

Markus Reischl*, Andreas Bartschat, Urban Liebel, Jochen Gehrig, Ference Müller and Ralf Mikut

ZebrafishMiner: an open source software for interactive evaluation of domain-specific fluorescence in zebrafish

Abstract: High-throughput microscopy makes it possible to observe the morphology of zebrafish on large scale to quantify genetic, toxic or drug effects. The image acquisition is done by automated microscopy, images are evaluated automatically by image processing pipelines, tailored specifically to the requirements of the scientific question. The transfer of such algorithms to other projects, however, is complex due to missing guidelines and lack of mathematical or programming knowledge. In this work, we implement an image processing pipeline for automatic fluorescence quantification in user-defined domains of zebrafish embryos and larvae of different age. The pipeline is capable of detecting embryos and larvae in image stacks and quantifying domain activity. To make this protocol available to the community, we developed an open source software package called „ZebrafishMiner“ which guides the user through all steps of the processing pipeline and makes the algorithms available and easy to handle. We implemented all routines in an MATLAB-based graphical user interface (GUI) that gives the user control over all image processing parameters. The software is shipped with a manual of 30 pages and three tutorial datasets, which guide the user through the manual step by step. It can be downloaded at <https://sourceforge.net/projects/scixminer/>.

Keywords: high throughput, zebrafish, image processing

<https://doi.org/10.1515/cdbme-2017-0042>

1 Motivation

The zebrafish (*Danio rerio*) has several features, for instance, its morphological and physiological similarity to mammals, its fertility, its transparency, and its ex utero development, making it a good model organism for genetic or toxicological high-throughput screens [1]. The demand for automated methods to evaluate microscopic images from zebrafish larvae is steadily increasing [2]. Image processing pipelines are mostly tailored to the scientific problem to be solved. Easy-to-use interfaces are hardly found, because mostly the biological question can be answered without the software being intuitive to further users. Consequently, the related software is not available or at least difficult to use for other groups and similar projects starts with a reinvention or reimplementing of the algorithms.

In [3], we proposed a robust framework for zebrafish analysis being fault-tolerant and self-adaptive. Its aim was to assign signals in a fluorescent channel to eight predefined domains in zebrafish (corresponding to the tissues notochord, spinal cord, heart, skin, eye, brain, midbrain-hindbrain, yolk). The latter are annotated in a so-called reference embryo which is a stereotype automatically generated by the software based on expert markings. To assign the fluorescent signal we used a brightfield and a reference fluorescent channel for segmentation, both consisting of multiple z-slices. The pipeline consists of the detection of the embryo in the image, the merging of all focus planes (extended focus), the adaptation of mathematical models to the embryo outline, an automatic check for plausibility, a warping of each fish to the reference and the quantification of domain activity. Its main limitations have been the fix definition of domains at a defined age of 32 hours and 48 hours post fertilization (hpf). Details and parameters are given in [3].

*Corresponding author: **Markus Reischl:** Institute for Applied Computer Science, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany, e-mail: markus.reischl@kit.edu

Andreas Bartschat, Ralf Mikut: Institute for Applied Computer Science, Karlsruhe Institute of Technology, Hermann-von-Helmholtz Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

Urban Liebel: 32dots GmbH & CO KG, Breitenbachstraße 37, 69234 Dielheim, Germany

Jochen Gehrig: ACQUIFER, Division of DITABIS, Digital Biomedical Imaging Systems AG, 75179 Pforzheim, Germany

Ference Müller: Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, University of Birmingham, B15 2TT, United Kingdom

This type of mathematical model-based operations can easily be implemented in MATLAB-based environments. This is a main advantage in contrast to other image processing frameworks that became popular for biological image processing in the last few years (see survey in [4]).

Especially, the identification of tissue domains (e.g. notochord, brain etc.) in zebrafish images and the assignment of fluorescent signals to a domain is not yet implemented in a standardized software package. Furthermore, the generalization of information of the same domains in different larvae is missing. There is no instruction about how a consistent image processing pipeline can be established and parameterized in a proper manner.

The main obstacles for the standardization of a computer vision routine are embryos and larvae of different ages, which have to be adjusted, dirt on the objectives or errors in automatic image acquisition, variations in injections (different brightness, etc.) and statistical inadequacy regarding alignment, domain detection and quantification.

In this article, we present a MATLAB-based software toolbox called ZebrafishMiner, which is able to quantify fluorescence in zebrafish domains. It guides the user through the processing pipeline by a set of easy to handle buttons in a Graphical User Interface (GUI). The user is able to apply the pipeline to a set of images or to adapt the software to its own needs by new references. The most important parameters can be set in the interface. Furthermore, the software package contains a manual and demonstrates all procedures by three tutorial examples consisting of real-world examples.

2 Material and methods

The ZebrafishMiner is built as an extension package for the MATLAB Toolbox SciXMiner [5], including manuals, benchmark data sets and further examples. Both packages can be downloaded at <https://sourceforge.net/projects/scixminer/>.

To use the ZebrafishMiner, it can be activated in the extension package manager (Extras - Choose application-specific extension packages). After a restart of SciXMiner, a new item “Zebrafishminer” is contained in the menu bar. The ZebrafishMiner is started by applying the menu item „Initialize”. The GUI switches to the start screen of the ZebrafishMiner.

The user interface shows two drop-down menus, two edit fields for paths and a set of buttons. The first drop-down menu specifies the ZebrafishMiner-GUI within SciXMiner and needs not to be changed. The second one specifies the mode of the ZebrafishMiner. Three modes are possible:

„Fluorescence evaluation“, „Reference creation” and „Parameter settings“. Each mode consists of an independent user interface.

Fig. 1 shows the graphical user interface (GUI) of the ZebrafishMiner as a part of the open source software package SciXMiner licensed under GNU-GPL. It consists of three main windows: Fluorescence evaluation, Reference creation and Parameter. ZebrafishMiner enables the user to comfortably browse through the functionality of all image processing pipelines. If deeper understanding is unwanted, all pipelines provide a „Do all“-button, which operate with a standard set of parameters and functions. The interfaces enable the user to evaluate fluorescence in a set of larvae with a given reference embryo or to generate a reference on his/her own.

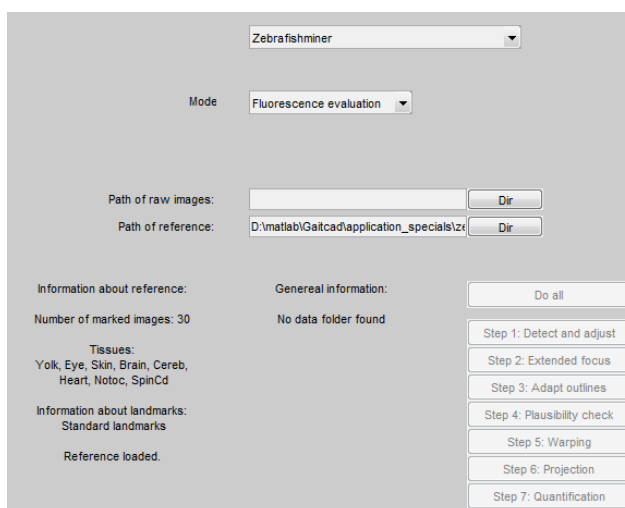


Figure 1: Graphical user interface for fluorescence evaluation

The automatic „Fluorescence evaluation” detects embryos in a stack of brightfield and fluorescent images and compares them to a reference embryo with known/marked domains. As an output, fluorescent activity in each domain is quantified. The software is explicitly not designed to morphology screens but for detection of changes of expression activity of reporters. A standard reference embryo aged 32 hpf is integrated in the software and is contained in a SciXMiner-subfolder.

If the ZebrafishMiner is started in the „Fluorescence evaluation” mode, data of raw images needs to be selected. The data is expected to be customized according to the Olympus microscopy file format, containing a number n of z -slices in three different modalities. The three modalities need to show a brightfield, a fluorescent reference image and a fluorescent readout image, with the endings for brightfield („-bf.tif“), a first fluorescent channel (here: cyan fluorescent

protein CFP) „-cfp.tif“, and a second fluorescent channel (here: yellow fluorescent protein YFP) „-yfp.tif“, altogether 3n images per embryo are expected - alternative modalities can be set in the parameter dialog.

To analyse a zebrafish embryo, seven steps can be executed one- by-one using buttons:

1. „Detect and adjust“ reads data, removes the dirt not connected to the larva, applies several filter routines and uses morphological operators followed by a detection of the biggest object in the image. The larva will be rotated such that the head is located left the tail upside. After cutting the image, the detection is complete. All results are stored in a folder called „extr“.
2. „Extended focus“ merges the stacked images to one image with maximum sharpness based on a wavelet transformation. The routine is applied to brightfield, CFP and YFP images independently. All results are stored in a folder called „efoc“.
3. „Adaptation of outlines“ detects characteristic points (landmarks, anchor points) in the images and adapts a mathematical function to the border of the larva consisting of geometric elements such as circles and straight lines. All results are stored in a folder called „regr“.
4. „Plausibility check“ is a control step. Based on a classifier generated in the reference procedure (see Section „Automatic referencing“), all embryos are evaluated using a priori knowledge about morphology constraints. Embryos not fitting the features are excluded to avoid a biased outcome.
5. „Warping“ uses the remaining landmarks to project all images contained in the „efoc“-folder onto the landmarks of the reference. The results are stored in the „proj“-folder.
6. „Maximum projection“ calculates an overlay image of all warped CFP and YFP images by maximum projection, to show the overall outcome within one single image.
7. „Quantification“ generates a table assigning values to the fluorescence of each domain.

All processing steps are stored in separate subfolders, to easily follow processing results and error messages.

To extend embryo evaluation to those younger or older than 32hpf, ZebrafishMiner offers the functionality „Reference Creation“ to interactively replace the reference larva. To adapt the ZebrafishMiner to a new class of images, three steps have to be modified: domains of interest have to be defined, the landmark calculation needs to be adapted and the plausibility check has to be updated with examples.

Domains can either be defined manually (drop down menu: „Reference by manual assignment“) or automatically by fluorescently labeled domains (drop down menu: „Reference by Fluorescence“). Both ways are implemented guiding the user step-by-step through the pipeline.

„Reference by manual assignment“ provides the user with a painting device called „Tissue manager“ to colour a tissue in embryo images (Fig. 2). It is possible to add and delete tissue domains, as well as to save the current selection in a file or to load a saved combination of domains. The mode „Reference by Fluorescence“ defines domains by imported raw data files. Here, the data files need to show fluorescence in the tissues which shall be added. In each mode, up to eight tissue domains can be imported. In both cases, a path to the raw data and a path to the new reference needs to be set.

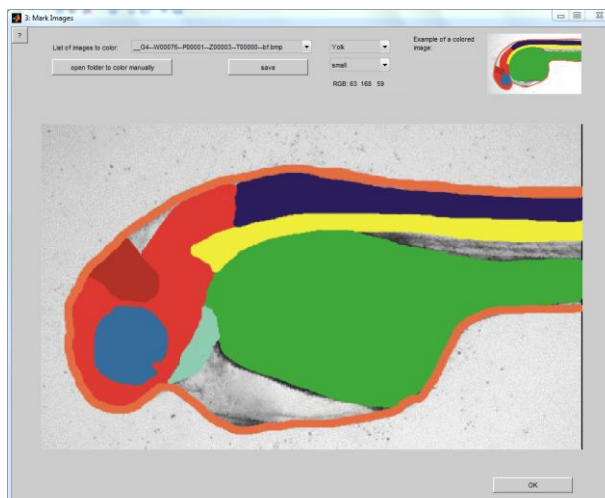


Figure 2: Tissue manager

To replace the standard reference, five steps are necessary. They can be executed stepwise using the buttons in the GUI. Alternatively, the button „Do all“ applies the complete algorithm to the selected data:

1. „Generate landmarks“ computes the first three steps of the zebrafish image processing („Detect and adjust“, „Extended focus“, „Adapt outlines“).
2. „Select image“ opens a new window, which gives the user the possibility to select a representative set of images and thus exclude erroneous processing results from further processing.
3. By pushing the button „Plausibility check“, a classifier automatically generates a list of parameters that describes a well-processed image. All images not fitting within this scheme are excluded from being processed with the new reference larva.

4. In the case of „Reference by manual assignment“, the button „Select images to mark“ lets the user choose larvae to mark - domains may then be defined in the „Tissue manager“.
5. The button „Create Reference“ calculates the average and the deviation of all landmarks to create the landmarks of the reference. All domains of all chosen larvae are compared to create an average for each domain. These combinations of landmarks and domains build the new reference.

In case that the new reference does not match the user's preferences, there are also various possibilities to adapt image processing and landmarking. An easy implementation is given by the two buttons „Edit binary algorithm“ and „Reference preview“. The first one opens a source code editor where algorithms can easily be manipulated in MATLAB, the second shows the result of the binary image calculation of each domain. With this button, the binary image calculation can be checked. If further manipulation is needed, all source code files provide open access for modification.

The third option in the drop down menu („parameter handling“) opens a GUI where parameters of image processing and the user interface can be changed. E.g., there are possibilities to define pixel thresholds, to generate figures of all processing steps, to switch between automatic and manual z-slice focus, to determine the size of the reference larva, to manipulate normalization and filtering algorithms, etc. However, the setting of the parameters may require deeper knowledge of image processing.

3 Results

ZebrafishMiner enables the user to quantify tissue fluorescence in zebrafish embryos and larvae. As a standard, eight tissue domains are already implemented. ZebrafishMiner offers the possibility to interactively replace domains either by manual marking or by using larvae with fluorescently labelled tissues. All algorithms are suited for high-throughput screening and have already proven to be suitable [3]. Adaptation to modified test conditions like brightness, morphological changes etc. is possible by either setting parameters in the graphical user interface or by adapting the algorithms manually.

The data does not have to fit a certain microscope format, however, file names have to be customized according to the Olympus microscopy file format. Three data modalities

are needed for correct processing - brightfield, fluorescent reference and fluorescence channel.

The ZebrafishMiner is under further development, which makes documentation very important. We developed a user manual that currently contains more than thirty pages. Included are also three real-world tutorial datasets for „Fluorescence evaluation“, „Reference by manual assignment“ and „Reference by Fluorescence“, that can be downloaded in the benchmark section at <http://sourceforge.net/projects/SciXMiner/>. The processing of each dataset is described in detail, text and figures step-by-step, in the manual, such that users who are uncommon to MATLAB can easily process data on their own.

Our ongoing development deals with the generation of reference larvae based on a mixture of „Reference by manual assignment“ and „Reference by fluorescence“. Furthermore, we constantly accelerate the processing pipeline by using parallel processing. In mid-term we are going to publish a plug-in suited for further model organisms as well (e.g. Drosophila).

Author's Statement

Research funding: We would like to express our gratitude to DAAD, BIF-IGS and Helmholtz Association for funding and supporting this research work. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent is not applicable. Ethical approval: The conducted research is not related to either human or animals use.

References

- [1] Detrich HW, Westerfield M, Zon LI. The Zebrafish: Disease Models and Chemical Screens. 3rd Edition: Disease Models and Chemical Screens 3rd Edition. 2011.
- [2] Alshut R, Legradi J, Liebel U, et al. Methods for automated high-throughput toxicity testing using zebrafish embryos. *Lecture Notes in Artificial Intelligence* 6359: 219-226, 2010.
- [3] Gehrig J, Reischl M, Kalmar E, et al. Automated high throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nature Methods* 6: 911-916, 2009.
- [4] Eliceiri K, Berthold M, Goldberg I, Ibanez L, Manjunath B, et al. Biological imaging software tools. *Nature Methods* 9: 697-710, 2012.
- [5] Mikut R, Bartschat A, Doneit W, González Ordiano JÁ, Schott B, Stegmaier J, Waczowicz S, Reischl M. The MATLAB Toolbox SciXMiner: User's Manual and Programmer's Guide. arXiv:1704.03298, 2017