Automated Segmentation and Classification of Zebrafish Histology Images for High-Throughput Phenotyping

Brian Canada[1], Georgia Thomas[2], Keith Cheng*[,3], and James Z. Wang*[,4]
[1,3,4]Integrative Biosciences Program, [2]Department of Cell & Molecular Biology,
[3]Department of Pathology, [4]College of Information Sciences and Technology,
The Pennsylvania State University
{[1]canada, [2]gkt104, [3]kcheng, [4]jwang}@psu.edu

***Abstract*—Because of its small size and rapid development, the larval zebrafish is an ideal model organism for studying mutant phenotypes using "high-throughput" histological analysis. Although the preparation and subsequent digitization of zebrafish larval histology specimens can be conducted in parallel, the scoring and annotation of the resulting virtual slides is largely manual and therefore rate limiting, which motivates the development of systems for automated characterization of histology images. We present a prototype for automated segmentation and classification of histology images in animal models, with a pilot study focusing on larval zebrafish eye and gut images. We show that the segmentation of the images into regions of individual cell layers can be conducted with good precision using combinations of widely-used image processing operations, and that the resulting classification system, based on a decision tree algorithm, exhibits promising performance.**

I. Introduction

A. Use of the zebrafish for high-throughput phenotyping

The zebrafish has been shown to be an excellent model organism for vertebrate development and human disease, largely because its transparent, readily accessible embryo develops outside the mother's body, and most organ systems are well differentiated by 7 days post-fertilization [1], allowing mutant phenotypes to be readily identified in a relatively short amount of time. Histology, the microscopic study of biological tissues, is a highly sensitive method of characterizing such phenotypes.

Because of their small size, zebrafish larvae are amenable to "high-throughput" histology, a procedure developed by Cheng and colleagues [2, 3] in which arrays of up to 50 zebrafish larvae at different stages of development are fixed to preserve morphology, embedded in agarose gel, processed into paraffin blocks, and sliced into thin sections. The sections are individually mounted on glass microscope slides for staining and then scanned to produce virtual slides for scoring and annotation of phenotypes (see Fig. 1).

The phrase "high-throughput," as used in this context, does not currently apply to all steps in the cycle, however. While fixation, embedding, sectioning, staining, and digitization of an entire array of larvae may be conducted in parallel (in the sense that in each of these steps, multiple larvae are being processed together), the process of scoring each image is rate limiting. Comprehensive annotation of the phenotype requires scoring several levels (planar sections) of each larva. While this process is manageable when manually annotating a small number of larvae, the task becomes increasingly impractical as the number of larvae rises. Complicating the matter is the problem of subjectivity in the process of scoring and annotating histology images. While largely reliable and useful clinically, the qualitative aspects of current histological assessments result in intra- and inter-observer variability [4]. This variability can be due to differences in training, ability, timing, and experience. Therefore, in order to maximize the effectiveness of histological studies as applied to animal model systems, it is critical that some form of automatic, quantitative method be developed for the analysis of histology images.

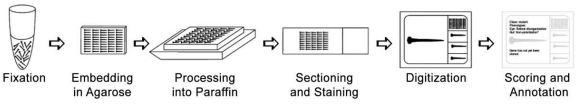

Fig. 1. Current laboratory pipeline for "high-throughput" histology. The steps up to and including digitization can be automated and/or conducted in parallel, but the overall process is rate-limited by the manual (and tedious) process of scoring and annotation.

B. Current efforts in zebrafish image analysis

Presently, the only major work we are aware of in the area of automatic zebrafish image analysis has come from Wong and colleagues, who developed a method [5] for quantitative neuronal phenotyping based on automatic analysis of light microscope images of live zebrafish embryos with altered expression of Alzheimer's disease (AD)-linked genes. In particular, this method allows for quantitative assessment of neuron loss, automated detection of somites (groups of mesoderm cells that form along the neural tube and go on to differentiate into skin, muscle, and vertebrae), and determination of gene expression levels in the affected embryos. In principle, automated methods such as this can be readily scalable to large-scale phenotyping studies in which manual image analysis is inefficient and variable [6]. To our knowledge, however, no computerized methods are available for the automated quantitative analysis of images of zebrafish

*K.C. and J.Z.W. have contributed equally to this work. B.C. is supported by the Penn State Academic Computing Fellowship. G.T. and K.C. are supported by NIH, the Life Sciences Greenhouse of Central Pennsylvania, and Pennsylvania Department of Health Tobacco Settlement Funds. J.Z.W. is supported by NSF.

978-1-4244-1813-8/08/$25.00 © 2008 IEEE

histology. To address this deficiency, we are developing SHIRAZ (System of Histological Image Retrieval and Annotation for Zoopathology), a novel computational framework for the automated, ontology-based annotation and retrieval of histological images for quantitative, "high-throughput" phenotyping in animal model systems. This framework is ultimately intended for use with histology images for a variety of animal models, but owing to our experience in zebrafish genetic studies and histological analysis, zebrafish serves as a compelling pilot model for prototype development. Here, we present our preliminary work on the segmentation, feature extraction, and classification methods that will form the foundation of the SHIRAZ framework.

II. METHODS

In order to demonstrate proof of principle, we have developed a working prototype using the technical computing language MATLAB. The segmentation and feature extraction algorithms, as well as classification models, have been developed to analyze images of larval zebrafish eye (retina) and gut (intestine). We have chosen to initially examine eye and gut because these tissues have an inherent polarity and "directional" organization that, when disrupted, results in mutant phenotypes that are relatively easy for human experts to detect.

A. *Slide preparation and image preprocessing*

Zebrafish larvae were collected, fixed, and embedded as described in [2, 3]. Eye and gut images were manually cropped from 40x virtual slides of hematoxylin & eosin (H&E)-stained zebrafish sections captured using an Aperio T2 slide scanner in the Penn State Zebrafish Functional Genomics Facility. In order to reduce the computational cost of image segmentation and feature extraction, the images of the zebrafish eye and gut used in our prototype were converted to grayscale color space and finally scaled down to a square 512 × 512 matrix. If cropping out the area of interest could not yield a square matrix, the image was padded with white pixels to form a 512 × 512 matrix.

B. *Eye segmentation and feature extraction*

The zebrafish eye consists of several distinct layers, including the lens, the ganglion cell layer (GCL), the inner plexiform later (IPL), the inner nuclear layer (INL), the outer plexiform layer (OPL), the photoreceptor layer (PRL), and the retinal pigmented epithelium (RPE) [1]. The segmentation of the zebrafish larval eye is carried out by a series of image processing operations including histogram equalization, connected component analysis, thresholding, edge detection, as well as several morphological operations such as dilation, erosion, closing, and opening. The parameters used for each of these operations were heuristically determined, but are a loose function of the skewness of the image histogram. Each cell layer in the zebrafish eye uses a different combination of these image processing operations to achieve the most robust detection of layer boundaries. The layers with the most distinct boundaries — the lens, the RPE, and the IPL — are detected first. The PRL and the INL are then segmented from the area bound by the IPL and the RPE (assuming the IPL exists, which is not the case in many severely disorganized mutants). Once the PRL and INL are found, the GCL is isolated from the remaining "un-segmented" area of the eye. It should be noted that at the low resolution and reduced color space used in the prototype application, the OPL is usually too thin to be detected with any degree of robustness, and so it remains unsegmented as part of either the PRL or INL. Fig. 2 shows an example of a mutant eye segmented by the SHIRAZ prototype into its constituent cell layers.

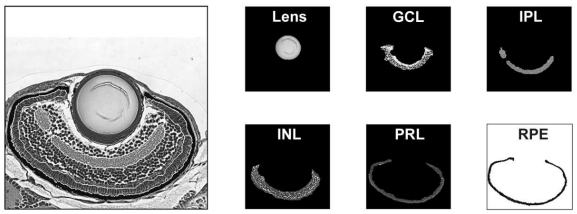

Fig 2. Results of automatic segmentation by SHIRAZ of a mutant zebrafish eye image into individual cell layers, with an exploded view on the right. The mutant phenotype of the disrupted inner plexiform layer (IPL; see upper right panel) is clearly detected.

Presently, the features extracted from each cell layer (when applicable) include:

- *filled area* - the area, or the number of pixels of an object, with all holes filled
- total number of pixels along the *perimeter* (both outer and inner perimeters, if applicable)
- *compactness* - the degree to which the shape of the layer resembles that of a perfect circle, computed as $[perimeter^2/(4\pi \times area)]$
- *eccentricity* - the distance between the two foci of the equivalent ellipse of the shape divided by the length of the major axis
- *extent* - the area of an object divided by the area of the object's bounding box
- *solidity* - area of an object divided by the area of the object's convex hull
- *fractal dimension* - sometimes called the *box-counting dimension*, which is the number of square boxes of a given side length that are required to cover the object, repeated for different side lengths (reported as a vector quantity, with one scalar for each box side length)
- seven *moment invariants*
- four *gray level co-occurrence features*

Moment invariants are dimensionless scalar descriptors that remain constant even when the object is translated, rotated, scaled larger or smaller, or mirrored (*i.e.*, "flipped" about an axis) [7]. An image's gray level co-occurrence matrix (GLCM) is one means of representing the frequencies with which different combinations of gray level intensities are observed in an image [8] and is often used to describe the "texture" of an image, employing features such as *contrast* (also known as *variance* or *inertia*), *correlation, energy* (also known as *uniformity*), and *homo-*

geneity. These texture features are computed from adjacent 8 × 8 pixel blocks inscribed within each segmented cell layer.

C. Gut segmentation and feature extraction

The zebrafish larval gut is characterized by several features, including the thickness and degree of folding of the epithelial lining, the polarity of the epithelial cells (*i.e.,* the position of the nuclei with respect to the basement membrane of the epithelium), the degree of folding of the lumen (which may be characterized by the number of distinct villi protruding into the lumen), as well as the amount and "granularity" of cellular debris and mucous that has sloughed off of the lining into the lumen. The segmentation procedure begins with the detection of the lumen, followed by an approximate segmentation of the epithelial lining, then identification of the boundaries of epithelial cell nuclei (and the degree to which the epithelial cells have been polarized) as well as the detection of cilia and any sloughing off of debris and mucous into the lumen. The feature extraction procedure for gut images is similar to that used for the eyes, but in our prototype only 30 features are extracted for each gut image, as opposed to 92 features for the eye images. The higher number of eye image features results mainly from the fact that similar sets of features are extracted for each segmented cell layer, and there are a greater number of such layers in the eye images than in the gut images.

D. Classification

The current version of the SHIRAZ prototype uses a tree-based algorithm (specifically, the Classification And Regression Tree or "CART" algorithm [9]) to classify the eye and gut images according to their degree of abnormality. The CART algorithm works by finding a hierarchy of features that yield the best discrimination among the desired set of classes. At the "root node" of the tree, which represents the full data set, the best degree of discrimination (or "best split") is computed for all features. Once a best split has been chosen for the root node, the data is partitioned into two "child" or "leaf" nodes. This process is repeated until a pre-chosen stopping criterion is reached, such as when the data has been partitioned enough so that all leaf nodes contain, at most, a specified maximum number of samples. After the tree is constructed, it may be found that "pruning" the tree back to a smaller number of terminal nodes may improve the overall classification accuracy.

While it may or may not necessarily yield the best accuracy as compared to other methods such as Gaussian mixture models, support vector machines (SVM), or artificial neural networks, the CART algorithm is especially useful because it uses a "white box" model, where the process of classification can be visualized as a tree or flowchart, and is therefore easy for most people to comprehend (see Fig. 3), which can help to provide a sense of objectivity and order to the practice of histological assessment, often a highly subjective process. One drawback of CART is that the splits can only be performed on one dimension at a time. For our application, however, the dimensionality of the data is relatively low compared to more complex datasets that may be more amenable to other classification methods, such as support vector machines, which make use of linear and nonlinear combinations of features [10]. The types of features currently used in our prototype application may not be particularly suited to nonlinear combinations, but even so, we are planning to explore SVM and other classification methods, perhaps in conjunction with CART, to see which yields the best accuracy.

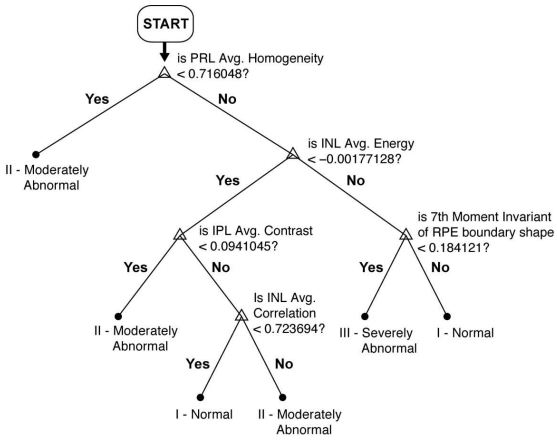

Fig. 3. An example decision tree (before pruning) as generated by the CART algorithm for classifying eye histology images.

III. Results

The SHIRAZ prototype has been tested for 79 images of the zebrafish larval eye and 87 images of the larval gut, ranging in age from 2.5 to 7 days post-fertilization (dpf). In general, the segmentations of both the eye and gut images produce excellent results, even for the most severe mutants. In some cases, artifacts resulting from errors in the histology slide preparation process may affect the quality of the segmentation, but continued enhancements to the segmentation algorithms are incrementally improving performance. In addition, the segmentation algorithms are quite robust to differences in image brightness, contrast, and sharpness that may result from variations in the image capture protocol. However, some images do present more of a problem than others, particularly in the case of detecting the boundary between the PRL and INL layers, which by virtue of their similarity in staining color are somewhat difficult to discriminate. Despite this obstacle, the accuracy of the PRL-INL boundary detection procedure—which now relies on the differences in the shapes and orientations of each layer's cells as opposed to color and texture—has improved considerably since development began, and continues to improve.

Preliminary classification tests were performed using three modes of classification, including *binary* (*i.e.,* classification as normal/wild-type vs. abnormal/mutant), *three-class* (normal, moderately abnormal, or severely abnormal), and *five-class* (normal vs. one of four possible mutant classes of increasing degrees of abnormality). Because of the relatively small number of images available at the time, we chose not to have separate data sets for training and testing of the clas-

sification model, and we instead used ten-fold and leave-one-out cross-validation. Classification accuracies were computed as the percentage of the images in the testing partition whose automatic classification matched the ground truth classification as determined by prior human evaluation. As shown in Table I, the binary classification performed well for both eye and gut images, with accuracies ranging from 86% to 90%, but as the number of classes increase, the classification accuracy correspondingly decreases, particularly for the gut images.

TABLE I
SUMMARY OF ZEBRAFISH HISTOLOGY IMAGE CLASSIFICATION RESULTS USING 10-FOLD AND LEAVE-ONE-OUT CROSS-VALIDATION

# of classes	Eye Images		Gut Images	
	10-fold CV	Leave one out	10-fold CV	Leave one out
Binary	90%	87%	86%	86%
Three classes	85%	85%	72%	71%
Five classes	72%	70%	56%	55%

IV. DISCUSSION

These results are quite encouraging (although more so for the eye images), especially in light of the fact that all organ-specific images have been grouped together for classification, regardless of the age of the larva shown in a given image. Generally, it is difficult to distinguish histological abnormalities in younger (< 5 dpf) larvae because cell layer differentiation is still progressing at this stage. This may, in part, explain the disparity between the eye and gut error rates for multi-class characterization. Ideally, images of larvae of different ages should be classified separately, but during this initial phase of SHIRAZ development, not enough normal and mutant images for any given age were available to yield strongly conclusive classification results, since preprocessing is currently a tedious process. Consequently, all ages of larvae were classified together. We are in the process of preparing a much larger set of images so that we can conduct individual classification tests for specimens at specific ages (*e.g.,* all images of 3dpf larval eyes will be classified in one test, all 5dpf larval eyes will be classified in another test, and so forth) and also so that we can have enough specimens representative of each class (*i.e.,* normal vs. varying degrees of abnormality).

V. CONCLUSIONS AND FUTURE WORK

We have developed a prototype system for the automated segmentation and classification of zebrafish histology images, which to the best of our knowledge is the first such published work. However, this is merely the starting point of our research in the automated characterization of animal histology, and while our preliminary results are promising, the prototype has several opportunities for improvement that we are currently addressing. For example, image segmentation is not always accurate, such as with respect to the detection of the PRL-INL border in the eye images. In addition, the prototype currently utilizes low-resolution grayscale images manually extracted from the original color virtual slides. While this may be advantageous from a computational cost standpoint, we may be ignoring important image details (such as features of individual cell nuclei) that could only otherwise be detected using high-resolution color images. We will therefore determine the optimum extent of preprocessing necessary to maximize segmentation and feature extraction accuracy while minimizing both manual effort and computational cost. To this end, we are developing methods to automatically extract the eyes and gut from the original virtual slides, so that manual image cropping is unnecessary.

Classification accuracies are expected to improve as the set of available images grows to include more representatives of all ages and classes of histological specimens. We also recognize that in spite of the usefulness of a "white box" classifier, such as the CART algorithm used here, classification accuracies may also be improved by the use of support vector machines, Gaussian mixture models, or artificial neural networks (or a combination of these, perhaps integrated with CART), which we will explore.

We plan to further develop our SHIRAZ prototype from a zebrafish image segmentation and classification system into a parallelizable, ontology-based framework for automatically annotating images of animal histology as well as providing a means for retrieval of images and quantitative phenotype information from an online database. Such a system will help to bring the scientific community closer to truly "high-throughput" phenotyping and also to assign possible functions to previously uncharacterized genes. The phenotype information stored in the database will be based on the Phenotype and Trait Ontology (www.bioontology.org/wiki/index.php/PATO:Main_Page) to ensure inter-compatibility with other ontology-based biological data repositories and facilitate interactions among animal model researchers.

REFERENCES

[1] C. Nusslein-Volhard, and R. Dahm (eds.), *Zebrafish: A Practical Approach,* Oxford: Oxford University Press; 2002.

[2] G.S. Tsao-Wu, C.H Weber, L.R. Budgeon, and K.C. Cheng, Agarose embedded tissue arrays for histologic and genetic analysis, *Biotechniques,* vol. 25, 1998, pp 614-618.

[3] N.A. Sabaliauskas, C.A. Foutz, et al., High-throughput zebrafish histology. *Methods,* vol. 39, 2006, pp 246-254.

[4] P. Colquhoun, J.J. Nogueras, et al., Interobserver and intraobserver bias exists in the interpretation of anal dysplasia. *Dis Colon Rectum*, vol. 46, 2003, pp 1332-1338.

[5] T. Liu, J. Lu, et al., Computerized image analysis for quantitative neuronal phenotyping in zebrafish. *J Neurosci Methods,* vol. 153, 2006, pp 190-202.

[6] T. Liu, A quantitative zebrafish phenotyping tool for developmental biology and disease modeling*." IEEE Signal Processing Magazine,* vol. 24, 2007, pp 126-129.

[7] M. Hu, Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory,* vol. 8, 1962, pp 179-187.

[8] R. M. Haralick, K. Shanmugam, I. Dinstein, (1973). *"*Texture features for image classification*." IEEE Transactions on Systems, Man, and Cybernetics*. vol. 3, 1973, pp 610-621.

[9] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and Regression Trees,* Belmont: Wadsworth International Group; 1984.

[10] T. Hastie, R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning,* New York: Springer; 2001.