Embryo zebrafish segmentation using an improved hybrid method

Microscopy

TAO WU*, JIANFENG LU*, YANTING LU*, TIANMING LIU† & JINGYU YANG*

Journal of

* School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

†Computer Science Department, University of Georgia, Athens, GA, U.S.A.

Key words. Active contour model, edge stopping function, expectation maximization algorithm, image segmentation, zebrafish embryo image.

Summary

Zebrafish is an invaluable vertebrate model in life science research and has been widely used in biological pathway analysis, molecular screening and disease modelling, among others. As a result, microscopic imaging has become an essential step in zebrafish phenotype analysis, and image segmentation thus plays an important role in the zebrafish microscopy analysis. Due to the nonuniform distribution of intensity and weak boundary in zebrafish microscope images, the traditionally used segmentation methods may lead to unsatisfactory result. Here, a novel hybrid method that integrates region and boundary information into active contour model is proposed to segment zebrafish embryos from the background, which performs better than traditional segmentation models. Meanwhile, how to utilize the gradient information effectively in image segmentation is still an open problem. In this paper, we propose to improve the aforementioned hybrid method in two aspects. Firstly, the mean grey value of background is estimated by the expectation maximization (EM) algorithm to constrain the active curve evolution. Secondly, an edge stopping function sensitive to gradient information is designed to stop curve evolution when the active curve reaches the embryo boundary. Experimental results show that the proposed methods can provide superior segmentation results compared to existing algorithms.

Introduction

The zebrafish (*Danio rerio*) is an invaluable vertebrate model system in life science research, and computerized characterization of zebrafish morphology is significantly important for biological pathway analysis, disease modelling and drug discovery (Zhang *et al.*, 2004; Stern & Zon, 2003),

Correspondence to: Jianfeng Lu, School of Computer Science, Nanjing University of Science & Technology, Xiaolingwei 200#, Nanjing, Jiangsu province, China, 210094. Tel: 86-25-84315751-824; fax: 86-25-84315960; e-mail: lujf@njust.edu.cn

among others. In recent years, many researchers have been using zebrafish for answering basic biological questions, for toxicology research and for environmental monitoring (Schier & Talbot, 2005). Particularly, recently available highthroughput zebrafish bioimaging systems, such as the GE In Cell 1000/3000 systems, can produce dozens of thousands images (data up to gigabytes) at a time, which is practically impossible to handle them manually by average biologists. If the image data can be processed and analysed by computerized algorithmic pipeline, the results not only will be more objective, but also can be fed to structured database for further statistical analysis and data mining (e.g., Lu *et al.*, 2006, 2011, 2012; Liu *et al.*, 2006, 2008, 2008b).

Essentially, automatic image segmentation, which is very important for the subsequent steps of image registration, shape analysis and morphological analysis and understanding, is a fundamental problem in computerized zebrafish microscopy image analysis. Accurate zebrafish embryo segmentation contributes to not only biological research of zebrafish development, but also to the subsequent zebrafish phenotype analysis, e.g. curvature quantization of zebrafish tail and corresponding classification, zebrafish shape registration and zebrafish image retrieval, and so on.

Image segmentation can be done using various methods such as threshold-based segmentation and area growth. However, the intensity distribution of embryo is typically nonuniform in many zebrafish bioimages. For instance, in some zebrafish images, the grey value of the head area is less than that of the background, whereas the grey value of embryo tail is greater than that of the background. The Otsu algorithm (Otsu, 1979), among the most popular thresholding methods, can obtain satisfactory segmentation results under the assumption that the distributions of target and background appear as a two-peaked shape. However, in the case of the overlapping intensity distributions between the object and the background, this method may have difficulties. For instance, in our case, part of the zebrafish embryo information will be lost after the segmentation by the Otsu algorithm. Meanwhile, edge detection plays an important role in image segmentation. In recent years, several new edge detection approaches have been presented with the development of artificial intelligence and mathematics, such as wavelet transform method, neural network method, morphological method and so on. In general, all these methods need to obtain gradient information first. However, accurate target boundary cannot be obtained easily based on the gradient information only. Hence, for these edge-based methods, not only the earlier errors might be passed on to the subsequent steps, but also the final edge may not be a closed contour. As the boundary between the zebrafish embryo and the background is weak, edge detection algorithms may produce a contour that is incomplete or has some background pixels misclassified as the zebrafish target.

In recent years, partial differential equations (PDE) based on active contour models are widely used in image segmentation methods. Active contour model such as the snake method, initially proposed by Kass, Witkin and Terzopoulos (Kass, 1988; Terzopoulos, 1988), is a popular model for image segmentation. Essentially, the active contour model overcomes such drawbacks as nonclosed boundary and susceptibility to noise in previous image segmentation methods, and can achieve more natural segmentation result. However, the initial snake method needs to set many parameters and has high computational costs. Therefore its utilization is limited by problems associated with the initialization and poor convergence to boundary concavities. Since the traditional geometric active contour model (Caselles, 1997) depends only on image edge, which, in the case of zebrafish embryo, is blurry and discrete, it is difficult to obtain the ideal result. Afterwards, the simplified Mumford-Shah model proposed by Chan & Vese (2001) can segment target without edge information, and is a typical method used for segmenting target with blurry boundary. The model divides image into two parts with different mean grey levels, associated with the foreground and background respectively. However, due to the large variation in intensity distribution, the method's reliance on global information of the homogeneous regions may lead to many missing parts in the final segmentation of the zebrafish embryo.

Generally speaking, hybrid methods can effectively integrate the region and boundary information into active contour model to achieve more accurate segmentation result (Zhang, 2008). The boundary information can help to detect the precise location of the target object, and the region information can help to prevent the boundary leakage problem. However, this hybrid method assumes that the prior knowledge of the grey-level of background is available, and this condition cannot be satisfied for most cases in our application. By contrast, for images with weak target boundary, the boundary leakage problem still persists when executing segmentation using this method. Therefore, how to utilize the gradient information effectively to overcome the boundary leakage is still an open problem. In response to the above problems, in this paper, we propose a novel algorithmic pipeline to improve the aforementioned hybrid method in two aspects. Firstly, the mean grey value of background is estimated by the expectation maximization (EM) algorithm in order to constrain active curve evolution. Secondly, an edge stopping function sensitive to the gradient information is designed to stop curve evolution when active curve reaches embryo boundary. In addition, the numerical scheme can adopt additional operator splitting algorithm (Weickert, 1998) to reduce computational costs. In short, combined with the level set approach, the active curve automatically, and then reduce the interference of image noise.

Method

Dataset

Embryos were obtained from natural spawning of wild-type (Tubingen long fin [TLF] strain) adults. They were raised and staged according to Kimmel *et al.* (1995), and images of zebrafish embryos at 48 h postfertilization were acquired using IN CELL Analyzer 1000 (GE Amersham, Buckinghamshire, UK) with 96-well plates (Lu *et al.*, 2006). More details of the imaging settings and data acquisition can be found in previous publications (Lu *et al.*, 2006, Liu *et al.*, 2008).

Level set method

One of the problems with active contour model is that it cannot deal with topological changes naturally during the curve evolution. To address this issue, we apply the level set method to the original active contour model, in order to allow natural topological evolution of the model into the target boundaries. In this section, the level set method is briefly described first, and then the traditional classic models (e.g. active contour model without edges) and their disadvantage will be presented. The hybrid model that can segment target more accurately will be designed at last.

The original active contour model has relatively high time complexity, and moreover, it cannot deal with topological changes of the active curve and nor it can merge or split the curves in the case of multiple targets. The level set method later proposed by Osher & Sethian (1988) can solve the numerical curve evolution problem effectively and detect edges of multiple objects, which paves the way for image processing based on partial differential equation. The variational level set method is usually used in the active curve evolution, and the Heaviside function (Zhao, 1996) is used to achieve the goal, which is defined as follows:

$$H(u) = \begin{cases} 1, & u \ge 0\\ 0 & u < 0 \end{cases}$$
(1)

In practice, the regularized Heaviside function

$$H(u) = \frac{1}{2} \left[1 + \frac{2}{\pi} \arctan\left(\frac{u}{\xi}\right) \right]$$
(2)

is adopted, whose derivative function is

$$\delta_{\xi}(u) = \frac{1}{\pi} \frac{\xi}{\xi^2 + u^2},$$
(3)

where the parameter ξ controls the change rate of *H* varying from 0 to 1.

Typically, u(x, y) is defined to be the signed Euclidean distance function from the point (x, y) to the curve *C*, and it is negative when pointing outside the curve and is positive when inside the curve. Based on this definition, the active curve can be located by zero level set and *u* is changing constantly by the evolution equation. As a result, the topology changes can be easily dealt with by pushing the zero level set to the boundary of target image, and the evolution will automatically stop at the boundary. The convergence criterion of the curve evolution is as follows: $\frac{\sum_{|u_{i,j}^n| < c} |u_{i,j}^{n+1} - u_{i,j}^n|}{M} \le h^2 \Delta t$, where *h* denotes space step with its unit grid point and Δt denotes time step, *M* is the number of grid points satisfying the condition $|u_{i,j}^n| < c$ and the parameter *c* is a constant number $(h \le c \le 2h)$.

Brief introduction of the Chan–Vese model

If the mean intensity of the zebrafish target is quite different from that of the background, the image can be divided into two parts by the closed curve, and the region enclosed by the curve corresponds to the target. Then, the curve is regarded as the target contour. Based on this idea, the energy function

$$E(c_1, c_2, C) = \mu \oint_C ds + \lambda_1 \oint_{\Omega_1} (I - c_1)^2 dx dy + \lambda_2 \oint_{\Omega_2} (I - c_2)^2 dx dy$$
(4)

was proposed by Chan & Vese (2001), called Chan–Vese model here, where Ω_1 and Ω_2 denote regions inside and outside active curve respectively and *C* in Eq. (4) denotes the curve.

The equation can be derived by the variational level set:

$$E(c_1, c_2, C) = \mu \oint_{\Omega} \delta(u) |\nabla u| dxdy + \lambda_1 \oint_{\Omega} (I - c_1)^2 H(u) dxdy + \lambda_2 \oint_{\Omega} (I - c_1)^2 [1 - H(u)] dxdy.$$
(5)

When *u* is fixed, minimizing *E* relative to c_1 and c_2 will derive the parameters:

$$c_{i} = \frac{\iint_{\Omega_{i}} Idxdy}{\iint_{\Omega} Idxdy}, \quad i = 1, 2$$
(6)

and the parameters c_i corresponds to the mean grey level of different areas. The details of the numerical solution can be found in the reference by Chan & Vese (2001). In the experiment, $|\nabla u|$ is used instead of δ_{ξ} (Zhao, 1996) in order to expand the variation range of embedding function *u*. The property of the signed distance function gives $|\nabla u| = 1$, so the iteration equation can be simplified as

$$\frac{\partial u}{\partial t} = \mu \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) - (I - c_1)^2 + (I - c_2)^2.$$
(7)

When the curve is initialized around the target, better segmentation result can be achieved. When there is no complex background, the initial curve can be located at any position. For the cases with complex background or targets with nonuniform grey intensity distributions, c_1 and c_2 as the global properties will not produce desired results, and only part of the target with low intensity can be segmented correctly. But the remaining part, which has a similar grey value as background like the tail part of zebrafish, will be segmented as background, so the zebrafish embryo could not be completely extracted based on this model.

Hybrid model with integrated region and boundary information

For the Chan-Vese model, when the difference between the intensity mean of the background and that of zebrafish embryo is weak, the segmentation result may be incomplete. A hybrid model (Zhang, 2008) was thus proposed to achieve robust and accurate segmentation results, which not only uses the boundary information to detect the accurate location of the target, but also uses the area information to solve the boundary leakage problem during the evolution. Because the mean grey level of background is larger than most grey levels of zebrafish embryo area, the mean grey level of background can be used as a threshold to divide the image into two parts. If the threshold is not accurate or the edge stopping function cannot stop the evolution of active curve accurately in the hybrid model, the zebrafish embryo segmentation result will not be ideal. In view of uniform grev level of background in zebrafish embryo images, we estimate the threshold by the EM algorithm to improve the segmentation result, and we design a gradient-sensitive edge stopping function to solve boundary leakage problem and obtain accurate contour of zebrafish embrvo.

There are three steps in our hybrid model as follows. Firstly, the mean grey level of background area will be estimated by the EM algorithm, which is used to guide the active curve evolution to the boundary of the area with larger intensity. Next, the energy function with region and boundary information is constructed and the corresponding iterative equation is formulated. Finally, the active curve will be evolved by the iterative equation and stops when the curve reaches the boundary of zebrafish embryo by the proposed edge stopping function. The EM algorithm (Dempster, 1977) searches the parameters by maximum likelihood estimation from the probability model that depends on the hidden variables. This approach attempts to optimize the fitting between the given data and the mathematical model assumed, and some parameters can be estimated based on the potential joint probabilistic distribution that fits the data best. The EM algorithm is relatively easy to implement and converges quickly. The grey levels of pixels of background are assumed to follow Gaussian distributions. We choose to estimate the grey level mean of background by the EM algorithm. Because the grey level of the background tends to be uniform that accords with Gaussian distribution, the grey level mean of background will be estimated well.

Due to the blurry boundaries in zebrafish image, an edge stopping function is then used to terminate the curve evolution towards inner target when the gradient is up to some extent, which can avoid boundary leakage. In this paper, the edge stopping function (Weickert, 2001) is defined as follows:

$$g(r) = 1 - \exp\left[-\frac{c}{\left(r/K\right)^{p}}\right],$$
(8)

where *r* is the gradient of zebrafish image smoothed by a Gaussian filter and *K* is a contrast parameter to control the variation range of *g* from 0 to 1. The constant parameter *c* and *p* in formula (8) are 3.315 and 8 respectively as suggested in reference by Weickert (2001). If *r* is slightly smaller than *K*, $(r/K)^p$ will be very small so that the exponential term is almost zero and *g* is about 1; conversely, *g* is about 0 and the evolution will stop.

Compared with the Chan–Vese model, the hybrid model has better robustness in the case of complex background or target. The energy function of hybrid model is as follows:

$$E(u) = -\alpha \iint_{\Omega} (I - \mu) H dx dy + \beta \iint_{\Omega} g |\nabla H| dx dy, \quad (9)$$

where α , β are weighting parameters of the area and length terms respectively, μ is the threshold that corresponds to the evaluated mean grey level of the background, *g* is the edge stopping function defined in Eq. (8), and the gradient descent flow by minimizing the energy function is

$$\frac{\partial u}{\partial t} = \delta_{\xi} \left[\alpha \left(I - \mu \right) + \beta \operatorname{div} \left(g \frac{\nabla u}{|\nabla u|} \right) \right].$$
(10)

Then, its simplified form is

$$\frac{\partial u}{\partial t} = \alpha \left(I - \mu \right) + \beta \operatorname{div} \left(g \frac{\nabla u}{|\nabla u|} \right), \tag{11}$$

which corresponds to the evolution equation

$$\frac{\partial C}{\partial t} = \alpha \left(I - \mu \right) \vec{N} - \beta \operatorname{div} \left(g \frac{\nabla u}{|\nabla u|} \right) \vec{N} = \alpha \left(I - \mu \right) \vec{N} - \beta g \operatorname{div} \left(g \frac{\nabla u}{|\nabla u|} \right) \vec{N} + \left(\nabla g \cdot \frac{\nabla u}{|\nabla u|} \right) \vec{N}, \quad (12)$$

© 2013 The Authors

Journal of Microscopy @ 2013 Royal Microscopical Society, **250**, 68–75

where \cdot denotes inner product operation and is the unit normal whose direction points to the inside of the curve.

Then, the evolution equation of Chan–Vese model can be adjusted as

$$\frac{\partial C}{\partial t} = \lambda [(I - c_1)^2 - (I - c_2)^2] N - \beta \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) N$$
$$= 2\lambda (c_1 - c_2) \left(I - \frac{c_1 + c_2}{2}\right) N - \beta \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) N,$$
(13)

when the difference of grey-level means between the embryo and the background is small, the weight $2\lambda(c_1 - c_2)$ in the first term is small and so is $(I - \frac{c_1 + c_2}{2})$, and then the first term has little impact on the segmentation result. The hybrid model extracts the area with the intensity larger than μ , but the mean intensity of the zebrafish embryo body is lower, and the area of embryo body is expected to be encircled by a closed curve, so the grey-level inverse technique is applied to make the mean intensity of embryo body larger than that of background. Then, the mean grey values of the embryo and the background are estimated by the EM algorithm, the mean grey value of the background is for the parameter μ in formula (9) and (10), and the gradient information is used to constrain the curve to avoid its moving towards the inner embryo.

Results and discussion

Experimental results and analysis

In our experiments, about 100 zebrafish embryo images acquired by the GE IN Cell 1000 machine are used, and each one contains a whole embryo. The size of each image is about 320^*240 . The algorithm is implemented by Matlab 2009, and the hardware environment is P4 2.1G with 4G RAM. Based on the two parameters c_1 and c_2 with global property in the Chan–Vese model, the image is segmented into two parts by performing a *k*-means clustering that considers parts outside the curve as the background.

Experiments on the Chan–Vese model and the hybrid model were conducted to compare their performances. The initial curve must be closed and can be in arbitrary shape, and the shape of all initial curves is set to be circular in our experiments. Fig. 1 shows segmentation results of these two models by the single initial circle curve, and Fig. 2 shows results by multiple initial active curves. The positions of the single circular curve or multiple circular curves in Fig. 1 and Fig. 2 indicate the locations of zero level set. With the embedding function changing, the zero level sets produced by the new closed curve emerge, whereas some curves disappear. For the Chan-Vese model, different locations of initial curves will lead to different segmentations for the image with nonuniform intensity distribution. When mean intensities of the areas inside and outside the curve both change very little during several sequential iterations, the curve evolution will stop.



Fig. 1. The first row: original images with a circular initial curve; the second row: the evolution results by the Chan–Vese model; the third row: the evolution results by our hybrid model.

From the evolution results by the Chan–Vese model in Fig. 2(c), (d), it can be found that some zero level sets emerge far away from the initial curve evolution result, so it is $|\nabla u|$ but not δ_{ξ} that affects the variation range of the embedding function u from the local area around active curve to full scope and segment the image better. The inner areas of all closed curves are considered as zebrafish embryo, and other areas are regarded as the background. In the Chan–Vese model, the curve evolution will stop when the sum of two within-class variance reaches the minimum value. As shown in Fig. 1(c), 1(d) and Fig. 2 (d), due to the grey level range overlapping between the zebrafish embryo and background, the segmentation results are not satisfied. Even if multiple closed curves are used as the initial active curves (Fig. 2(d)), the segmentation results are still unsatisfactory.

By the hybrid model we proposed, the zebrafish embryo, whose grey level is larger than that of background, can be enclosed inside the result curve (see Fig. 1 (e), (f) and Fig. 2(e), (f)). Compared with the results of the Chan–Vese model and the hybrid model in Fig. 1 and Fig. 2, it is apparent that the hybrid method achieved much better results in terms of the segmentation accuracy. However, to make the segmentation results more independent of and less sensitive to the initialization, multiple initial curves can be used, see Fig. 2 for example. Part of background is with lower intensity than that of zebrafish embryo, and some small closed curves corresponding to zero level sets are not what we desire to have

Fig. 2. (a) and (b): original images with many circular initial curves; (c) and (d): segmentation by the Chan–Vese model with many initial curves, (e) and (f): evolution results by our hybrid model; (g) and (h): results of (e) and (f) after morphological processing.

(g)

(h)

and thus can be removed by area threshold so as to obtain the real embryo contour only. Due to the weak and blurry boundary, the obtained contour is not smooth enough, so morphological operations are applied to smooth the contour. The disk with radius 5 is chosen as the structure element, and pairwise morphological open operation and close operation are carried out. Examples of the final results are shown in Fig. 2(g), (h).

Fig. 3 presents ten zebrafish images with automatic and manual segmentation results. To quantify the automatic segmentation performance, the commonly used area agreement is defined as $R(T_i, T_j) = \frac{V(T_i \cap T_j)}{(V(T_i) + V(T_j))/2}$, where V(.) indicates the area of the map, and T_i , T_j are automated and manual segmentation maps respectively, and the \cap operator takes the intersection of two maps. The area overlap definition is akin to that in reference by Cardenas *et al.* (2001), and the definition implies that more overlap area means a higher area agreement. Table 1 provides the results of 10 zebrafish images

EMBRYO ZEBRAFISH SEGMENTATION USING AN IMPROVED HYBRID METHOD 73



Fig. 3. The comparison between manual and automated segmentation: the red curves indicate automated segmentation results and the blue curves indicate manual segmentation results.

that are selected randomly. The corresponding images and segmentation results are shown in Fig. 3, and the standard deviation of area agreement results is 0.016, the mean of area agreement for all images is about 93.1%. These results indicate that the automated segmentation is comparable to that by manual segmentation.

The contrast parameter K for edge stopping function in Eq. (8) can affect the segmentation result. The larger the K the smaller the value of edge stopping function g for the same gradient value. As a result, it leads to shorter and smoother active curve. By contrast, active curve evolution may stop when the gradient value is small due to very small weighting parameter g, so that segmentation may miss some parts of target when curve is inside of target, but can obtain accurate

result at concave boundary. These two points have been observed by experiments setting with different *K* values.

Image segmentation result typically depends on the tradeoff of some parameters, and these parameters include the scale of Gaussian filter in Eq. (8), the weighting parameter β and the time step. If the scale of Gaussian filter is larger, *K* will tend to be smaller so that more accurate segmentation result can be obtained. If the time step is larger, convergence speed will be faster at the cost of boundary accuracy. At the same time, the influence of edge stopping function is controlled by weighting parameter β .

Finally, the time cost of our zebrafish embryo segmentation method is detailed. For the segmentation with single initial curve, the time costs of the EM step for Fig. 1(a) and Fig. 1(b)

Table 1.	Area	agreemen	t for	10	zebra	fish	images
----------	------	----------	-------	----	-------	------	--------

Index of Figure	ABMS	Area by ours	OA by ours	AA by ours	Time by ours (s)	Area by CV	OA by CV	AA by CV	Time by CV (s)
a	15043	15 666	14 731	0.959	5.227	12 316	12 139	0.8874	5.698
b	13891	14198	13464	0.959	3.783	9 605	9 546	0.8126	2.394
с	16410	17 538	15764	0.929	5.678	13 562	13 238	0.883	5.816
d	16441	17 270	15812	0.939	6.271	11 437	11079	0.7948	6.790
e	30105	26 037	25 625	0.913	15.967	17 324	17074	0.7200	10.544
f	14 512	15 415	14 223	0.951	8.289	10 377	9 6 7 1	0.7771	7.421
g	18672	19 602	17985	0.940	11.950	15 338	14 298	0.8408	12.590
h	38975	36 305	34925	0.928	10.640	33 486	32 372	0.8935	18.225
i	13 701	14 265	13105	0.937	4.825	10 355	9 807	0.8153	6.343
i	12997	12 579	11760	0.920	3.957	12 960	12 149	0.9361	3.099
Mean				0.938				0.836	

ABMS: area by manual segmentation, OA: overlap area, AA: area agreement.

are 4.553s and 3.017s respectively, and the total costs are 15.922s and 11.234s. For the segmentation with multiple initial curves, the time costs of the EM step for Fig. 1(a) and Fig. 1(b) are 4.476s and 3.102s respectively, and the total costs are 9.609s and 6.672s. It can be seen that if there are multiple initial closed curves at the beginning, the curve will reach the boundary faster and the result is not influenced by the initial position of the curve.

Discussions and conclusions

In microscopic zebrafish images, the grey level of background tends to be uniform and is larger than that of zebrafish embryo on the whole, so the mean grey level of background can be evaluated by the EM algorithm preferably. But if the background or target has wide grey level range that has overlapping distributions, the threshold evaluation will be inaccurate and the segmentation result will be unsuccessful due to undesirable constraint condition.

The choice of an appropriate value for the parameter *K* is based on the information of low gradient around the boundary, whereas how to set the parameter *K* automatically should be studied further. Furthermore, an appropriate edge stopping function plays an important role for accurate segmentation, and other edge stopping functions need to be investigated and applied properly, e.g. $g(r) = \frac{1}{1+(r/K)^p}$ in reference by Weickert (1998). Our experiments show that an appropriate contrast parameter *K* not only contributes to constrain active curve close to boundary but also form smooth contour.

In our method, the EM algorithm, which accounts for a considerable amount of computing time, should be improved in the future. In this paper, the mean grey level of the background estimated by the EM algorithm is used as the threshold, and of course other methods should be studied to estimate this threshold better and/or faster. The hybrid model firstly uses this threshold to obtain most target area, and then utilizes edge stopping function to locate zebrafish embryo contour well. If the first step can reduce the occurrences of undesirable results, the final segmentation can see more improvements, so how to obtain an appropriate threshold is still an important problem and worth being investigated further.

Based on the boundary information and the region information, an active contour model is used to evolve the curve towards the boundary automatically and achieve a complete contour, which can be applied to the segmentation of zebrafish image with weak boundary and thus obtain desirable results. Compared with the Chan–Vese model, the novel hybrid model in this paper combines the expectation maximization algorithm to evaluate the mean grey level of the background that is used to guide image segmentation, and makes use of the gradient information as a constraint to deal with blurring and weak boundary. Consequently, our hybrid model achieves better segmentation results than other existing methods. In the future, we plan to integrate the algorithmic pipeline into a software filer that can be plugged into the ZFIQ software package (Liu *et al.*, 2006) and be distributed for wider applications.

Acknowledgements

This work is funded by National Natural Science Foundation of China (NSFC) under Project 60775015, the Priority Academic Program Development of Jiangsu Higher Education Institutions and QingLan Project. Certain zebrafish images were obtained from our previous collaborations with Dr. Weiming Xia, Dr. Stephen Wong, Dr. Scott Holley and Dr. Jarema Malicki. We also would like to thank Dr. Jin Gu and Yuting Sheng for their helpful editing on this manuscript.

Conflict of interests

The authors state no conflict of interest.

References

- Cardenas, V., Ezekiel, F., Di S.V., Gomberg, B. & Fein, G. (2001) Reliability of tissue volumes and their spatial distribution for segmented magnetic resonance images. *Psychiatr. Res.* 106(3), 193–205.
- Caselles, V., Kimmel, R. & Sapiro, G. (1997) Geodesic active contours. *Int. J. Comput. Vis.* 22(1), 61–79.
- Chan, T. & Vese, L. (2001) Active contours without edges. *IEEE Trans. Image Process* **10**(2), 266–277.
- Dempster, A., Laird, N. & Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm. J. R. Statist. Soc. B 39(1), 1–38.
- Kass, M., Witkin, A. & Terzopoulos, D. (1988) Snakes: active contour models. Int. J. Comput. Vis. 1(4), 321–331.
- Kimmel, C., Kimmel, B., Ullmann, B. & Schilling, T. (1995) Stages of embryonic development of the zebrafish. *Dev Dyn.* 203(3), 253–310.
- Liu, T., Lu, J., Wang, Y., Campbell, W., Huang, L., Zhu, J., Xia, W. & Wong, S. (2006) Computerized image analysis for quantitative neuronal phenotyping in zebrafish. *J. Neurosci. Methods* 153(2), 190–202.
- Liu, T., Nie, J., Li, G., Guo, L. & Wong, S. (2008) ZFIQ: a software package for zebrafish biology. *Bioinformatics* 24(3), 438–439.
- Liu, T., Li, G., Nie, J., *et al.* (2008b) An automated cell detection method for cell detection in zebrafish. *Neuroinformatics* 6(1), 5–21.
- Lu, J., Liu, T., Ding, J., Zhu, J., Yang, J., Xia, W. & Wong, S. (2006) Automated quantization of zebrafish somites in high-throughput screens. Proceedings of IEEE/NLM International Workshop on Life Science Systems and Applications. pp. 34–35.
- Lu, Y., Lu, J., Liu, T. & Yang, J. (2011) Automated gene oscillation phase classification for zebrafish presomitic mesoderm cells. *Cytometry* A 79A(9), 727–735.
- Lu, J., Wu, T., Liu, T., Chen, C., Zhao, C. & Yang J. (2012) Automated quantification of zebrafish somites based on PDE method. *J. Microscopy* 248(2), 156–162.
- Osher, S. & Sethian, J. (1988) Fronts propagating with curvaturedependent speed: algorithms based on Hamilton-Jacobi Formulations. *J. Comput. Phys.* **79**(1), 12–49.
- Otsu, N. (1979) A threshold selection method from gray-level histograms. *IEEE Trans. SMC* 9(1), 62–66.

- Schier, A. & Talbot, W. (2005) Molecular genetics of axis formation in zebrafish. Annu. Rev. Genet. 39, 561–613.
- Stern, H. & Zon, L. (2003) Cancer genetics and drug discovery in the zebrafish. *Nat. Rev. Cancer* **3**(7), 533–539.
- Terzopoulos, D., Witkin, A. & Kass, M. (1988) Constraints on deformable models: recovering 3D shape and nonrigid motion. *Artif. Intelligence* 36(1), 91–123.
- Weickert, J. (2001) Application of nonlinear diffusion in image processing and computer vision. *Acta Math. Univ. Comenianae* **70**(1), 33–50.
- Weickert J. (1998) Anisotropic Diffusion in Image Processing. (ed. By B. G. Teubner) Stuttgart, Germany.
- Zhao, H., Chan, T., Merriman, B. & Osher, S. (1996) A variational level-set approach to multiphase motion. *J. Comput. Phys.* **127**(1), 179–195.
- Zhang, Y., Matuszewski, BJ., Shak, L. & Moore, CJ. (2008) Medical image segmentation using new hybrid level-set method. Proceedings of Fifth International Conference Biomedical Visualization: Information Visualization in Medical and Biomedical Informatics, pp. 71–76.
- Zhang, L., Zhou, H., Su, Y., *et al.* (2004) Zebrafish Dpr2 inhibits mesoderm induction by promoting degradation of nodal receptors. *Science* **306**(5693), 114–117.