Proceedings of the 9th International Conference on Information Technology and Applications in Biomedicine, ITAB 2009, Larnaca, Cyprus, 5-7 November 2009

Automatic Standardisation of a Zebrafish Embryo Image Database

Fernando Boto, Céline Paloc, Alexis Verbeke, Carles Callol, Ainhoa Letamendi, Izaskun Ibarbia

Abstract— **Recently, there has been an increasing interest to propose computational approaches based on image processing to automate the comparison of spatial gene expression patterns contained in transgenic embryo images. The first step is typically to classify the images between lateral and dorsal/ventral views, and then to align them along the anterior and posterior ends. While some methodologies have been recently proposed for standardizing images in Drosophila database, the case of zebrafish embryos have never been tackled. In this paper, we propose a standardisation methodology for building a zebrafish embryos image database. Details of our approach and the results of using this approach on a pilot dataset are presented. A quality test first allows rejecting images presenting artefacts and which cannot be processed. The retained images are then classified, achieving a success rate of 93%.**

Index Terms— *microscopic image processing, zebrafish embryos, classification, database standardization.*

I. Introduction

While tremendous advances in imaging hardware make now possible the rapid acquisition of thousands of microscopic images of small-animal models (embryos of drosophila, c-elegans, chicken, zebrafish...), the current method of analysing these images is still mainly the manual and visual inspection to infer potential genetic or environmental interactions by identifying functional or morphological similar characteristics [1-3]. Recently, there has been an increasing interest to propose computational approaches based on image processing to automate the comparison of spatial patterns contained in such images (Figure 1). However, scientists publish digital images in different sizes and orientations, and under different illumination conditions, which makes similarity-based image classification particularly challenging. Bringing images into standardisation is a critical first step in making a meaningful comparative analysis of spatial patterns. While some methodologies have been recently proposed for standardizing images in Drosophila database, the case of zebrafish embryos has never been tackled.

II. State of the Art

Despite the fact that there exist many methods for general image retrieval, there is little work in automatic analysis and comparison of embryo images. Some very recent work has focused on embryogenesis staining pattern images of the Drosophila. To meaningfully compare gene expression patterns computationally, it is important to first standardize the embryo images. Standardization involves the following: removing the background information, isolating the embryo, aligning the embryo in the standard orientation used by embryologists (anterior pole to the left and dorsal side up), uniforming illumination and normalising the size of the resulting images.

In [4], the images that were used for gene expression pattern comparison were already segmented and aligned. The standardization procedure only involved an edge-fitting algorithm to crop the Drosophila embryo into a bounding box and a size standardization to normalize the resulting images.

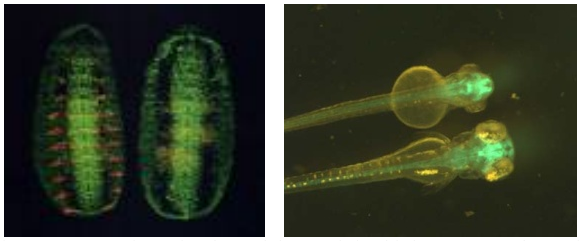


Fig.1. Transgenic animal models used in biology research. Left: Drosophila embryos – Right: Zebrafish embryos.

The segmentation and alignment method was described in [5-7]. A single threshold on image-pixel intensity is insufficient to extract the embryo because of background shadows and buzzy embryo boundary. As the embryonic region has much richer texture information than the image background, the embryo is segmented by thresholding the local variance of a small region (e.g. 3×3 pixels) around each pixel. For a segmented embryo, the principal direction along which the variation of all embryonic pixel-coordinates is the greatest is considered as the anterior-posterior axis of the embryo. The image is then rotated to make this axis horizontal, and the embryonic region is cropped and rescaled.

In [5], a visual inspection confirms that the above method is effective in 4400 images provided by [1]. However, the process does not differentiate the anterior – posterior or dorsal - ventral orientation of the embryo. In [6], 30,000 images of [1] were processed and 67% were successfully standardised as described. However, only lateral views were used. In [7], another processing step based on watershed was added to extract the main embryo from others that might appear in the experimental image. [7] claims a success rate

Manuscript received July 19, 2009.

F. Boto, C. Paloc and A. Verveke - Vicomtech, Paseo de Mikeletegi 57, 20009 San Sebastian, Spain. Corresponding author: fboto@vicomtech.org.

C. Callol, A. Letamendi and I. Ibarbia - Biobide, Paseo de Mikeletegi 58, 20009 San Sebastian, Spain. Corresponding author: ibarbia@biobide.es

978-1-4244-5379-5/09/$26.00 ©2009 IEEE

of nearly 78% using 8,566 ISH images of [1].

In [8], some more processing tasks are described - differentiating between dorsal and lateral images, and aligning anterior and posterior ends of the images. They propose to employ the curvature factor of the convex hull edges to differentiate between the dorsal and the lateral views, as well as between the anterior and posterior ends of the image. Such a methodology was tested on 30 images of [1] and a correct classification rate of 70% was obtained for lateral images, while 60% was obtained for dorsal images. The algorithm proposed for aligning the anterior and posterior ends was successful in aligning 85% of 40 images.

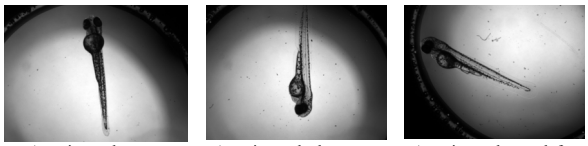
While the previous work deals with Drosophila embryo, [9] address the case of Drosophila imaginal discs. Such structures have significant intra and inter-class variability in size, shape and stain patterns, and it is difficult to make any parametric or model based assumptions. [9] proposes a method based on unsupervised shape learning algorithms.

In the case of zebrafish embryo, the standardisation problem has never been addressed. Zebrafish embryos have different shape of drosophilia embryos, as shown in Figure 1. Shape representation and extracted features should therefore be specifically defined for zebrafish embryos.

III. ZEBRAFISH EMBRYO SHAPE CHARACTERISTICS

[10] describes a series of stages for development of the embryo of the zebrafish. The stages are based on morphological features, identified by examination of the live embryo with the microscope. From day 2 post-fertilization (dpf), once escaped the chorion, the embryo presents a unique and rigid 3D representation (Figure 2). When the embryo lies on the bottom of a plate, its movements are restricted - the embryo can only move in the plane of the plate, and rotate along the two axes parallel to the plate.

Thanks to these characteristics, the microscopic images present a limited number of views. For each image, the expert characterizes the position of the embryo, using standard directional terms commonly used in biological sciences for organisms (Figure 2). When building a zebrafish database, it is important to differentiate between lateral and ventrodorsal images, because these views show different gene expression patterns.



Anterior to the top - Ventral. | Anterior to the bottom – Left. | Anterior to the top-left - Right.

Fig. 2 Some examples of zebrafish embryo images. Associated orientation / position using standard directional terms.

We use microscopic images of transgenic zebrafish at 2dpf expressing green fluorescent protein (GFP) in the myocardium. The view is arbitrary. Each image contains the brightfield image of the whole embryo superimposed by the fluorescent expression of the GFP (Figure 3).

IV. METHODOLOGY

The automatic standardisation of the zebrafish images was performed using custom software including the main following processing tasks: isolating the embryo from the background and recognising the embryo orientation using shape representation and classification.

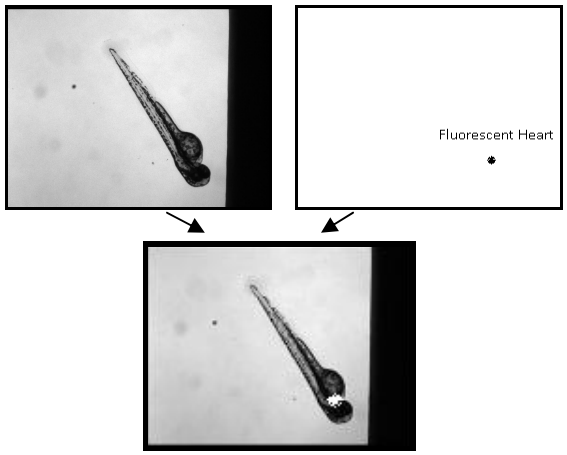


Fig. 3 Image pair – Top left: brightfield image of embryo. Top right: Fluorescent image of the same embryo; only the heart is expressing. Bottom: superimposed images of the embryo.

A. Shape representation

In order to recognize the embryo orientation, the image of the embryo has to be described or represented by certain features. Since shape is one of the primary low level features used in image classification, shape representation is a fundamental issue in this type of applications [11]. The main objective of shape description is to measure geometric attributes of an object that can be used for classifying, matching and recognizing the object. The main tasks of shape representation are the segmentation of the object shape from the background, and the choice of the description features. In our application, we chose to use the convex hull of the extracted shape, which is known to be an efficient shape representation [8].

1) Pre-Processing

Our images are facing uneven illumination at the edges of the image (Figures 2, 3), a common problem in microscopy irrespective of the type of camera and method of microscope attachment. This may be attributed to multiple factors from the illumination filament, the design of the light path between the camera and the microscope, or the behaviour of the imaging device.

White shading correction is traditionally used to equalise uneven background illumination. Unfortunately, a white shading correction image must be acquired for each objective magnification and recalculated if the microscope illumination settings are altered. Such method should be prohibited in a HTS context, where equipment is subject to important variations, and might rapidly suffer from

calibration deviation.

White shading correction is traditionally used to equalise uneven background illumination. Unfortunately, a white shading correction image must be acquired for each objective magnification and recalculated if the microscope illumination settings are altered. Such method should be prohibited in a HTS context, where equipment is subject to important variations, and might rapidly suffer from calibration deviation.

We therefore use the alternative shading correction method proposed by [12] that is based upon the intrinsic properties of the image, and that does not require a reference white image. Instead, it extracts the underlying illumination pattern and uses this to correct the image.

The image is first convolved with a Gaussian kernel filter. The resulting image is then inverted, merged with the original image, and normalized (Figure 4).

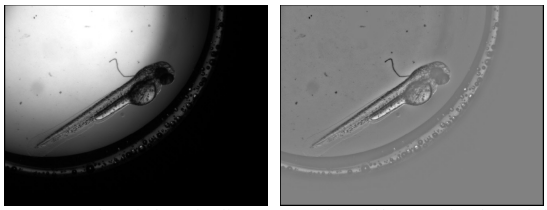


Fig. 4 Example of the Shading Correction Filter. Left: the original image; Right: the processed image

2) Segmentation

We use a seeded region growing algorithm to extract the shape of the embryo from the pre-processed brightfield image. More specifically, the brightfield image is first severely thresholded to obtain binary group of pixels or subregions (BLOB) for the region growing procedure, each BLOB is defined by its contour. Those BLOB are expected to be mainly part of the embryo shape. We use the Otsu's method [13], which finds a threshold on statistical basis and allows to threshold the images automatically without a priori information.

Such method requires the seed as additional input. This is where we use the fluorescent image of the GFP expression. If the microscope is properly calibrated and the two images are correctly registered, the fluorescence will always be located inside the embryo and can be used as the starting region of the algorithm. We convert the fluorescent image to a binary image using also the Otsu's method and the mayor BLOB is considered as the seed to start processing the region growing. The *convex hull* is computed for the first seed BLOB and is updated at each iteration of the process, to obtain the minimal convex polygon structure.

Starting with the seed convex polygon, this polygon is iteratively grown by comparing all unexplored contour BLOBs. The distance between the BLOB contour (point by point) and the convex hull (line by line) is used as criteria to join the BLOB to the final region. At each step the convex hull is recomputed and the process is repeated iteratively. The values of the stopping criteria have been defined empirically. The process continues until all BLOBs satisfying the criteria are allocated to the region.

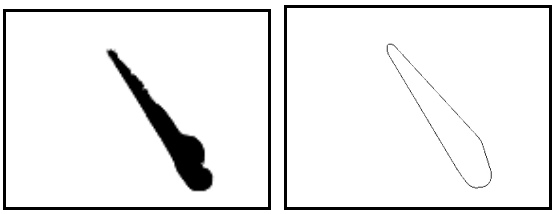


Fig. 5– Shape representation extraction. Left: binary mask obtained from segmentation procedure; Right: corresponding convex hull.

3) Quality test

Images presenting artefacts, noise, dust, strong illumination etc... and cannot be classified even by an expert should be rejected. A quality test was integrated into the segmentation process. The method rejects the segmentation process when the Otsu's binarization process produces large BLOBS and image is qualified as noisy and the classification process is stopped.

B. Features extraction

The next step is to represent the shape boundary information for efficient classification into lateral o dorsal/ventral view. We started from the observation that the convex hull is nearly symmetric in the dorsal view, whereas this is not the case in the lateral view. We proposed to employ the symmetry feature expressed in the form of curvature of the edge to differentiate between the dorsal and the lateral views. In our application the images produced highly similar anatomical shapes. In order to increase robustness of the classifier, topological shape features were added, using distances between the shape boundary and the seed location (Figure 6). A decision rule of the classifier was then empirically developed, making use of the set of extracted features to decide on the orientation of the embryo.

C. Normalization

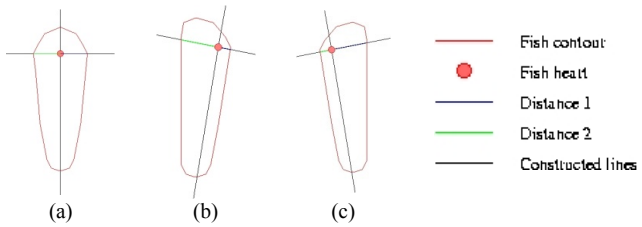


Fig..6 - Schemas of the analysis of the shape of the embryo. (a) : "Ventral only" position; (b) : "Lateral left" position ; (c) : "Lateral right" position..

The initial image is rotated using the orientation angle of the extracted shape with the horizontal – the principal axis of the convex hull is considered as the anterior-posterior axis of the embryo.

V. Results and discussion

We used a database of 228 image pairs (brightfield and fluorescent images). The images were acquired arbitrarily in 9 sessions, with different lighting parameters and specimens. No special consideration for avoiding artefacts was taken. Each brightfield image was classified manually by expert: ventrodorsal or lateral left/right. The results are summarised in the next Table.

TABLE I
SUMMARY OF RESULTS

Rejected Images	Lateral	Dorsal/Ventral	Both
54%	99%	45%	93%

We first evaluated the efficiency of our quality assessment methodology. 54% of the images were rejected. We visually inspected those images. Most of them are clearly low-quality images, so that it is almost impossible to segment the embryo, even manually. Our procedure is successfully assessing image quality and selecting images suitable for classification.

Among the images which were accepted for orientation classification, a success rate of 99% was obtained for lateral images, while 45% was obtained for dorsal images. Overall a success rate of 93% was reached. The success difference between lateral and dorsal is probably due to the fact that the zebrafish embryo tends to lie on its lateral side, and the amount of images showing lateral embryos is much higher than the ones showing dorsal embryos (in our experiment, the 228 images include 201 lateral and 27 dorsal).

Our procedure proposed for aligning the anterior and posterior ends was successful in aligning 100% of the 105 retained images (independently of whether it was successfully classified). In [8], the algorithm achieves a successful alignment for 85% of 40 images. Some of the normalized embryos are shown in Figure 7.

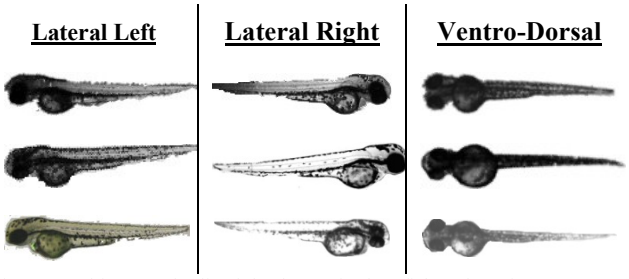


Fig. 7 Resulting standardized database of zebrasfish embryo images.

Although both methods are based on a convex-hull shape representation, the morphologies of drosopholia and zebrafish embryos are different, and the results can therefore not being directly compared. The Drosophila shape is nearly symmetric along the anterior-posterior axis as well as the ventral-dorsal axis, which makes it more difficult to extract significant features, and can explain a lower classification rate than for the zebrafish embryo.

VI. Conclusion

We have developed the first automatic standardization methodology for microscopic images of transgenic zebrafish at 2dpf.

Such problem can become particularly challenging, as the images may be contaminated by a variety of noise sources, due to illumination conditions, presence of foreign elements in the image, shadows… We propose to deal with noise by defining discriminating features of noisy images. Our method performs well, as the images were successfully rejected when unsuitable for analysis.

We then isolate the embryo from the background and recognizing the embryo orientation using shape representation and classification. A success rate of 93% was reached, compared to a 70% in early work for Drosophilia embryos. We are then able to align the images along the anterior and posterior ends.

The problem of standardizing zebrasfish embryos images has never been tackled. However, the high successful rate obtained in this first study is encouraging, and can drastically reduce expert effort in aligning images for spatial gene expression comparison, although double check is still required.

References

[1] Berkeley Drosophila Genome Project http://www.fruitfly.org
[2] GEISHA Gallus database http://geisha.arizona.edu/geisha/
[3] ZFIN The Zebrafish Database Project http://zfin.org.
[4] S. Kumara, K. Jayaraman, S. Panchanathana, and R. Gurunathana. "BEST: A Novel Computational Approach for Comparing Gene Expression Patterns From Early Stages of Drosophila melanogaster Development," *Genetics*, vol. 162, pp. 2037-2047, Dec 2002.
[5] H. Peng, and E.W. Myers, "Comparing In situ mRNA Expression Patterns of Drosophila Embryos," *8th Conf. on Computational Molecular Biology*, pp. 157-166, April 2004.
[6] H. C. Peng, F. H. Long, J. Zhou, G. Leung, M.B. Eisen, and E.W. Myers. "Automatic image analysis for gene expression patterns of fly embryos," *BMC Cell Biology*, vol. 8, no. sup. 1, July 2007.
[7] J. Y. Pan, A. Balan, E. P. Xing, A. Traina, and C. Faloutsos, "Automatic mining of fruit fly embryo images," *Proc 12th ACM SIGKDD*. 2006.
[8] P. Kuchi, S. Kumar, and S. Panchanathan, "Classification of Stained Embryonic Images of Drosophila," *Signal and Image Processing (SIP)*, August 2003.
[9] P. Ahammad, C. L. Harmon, A. Hammonds, S. Shankar, and G. M. Rubin, "Joint nonparametric alignment for analyzing spatial gene expression patterns in Drosophila imaginal discs," *Proc IEEE CVPR*, pp. 755–760, 2005.
[10] C. B. Kimmel, W.W. Ballard, S. R. Kimmel, B. Ullmann, and T.F. Schilling, "Stages of embryonic development of the zebrafish," *Developmental Dynamics*, vol. 203, no. 3, pp. 253-310, July 1995.
[11] M. João Fonseca, A. Ferreira, and J. A. Jorge, "Generic Shape Classification for Retrieval," *Proc. of the 6th Int. Workshop on Graphics Recognition*, pp. 291-299. 2006
[12] F. J. Leong, M. Brady, and J. McGee, "Correction of uneven illumination (vignetting) in digital microscopy images," *Journal of Clinical Pathology*, vol. 56, no. 8, pp. 619-621, 2003.
[13] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Transactions on System Man Cybernetics*, Vol. SMC-9, no. 1, pp. 62-66, 1979.