standard microscopy; (iii) its external rapid embryonic development; and (iv) its demonstrated molecular similarities to mammalian models and – most significantly – humans.

Transgenic zebrafish – expressing fluorescent proteins in specific tissues or organs – are used for the screening of chemicals. In these fish, the changes in green fluorescent protein (GFP) expression upon chemical exposure are measured to determine a given chemical's toxicity. We have used a transgenic zebrafish (HGn50D, [3][4]), which expresses GFP specifically in the yolk sac, eyes and heart to visualize the size of the yolk sac during normal and chemically perturbed yolk absorption. During their first week of life, the zebrafish embryos feed on their yolk. If the yolk for some reason cannot be absorbed, the embryos are affected by malnutrition resulting in small size and developmental defects, a condition termed embryonic malabsorption syndrome (EMS). EMS can result from different events, including mutations in genes that encode for proteins that are involved in lipid metabolism, or from exposure to certain drugs, such as clofibrate and gemfibrozil [5]. EMS may also be caused by exposure to environmental pollutants. However, a quantitative method for determining the extent of yolk resorption in zebrafish has not been developed. The objective therefore is to identify and study any discernable abnormality of the yolk sac. Consequently, on the fifth day of embryonic life, the fish is imaged using fluorescent microscopy. The acquired images depict the fish's yolk sac, eyes (lens) and heart, as the GFP is expressed in these regions. On each image, it is of paramount importance to accurately determine the number of pixels that make up the embryonic yolk, thus computing its size as well as accurately measuring the average distance between each eye and the heart. The average distance between each eye and the heart is computed as a measurement of the size of the fish, while the distance between the eyes is measured to determine whether the fish was imaged from a ventral view, or if it had tilted, which would affect the measurement of size of the yolk sac.

To reach sound biological conclusions, an accurate segmentation of the zebrafish's embryonic state is required; the more accurate the image analysis is, the more reliable the biological results become. Segmentation of such images is a challenging task because these images are contaminated with noise and artifacts, while they present significant variation in their illumination. Furthermore, the appearance of zebrafish may vary in intensity, texture, size, and orientation. In addition, skin pigmentation makes the discrimination of the regions in which the GFP is expressed difficult.

To the best of our knowledge, there is no previous work on the segmentation of fluorescent microscopy images depicting the yolk of transgenic zebrafish. Image segmentation techniques based either on thresholding or on determining the regional edges of images alone are not sufficient to analyze these images. Methods of deformable models [6][7][8] – defined as curves or surfaces that deform under the influence of internal smoothness or external image forces – may not be appropriate as they can be trapped to a local minimum by spurious edges or noise when they are initialized far from the object boundary. Finally, statistical shape and appearance models [9][10][11] provide inaccurate

*Research 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) and by the University of Houston (UH) Hugh Roy and Lillie Cranz Cullen Endowment Fund. The biological experiments were supported by a grant from the EPA (R834289). All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representative of the official views or policies of the sponsors.

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, and J.Å. Gustafsson are with the Department of Biology and Biochemistry, Center for Nuclear Receptors and Cell Signaling, University of Houston, Houston, TX 77204-5056, USA. (mbondessonbolin@uh.edu, jgustafsson@uh.edu).

results when applied to these images because these models need a large training dataset, which is not currently available.

In this paper, we propose a method for analyzing images depicting transgenic zebrafish expressing fluorescent proteins. First, we construct a deformable, geometric 2D atlas of the anatomical structures of the transgenic zebrafish expressing fluorescent proteins. This atlas is represented by a subdivision mesh. Subsequently, the subdivision mesh is deformed in order to fit optimally to each zebrafish depicted on each particular image. The fitting problem is expressed as an optimization problem, which is solved by using genetic algorithms.

II. Zebrafish Atlas

A 2D atlas, defined as a template shape of the anatomical structure of an object, can be represented as a planar mesh consisting of polygonal elements such as quadrilaterals. The atlas is constructed by determining the quadrilaterals' vertices of a coarse mesh M^0 or, in other words, its control points. Then a fractal-like process, called *subdivision*, uses the coarse mesh M^0 to generate a sequence of increasingly fine meshes, which converge to a smooth mesh M^∞ at the limit. The limit mesh M^∞ represents optimally the anatomical boundaries of the object depicted. In practice however, the quality of the limit mesh is approximated by a high resolution smooth mesh M^k, where k denotes the highest subdivision level [12].

In this work, a deformable, geometric 2D atlas of the anatomical structures of the transgenic zebrafish expressing fluorescent proteins has been constructed. The 2D atlas partitions the zebrafish into the following four key regions: eyes (2 regions), heart, and yolk. The 2D atlas was represented as a Catmull-Clark subdivision mesh [13]. The constructed 2D atlas as well as the sample image used for its construction are depicted in Figure 1. Note that the coarse mesh M^0 has in total 30 control points.

III. Deforming the Atlas

'Fitting' an initial coarse mesh-model M^0 to the image data means to move its vertices locally in order to form a new subdivision mesh ($\widehat{M^0}$) whose associated limit mesh ($\widehat{M^\infty}$) accurately fits the object in the given image I. The automatic fitting of a high resolution mesh-model $\widehat{M^\infty}$ to a particular object depicted on the image I can be viewed as an optimization problem which is tackled by using the proposed genetic algorithm that determines the optimal positions of the control points of the deformable coarse mesh $\widehat{M^0}$ for a given image I.

A. Chromosome

The chromosome m represents a high resolution deformable mesh $\widehat{M^\infty}(m)$. Due to the inherent multi-resolution structure of the subdivision mesh, only the control points of the deformable coarse mesh $\widehat{M^0}(m)$ are needed to generate the deformable mesh $\widehat{M^\infty}(m)$. Therefore, the chromosome m encodes, the displacement vectors $\vec{d_i}$, i=1,...,N of the N control points v_i of the M^0 mesh. In order to maintain the overall shape of the mesh, a constraint is applied to the displacement vector $\vec{d_i}$ of each control point v_i. This constraint is described next.

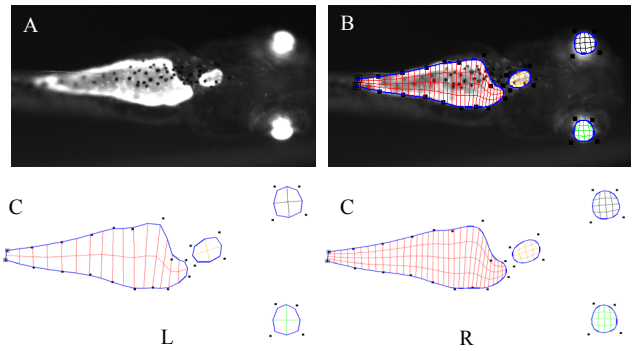

Figure 1. (A) Sample image used in constructing the 2D atlas of the Zebrafish. (B) Atlas overlaid on the sample image. (C) 2D atlas at the coarsest level of subdivision (L) and after being subdivided two times (R). In these images, the control points are depicted by large dots. The region boundaries are depicted in blue. (C)

Let us consider a pair of adjacent quadrilaterals – of the initial coarse mesh M^0 – consisting of the vertices $\{v_i,v_{i-1},v_{j-1},v_j\}$ and $\{v_{i+1},v_i,v_j,v_{j+1}\}$, sharing a common edge $\{v_i,v_j\}$. Figure 2(a) depicts these vertices. The vertex v_i is located above the lines $L_{j,j-1}$, $L_{j,j+1}$, right of the line $L_{i-1,j-1}$ and left of the line $L_{i+1,j+1}$ ($L_{m,n}$ denotes a line passing through vertices v_m and v_n). In order for the deformable coarse mesh $\widehat{M^0}(m)$ to maintain the overall shape of the mesh, the relative positions of the deformed vertex $\hat{v}_i$ with respect to the lines $\hat{L}_{j,j-1},\hat{L}_{j,j+1},\hat{L}_{i-1,j-1},\hat{L}_{i+1,j+1}$, should be identical to the relative positions of the vertex v_i with respect to the lines $L_{j,j-1}$, $L_{j,j+1}$,$L_{i-1,j-1}$,$L_{i+1,j+1}$ ($\hat{L}_{m,n}$ denotes a line passing through vertices $\hat{v}_m$ and $\hat{v}_n$).

An example of an acceptable and unacceptable deformable coarse mesh $\widehat{M}^0(m)$ is depicted in Figs. 2(c) and 2(d), respectively. As one may observe in Fig. 2(c), the $\hat{v}_i$ vertex is located below the $L(\hat{v}_{j-1},\hat{v}_j)$ while the corresponding vertex v_i of the initial coarse mesh M^0 (Fig. 2(a)) is located above the $L(v_{j-1},v_j)$ line.

B. Fitness Function

Each chromosome m in every population is evaluated using a fitness function, $F(m)$, which assigns to it 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. For the fitting optimization problem, the chromosome evaluation contains the following two objectives: (i) maximization of the total sum of image's intensities inside the deformed mesh $\widehat{M^\infty}(m)$ and simultaneous minimization of the total sum of image's

intensities outside the $\widehat{M^{\infty}}(m)$, and (ii) maximization of the total sum of image's intensities inside the $\widehat{M^{\infty}}(m)$ as well as near its perimeter and simultaneous minimization of the total sum of image's intensities outside the $\widehat{M^{\infty}}(m)$ as well as near its perimeter.

In the case when the intensity values of the object resemble the background (e.g., tail of the yolk sac), instead of maximizing/minimizing the total sum of image's intensities we maximize/minimize the total sum of edge's intensities.

It is worth pointing out that rather than computing the aforementioned objectives for the total area inside and outside the mesh, we compute them for four quadrilaterals Q_{i1},Q_{i2},Q_{i3},Q_{i4} each defined near a vertex $\hat{v}_i$. Figure 3 depicts Q_{i1},Q_{i2},Q_{i3},and Q_{i4} quadrilaterals with red, gray, yellow and green color. For instance, Q_{i1} is the area of the image which contains the pixels p having the following attributes: (i) their distances from the lines $L_{i,\mathrm{ia}}$, $L_{i,\mathrm{id}}$ are higher than w_1 and less than w_2, (ii) their distance from the line $L_{i,\mathrm{ic}}$ is less than w_3. The remaining quadrilaterals are similarly defined.

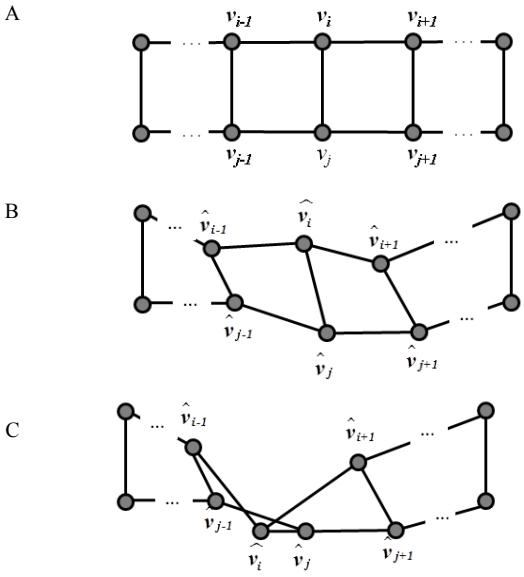

Figure 2. Example of the initial mesh (A), an acceptable deformation (B) and an unacceptable deformation (C).

Hence, the above objectives can be redefined as follows: To maximize, for each vertex $\hat{v}_i$, the total sum of image-intensities (or edge-intensities) inside the Q_{i1},Q_{i2} while simultaneously minimizing the total sum of image-intensities (or edge-intensities) inside the Q_{i3} ,Q_{i4} .

The fitness function of a chromosome m is defined as:

$$F(m) = \sum_i \sum_q E_{iq}^2 \qquad (1)$$

where E_{iq} denotes the total sum of image-intensities (or edge-intensities) of the Q_{iq} quadrilateral in the case when q=1,2, or the total sum of their complements in the case when q=3,4.

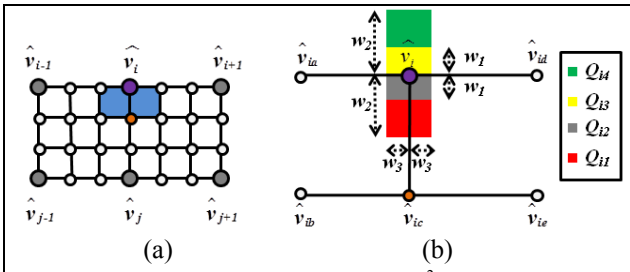

Figure 3. (a) Part of a deformable $\widehat{M}^2$ mesh. The vertices colored in gray and purple correspond to control points. (b) Enlargement of the blue region of (a) where the Q_{i1},Q_{i2},Q_{i3}, and Q_{i4} quadrilaterals are depicted in red, gray, yellow, and green color.

C. Evolutionary circle and Termination criterion

The genetic algorithm searches for the optimal deformable mesh 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. Thereafter, 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 [14], and (iii) the Wavelet mutation [15]. 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 has remained unchanged.

IV. Results

Several experiments were performed to evaluate the accuracy of the proposed approach on the segmentation of images depicting the anatomical structures (yolk sac, eyes, and heart) of transgenic zebrafish expressing GFP. The images used for validation of the proposed method – 50 in total – were obtained by inverted fluorescent microscopy; the fish were imaged from their ventral side using an inverted fluorescent microscope, on the 5th day post fertilization (dpf).

During the experiments the parameter values w_1=5, w_2=10, and w_3=1 were determined experimentally. Furthermore, G_F was set equal to 500, when segmenting the fish's yolk sac, and 200 when segmenting the fish's heart and eyes.

The evaluation of the proposed approach was based on a spatial overlap measure, the Dice Similarity Coefficient (DSC) [16], which is defined as:

$$DSC(A,B) = 2\frac{|A \cap B|}{|A| + |B|} \qquad (2)$$

where A, B denote two regions, |A| and |B| denote their areas, while $|A \cap B|$ denote the area of the shared region. A value of 0 indicates no overlap while a value of 1 indicates perfect agreement.

The evaluation results of the proposed method are presented in Table I. In this table, the mean and standard deviations of DSC are shown. Note that the DSC was

computed for each anatomical region of the deformed meshes. The ground truth was provided by expert biologists. The results indicate that the proposed method efficiently segments the yolk sac, eyes, and heart. Furthermore, we have compared our method with the AAM [9][17]. The model used in AAM consisted of the four anatomical regions of ZF expressing GFP and was built using a dataset of 25 images. The results of this method are presented on Table I. The proposed method outperforms the AAM method by a significant margin.

Three segmentation examples are depicted in Fig. 4. Note that the segmentation results of the proposed method are adequate even if the yolk's size and shape varies significantly. Furthermore, the proposed algorithm segments the four regions even when they contain dark areas caused by the fish's pigmentation.

V. Conclusion

In this paper, a method for the segmentation of images depicting transgenic zebrafish is presented. The proposed approach is based on the optimal deformation of a geometric 2D atlas to optimally fit the anatomical structures of the transgenic zebrafish expressing fluorescent proteins.

Acknowledgment

The authors would like to thank Dr. Kawakami, the National BioResource Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan for the transgenic fish HGn50D.

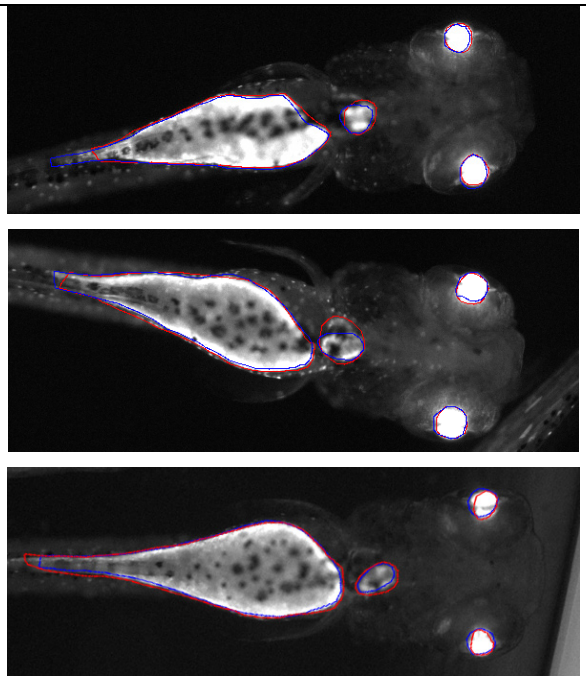

Figure 4. Comparison of the mesh boundaries obtained manually (blue) and by our proposed method (red) overlaid on transgenic zebrafish images.

TABLE I. Proposed Method vs Shape Appearance Model

		DSC (%)			
		R_1	R_2	R_3	R_4
Proposed Method	***Average***	93.47	84.55	85.47	70.52
	Standard deviation	1.56	5.03	6.66	11.4
Active Appearance model (AAM)	***Average***	81.47	63.70	67.28	62.52
	Standard deviation	11.61	24.86	23.37	20.11

Comparison of mean and standard deviations of DSC computed for the four anatomical regions for the proposed method and the AAM. The regions IDs "R_1","R_2","R_3","$R4$" correspond to the anatomical regions of the yolk, right eye, left eye, and heart, respectively.

References

[1] 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.

[2] A.J. Hill, H. Teraoka, W. Heideman and R.E.Peterson, "Zebrafish as a model vertebrate for investigating chemical toxicity," *Toxicological Sciences*, vol. 86, no.1, pp 6-19, 2005.

[3] K. Asakawa et al. "Genetic dissection of neural circuits by Tol2 transposon-mediated Gal4 gene and enhancer trapping in zebrafish," Proc. Natl. Acad. Sci., USA, vol. 105, pp. 1255-1260, 2008.

[4] S. Nagayoshi et al. "Insertional Mutagenesis by the tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: TCF7 and synembryn-like," Development, vol. 135, no. 1, pp 159-169, 2008.

[5] D. Raldúa, M. André and P.J. Babin, "Clofibrate and gemfibrozil induce an embryonic malabsorption syndrome in zebrafish," *Toxicology and Applied Pharmacology* vol. 228, pp.301–314, 2008.

[6] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," *Int. Journal of Computer Vision*," vol. 1, pp. 321-331, 1987.

[7] L.H. Staib and J.S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1061-1075, Nov. 1992.

[8] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Proc. Fifth IEEE Int. Conf. Computer Vision*, pp. 694-699, 1995.

[9] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models—Their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.

[10] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance Models," *in Proc. Fifth European Conf. Computer Vision*, vol. 2, pp. 484- 498, 1998.

[11] M.E. Leventon, E.L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours," in *Proc. IEEE Conf. Computer Vision and Pattern Recogn*, vol. 1, pp. 1316-1323, 2000.

[12] M. Bello, T.Ju, J.P.Carson, J. Warren, W. Chiu and I.A. Kakadiaris, "Learning-based segmentation framework for tissue images containing gene expression data," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 728-744, 2007.

[13] E. Catmull and J. Clark, "Recursive generated B-splines superfaces on arbitrary topological meshes," *Computer Aided Design*, vol. 10, no. 6, pp. 350-355, 1978.

[14] 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.

[15] 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.

[16] L.R. Dice, "Measures of the amount of ecologic association between species," *Ecology* vol. 26 no. 3, pp. 297–302, 1945.

[17] Active appearance models : http://www.mathworks.com/matlabcentral/fileexchange/26706 [available online]