Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2013 ; : 902–905. doi:10.1109/ISBI.2013.6556621.

AUTOMATED QUANTIFICATION OF ZEBRAFISH TAIL DEFORMATION FOR HIGH-THROUGHPUT DRUG SCREENING

Omer Ishaq^{*}, Joseph Negri[†], Mark-Anthony Bray[†], Alexandra Pacureanu^{*}, Randall T. Peterson^{†,‡}, and Carolina Wählby^{*,†}

^{*}Centre for Image Analysis/SciLifeLab, Uppsala University, Sweden

[†]Broad Institute of Harvard and MIT, Cambridge, MA

[‡]Massachusetts General Hospital, Harvard Medical School, Department of Medicine, Charlestown, MA

Abstract

Zebrafish (*Danio rerio*) is an important vertebrate model organism in biomedical research thanks to its ease of handling and translucent body, enabling *in vivo* imaging. Zebrafish embryos undergo spinal deformation upon exposure to chemical agents that inhibit DNA repair. Automated image-based quantification of spine deformation is therefore attractive for whole-organism based assays for use in early-phase drug discovery. We propose an automated method for accurate high-throughput measurement of tail deformations in multi-fish micro-plate wells. The method generates refined medial representations of partial tail-segments. Subsequently, these disjoint segments are analyzed and fused to generate complete tails. Based on estimated tail curvatures we reach a classification accuracy of 91% on individual animals as compared to known control treatment. This accuracy is increased to 95% when combining scores for fish in the same well.

Index Terms

Curvature extraction; high-throughput screening; quantitative microscopy; zebrafish (*Danio rerio*)

1. INTRODUCTION

High-throughput screening (HTS) is a technique for searching large libraries of chemicals to better understand treatment of disease and disease pathways. Quantitative microscopy has emerged as one of the most powerful and informative ways to analyze cell-based screens. However, many diseases and biological pathways can be better studied in whole animals — particularly diseases that involve organ systems and multicellular interactions.

Zebrafish (*Danio rerio*) is an important model organism in many fields of research, including organismal development, cancer, and neurobiology [1, 2, 3]. Many human disease models have been established in zebrafish, and this along with its translucent body makes it an attractive candidate for large-scale whole-organism image-based assays for early-phase drug discovery campaigns and for identifying biologically active compounds [4, 5]. However, a current limitation is the availability of widely accessible high-throughput image analysis platforms. Such platforms will improve the data analysis throughput, robustness of phenotype scoring and reliability of statistical metrics for evaluating assay performance.

In this study we focus on drug-induced neuronal damage, such as in response to Camptothecin (Cpt). This is expressed as a severe curling of the fish tails due to the inhibition of the DNA repair [6]. Methods for accurate, robust and automated quantification of this gross deformation, in microscopic images, would be beneficial to a variety of high-

content screens in zebrafish. However, the zebrafish larval tails are almost transparent which confounds their delineation and subsequent curvature extraction. In addition, the task is significantly impeded by the presence of multiple, sometimes overlapping fish in each micro-plate well. On the other hand, placing multiple fish in each well increases the statistical value of the screen while limiting the use of valuable chemical libraries.

Previous work on image-based analysis of zebrafish has either focused on analysis of structure and function at the cellular level [7] or has been limited to analysis of single-fish wells [8, 9]. Previous analysis of multi-fish wells has been constrained to non-overlapping fish [10]. To the best of our knowledge, the method we propose here is the first to cope with zebrafish curvature extraction from multi-fish wells with overlapping fish.

2. METHOD

Typically, object feature extraction is handled as a sequential pipeline comprising object identification, subsequent delineation and finally feature extraction. A drawback of such an approach is the propagation and compounding of errors through the pipeline. To address this we propose a different, yet simple, bottom-up approach requiring limited parameter tuning. The method has four steps: (i) Image pre-processing and foreground/background segmentation (Sec. 2.1); (ii) Generation of branch-free refined medial axes for all “tail-like” foreground regions (Sec. 2.2); (iii) Geometry-based selection and fusion of these candidate tail-segments into complete fish tails (Sec. 2.3); (iv) Computation of the tail curvature (Sec. 2.4). In other words, we are not trying to explicitly identify individual fish, but search directly for the tails on which we make feature measurements.

2.1. Pre-processing

As shown in Figure 1a, the microscopy images suffer from inhomogeneous illumination. This inhomogeneity is estimated through a convex-hull approach and corrected through pixel-wise subtraction (Fig. 1b) [11]. Subsequently, the image is grayscale inverted to yield bright fish on a dark background. Next, we apply Gaussian smoothing and Otsu thresholding to yield the fish as the foreground of a binary image (Fig. 1c).

2.2. Refined medial axis generation

Next, we want to extract the medial axis for each “tail-like” foreground segment (i.e., elongated foreground region with upper and lower bounds on its width), as shown in Fig. 1e. Blum’s Medial axis transformation (MAT) [12] often results in branched skeletons requiring pruning. Instead, we use a modified branch-free approach with variable smoothness and sub-pixel resolution. For clarity we will henceforth refer to this approach as the refined medial axis (RMA), in contrast, we will refer to the output of a traditional MAT as the medial axis (MA). For details of RMA generation and propagation please refer to our previous work [13]. However, our previous work was constrained to a single RMA per image; therefore, it utilized a naive seeding strategy for RMA initialization. At present, we aim to measure the curvature of multiple objects in an image; therefore, we propose an efficient multi-seeding strategy. First, we compute a MAT of the image foreground (green curves in Fig. 1d) and select a random set of seed-points situated on these MAs (red markers in Fig. 1d). Subsequently, we remove a subset of these seed-points to decrease the computational complexity of RMA generation in a two-step process: (i) Any seed which occurs close to another seed is removed to ensure a more uniform spatial distribution of the seed-points; (ii) Any seed which occurs close to an already generated RMA is removed as it is likely to result in the same RMA. Each RMA consists of a set of ordered 2D points, each point P_k encodes both the position and the local width: $P_k = \{x, y, w\}$. An RMA terminates when either one of these three conditions is satisfied: (i) Intersection with a foreground boundary; (ii) Increase

in the object width beyond a threshold; (iii) An increase in the width of a tapering object (this typically occurs when a narrow tail is occluded by a wider yolk-ball).

2.3. RMA fusion

As mentioned earlier, an RMA may represent only a part of a zebrafish tail; therefore, multiple RMAs may need to be joined/fused together (i.e., through interpolation) to yield a complete tail. Before the fusion, the segments (i.e., represented by the RMAs) shorter than a lower threshold are removed and segments with length situated in an intermediate interval are retained only if they can be fused with longer segments. A fusion occurs when two physically disjoint RMAs are almost collinear (i.e., similar slopes and intercepts) and situated close to each other. RMA fusion is instrumental for recovering the tail curvatures for fish that overlay, and therefore occlude, each other. Figure 1f displays an example of fusion: the red lines represent tail-segments fused together to yield the complete zebrafish tails (shown in yellow).

2.4. Curvature extraction

An RMA comprises an ordered set of 2D points and is regularized through B-spline fitting. The RMA curvature is defined as the integral of the local curvatures along its length. We calculate these local curvatures for each triplet of neighboring points, that is, we compute $n - 2$ local curvatures for a RMA comprising n points. Given any three neighboring points P , C and N (i.e., the previous, the current and the next point), a pair of vectors V_1 and V_2 are generated as the difference of the corresponding components of the point pairs $P-C$ and $C-N$ (Fig. 2a). The angle α between V_1 and V_2 (i.e., the local curvature at C) is calculated as,

$$\alpha = \arccos \frac{\langle \vec{V}_1, \vec{V}_2 \rangle}{\|\vec{V}_1\| \cdot \|\vec{V}_2\|}. \quad (1)$$

The RMA direction (i.e., either left or right curving) is determined through the sign of the normal-component (i.e., the $\hat{z}$ component) of the cross product $V_1 \times V_2$, defined as,

$$\begin{vmatrix} c_y - p_y & 0 \\ n_y - p_y & 0 \end{vmatrix} \hat{x} - \begin{vmatrix} c_x - p_x & 0 \\ n_x - c_x & 0 \end{vmatrix} \hat{y} + \begin{vmatrix} c_x - p_x & c_y - p_y \\ n_x - c_x & n_y - c_y \end{vmatrix} \hat{z}. \quad (2)$$

We use the fan-beam width [14] as a descriptor for overall tail-bend. The fan-beam width is a measure of the plane angle (Fig. 2b) and is measured with reference to the anterior-most point on the RMA. We distinguish between the anterior and the posterior ends of a tail-segment through analysis of its width; an anterior end is wider than a posterior one.

3. RESULTS

3.1. Image data

The image data consists of two replicates (i.e., data sets) containing 67 and 72 images, respectively. The images were acquired through bright-field microscopy in round-well 96-well plates with typically 5 zebrafish per well. In the first data set, 24 wells were untreated (i.e., negative control), remaining groups of 19 and 24 wells were treated with 100 nM and 200 nM of Cpt, respectively. The second data set comprised 36 wells treated with 100 nM Cpt and 36 untreated wells.

3.2. Classification of the micro-plate wells

Figure 3 presents the tail curvatures for the two datasets. The curvatures are presented on a per-fish basis (as opposed to a per-well approach), that is, each datapoint in a curvature distribution represents a single fish. The medians and the 25th and the 75th percentile cut-offs are shown.

We classify the micro-plate wells as either treated (i.e., we consider the 100 and 200 nM treated fish as belonging to a joint superclass) or control. We use four different classifiers: (i) Naive Bayesian; (ii) Support vector machines (SVMs) with a linear kernel; (iii) SVMs with radial basis functions (RBFs); (iv) AdaBoost ensemble classifier. For each zebrafish, we use a four element feature vector comprising tail curvature, normalized tail curvature (normalized by the tail length), fan-beam width and normalized fanbeam width.

The micro-plate wells were classified on both per-well and per-fish basis: (i) In the per-well approach, the feature vectors for all fish within a well are added (i.e., element-wise addition) to yield a resultant feature vector for the well which is then used for classification; (ii) In the per-fish approach, each fish in a well is classified independently on the basis of its feature vector. Subsequently, the classification labels for the fish in a well are combined in a simple voting scheme to generate a classification label for the well. All system parameters were tuned on one of the two data sets, while tested on the other. The same approach was used for classifier training.

We observe a maximum 95% per-well and 91% per-fish classification accuracy using RBFs-based SVMs; the best out of the four tested classifiers. Fig. 4 shows a set of in/correctly classified wells using the per-fish approach. Observe that some misclassifications (e.g., Fig. 4d) result from debris in the micro-plate wells.

4. CONCLUSIONS

We have proposed an automatic and robust method for high-throughput extraction of the zebrafish tail curvature in multi-fish wells. It is a general algorithm and relatively independent of the number and posture of the fish in the well. Our results indicate class separation between the curvatures for the un/treated zebrafish. Future work includes the application on larger datasets. Some steps of the analysis pipeline have been implemented as part of the open source CellProfiler project (www.cellprofiler.org). We aim to make the full algorithm available to the zebrafish screening community.

Acknowledgments

This work was supported in part by grant number R01 GM095672 from NIH and SciLifLab Sweden.

References

1. Lieschke G, Currie P. Animal models of human disease: zebrafish swim into view. *Nat Rev Genet.* 2007; 8(5):353–367. [PubMed: 17440532]
2. Feitsma H, Cuppen E. Zebrafish as a cancer model. *Mol Cancer Res.* 2008; 6:685–694. [PubMed: 18505914]
3. Rinkwitz S, Mourrainb P, Beckera T. Zebrafish: An integrative system for neurogenomics and neurosciences. *Prog Neurobiol.* 2011; 93(2):231–243. [PubMed: 21130139]
4. Ignatius M, Langenau D. Fluorescent imaging of cancer in zebrafish. *Method Cell Biol.* 2011; 105:437–459.
5. Peal D, Peterson R, Milan D. Small molecule screening in zebrafish. *J Cardiovasc Transl Res.* 2010; 3:454–460. [PubMed: 20680709]

6. Langheinrich U, Hennen E, Stott G, Vacun G. Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling. *Curr Biol*. 2002; 12(23):2023–2028. [PubMed: 12477391]
7. Megason, SG. Current challenges in image analysis for in toto imaging of zebrafish. *Biomedical Imaging: From Nano to Macro*, 2011 IEEE International Symposium on; 2011. p. 1730-1732.
8. Tianming L. A quantitative zebrafish phenotyping tool for developmental biology and disease modeling. *IEEE Signal Proc Mag*. 2007; 24(1):126–129.
9. Wu T, Lu J, Lu Y, Yang J. Research on segmentation of dorsal diencephalon and ventral midbrain of zebrafish embryo based on active contour model. *Int J Comput Biol Drug Des*. 2012; 5(1):3–15. [PubMed: 22436295]
10. Burgess HA, Granato M. Modulation of locomotor activity in larval zebrafish during light adaptation. *J Exp Biol*. 2007; 210:2526–2539. [PubMed: 17601957]
11. Wählby C, Kamensky L, Liu ZH, Riklin-Raviv T, Conery AL, O'Rourke EJ, Sokolnicki KL, Visvikis O, Ljosa V, Irazoqui JE, Golland P, Ruvkun G, Ausubel FM, Carpenter AE. An image analysis toolbox for high-throughput *C. elegans* assays. *Nat Methods*. 2012; 9(7):714–716. [PubMed: 22522656]
12. Blum, H. *Models for the Perception of Speech and Visual Form*. MIT Press; 1967. A transformation for extracting new descriptors of shape; p. 362-380.
13. Wells D, French A, Naeem A, Ishaq O, Traini R, Hijazi H, Bennett M, Pridmore T. Recovering the dynamics of root growth and development using novel image acquisition and analysis methods. *Phil Trans R Soc B*. 2012; 367:1517–1524. [PubMed: 22527394]
14. Bracewell RN, Riddle AC. Inversion of fan-beam scans in radio astronomy. *Astrophysical Journal*. 1967; 150:427–434.

Fig. 1. Steps for curvature extraction: (a) An input image; (b) After illumination correction; (c) Binary image after smoothing and thresholding; (d) Computed medial axes (highlighted in green) and seed-points (highlighted in red); (e) Refined medial axes (highlighted in yellow); (f) Medial axis fusion: the red lines represent tail-segments fused together to yield the complete tails (shown in yellow).

Fig. 2.

(a) Curvature estimation for three neighboring points. (b) Fan-beam width for a zebrafish tail (i.e., measured with reference to the anterior-most point on the RMA).

Fig. 3. Medians and the 25th and the 75th percentile curvatures for the two replicates R_1 and R_2 , where M = number of fish.

Fig. 4.

Three correctly classified (images a–c) and three incorrectly classified (d–f) wells. Fish with red and blue tracks have been classified as treated and untreated, respectively. The image borders have been color-coded according to the groundtruth; red and blue borders represent treated and untreated wells, respectively.