

AUTOMATED SEGMENTATION, ANALYSIS AND CLASSIFICATION OF ZEBRAFISH VASCULATURE

A Dissertation

Presented to

the Faculty of the Department of Computer Science

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Charu Hans

May 2015

AUTOMATED SEGMENTATION, ANALYSIS AND CLASSIFICATION OF ZEBRAFISH VASCULATURE

Charu Hans

APPROVED:

Shishir Shah, Chairman
Dept. of Computer Science

Edgar Gabriel
Dept. of Computer Science

Guoning Chen
Dept. of Computer Science

Fatima Merchant
Dept. of Engineering Technology,
College of Technology

Maria Bondesson
Dept. of Biology and Biochemistry

Dean, College of Natural Sciences and Mathematics

AUTOMATED SEGMENTATION, ANALYSIS AND CLASSIFICATION OF ZEBRAFISH VASCULATURE

An Abstract of a Dissertation
Presented to
the Faculty of the Department of Computer Science
University of Houston

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

By
Charu Hans
May 2015

Abstract

An important factor facilitating the application of zebrafish in biomedical research is high throughput screening of vertebrate animal models. Presently, the zebrafish has become a viable model for various research including vertebrate development, gene expression analysis, human diseases modeling, drug screening and toxicology analysis. Compared to other models, the zebrafish embryo is an attractive model due its fast development, small size, transparent embryos that develop outside the mother, and the availability of a large selection of mutant strains. However, the lack of tools for automated analysis of complex images is a huge bottleneck in utilizing the zebrafish to its full potential.

Zebrafish have a closed circulatory system, and the mechanism of vessel formation are highly similar to those in humans. Being able to model the growth of blood vessel in the vasculature system of zebrafish is interesting for understanding both the circulatory system in humans, and for facilitating large scale screening of the influence of various chemicals on vascular development. Compared to other models, the zebrafish embryo is an attractive alternative for environmental risk assessment of chemicals since it offers the possibility to perform high throughput analysis in vivo. Intersegmental vessels (ISV) and caudal vein plexus (CVP) undergoes active development via angiogenesis. Hence, providing excellent models to study vasculature system.

Most of the current research based on ISV observe the presence or absence of ISVs or perturbation of ISV morphology but do not quantify growth dynamics. Moreover, these analysis are done manually hence it is tedious and expensive. All of these factors facilitate the need for automated image processing methods to quantitatively

analyze the image dataset. In this work, we have focused on developing an image processing algorithm to automatically segment and quantify ISVs of zebrafish embryos that have been treated by various toxins. We tested the algorithm on images of zebrafish embryos obtained from screening compounds that may act as an ISV disruptor. The efficiency of segmentation approach is demonstrated by our experiments of the entire zebrafish vasculature recorded from the fluorescence microscope. The experiments also demonstrate that automated segmentation of ISV is comparable to that of manual segmentation. The quantified features are used to train a linear SVM classifier to identify morphological changes in a dataset consisting of ISV zebrafish embryo images.

Recently, studies have indicated that zebrafish CVP also undergoes active development, hence providing an additional measure for studying vascular development. In this work, we have presented an approach to segment and detect abnormalities in the CVP region of zebrafish embryos due to exposure to toxins. Morphological changes due to toxin exposure is modeled based on the proposed gradient weighted co-occurrence histogram of oriented gradients (gCo-HOG). These features are compared to more commonly used histogram of oriented gradients (HOG) features and Co-HOG features that utilizes spatial distribution of neighboring pixels to capture spatial structure. The features are used to train a linear SVM classifier to identify morphological changes in a dataset of region of CVP zebrafish embryo images.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Zebrafish Image Analysis	4
1.3	Zebrafish Vasculature Anatomy	6
1.4	Challenges	8
1.5	Research Goal	9
1.6	System Overview	13
1.7	Summary of Contribution	16
1.8	Thesis Outline	17
2	Related Work	19
2.1	Manual Analysis	19
2.2	Semi-Automated Analysis	22
2.3	Automated Analysis	24
2.3.1	Template Based	26
2.3.2	Learning Based	27
2.3.3	Contour Based	29
2.3.4	Curvature and Edge Based	29
2.3.5	Tracking Based	30

3	Zebrafish Segmentation	34
3.1	Embryo Segmentation	34
3.2	Region of Interest Detection	38
3.3	Intersegmental Vessel Segmentation	40
3.3.1	Eigen Analysis	43
3.3.2	Adaptive Niblack thresholding	47
3.3.3	Linking vessels	48
3.3.4	Post Processing	49
3.4	Caudal Vein Plexus Segmentation	50
4	Feature Extraction	52
4.1	Intersegmental Vessel Feature Extraction	52
4.2	Caudal Vein Plexus Feature Extraction	54
4.2.1	Gray Level Co-occurrence Matrix	54
4.2.2	Histogram of Oriented Gradients	55
4.2.3	Co-occurrence of Histogram of Oriented Gradients	56
4.3	Vasculature Time-lapse Features	58
4.3.1	Vasculature Registration	59
4.3.2	Vasculature Morphology	62
5	Materials	64
5.1	Two Dimensional Data	65
5.1.1	Chemical Treatments	65
5.1.2	Imaging	65
5.2	Three Dimensional Time - Lapse Data	66
5.2.1	Chemical Treatment	66
5.2.2	Imaging	67

5.2.3	Time-lapse imaging	68
6	Zebrafish Analysis	70
6.1	Intersegmental Vessel Analysis	71
6.1.1	Segmentation Analysis	71
6.1.2	Toxicity impacts Intersegmental Vessels	76
6.2	Toxicity impacts Caudal Vein Plexus	79
6.2.1	Gray Level Co-occurrence Matrix Analysis	80
6.2.2	Histogram of Gradient Analysis	81
6.2.3	Co - occurrence of Histogram of Gradient Analysis	81
6.2.4	Gradient Co - occurrence of Histogram of Gradient Analysis .	82
6.3	Toxicity Screening with ISV, CVP	83
6.4	Vasculature Time - Series Analysis	88
7	Conclusions and Perspectives	91
7.1	Discussion and Limitations	92
7.2	Future Work	96
	Bibliography	97

List of Figures

1.1	Zebrafish embryo anatomy	7
1.2	System Overview	15
2.1	Overview of an automated/semi-automated image analysis methods. .	25
2.2	Overview of vasculature image analysis methods	31
2.3	Overview of an automated vasculature image analysis methods	31
3.1	Zebrafish embryo extraction procedure	35
3.2	Intensity profile of ISV + DLAV region compared to tail + head region	39
3.3	Procedure for finding head yolk position	40
3.4	ROI Detection	43
3.5	ISV principal direction estimation	48
3.6	CVP Segmentation	50
4.1	ISV skeleton	53
4.2	HOG features from a CV image	56
4.3	Co-HOG features from a CV image	58
4.4	Quantification Procedure	62
5.1	Zebrafish full anatomy 2d dataset	66
5.2	Maximum intensity projection	68
5.3	Zebrafish vasculature dataset	69

6.1	ISV segmented images	73
6.2	Comparison between the automated segmentation and manual segmentation of ISV.	75
6.3	Effect of toxins on ISV	76
6.4	Effect of toxins on ISV morphology	78
6.5	Healthy and treated CVP	79
6.6	Effect of varying dosage on ISV	85
6.7	Effect of varying dosage on CVP	85
6.8	Safe Dosage for ISV and CVP	87
6.9	Average comparison between healthy and treated	90
6.10	Individual comparison between healthy and treated	90
7.1	Images with the blurred and noisy ISV	94
7.2	Images with the blurred and noisy CVP	95

Chapter 1

Introduction

1.1 Motivation

Zebrafish has recently emerged as an invaluable model for biomedical research. Enormous amounts of data is being generated from imaging zebrafish embryos. Manual image analysis is cumbersome and subjective, there by increasing the need for automated image analysis tools. Lack of automated image processing tools for zebrafish analysis is a huge bottleneck in realizing zebrafish application to its full potential.

The growing utility of zebrafish as a model for biomedical research is attributed to its properties such as transparency, reproducibility, growth rate, and small size. The transparency of zebrafish embryo makes it a good model for direct observation of the development of organs and tissue. For this reason, zebrafish has been widely

utilized to study neural developmental [11], and vasculature development [23]. It enables researchers to study different development stages of zebrafish to adulthood [31]. Transparent zebrafish promotes relatively easy modification of genetic expressions to mimic various diseases including Cancer [2], Alzheimer [45], and Parkinsons [8]. Neuronal structures and blood vessels can be directly observed, providing an opportunity to study abnormal neuronal and vascular patterns due to gene manipulation, drug screening and other environmental cues.

Its small size and high reproduction rate allows researchers to image a large number of zebrafish in a short period of time [37]. Advantages of zebrafish as a vertebrate animal model make it suitable for screening chemicals in an effective manner [30]. Studies are being conducted to evaluate the embryotoxic effect of chemicals [60]. Transgenic zebrafish has been developed to study the effects of chemicals on blood vessel development [54]. The availability of a large number of zebrafish embryos makes them suitable for characterizing gene expressions [19].

Advances in imaging technique combined with various advantages of zebrafish, is another reason driving their wide applicability in various research areas [37]. Imaging zebrafish is fast and cost-effective as compared with other animal models. Thus, zebrafish is becoming one of the most popular species in high-throughput screening. In most of the above work, data is generated in large amounts and it is difficult to analyze it manually. Manual analysis is time-consuming, suffers from inter class variability and is not scalable for high throughput analysis. With rapid advances in science and environmental pollution, the need to screen huge numbers of compounds during the drug screening and toxicology analysis has been growing. At present, there

is substantial effort required to validate assays for such use. A key challenge has been the automated assessment of imaging data because there are few image analysis tools capable of performing analysis in automated manner. In the past few years, there has been a significant increase in the number of research efforts focusing on analyzing behavioral and morphological features in adult zebrafish [40]. Morphological analysis methods have been developed to extract features of neurons [31] and vasculatures [20] from zebrafish. Methods have also been developed where zebrafish swimming pattern and behavior are recorded by video equipment for analysis in response to stress, predators, alarms, and drugs [5].

In this thesis, we focus on developing image processing algorithms for segmentation, analysis, and classification of zebrafish under the influence of various chemicals. The interpretation and analysis of toxicology screens consumes about 70% of analysis time [1]. Due to industrialization, we need to regularly perform a risk assessment on new chemicals, pesticides, and drugs. Frequency with which new hazardous chemicals are being identified further emphasizes the magnitude of the problem. Concurrent with the increasing levels of environmental pollution, accumulating evidence shows that there is a vital need to study effects of exposure levels on human and animal health [43]. Research community is still trying to understand the effects of exposure of chemicals on organism. There are many questions that have to be answered. Firstly, what is the nature of effect, and to what extent? Secondly, after what dosage to study chemicals do the hazardous effects becomes threatening? Is there a way to establish predictive toxicity testing of chemicals? zebrafish as the model of toxicology has been well established. Emphasis is on creating tools for toxicity assessment

through qualitative or quantitative screening for varying exposure levels of toxicants.

1.2 Zebrafish Image Analysis

Over the past several years, efforts have been dedicated to developing methods to process, analyze, and quantify zebrafish images. Over the years researchers have dedicated efforts to develop image processing algorithms to analyze behavior of zebrafish. The first step in image processing for behavioral analysis is usually the continuous tracking and segmentation of the whole zebrafish from digital images. Bang et al. developed an automated screening assay to detect hearing effects in zebrafish by monitoring their behavior after receiving a loud sound burst [5]. Kato et al. proposed to study swimming pattern and schooling behavior for analysis [29].

Various neurobiological analysis algorithm for quantifying changes in zebrafish has been proposed. Xu et al. [59] proposed an algorithm focused on detecting and quantifying pigments in zebrafish embryos. They automatically identify torso area through series of image processing steps. Liu et al. [40] have proposed to automatically quantify the neurons and somites in a large number of zebrafish images, as well as quantitative measurement of gene expression levels in zebrafish embryos. Chen et al. [13] proposed a robust automatic segmentation to identify region of interest (ROI) for gene expression quantification.

Image processing methods have also been developed to analyze morphological features in zebrafish. Stern et al. [52] focuses on automatically detecting specific interest points in microscopy images. Alshut et al. [1] proposed learning based

classification approach, which detects the embryo in the well, classifies its status as dead, alive or unknown, and derives characteristic parameters for the regions. Liu [39] proposed a recognition model for high throughput screening of toxicity based on image descriptors based on color and texture combined with a support vector machine to recognize three basic phenotypes (hatched, unhatched, and dead).

In addition to neuronal structure, blood vessels is another structure that is often used in zebrafish-based studies. These studies are particularly interesting since, the cardiovascular system is one of the first organ to emerge during embryonic development. The vascular system is a vital component of all vertebrate animals, supplying oxygen and essential nutrients to every tissue and organ. Zebrafish have a closed circulatory system, and the mechanism of vessel formation are highly similar to those in humans. A wide range of congenital diseases are associated with blood vessel formation [21], development of the cardiovascular system and associated vascular defects [50], and the adverse effects of exposure to toxic elements during development [32]. Some previous studies have already employed image processing methods to studying vasculature development. Tran et al. [54] proposed an algorithm using the Discovery-1/MetaMorph software to analyze blood vessel images. Vogt et al. [55] implemented a user guided image interpretation tool to generate rule-based hierarchical image segmentation and blood vessels quantification. Feng et al. [20] developed a 3D attributed vessel representation graph (AVRG) approach to reconstruct caudal vasculature of zebrafish embryo. The whole vascular structure is reconstructed and then utilized for quantification of the number and connectivity of the vessels, their size, length, and volume, as well as the distance between any two vessels. Existing

method perform well with uniformly illuminated zebrafish images, but resolve to manual analysis under varying illumination [54]. Automated analysis of vasculature is still a budding field. Feng et al. [20], and Vogt et al. [55] are the only studies dedicated to automated analysis of vasculature blood vessel. The work presented in the above mentioned has been performed on limited number of screens; there is still a question about its scalability.

In this work, we focus on developing an image processing algorithms for the high-throughput screening (HTS) of assays of compounds for toxicology analysis. In particular, we will focus on the vasculature system of zebrafish. The methodology presented in this work is quantitative, can be utilized in wide varieties of toxin treated zebrafish, and capable of quantifying changes in ne structure not quantiable by the human eye. Based on existing methods, automatic identification of the applicable ROI, and accurate quantification of ROI features has been very challenging.

1.3 Zebrafish Vasculature Anatomy

The zebrafish embryo (fig. 1.1) is an appropriate model to study vascular morphogenesis in vivo. The formation of vertebrate blood vessels is subdivided into two distinct morphogenetic processes, called vasculogenesis and angiogenesis. The earliest endothelial cells differentiate and aggregate into vascular cords by the mechanism of vasculogenesis. The subsequent vessels develop from existing vessels by the process of angiogenesis. First embryonic vessels to appear by vasculogenesis are the dorsal aorta (DA) and the posterior cardinal vein (PCV) [27]. The intersegmental vessels

(ISV) of the trunk are the first angiogenic vessels to form in all vertebrates. ISV development is two step process: (1) A set of new sprouts emerges from the dorsal side of the DA at 22 hour post fertilization (hpf) and these sprouts grows dorsally and connect to form the future dorsal longitudinal anastomotic vessel (DLAV); (2) Second step starts at 32 hpf, involves sprouts from the PCV [61]. These sprouts will either connect to an existing vessels, or alternatively, they will grow up to the level of the horizontal myoseptum. Just like ISV, caudal vein plexus (CVP) angiogenesis occur in two waves [15]: (1) Sprouting occur in the caudal vein region between 25 hpf and 30 hpf to form the primordial CVP. Soon after that, sprouting angiogenesis in this region slows down, but the primordial CVP continues to mature; (2) Second wave of caudal vein morphogenesis begins at around 48 hpf.

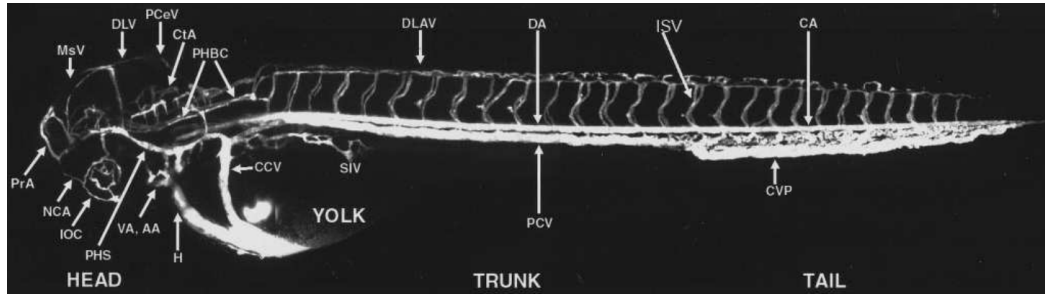


Figure 1.1: Zebrafish embryo anatomy [27].

The morphogenesis of the ISV, and CVP involves sprouting, migration, and pruning and can therefore serve as an excellent model for studying angiogenesis.

1.4 Challenges

Current methods analyze images based on intensity and thresholding. These methods perform well on clearly resolved objects against a uniform background but often resort to manual analysis for images that possess information at multiple sizes and heterogeneity across images. To add to the challenge, the images are affected with noise and artifacts. Furthermore, the appearance of zebrafish may vary in intensity, texture, size, and orientation. This problem was brought to attention in a work where zebrafish vasculature became quantifiable only after manual segmentation of the trunk [54].

Moreover, there is a growing requirement for automated image analysis techniques to process, analyze, and quantify zebrafish images. Meeting such objectives is essential for zebrafish research to reach its full potential in discovery and understanding. Despite the potential of the zebrafish as a model, quantification is still a fledgling field in zebrafish research, primarily due to the lack of tools that can yield objective and quantitative measurements from imaging. From an image processing perspective, a successful algorithm must take into consideration the objectives of the HTS and the content of the images.

All these factors advocate, for automation of as many steps of the analysis as possible. There is an increasing demand for automated image processing to generate quantitative results and to avoid time-consuming manual analysis. According to [44], the development of dedicated image processing methods has become a serious bottleneck in the full exploitation of the information contained in the acquired

image sets. Many custom-made and nongeneric solutions have been developed to answer specific questions. These solutions have an enormous potential to support the analysis of current and future image-based phenotype studies and to avoid parallel developments that tend to “reinvent the wheel”. We believe that development of a computerized data processing pipeline would be a significant step towards reproducible quantification of phenotypes in large scale or high throughput imaging studies.

Although methods have been developed to process zebrafish images, as the applications of the zebrafish model expands, there is a concurrent demand for a variety of image processing methods. In this work we will focus on the vasculature system of zebrafish, segmenting and developing algorithms for blood vessel development. The vasculature structure is composed of arteries and veins. Blood vessels are highly diverse in length, width, and branching pattern. Significant challenges are associated with blood vessel extraction including, noisy signal, drift in image intensity and poor image contrast.

1.5 Research Goal

This work presents a framework of image processing and analysis algorithms that consists of extracting zebrafish, realigning zebrafish embryo images of different orientations, and segmenting, quantifying, and classifying intersegmental vessels (ISV) and caudal vein plexus (CVP).

Following initial formation of the primitive vasculature by vasculogenesis, most

subsequent vessel formation during development takes place via angiogenesis and includes the formation of new vessels by budding growth from, or remodeling of, preexisting vessels. The formation of the ISV and CVP, provides an excellent model for angiogenesis.

ISVs of the trunk are among the first angiogenic vessels to form in all vertebrates. ISVs (fig. 1.1) are essential for development and nutrition of the zebrafish embryo. Unlike other vessels, the ISVs are interesting because of the patterned appearance and easy accessibility. Tran et al. [54] has shown that treatment of zebrafish at this early time point had a very strong antiangiogenic effect, causing nearly complete absence of intersegmental vessel growth. The vasculogenic dorsal aorta and posterior cardinal vein were unaffected. Since the intersegmental vessels in zebrafish form by sprouting from the preformed vasculogenic vessels, any compound that disrupts vasculogenesis will also inhibit the growth of angiogenic vessels [54]. All these properties makes ISV one of the most highly investigated vessel in the zebrafish.

In this work, we have focused on developing an image processing algorithm to automatically segment and quantify ISVs of zebrafish embryos that have been treated by various toxins. The processing pipeline consists of Segmentation, Region Detection, ISV Extraction, ISV refinement, feature quantification and classification. The efficiency of segmentation approach is demonstrated by our experiments of the entire zebrafish vasculature recorded using fluorescence microscopy. The experiments also demonstrate that automated segmentation of ISV is comparable to that of manual segmentation. The quantified features are used to train a linear SVM classifier to identify morphological changes in a dataset consisting of ISV zebrafish embryo

images.

Recently, studies have indicated that CVP also undergoes active development, hence providing an additional measure for studying vascular development [12]. The study of the zebrafish CVP and its utilization as a screening assay has not been as prevalent as the ISV, and very few studies have attempted to identify regulators of CVP development. In this work, we study the impact of toxicity on the shape of CVP. Instead of the fine mesh-work of the CVP observed in healthy embryos, treated embryos exhibit the formation of loops in the CVP. Previous research has primarily focused on using color, texture [54] and morphological changes [20] for toxicity analysis. Shape information, however, can be another attribute that can be evaluated. A shape descriptor can capture the outline of the shape that other color or texture-based descriptors may not be able to capture.

Hence by quantifying the shape of CVP, and comparing it against that of a healthy embryo, we can identify changes in CVP due to exposure to toxins. Morphological changes due to toxin exposure is modeled based on the proposed gradient weighted co-occurrence histogram of oriented gradients (gCo-HOG). These features are compared to more commonly used histogram of oriented gradients (HOG) features and Co-HOG features that utilizes spatial distribution of neighboring pixels to capture spatial structure. The features are used to train a linear SVM classifier to identify structural changes in a dataset of region of CVP zebrafish embryo images.

We have also presented a study for analyzing effects of Arsenic on overall vasculature development of zebrafish in time-lapse confocal images. We use a transgenic zebrafish that expresses green fluorescent protein (GFP) in the vascular system Tg(Flk1:GFP) to visualize vessel growth in the fish embryo. This transgenic zebrafish line expresses GFP in vascular endothelial cells, which permits real-time imaging of the formation and growth of blood vessels. Time-lapse confocal imaging of embryonic vasculature in the zebrafish is used in conjunction with digital image analysis to monitor and quantify the effect of toxins on vascular development. Imaging captures the dynamics of blood vessel formation over time. In order to quantify these morphological changes, i.e. how much vascular structure has changed over a period of time, it is necessary to compensate for any movements caused by the growing embryo. Thus, in order to record the temporal changes occurring due to vessel growth, it is necessary to establish spatial correspondence between blood vessels that may appear displaced due to embryo movement. Thus quantification of temporal vascular growth can be seen as a problem of image registration.

The goal of image registration is to align two images, so that common features overlap and differences are emphasized. Image registration has been used widely in the medical field to quantify the influence of changes over time. As an example, registration is required in medicine for comparing computer tomography of patients scan [7], aligning images from various different modalities to diagnose diseases, etc. Recently, image registration has found application in growth monitoring of tumors and bone [10], [34]. We use a non-rigid registration approach to align images. Non-rigid mapping is based on complete correspondence of images and includes a deformation

model as the underlying transformation. We utilize free-form deformation based on B-splines for growth monitoring, and use intensity differences as a similarity measure.

Overall, the methodology presented in this work is quantitative, can be utilized with wide varieties of toxin treated zebrafish, and capable of quantifying changes in fine structure not quantifiable by the human eye. The algorithm automatically processes multiple image files, saves the intermediate image processing results, writes the results in a excel/text format for further statistical analysis. We believe that development of a automated image processing pipeline would be a significant step towards reproducible quantification of zebrafish analysis for HTS.

1.6 System Overview

Zebrafish is a versatile model for vasculature analysis. As its application continues to grow, it is imperative to have an automated image analysis system (fig. 1.2) in place to be able to analyze high throughout data. Obtaining region of interest is the first step for any type of analysis. Multiple zebrafish are imaged in one dish. It is essential to extract individual zebrafish from image. Our system extracts relevant zebrafish embryo from images. Further, we segment ISV and CVP from embryos. Automated segmentation of ROI opens new opportunities for diversified analysis on zebrafish.

ISVs are widely used to study toxicology, development, and disease modeling. ISV segmentation is based on multiscale analysis combined with directional information. Multiscale analysis, makes it possible to capture responses from vessels of

varying diameter and direction information aids to isolate ISV from other vessels in the region. CVP with its easy accessibility during imaging process, is a potential candidate for research. Realizing its plausible benefits, we have proposed segmentation algorithm based on curvature properties of zebrafish tail.

Although many types of analysis and quantification can be performed on segmented region. Our framework computes morphological properties for quantification of ISV. Shape descriptor is proposed for CVP analysis. Shape descriptor is inspired from Histogram of Gradients. Co-Occurrence of histogram of gradient is combined with gradient information to obtain discriminative representation for CVP region.

Lastly, classification component completes the system. We treat classification as two class binary problem. This system is indicative of healthy embryo vs defective embryo. We demonstrate the validity of our system by analyzing toxin treated data. We used the model to estimate the safe dosage for various chemical. This system can be easily trained to be used for drug modeling, gene expression, etc.

System proposed in this thesis is robust in the manner that each step can be isolated and easily merged with existing or future research. For example, extraction of embryo, ISV and CVP can be effortlessly combined with varied feature extraction methods.

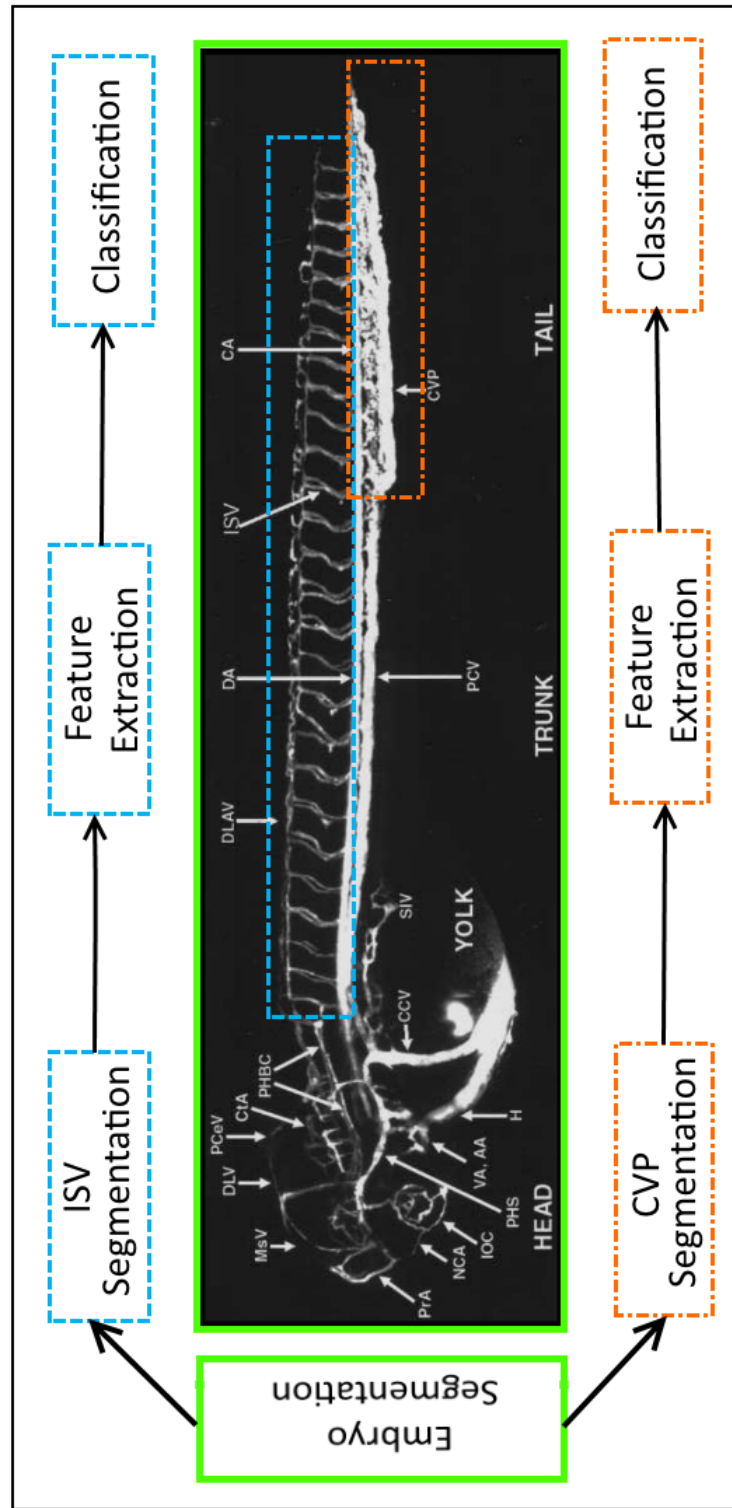


Figure 1.2: Zebrafish segmentation, feature analysis, classification system

1.7 Summary of Contribution

This thesis studies the image analysis of zebrafish embryo. It makes seven major contributions to this goal:

- Foremost contribution is the ISV segmentation. There has been many studies indicating viability of using ISV as a model for various application. To our knowledge, automated segmentation of ISV is still largely unstudied problem. In this work we have proposed use of multiscale response and direction information to segment blood vessels.
- Next is quantification of ISV. We have proposed feature extraction method for ISV morphology, and shown applicability to quantify ISV. Further, we have shown how the features are used for toxicology analysis.
- Next contribution is segmentation of CVP. CVP is relatively new model as compared to ISV for studying zebrafish. There has been studies showing active molecular development for CVP. Its importance as a model is gaining popularity. CVP segmentation can greatly simplify analysis for HTS. We have proposed an algorithm for CVP segmentation.
- In order to utilize shape property of CVP, we have proposed a shape descriptor based on gradient weighted co-occurrence of histogram of gradient. We have proposed to use these descriptors to be able to classify healthy zebrafish embryo, against unhealthy embryos.
- We present a toxicology screening system based on lethality rate for ISV and

CVP analysis. We used the models derived from morphological properties of ISV and shape properties of CVP to study the effect of increasing dosage on zebrafish in terms of number of zebrafish being impacted with chemicals. Further, this helps us to establish the safe dosage for various chemicals.

- We studied the vasculature development for time lapse data. This work focuses on capturing temporal vasculature changes in zebrafish. This can work as an initial pass for studying changes in zebrafish development under the influence of chemicals. We have shown utilization of developed algorithms to study arsenic treated zebrafish embryo.
- Lastly, the work-flow presented in this thesis is fully automated. The segmentation and analysis of the imaging data is one of the most challenging tasks in automation of the zebrafish applications. Due to lack of robust solutions for this problem, most of the analysis is currently being performed manually. This serves as a huge bottleneck for interpreting data from HTS.

1.8 Thesis Outline

The remaining chapters are organized as follows:

- Chapter 2 gives the detail about related work. Related work is clustered according to type of analysis (automated vs manual).
- Chapter 3 presents segmentation algorithm for zebrafish embryo, ISV and CVP.

Embryo segmentation is based on region and moment analysis. ISV segmentation uses multiscale property with directional information. CVP segmentation is based on curvature analysis.

- Chapter 4 describes the features used for quantification of ISV and CVP. We describe our proposed algorithm utilizing gradient information with co-occurrence matrix of histogram of gradient. Also, we will discuss time lapse vasculature quantification.
- Chapter 6 summarizes approach and our results. We will present results for ISV segmentation, comparing proposed segmentation with manual segmentation. We will also present a classification system for ISV (healthy vs unhealthy), based on its morphological properties. We will also present result for classification system based on shape properties of CVP. We will describe the toxicology screening system based on modeling ISV and CVP, and how it can be used to derive safe dosage.
- Lastly, chapter 7 concludes the thesis. We discuss limitations of the work and give guidelines for future work.

Chapter 2

Related Work

Utilization of zebrafish as a model has received lots of attention in recent year. Advances in imaging technologies and growing number of applications is pressing the need for automated image analysis methods. Table 2.1 presents the summary of past work for zebrafish analysis. We divided the related work into three sections on the basis of analysis: manual; semi-automated and automated. We further grouped the automated methods based on the type of methods used for analysis.

2.1 Manual Analysis

Arslanova et al. [4] developed a method to examine live zebrafish that were each treated with gamma-secretase inhibitors (GSI), DAPT N-[N-(3,5-difluorophenacetyl-L-alanyl)]-S-phenylglycine t-butyl ester, Gleevec, or fragments of Gleevec in a zebrafish model of Alzheimer's disease (AD).

Table 2.1: Summary Zebrafish Image Analysis

Method	Imaging Modality	Anatomy	Dimension	Application	Analysis Type	Year
Liu et al.	Digital Microscopy	Whole Organism	Gene Expression	2D	Automated	2006
Shirinifard et al.	Confocal	Vasculature	3D + t	Toxicology	Manual	2013
Chen et al.	n.s.	Vasculature	2D	Gene Expression	Manual	2005
Liu et al.	Bright Field	Whole Organism	2D	Toxicology	Automated	2012
Xu et al.	Bright Field	Torso	2D	Toxicology	Automated	2010
Arsanova et al.	Microscope	Whole Organism	2D	Gene Expression	Manual	2010
Tran et al.	Fluorescent	Vasculature	2D	Drug Discovery	Semi Automated	2007
Yozzo et al.	Fluorescent	Whole Organism	2D	Toxicology	Semi Automated	2013
Stern et al.	Microscope	Whole Organism	2D	Toxicology	Automated	2011
Kato et al.	CCD Camera	Whole Organism	2D + t	Behavior pattern	Automated	2004
Vogt et al.	Fluorescent	Vasculature	2D	Toxicology	Automated	2009
Feng et al.	Fluorescent	Vasculature	3D	Toxicology	Automated	2005
Chen et al.	Microscope	Whole Organism	2D	Drug Discovery	Automated	2011
Tal et al.	Confocal	Whole Organism	3D + t	Toxicology	Semi-Automated	2014
Kamali et al.	Confocal	Brainstem	3D	Disease Modeling	Automated	2009
Anilla et al.	Fluorescent	Whole Organism	2D	Disease Modeling	Semi - Automated	2013
Mccollum et al.	n.s	Whole Organism	2D	Toxicology	n.s.	2011
Alshut et al.	Microscope	Whole Organism	2D	Toxicology	Automated	2010
Peravali et al.	Microscope	Brain	2D	n.s.	Manual	2011
Bang et al.	CCD Camera	Whole Organism	2D + t	Behavior pattern	Automated	2002

Several GSI are in clinical trials for the treatment of AD. Brain regions (ROI) from the dorsal view images are manually segmented. The intensity of each ROI is quantified. To measure the gene expression, a manual approach is not only time-consuming but also lack objectivity due to inter-observer variations.

Shirinifard et al. [51] discussed quantitative analysis of growth dynamics for characterization of both normal and perturbed growth for ISV. Their method is based on analyzing ISV trajectories (sequences of successive 3D locations over time). Manual Tracking plugin in FIJI ImageJ is used to track 3D position of ISV sprout bases and tips over time (x, y, z coordinates over time). ISV sprout base position is the midpoint of the sprout where it intersects with the DA. ISV tip position is the farthest point on the ISV from the point of attachment on DA. Displacement and angle between the ISV and the DA is used to correct for zebrafish embryos twitching. Average trajectories are calculated for control and for arsenic treated embryo. Average trajectories are fitted with quadratic curve to produce a canonical ISV trajectory. Curvature, average directed migration speed and directionality were computed from canonical trajectories. Curvature, average directed migration speed and angle between the ISV and DA were different for arsenic treated vs untreated.

Chen et al. [12] proposed to use caudal vein to study the effects of modifications of sulfate 6-O sulfotransferase (HS6ST) genes on zebrafish development. Authors observed formation of large loops with high penetrance for the caudal vein. Quantification is done on the basis of number of embryos showing abnormal caudal vein development.

In the review done by Mccollum et al. [42] explores the questions of using zebrafish as a screening tool for human risk assessment. Toxicity effects on the zebrafish development are reviewed from existing literature. Toxic effects are reported on cardiovascular, somite, muscular, skeletal, and neuronal systems. They also report abnormal behavior, changes in gut, pancreas, liver, and kidney development, as well as toxic effects on the immune system and lipid metabolism.

Cheng et al. [14] studied the effects of cadmium on cardiovascular development in zebrafish embryos. They concluded that exposure to cadmium could affect the morphogenesis of the vasculature. In embryos with visible abnormalities, the vasculature in the malformed region correlated well with defective vascular patterning. There is a significant reduction in number of branches for cadmium exposed embryos as compared to health embryos.

2.2 Semi-Automated Analysis

Tran et al. [54] proposed an interactive algorithm using the Discovery-1/MetaMorph software to rapidly analyze each image. ISV is isolated from rest of images, by manually removing trunk from the fluorescent image. ISV is separated from rest of image using mask. The software counts the number of intersegmental vessels and branching arteries in the isolated trunk of the embryo. Manual segmentation of trunk restricts its usage for HTS. As shown in our work certain compounds do not alter ISV count but, end up changing other dynamics of ISV.

Yozzo et al. [62] screened 10 known cardiovascular toxicants through an image

analysis pipeline that included ISV sprout length quantification. Semi-automated methods are used to isolate regions of interest and quantify heart rate, circulation, pericardial area, and intersegmental vessel area, whereas a fully automated method is used to quantify body length.

Tal et al. [53] tested the effect of environmental chemicals on formation of the vascular system. They developed a quantitative assay in transgenic zebrafish and evaluated the assay using angiogenesis inhibitors that target VEGFR2 (PTK787) or EGFR (AG1478). Both PTK787 and AG1478 exposure impaired ISV sprouting, while AG1478 also produced caudal and pectoral fin defects at concentrations below those necessary to blunt ISV morphogenesis. The functional consequences of vessel toxicity during early development included decreased body length and survival in juvenile cohorts developmentally exposed to inhibitor concentrations sufficient to completely block ISV sprouting angiogenesis. Vascular growth depicted in time-lapse image stacks is quantified in automated manner where as ISV length is manually observed.

ZebIAT is image analysis tool developed by Anilla et al. [3] that allows both automatic or semi-automatic registration of the outer contour and inner organs of zebrafish embryos. ZebIAT provides a registration at different stages of development and an automatic analysis to study cancer progression. The user manually marks the organs or other areas of interest in the reference zebrafish embryo image. Outline of the zebrafish is obtained by segmentation and edge detection. Landmark based registration is applied to register images to reference image. Labeled cancer cells are detected using a method of multiscale product of wavelet and the mean fraction of

the area of each organ that exhibited cancer spots in the embryos is examined.

To enable large scale observations Peravali et al. [49] developed an algorithm to identify head region from low resolution image based on template matching. The regions of interest are subsequently imaged automatically at high magnification, enabling rapid capture of cellular resolution data. The pixel location in the source image yielding the largest normalized cross-correlation measure was considered the best match between the template and the source image, and is chosen as the center of the field of view for subsequent high-resolution imaging.

2.3 Automated Analysis

This section discusses automated techniques for zebrafish image analysis. Review is grouped into various subsection, based on type of method used for image analysis. Figure 2.1 gives an overview of algorithm utilized for analysis. It also includes, few of the methods described in section 2.2.

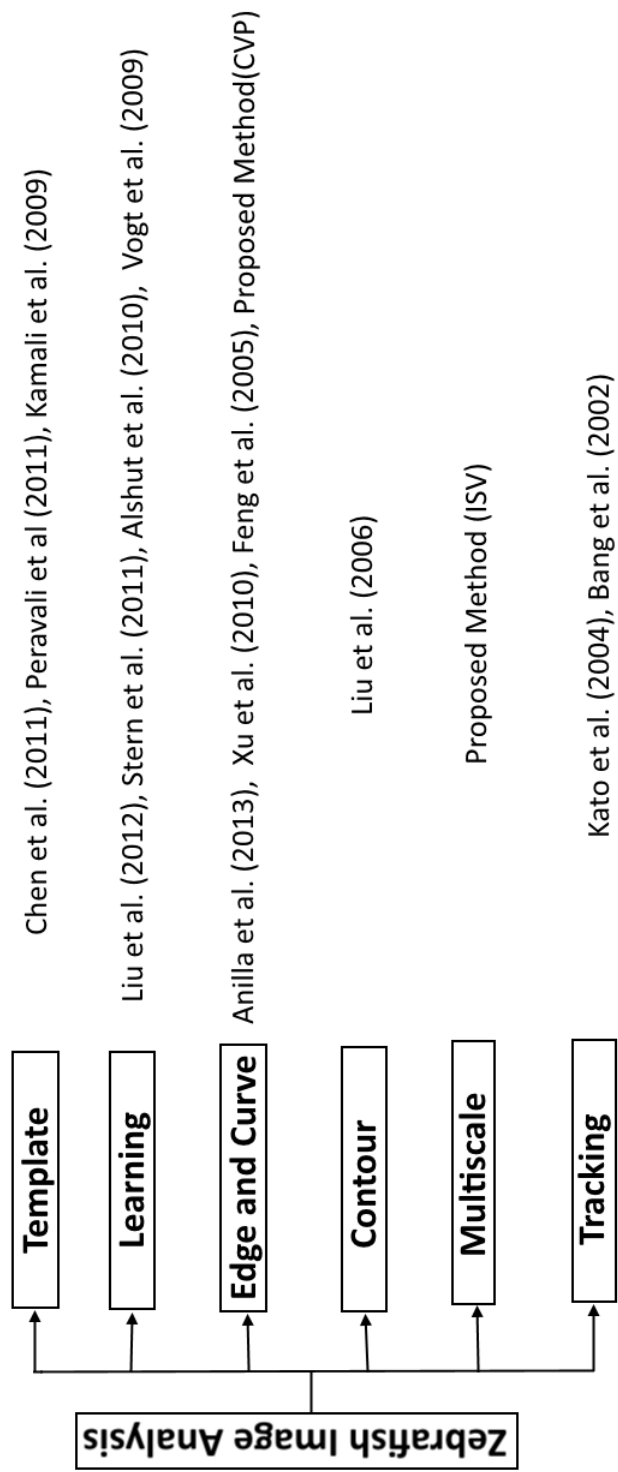


Figure 2.1: Overview of an automated/semi-automated image analysis methods.

2.3.1 Template Based

Chen et al. [13] proposed a robust automatic segmentation to identify ROIs for gene expression quantification. Their pipeline consisted of image alignment, segmentation, extraction and quantification of 15 ROIs. The telencephalon, left and right (L&R) dorsal diencephalon, L&R olfactory vesicles, ventral midbrain, L&R retina, L&R branchial arches, ventral hindbrain, L&R dorsal hindbrain, and L&R pectoral fin is quantified to evaluate gene expression. Size of each of the ROI is calculated for quantification.

Kamali et al. [28] developed a 3D template-based tracing algorithm to localize and differentiate the fluorescent neurons. A control template consisting of 28 zones with 14 zones on each side of the brainstem is created as a representation of 300 neurons that descend from the larval zebrafish brain into the spinal cord. The template is initialized by registering neurons manually identified in the different zones. After the creation of the template, image processing steps are applied to detect neurons and assign them to the template. First process is image registration of confocal z-stacks into a normalized space. User select three points on a MIP projection of confocal z-stack and corresponding points on template. Affine transform is applied to the 3D image stack to register it to the template. Second process of segmentation involves two steps: (1) neurons are segmented on each xy - plane to find a set of intensity contours with appropriate size and shape so as to reveal each neuronal cell body present in the plane (2) segmented results over consecutive xy-plane images are associated with the same neurons based on the assumption that each neuron spans several images along the z-direction. Lastly, for cell identification template is used

in conjunction with the segmented boundary contours to assign detected neurons to specific zones and sub-zones of the brainstem.

2.3.2 Learning Based

Liu et al. [39] proposed a recognition model for high throughput screening of toxicity based on image descriptors based on color and texture combined with a support vector machine to recognize three basic phenotypes (hatched, unhatched, and dead). The best performing model is attained with three image descriptors (color histogram, representative color, and color layout) identified as most suitable from an initial pool of six descriptors. The SVM classification model is developed using the three descriptors: (a) Global and Semi-global Edge Histogram Descriptor (GSEHD), (b) Representative Color Descriptor (RCD), and (c) Color Histogram Descriptor (CHD). They reported an average classification accuracy of 97.400.95% in a 10-fold cross-validation and 93.75% classification accuracy for a stress test with zebrafish images of low quality. This system can be used to identify extreme cases of toxicity, but further analysis is needed to identify less extreme cases.

Alshut et al. [1] proposed learning based classification approach, which detects the embryo in the well, classifies its status and derives characteristic parameters for the regions. Expert labeled the data to the classes dead, alive or unknown. Chorion is segmented from background using adaptive threshold and features are derived based on intensity edge filter, and chorion size. Two most discriminative features are selected by MANOVA (multivariate analysis of variances). Using these two features,

the classification is performed utilizing a Bayesian classifier.

Stern et al. [52] focused on automatically detecting specific interest points in microscopy images. Images are manually annotated by expert to identify interest points coordinates and to train models able to predict those interest points in new, unseen images. The approach first extracts sub-windows (or patches) around points of interest and at other randomly chosen positions within images. The patches are described by various visual features, RGB, HSV, Grayscale, Edge strength, Local Binary Pattern. A classification or a regression model is built using these features. In the classification scheme, the model is trained to predict whether the central pixel of a sub-window is an interest point or not (a binary classification problem). In the regression scheme, the model predicts the distance between the central pixel of a sub-window and the interest point.

Vogt et al. [55] implemented a user guided image interpretation tool to generate rule-based hierarchical image segmentation. The zebrafish embryos are first segmented, and the large vessels and head structures are identified based on pixel values. The algorithm then removes small and isolated objects. Next, the remaining regions are re-processed to identify the head, dorsal aorta, and posterior cardinal vein. These pre-processed images were then used for blood vessel quantification. Numerical measurements of blood vessels development such as area, length, and shape are captured.

2.3.3 Contour Based

Liu et al. [40] have proposed to automatically quantify the neurons and somites in a large number of zebrafish images, as well as quantitative measurement of gene expression levels in zebrafish embryos. Neurons are segmented from edge image using circular Hough transform. RohonBeard (RB) sensory neurons are segmented using method based on intensity projection. Somites are recognized using gray level co-occurrence matrix, and edge direction histogram. ROI for genes quantification is established with region growing segmentation. Neurons quantification is based on its count. There is neuron loss in zebrafish embryos due to knockdown of AD-linked genes. Somites are quantified using features extracted from gray level co-occurrence matrix for detection of defective somites.

2.3.4 Curvature and Edge Based

Xu et al. [59] proposed an algorithm focused on detecting and quantifying pigments in zebrafish embryos. They automatically identify torso area through series of image processing steps. Steps include, finding single zebrafish in an image and then rotating the images so the embryo is positioned with its head region to the upper right corner and with its back pointing upward and abdomen downward to allow the algorithm to search the abdominal region based on the zebrafish anatomy. A watershed method is used to remove the head region from the torso that contains the pigmentation. Otsus method is applied on the torso for segmentation, which is quantified and used for toxicology.

Feng et al. [20] developed a 3D attributed vessel represent graph (AVRG) approach to reconstruct caudal vasculature of zebrafish embryo. The major steps of the framework include pre-processing, vasculature reconstruction, vascular structure quantification, and presentation. Pre-processing steps for vessel segmentation include: (1) Image enhancement; (2) Adaptive thresholding; (3) Vessels registration; (4) Edge tracking and curve fitting. The cross-sections of a segmented vessel are represented by ellipses. The surface of a blood vessel is represented by triangle meshes. Vasculature structure is reconstructed by locating junction points at the intersections of two or more vessels. Reconstructed vasculature structure is then utilized for quantification of the number and connectivity of the vessels, their size, length, and volume, as well as the distance between any two vessels.

2.3.5 Tracking Based

Bang et al. [6] developed an automated screening assay to detect hearing defects in zebrafish by monitoring their behavior after receiving a loud sound burst. The image was acquired immediately before the tone burst was subtracted from the image acquired immediately after the tone burst. If the zebrafish did not respond to the tone burst, due to hearing defects, the subtraction of its images produced an image approximating zero. If the zebrafish responds to the tone burst by moving to a different position, the subtraction result contains both positive and negative pixel values that can be detected by taking the absolute value of the resulting image and segmenting it by an appropriate threshold.

Kato et al. [29] developed computer image processing system for quantifying zebrafish behavior. Zebrafish images were extracted in real time from the original image as a binary image by removing a background of the aquarium before the fish are introduced. To maintain an effective frame rate that is high enough to capture the movement of the zebrafish, skipping search method is applied on data. In each frame, only one fourth of the pixels are “skippingly” examined, and then only the areas of interest that are recognized as zebrafish are analyzed in more detail in the two diagonal directions and in two crosswise directions with a continuous search. Chasing behavior of pairs of fish are quantitatively analyzed based on the positional coordinates of their center of gravity.

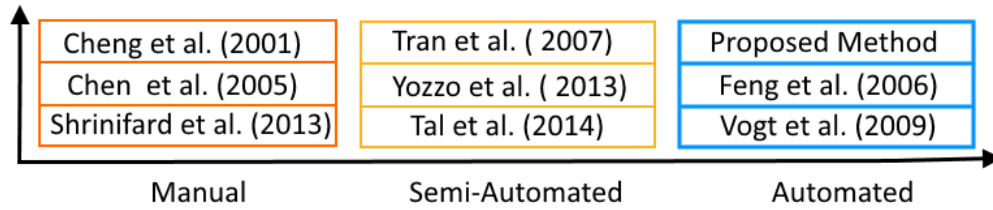


Figure 2.2: Overview of vasculature image analysis methods.

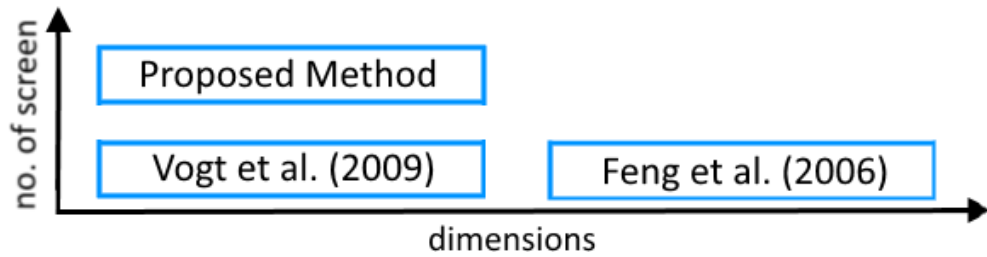


Figure 2.3: Overview of an automated vasculature image analysis methods.

Automated analysis of vasculature is still a budding field. Feng [20], and Vogt

[55], and proposed method are the only studies dedicated to automated analysis of vasculature blood vessel (fig. 2.2). The work presented in their papers has been performed on limited number of compounds (fig. 2.3); there is still a question about its scalability. Feng [20] performed 3d reconstruction of few vessels, where in our work we capture the dynamics of all ISVs. In this work, we have proposed a segmentation and quantification algorithm for ISV and CVP. Segmentation is a very crucial and important step for any kind of analysis. Interestingly, an automatic segmentation of ISV and CVP is still an untouched problem. Segmentation based either on thresholding or on edge detection of images are not sufficient to analyze ISV. ISV are highly diverse; they differ in size, shape, intensity, and occluded with noise. The orientation and amount of details of each embryo are different. One image has multiple embryos placed in various orientations. To quantify different regions, manual analysis requires one to draw ROIs for every image and then measure the features in each ROI. Often a human observer needs to rotate the images to the same orientation before drawing ROIs and quantifying pixel intensity, making their laborious work more tedious. In addition, manual analysis is subject to inter-observer variation and lacks repeatability.

In summary, although methods have been developed to process zebrafish images, as the applications of the zebrafish model expands, there is a demand for a variety of automated image processing algorithm. In this work we have presented a techniques and a framework for segmentation, representation, quantification and classification of the zebrafish ISV and CVP from images. The methodology presented is quantitative, can be utilized with wide varieties of toxin treated zebrafish, and capable of

quantifying changes in fine structure not quantifiable by the human eye. Relevance of our work is further amplified, by having a model to be able to determine safe dosages for assays of compounds. In next few chapters we will present detailed description of our approach.

Chapter 3

Zebrafish Segmentation

Zebrafish vessel segmentation can be utilized for delineation of morphological attributes of blood vessels, such as length, width, area and/or angles for disease modeling, drug screening, toxicology, and gene expression evaluation. Automated segmentation can help screen larger populations for vessel abnormalities. In this chapter we will describe segmentation algorithm for zebrafish embryo, its ISV and CVP.

3.1 Embryo Segmentation

An image may have multiple zebrafish embryos as shown in fig. 3.1A. Our aim is to capture the complete anatomic structure of each zebrafish. Due to the limited resolution of a microscopic image, adjacent objects may appear to be touching, each other or boundaries. The algorithm presented detects the valid zebrafish embryos in an image and excludes rest. For, this we smooth the image with Gaussian filter.

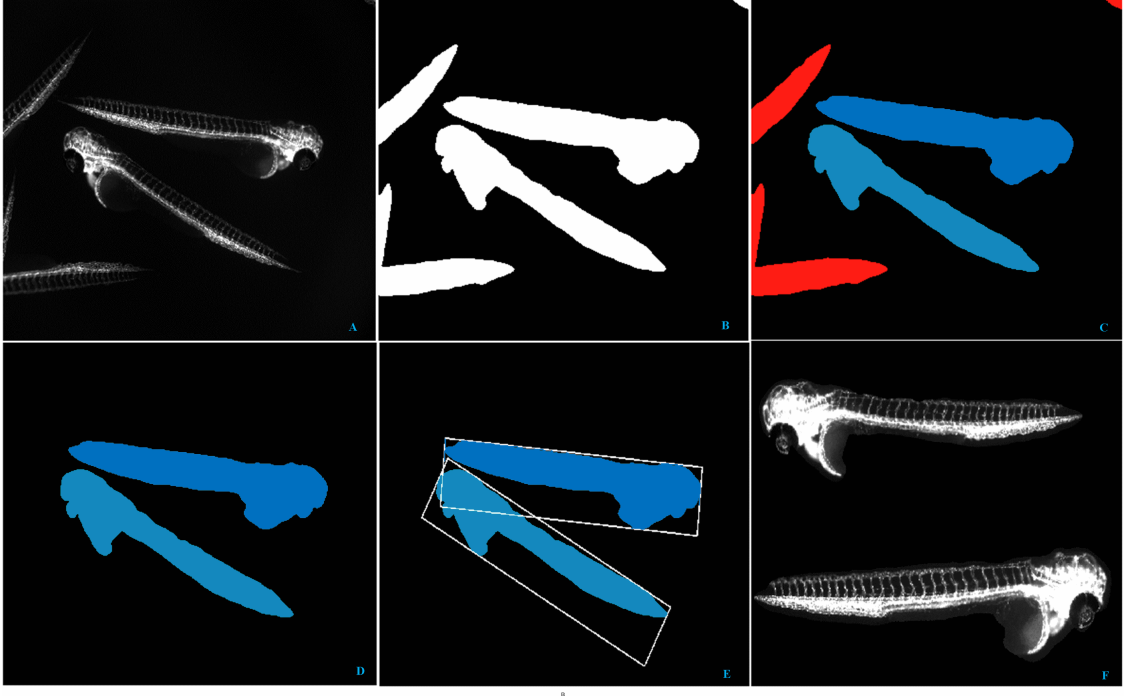


Figure 3.1: Figure 3.1A, 3.1B shows multiple embryos in an image, and corresponding segmented image. In fig. 3.1C embryos shown in red are outliers and in blue are valid embryos. After extracting the relevant embryo, we rotate each embryo by an angle that major axis of bounding box makes with the horizontal axis. If bounding box intersect, we choose the largest connected component.

Smoothing reduces the finer details of an embryo, hence producing a uniform foreground (zebrafish embryo) against uniform background. The background and the foreground are separated, hence we can perform thresholding with triangle method [63]. The method is based on a histogram of image intensities. The triangle method constructs a line between the histogram peak and the farthest end of the histogram. The threshold is the point of maximum distance between the line and the histogram.

Thresholding is followed by connected component labeling for extracting the each

of the zebrafish embryos. Let $C_i, i = 1, 2, \dots, n$ be the labeled component sets in an image I with size $w \times h$. In each connected set, let the labeled pixels be given by $C_i = ((x_{ij}, y_{ij}), j = 1, 2, \dots, m)$. Since in each image we are looking for a complete anatomic structure, we discard labeled component C_i in I under 3 conditions (fig. 3.1(c)):

- (i) if blobs are touching the boundaries of image.
- (ii) if number of labeled pixels in connected set is above certain threshold
- (iii) if number of labeled pixels in connected set is below certain threshold.

$$C_i = \begin{cases} 0, & \text{if } (x_{ij} == w || y_{ij} == h || x_{ij} == h || y_{ij} == w) j \in (1, 2, \dots, m) \\ 0 & \text{if } size(C_i) > upper \\ 0, & \text{if } size(C_i) < lower \\ 255, & \text{if } otherwise. \end{cases} \quad (3.1)$$

where, upper and low are calculated based on the dataset. Values for upper and lower is easy to estimate. Upper is approximately 1.5 times, and lower is $\frac{1}{3}$ of the average size of zebrafish embryo.

As the orientation of zebrafish embryo varies in the image, the next step of extraction will use the shape information of the embryo. The position of the zebrafish embryo in the image is normalized, to place longest axis of fitted ellipse parallel to horizontal axis. Moment invariants allow us to find the best fitting ellipse for a target object. For an image I , the moment of order $(p + q)$ is defined as:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q I(x, y) dx dy \quad (3.2)$$

$$p, q = 1, 2, \dots$$

We can simplify Eq. (3.2) further,

$$m_{pq} = \sum_p \sum_q x^p y^q \quad (3.3)$$

$$p, q = 1, 2, \dots$$

To further reduce Eq. (3.3) over a connected region C , in an image:

$$m_{pq} = \sum_C x^p y^q \quad (3.4)$$

$$p, q = 1, 2, \dots$$

Let (x_c, y_c) be the centroid of region C . The central moments are defined as:

$$\mu_{pq} = \sum_C (x - x_c)^p (y - y_c)^q \quad (3.5)$$

From the eq.(3.4), m_{00} , gives the area of C . The centroid $c = (x_c, y_c)$ and angle θ between the largest axis and the x-axis can be calculated as follows:

$$x_c = \frac{m_{10}}{m_{00}}$$

$$y_c = \frac{m_{01}}{m_{00}}$$

$$\theta = \frac{\arctan(\frac{b}{a-c})}{2} \quad (3.6)$$

where a, b, c cane be defined as:

$$a = \frac{m_{20}}{m_{00}} - x_c^2$$

$$b = 2(\frac{m_{11}}{m_{00}} - x_c y_c)$$

$$c = \frac{m_{02}}{m_{00}} - y_c^2$$

Connected components are rotated around its $c = (x_c, y_c)$, with the angle θ (fig. 3.1D). Since the mask obtained likely contains small connected components, we find the relevant embryo by selecting the largest connected component in the mask. The steps are exemplified in fig. 3.1D, 3.1E.

3.2 Region of Interest Detection

After isolating zebrafish embryo's complete anatomical structure, we proceed to detect region of interest (ROI). We are interested in ISV and CVP region. We bisect zebrafish into ISV + DLAV region and tail + head region. The vessels dorsal aorta, tail and head structure have high intensity value, as they are large in size, and tightly

knit together. On the other hand, ISV and DLAV are very thin and located at distance, this results in ISV and DLAV to having low intensity as compared to large vessels as depicted in fig. 3.2.

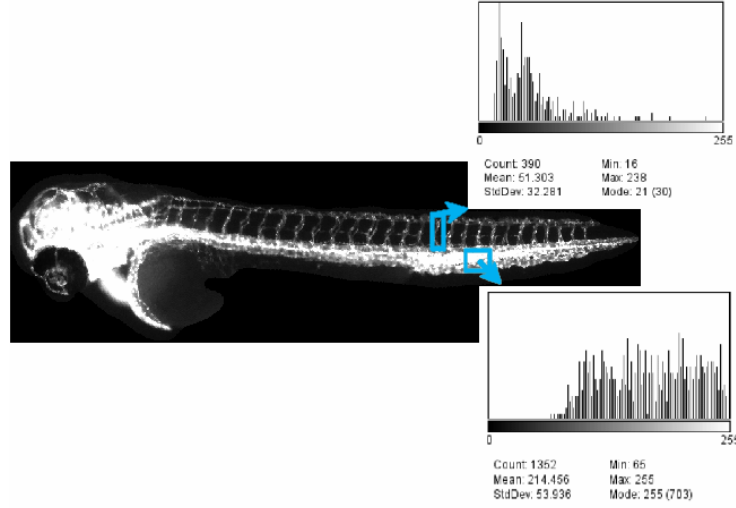


Figure 3.2: The difference in intensity profile of ISV + DLAV region compared to tail + head region. The difference in intensity profile average value is more than 150.

We enhanced contrast, to further exemplify differences in their intensity value. Intermodos threshold based on bimodal distribution is well suited for separating ISV + DLAV from rest of embryo. The histogram is smoothed using a running average of size 3, until there are only two local maxima. The threshold value is the average of two peaks.

Although above step separates zebrafish embryo into two region. But there is still some isolated portion of head, and yolk region left to be masked out. Specifically for tail + head region, we want to mask out only tail region. Skeleton of segmented image is used to find head, and yolk position. We can have two possibilities for head



Figure 3.3: Procedure for finding head yolk position. Figure 3.3B is obtained by masking out ISV + DLAV region from fig. 3.3A. Skeleton (fig. 3.3C) of mask is used to find the position of head and yolk and point of intersection of head and yolk region, and hence remove the rest of data.

and yolk. Head can be on right side of image (fig. 3.3D) or left side of image. Yolk can face up or upside down (fig. 3.3D). We scan the skeleton image from top (from origin) along image width and store pixel values in pixelLocUp. Similarly, we scan the skeleton image from bottom (from height) along image width and store pixel values in pixelLocDown. Next we compute the euclidean distance between each pair of entries in pixelLocUp, and pixelLocDown as described in algorithm 1. Next step is to find the position of yolk, if yolk is facing up or upside down (algorithm 2). Lastly for finding head location we use steps described in algorithm 3.

After eliminating yolk and head region, we are confined with two ROI, one enclosing ISV and other enclosing tail (fig. 3.4E, 3.4D).

3.3 Intersegmental Vessel Segmentation

Next we need to extract ISV from image. Extraction of ISV is challenging since it varies in size, local contrast is unstable, high curvature, and noisy. The ability to segment and quantify vessels, especially with smaller diameter, is limited by noise and contrast.

Algorithm 1 Algorithm for finding distances between skeleton point from top and bottom.

```
for  $i = 1$  to  $height$  do
  for  $j = 1$  to  $width$  do
    if ( $Image(i, j) == 255$ ) then
       $pixelLocUp = [pixelLocUp; ij]$ 
      break
    end if
  end for
end for
for  $i = height$  to 1 do
  for  $j = width$  to 1 do
    if ( $Image(i, j) == 255$ ) then
       $pixelLocDown = [pixelLocDown; ij]$ 
      break
    end if
  end for
end for
▷ pairwise Euclidean distance for  $pixelLocUp$ ,  $pixelLocDown$ 
 $pixelLocUpDist = dist(pixelLoUp, pixelLocUp)$ 
 $pixelLocDownDist = dist(pixelLocDown, pixelLocDown)$ 
```

Algorithm 2 Algorithm for finding yolk position, and cleaning region containing yolk

```
▷ Yolk is facing up
if ( $\max(pixelLocUpDist) > \max(pixelLocDownDist)$ ) then
     $top = true$ 
     $location = find(pixelLocUpDist == \max(pixelLocDist))$ 
    ▷ Yolk is facing down
else
     $top = false$ 
     $location = find(pixelLocDownDist == \max(pixelLocDownDist))$ 
end if
▷ clean image above pixelLocUp
if ( $top == true$ ) then
    for  $i = 1$  to  $pixelLocUp.size$  do
        for  $j = 1$  to  $pixelLocUp[i][1]$  do
             $Image(j, pixelLocUp[i][2]) = 0$ 
        end for
    end for
    ▷ clean image below pixelLocDown
else
    for  $i = 1$  to  $pixelLocDown.size$  do
        for  $j = height$  to  $pixelLocDown[i][1]$  do
             $Image(j, pixelLocDown[i][2]) = 0$ 
        end for
    end for
end if
```

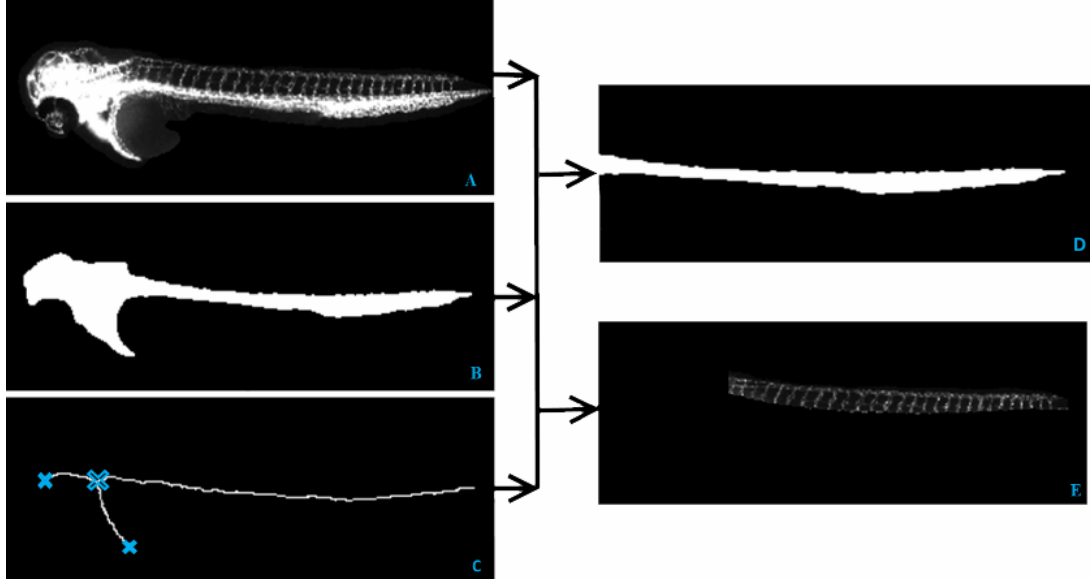


Figure 3.4: Procedure for extracting ISV + DLAV region and tail region. Figure 3.4D is obtained by masking segmented region shown in fig. 3.4b and removing head region. Figure 3.4E is obtained by masking out segmented region fig. 3.4B from 3.4A. Figure 3.4C is used to remove isolated region from head and yolk.

3.3.1 Eigen Analysis

The eigenanalysis of Hessian matrix is a crucial method for vessel detection. Hessian matrix components approximate 2nd order derivatives, and hence represents the shape information in terms of both, quality and quantity. Particularly interesting are the eigenvalues and eigenvectors of Hessian matrix. In this work, we are presenting the ISV detection method based on eigenanalysis of image Hessian matrix combined with multiscale image analysis. A segmentation method incorporating vessel direction and the eigenvector of the Hessian matrix is used for vessel detection and to obtain a segmented vessel tree.

The common approach to analyze local shape of a $2D$ image I is to consider its

Algorithm 3 Algorithm for finding head position and clean image containing head region

```

▷ head is on right side
if ( $dist(location, 0) > dist(location, width)$ ) then
  for  $i = 1$  to  $height$  do
    for  $j = location.y$  to  $width$  do
      ▷ clean image from head location to width
       $Image(i, j) = 0$ 
    end for
  end for
▷ head is on left side
else
  for  $i = 1$  to  $location.x$  do
    for  $j = 1$  to  $location.y$  do
      ▷ clean image from origin to head location
       $Image(i, j) = 0$ 
    end for
  end for
end if

```

Taylor Expansion in the neighborhood of a point x_0 .

$$I(x + x_0) \approx I(x_0) + \Delta x^T \nabla I(x_0) + \Delta x^T \nabla H(x_0) \Delta(x) \quad (3.7)$$

where ∇I is the gradient vector and H denote Hessian matrix of second-order partial derivatives of an image I .

$$H = \begin{bmatrix} \frac{\partial^2 I}{\partial x^2} & \frac{\partial^2 I}{\partial x \partial y} \\ \frac{\partial^2 I}{\partial y \partial x} & \frac{\partial^2 I}{\partial y^2} \end{bmatrix} \quad (3.8)$$

For a given pixel of an image a Hessian matrix is composed of second order partial derivatives. Second order derivatives locally approximates the structure of an image.

To solve these differential operators in a well-posed fashion we use concepts of linear scale space theory [33]. In this framework differentiation is defined as a convolution with derivatives of Gaussian:

$$\frac{\partial I}{\partial x} = I(x) * \frac{\partial G(x, \sigma)}{\partial x} \quad (3.9)$$

where the Gaussian is defined as:

$$G(x, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (3.10)$$

where the parameter σ is the standard deviation of the Gaussian kernel, it controls the scale and the smoothing. With increasing the scale the image gets less detailed. The scale allows to search for objects of similar dimensions as the chosen scale. Let the eigenvalues of the Hessian matrix be λ_1 and λ_2 ($\lambda_1 \geq \lambda_2$) and corresponding eigenvectors be v_1 and v_2 . The eigenvalues λ_1 and λ_2 of the Hessian matrix can be calculated through the following equation:

$$\Delta(H - I * \lambda) = 0 \quad (3.11)$$

where I is an identity matrix and λ is set of eigenvalues.

The eigenvalue of Hessian matrix are called principal curvature. The eigenvalues of the Hessian matrix evaluated at a point encodes the local morphology in all direction in terms of curvature [17]. The highest eigenvalue λ_1 corresponds to maximum change in curvature, and corresponding eigenvector gives a direction of maximum change [24]. The eigenvector v_2 is orthogonal to the vector v_1 .

Since vessels appear in different sizes it is important to introduce a measurement scale which varies within a certain range, which matches the local feature size. Evaluating the eigenvalue in a scale-space shows local morphology change with scale. The scale-space theory introduced by Lindeberg [38] uses for the purpose of detail removal a convolution with a Gaussian kernel. We denote the eigenvectors of the Hessian, $H(x, \sigma)$, at sampling location x and scale σ as $v_1(x, \sigma), v_2(x, \sigma)$. The eigenvalues are denoted by $\lambda_1(x, \sigma), \lambda_2(x, \sigma)$.

ISVs are extracted by tracing the direction of maximum eigenvalue normal to x -axis. Our response function is calculated from eigenvalue evaluated by extracting the direction of maximum eigenvalue. For each location, we compute the angle between eigenvector corresponding to maximum eigenvalue and vector X in x (horizontal) direction.

$$h(v_1(x, \sigma), X) = \frac{v_1(x, \sigma) \cdot X}{\|v_1(x, \sigma)\| \|X\|} \quad (3.12)$$

We analyze the response from each scale within pre-determined range, and select the maximum eigenvalue based on h . Since ISV in images exists with varying size, it is logical to perform feature extraction at the scale which matches the ISV size. The strongest response (with respect to σ) over scales directly corresponds with width of object [38]. For this purpose, our parameter selection is based on prior measurement of width of objects. From our data set, we manually measure the width of ISV. We randomly select $\frac{1}{15}$ images from our dataset, and measure the width of vessel, and use the range of values for scale parameter. From our observation, we found radius

of ISV to be in range $[0.5, 3]$.

$$r(x) = \begin{cases} \max_{\sigma} \lambda_1(x, \sigma) & \text{if } \beta \leq h(x, \sigma) \leq \alpha \\ 0 & \text{if } otherwise \end{cases} \quad (3.13)$$

All pixels which satisfy (3.13) belongs to ISV. Parameter α is chosen to be 45 and β is chosen to be 135. We want to remove all the pixels which are parallel or approximately parallel to horizontal axis. For our purposes, tolerance of ± 45 works well. We binarize ISV (say b_r) based on value from proposed adaptive Niblack thresholding.

3.3.2 Adaptive Niblack thresholding

Original Niblack's algorithm [46] is a local thresholding method based on the calculation of the local mean and of local standard deviation. The threshold is calculated by the formula:

$$T = m + k * s \quad (3.14)$$

where, m and s are the local mean and standard deviation of pixel values in local neighborhood, respectively. The size of the neighborhood should be small enough to preserve local details, but at the same time large enough to suppress noise. k is the correction factor to adjust how much of the total object boundary is taken as a part of given object.

Instead of having fixed k , we use scale parameter which matches the ISV size. We can derive scale information from equation (3.13). We return for every pixel the value of the scale (σ) with the maximum eigenvalue from (3.13). We utilized window size of $[125, 25]$ approximately based on box enclosing ISV.

3.3.3 Linking vessels

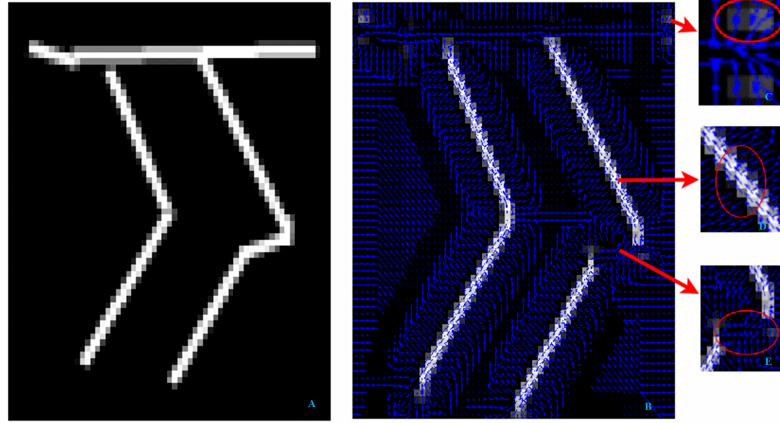


Figure 3.5: Figure 3.5A is used in place of original image to depict the structure similar to of ISV. 3.5A shows the response of maximum eigenvalue corresponding to $\text{scale} = 1$ for principal direction. In top right fig. 3.5C, shows few isolated regions from DLAV. These regions are outliers and will be removed. Figure 3.5D perfectly captures the response corresponding to principal direction. Figure 3.5E shows the region whose principal direction is lies outside the $[\alpha, \beta]$, hence no response is generated. These regions are merged.

Some ISV might appear disconnected, or some isolated response from DLAV region might be outliers as shown in fig. 3.5C. This happens when region along center of ISV, appears parallel to x-axis i.e. the angle of principal direction with horizontal axis does not fall in $[\alpha, \beta]$ as depicted in fig. 3.5E. We connect the disjoint

ISV's and remove outlier from DLAV based on the feature set extracted from each ISV region.

The broken ISV when region along center of ISV, appears parallel to x-axis i.e. the angle of principal direction with horizontal axis does not fall in $[\alpha, \beta]$ as depicted in fig. 3.5E. Here we use morphological operation to link the broken vessels segmented. We compute the maximum response from all vessels using (3.13). We binarize $r_a(x)$ based on value from Adaptive Niblack thresholding and call it b_{r_a} . Parameter k is the value of the scale (sigma) with the maximum eigenvalue from (3.15).

$$r_a(x) = \max_{\sigma} \lambda_1(x, \sigma) \quad (3.15)$$

At each potential vessel pixel from b_r , we apply morphological closing using linear structure. The direction of the linear structure is perpendicular to the x-axis direction. We set the length of the linear structure as 10 pixels in the experiments. We retain all those potential vessel pixel from b_r that belongs to b_{r_a} .

3.3.4 Post Processing

There are some isolated response from DLAV region which are outliers as shown in fig. 3.5C. We follow similar approach as for linking vessels, but instead we work with maximum response from vessels from DLAV region (3.16). We binarize $r_d(x)$ based on value from Niblack thresholding and call it b_{r_d} . Parameter k is the value of the scale(sigma) with the maximum eigenvalue from (3.16).

$$r_d(x) = \begin{cases} \max_{\sigma} \lambda_1(x, \sigma) & \text{if } \alpha < h(x, \sigma) > \beta \\ 0 & \text{if } otherwise \end{cases} \quad (3.16)$$

At each potential vessel pixel from b_{r_d} , we apply morphological closing using linear structure. The direction of the linear structure is parallel to the x-axis direction. We set the length of the linear structure as 10 pixels in the experiments. We remove all those potential vessel pixel from b_r that belongs to b_{r_d} .

3.4 Caudal Vein Plexus Segmentation

After extracting ROI enclosing CVP as shown in fig. 3.4d. We use structural properties of CVP for segmentation.

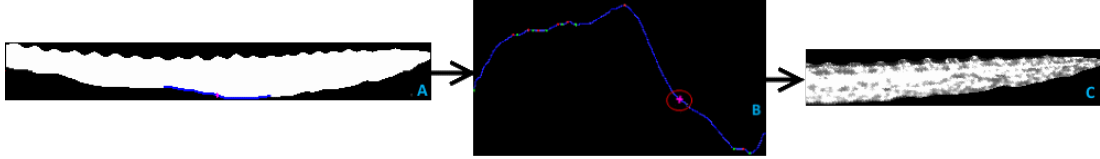


Figure 3.6: Procedure for CVP segmentation. Figure 3.6A is a region enclosing CVP. Curve C in blue represents lower boundary of tail region along midpoint of tail. Figure 3.6B C is smoothed using moving average filter. Pink points represents local maxima and green points represents local minima. Circled point represents the CVP region start location. CVP start position is the mid point of the local curve with largest steepness. Figure 3.6C is the segmented CVP obtained by masking out segmented region from fig. 3.3A with fig. 3.6A from start location of CVP.

The intuition behind our method is that there is a high gradient associated with transition of PCV region to CVP region in zebrafish. For this purpose, we trace the curve C along mid point m of lower boundary of the tail. Lets C be represented as:

$$C = ((x_i, y_i), i = 1, 2, \dots, n) \quad (3.17)$$

We compute the slope of the curve as:

$$S = (\frac{dy_i}{dx_i}, i = 1, 2, \dots, n) \quad (3.18)$$

S is smoothed using moving average filter. Let H and L denote the set of local maxima and local minima along S. Let $v = (h_{i0}, l_{j0})$ denote local maxima and local minima, respectively, such that:

$$\text{mod}(h_{i0} - l_{j0}) \geq \text{mod}(h_i - l_j), \forall h_i \in H, \forall l_j \in L, j - 1 \leq i \leq j$$

Point p where the CVP starts is given by:

$$p = \begin{cases} \frac{h_{i0} + l_{j0}}{2} & \text{if } H \neq \emptyset, L \neq \emptyset \\ m & \text{otherwise} \end{cases} \quad (3.19)$$

p signifies the mid point along the curve with maximum change in slope. CVP is obtained by masking region shown in fig. 3.3A from segmented region shown in fig. 3.6C from start location p of CVP.

Chapter 4

Feature Extraction

A toxicology screening can be observed as a typical two-class classification problem, consisting of feature extraction and classification process. In this chapter we will explain feature extraction algorithm for ISV and CVP.

4.1 Intersegmental Vessel Feature Extraction

ISV features are defined in terms ISV count; an average distance between ISV, total area of ISV; and an average ISV length. All terms were defined in pixels. Previous research have shown the quantification terms used in this work can use to analyze toxicity effect in ISV [20], [54], [55].

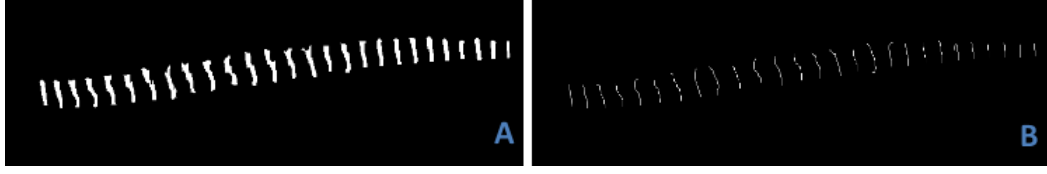


Figure 4.1: Skeleton of ISV, used for calculating length of ISV.

Our quantification procedure consists of reporting features: ISV count; average distance between ISV; total area of ISV; and average ISV length. We calculate these features for each of embryo. Features are obtained by performing connected component analysis. Number of connected component corresponds to ISV count. ISV distance is computed as the average of euclidean distance between ISV centroid, and the centroid of two adjacent neighboring ISV. Total area occupied by all components is ISV area. ISV length is computed by analyzing the skeleton of each of connected component. Image skeleton is obtained by thinning procedure explained in [36]. The general idea is to erode the ISVs surface iteratively until only the skeleton remains. The structure obtained is shown in fig. 4.1(b). We decompose each individual vessel segments as individual graph for calculating ISV length. We perform graph analysis on each vessel and compute end points (pixels with less than 2 neighbors). For calculating the length of ISV we find the number of pixels connecting the end points. Average ISV length is computed by averaging over all graphs. Feature vector is formed by concatenating ISV count; average distance between ISV; total area of ISV; and average ISV length.

4.2 Caudal Vein Plexus Feature Extraction

The study of the zebrafish CVP and its utilization as a model for screening assay has not been as prevalent as the ISV, and very few studies have attempted to identify regulators of CVP development. Previous research has primarily focused on using color, texture [54] and morphological changes [20] for toxicity analysis. Shape information, however, can be another attribute that can be evaluated. A shape descriptor can capture the outline of the shape that other color or texture-based descriptors may not be able to capture. Hence by quantification of the shape of CVP, and comparing it against that of a healthy embryo, we can identify changes in CVP due to exposure to toxins. This section will discuss utilization of gray level co-occurrence matrix (GLCM), histogram of oriented gradients (HOG), co-occurrence of histogram of oriented gradients gradient (Co-HOG) and weighted co-occurrence histogram of oriented gradients (gCo-HOG) for shape modeling.

4.2.1 Gray Level Co-occurrence Matrix

GLCM is a very well known texture analysis method [25], [26]. GLCM represents the angular and spatial over an image sub region. GLCM entries represents an estimate of the probability that two pixels with a specified displacement, d , and an angle, θ occurs in an image. Statistical measures are derived from the co-occurrence matrix. The features we used include: energy, homogeneity, correlation, and contrast. For varying choices of d and θ , we obtain a separate GLCM. The GLCM is implemented with certain degree of rotation invariance, which is achieved by combining the results

from various angles. Angles 0, 45, 90, 135 are considered as effective choices. We analyzed results with varying value of k and obtained best performance with 9 pixel displacement. The results is combined by averaging the GLCM for each angle and concatenating energy, homogeneity, correlation, and contrast into one feature vector.

4.2.2 Histogram of Oriented Gradients

HOG descriptor first proposed by Dalal and Triggs [16] has been used in many different problems in computer vision, such as pedestrian detection [56], face recognition [18], object recognition [9] and text recognition [57]. HOG features are extracted from image by first computing gradient orientation at every pixel. Orientations of gradients are quantized into histogram bins and each bin has an orientation range. Image is divided into blocks and in each block a histogram of oriented gradients is computed. HOG feature consists of concatenation of histogram of oriented gradients over all blocks.

In our case, HOG are computed on edge contours extracted using the canny edge detector (fig. 4.2(center)). The gradients are computed using a Sobel mask. The HOG descriptor is quantized into K orientation bins, each over an orientation range. The weight from each contour point depends on its gradient magnitude and is added to its orientation bin. Each bin in the histogram represents the sum of gradient magnitudes that have orientations within a certain angular range.

HOG are invariant to 2d rotation and illumination variations. On the other hand, HOG captures orientation of only isolated pixels, ignoring spatial relationship among

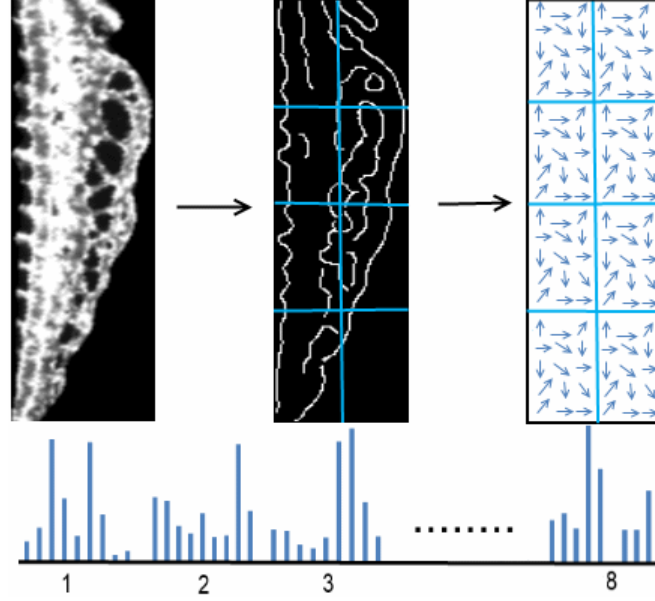


Figure 4.2: Extraction of HOG features from a CV image. (left) Original image. (center) Edge contours are extracted using an edge detector, image is divided into 8 blocks. (right) HOG vector is extracted from each sub region. (bottom) Concatenation of all the HOG vectors to obtain the HOG features for image.

neighboring pixels. Co-HOG captures spatial information and is more powerful in describing local structure. With spatial structure, more shape information of object can be captured.

4.2.3 Co-occurrence of Histogram of Oriented Gradients

Co-HOG is an extension of HOG. Co-HOG captures spatial information by measuring probability of oriented gradients between pairs of pixel. A pixel pair can be represented by an offset (x, y) , which captures the spatial relationship of the two points. As shown in fig. 4.3(left), we define 31 offsets including zero offset for a given point. The black pixel in the center is the pixel under consideration and the

neighboring blue pixels are with different offsets. Each neighboring pixel in blue color forms an orientation pair with the center black pixel and accordingly votes to the co-occurrence matrix. For each pixel in the image of size $M \times N$, the orientations ranging between $[0, 360]$ are quantized into eight orientation bins. Co-occurrence matrix at a specific offset (x, y) is defined as:

$$C_{x,y}(i, j) = \sum_p \sum_q \begin{cases} 1, & \text{if } I(p, q) = i, I(p + x, q + y) = j \\ 0, & \text{otherwise} \end{cases} \quad (4.1)$$

The Co-HOG descriptor is formed by concatenation of components of the co-occurrence matrix of each offset including offset 0. Co-HOG obtains 31 co-occurrence matrices. There are $8 \times 8 = 64$ elements in the co-occurrence matrix (fig 4.3(right)). The co-occurrence matrix calculated with zero offset has only 8 values.

Co-HOG extracts both local and global shape information, with varying offset sizes.

One potential limitation of co-occurrence histograms of oriented gradients is that both strong and weak gradients provide the same contribution in representing the spatial structure. To address this limitation, we investigate the inclusion of gradient strength in the generation of the histogram.

As shown in [47], we treat gradient as forces and use vector addition to combine forces using:

$$C_{x,y} = C_{x,y} + \|g_1 + g_2\| \quad (4.2)$$

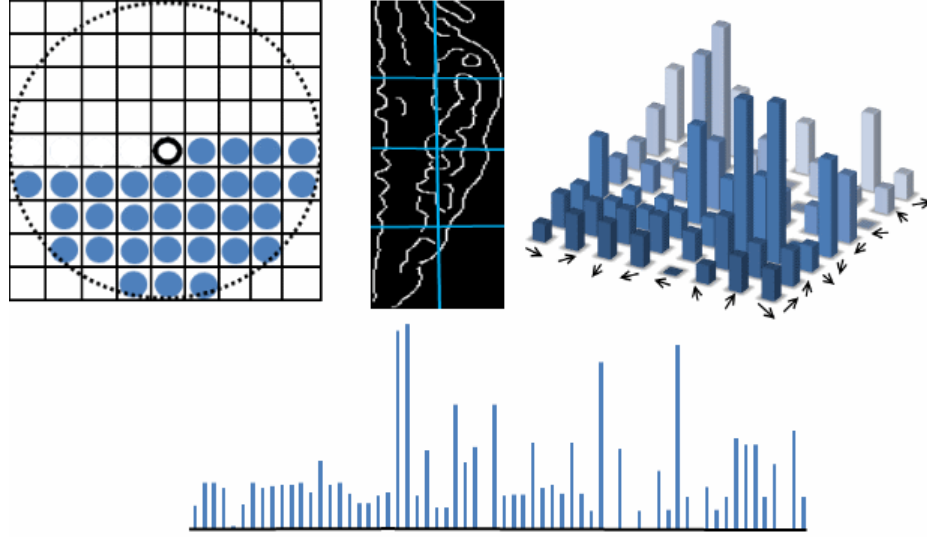


Figure 4.3: Extraction of Co-HOG features from a CV image. (left) Pixel offset. (center) Edge contours are extracted using an edge detector, image is divided into 8 blocks. (right) Co-HOG vector is extracted from one region. (bottom) concatenation of all the Co-HOG vectors to obtain the Co-HOG features for image.

where C is the co-occurrence matrix at a specific offset (x, y) as defined in equation 4.1, g_1 is the gradient magnitude at location (p, q) , and g_2 is the gradient magnitude at location $(p + x, q + y)$. The gCo-HOG feature descriptor of the whole image can then be constructed by concatenating all the regions features. The gCo-HOG is normalized to sum to unity.

4.3 Vasculature Time-lapse Features

There are various challenges associated with live imaging of zebrafish. One such complication is, zebrafish embryos twitch, so that growth and motion of embryos are not uniform across the 3D field. Imaging captures the dynamics of blood vessel

formation over time. In order to quantify these morphological changes, i.e. how much vascular structure has changed over a period of time, it is necessary to compensate for any movements caused by the growing embryo. Thus, in order to record the temporal changes occurring due to vessel growth, it is necessary to establish spatial correspondence between blood vessels that may appear displaced due to embryo movement. Thus quantification of temporal vascular growth can be seen as a problem of image registration. we use a non-rigid registration approach to align images. Non-rigid mapping is based on complete correspondence of images and includes a deformation model as the underlying transformation. We utilize free-form deformation based on B-splines for growth monitoring, and use intensity differences as a similarity measure.

4.3.1 Vasculature Registration

The goal of image registration is to find an optimum mapping that aligns two images. Since deformations are important for quantification of changes in images, it is necessary to find a mapping between two time points as accurately as possible. In our case we need to quantify blood vessel growth independent of motion artifacts. Hence we need a registration approach that establishes vessel correspondence between successive time frames. Affine and rigid registration approaches are mainly based on local stretching of images, and hence do not adequately capture structural changes.

Many applications in medicine require that object is modified in global scale.

Therefore, we have used free-form deformation (FFD). Free form deformation deforms an object by warping the image geometry in which the object is localized. The nature of deformation varies widely across different time points; hence it is difficult to use traditional B-spline registration, which is based on many parameters. If we only select few parameters, an approximate match can be obtained, whereas many parameters incur added computational costs. Hence, we have used FFD based on hierarchical B-splines for multilevel nonlinear registration [35]. The underlying idea of FFD is based on deforming an object by manipulating a mesh of 2D points.

Let $\Omega = \{(x, y) \mid 0 \leq x \leq X, 0 \leq y \leq Y\}$ be the domain of $2d$ points. Let Ω denote a $n_x \times n_y$ mesh of control points with uniform spacing Δ . Let Ω_{ij} be the value of ij -th control point located at (i, j) . The FFD f can be written as:

$$f(x, y) = \sum_{m=0}^3 \sum_{l=0}^3 B_l(s) B_m(t) \Omega_{i+l, j+m} \quad (4.3)$$

where, $i = \left\lfloor \frac{x}{n_x} \right\rfloor - 1$, $j = \left\lfloor \frac{y}{n_y} \right\rfloor - 1$, $s = \frac{x}{n_x} - \left\lfloor \frac{x}{n_x} \right\rfloor$, $t = \frac{y}{n_y} - \left\lfloor \frac{y}{n_y} \right\rfloor$. B_l and B_m are uniform cubic B-spline basis function defined as:

$$B_0(s) = \frac{(1-s)^3}{6} \quad (4.4)$$

$$B_1(s) = \frac{(3s^3 - 6s^2 + 4)^3}{6} \quad (4.5)$$

$$B_2(s) = \frac{(-3s^3 - 3s^2 + 3s + 1)^3}{6} \quad (4.6)$$

$$B_3(s) = \frac{s^3}{6} \quad (4.7)$$

They weigh the contribution of each control point to $f(x, y)$ based on its distance to (x, y) . The problem of deriving function f is reduced to solving for the control points in Ω . The control points in Ω behave as parameters for transformation. The degree of non-rigid deformation depends highly on the size of control points. Dense mesh can increase number of degrees of freedom, hence providing more flexibility and consequently an increase in computational complexity. In order to achieve the tradeoff between degree of freedom and computational complexity, we use a hierarchical multiresolution [35] approach, in which the resolution of the control mesh is increased, along with the resolution of image in an iterative loop. Consider a hierarchy of control mesh $\Omega_0, \Omega_1, \dots, \Omega_h$ overlaid on domain Ω . For simplicity, we will assume decreasing space as we move from Ω_s to Ω_{s+1} . Similar to above, we will have FFD $f_s(x, y)$ for each control mesh. Their sum defines the overall transformation model.

$$f(x, y) = \sum_{s=1} f_s(x, y) \quad (4.8)$$

Calculation of FFD for each control lattice introduces a significant overhead. To avoid this overhead, B-spline refinement can be applied hierarchically to control mesh. This can be achieved by B-spline subdivision algorithm. In this case, the control point mesh at level s is generated by using the mesh control points from level $s - 1$. More details can be found in [22]. To achieve correspondence between two images acquired at different time points, we have used a similarity criterion based on sum of squared difference (SSD):

$$SSD = \frac{1}{n} \sqrt{(\sum (I(t_0) - T(I(t))))^2} \quad (4.9)$$

4.3.2 Vasculature Morphology

Blood vessel growth is quantified by recording the temporal occurrence of differences in pixel intensities along registered vascular structures. Registration is performed on pair of consecutive images (I_t, I_{t+1}) . I_t is warped to I_{t+1} to give I_{reg_t} . Measuring the difference (pixel changed) between I_{t+1} and I_{reg_t} provides a standard for change quantification. This protocol is shown in fig. 4.4

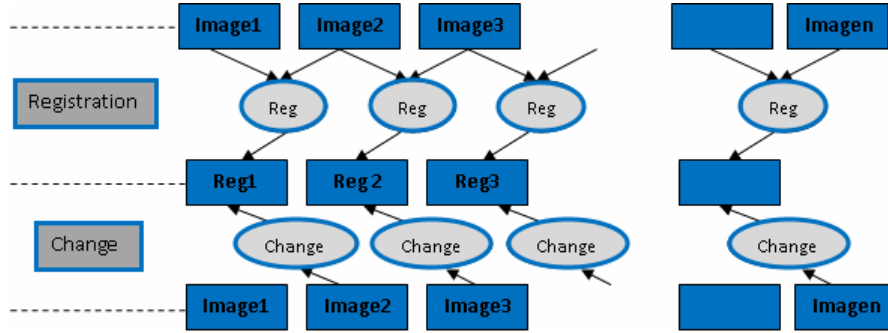


Figure 4.4: For a given sequence of images, image i is registered to $(i + 1)$. Measuring the difference between registered images i and $(i + 1)$, provides the change in number of pixels.

, where Δ_t is the change in number of pixels at time t .

$$\Delta_t = I_{t+1} - I_{reg_t} \quad (4.10)$$

The change measured over the entire imaging cycle can be characterized by the

function $f(\Delta_t)$ given by:

$$f(\Delta_t) = \{\delta_t\}_{t=1}^m \tag{4.11}$$

In next chapter, we will give details corresponding to zebrafish data used for experiments.

Chapter 5

Materials

Zebrafish embryo images are collected at Center for Nuclear Receptors and Cell Signaling, Department of Biology and Biochemistry, University of Houston. Zebrafish (*Danio rerio*) are reared and maintained at 28.5° C as previously described [58], and in accordance to the standard operating protocols approved by the Institutional Animal Care and Use Committee at University of Houston. A stable line of Tg(kdrl:EGFP)*mitfa*^{b692} is generated by crossing Tg(kdrl:EGFP) with *mitfab*^{692/b692} (Zebrafish International Resource Center, Eugene, OR) to facilitate GFP visualization without obstruction from melanophores. Embryos are collected from natural mating and staged accordingly [31].

Two dimensional data is acquired for ISV and CVP analysis. Three dimensional time lapse data is acquired to study overall growth in vasculature of zebrafish.

5.1 Two Dimensional Data

Zebrafish embryos are treated with chemicals selected from phase I of ToxCast chemical library (<http://www.epa.gov/ncct/toxcast/chemicals.html>) and solubilized in dimethyl sulfoxide (DMSO).

5.1.1 Chemical Treatments

Tg(kdrl:EGFP) *mitfa*^{b692} embryos are harvested in a petri dish after mating. At approximately 3 hpf, embryos are sorted and placed in 6-well plates (n = 20), followed by a single chemical treatment without renewal at 100 nM, 250 nM, 500 nM, 1 μ M, 10 μ M and 20 μ M, unless otherwise noted. Working dilution stocks of all chemicals are made such that the final concentration of the vehicle DMSO is at 0.1%. Each well contained a final volume of 3 mL of embryo medium, E3 (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄). Control embryos are treated with 0.1% DMSO. The embryos are incubated at 28.5° C until 72 hpf, during which they are assessed for vascular perturbations.

5.1.2 Imaging

At 72 hpf, control and treated embryos are manually dechorionated, if necessary, and anesthetized with 0.04% MS-222 (Pentair Aquatic Eco-Systems, Apopka, FL). Embryos are manually oriented and imaged using a 4X objective on an Olympus IX51 fluorescence microscope equipped with an Olympus XM10 camera and cellSens

Dimension software (Olympus, Center Valley, PA).

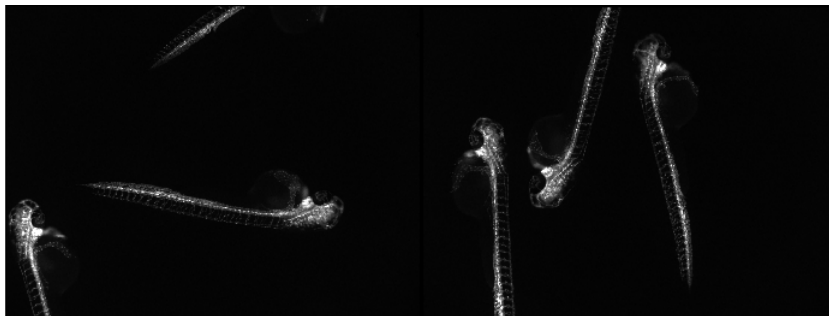


Figure 5.1: Original 2d images of zebrafish capturing zebrafish full anatomy.

5.2 Three Dimensional Time - Lapse Data

Zebrafish embryos are treated with Sodium (meta) arsenite (NaAsO_2) with $400 \mu\text{g}$ dosage.

5.2.1 Chemical Treatment

Sodium (meta)arsenite (NaAsO_2) is purchased from Sigma-Aldrich (St. Louis, MO) and dissolved in ultra pure deionized water(vehicle). $\text{Tg(kdrl:EGFP)}mitfa^{b692}$ embryos are harvested in a petridish after mating. Then, they are sorted and placed in 6-wellplates ($N = 10\text{-}30$) in 3 mL of embryo medium, E3 (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl_2 , 0.33 mM MgSO_4), followed by arsenite treatment at $400 \mu\text{g/L}$ (3.08 mM) without renewal. The embryos are incubated at 28.5°C until 72 hpf, at which

they are manually dechorionated, if necessary, and assessed for vascular perturbation and other developmental malformations. For determining the window of effect, embryos are treated with arsenite at 400 $\mu\text{g/L}$ at 0-24 hpf, 24-48 hpf, or 48-72 hpf. After exposure time is complete, embryos are washed multiple times and allowed to continue to develop in E3 at 28.5°C until assessment at 72 hpf. For RT-qPCR, 30 embryos are pooled as one biological sample in 3 mL of E3 and treated with arsenite at 10 $\mu\text{g/L}$, 50 $\mu\text{g/L}$, 100 $\mu\text{g/L}$, 200 $\mu\text{g/L}$, or 400 $\mu\text{g/L}$ up to 72 hpf. To examine RNA levels at different time points via RT-qPCR, embryos are treated with arsenite at 400 $\mu\text{g/L}$ up to 18 hpf, 20 hpf, 24 hpf, 28 hpf, or 48 hpf. Control embryos are treated with vehicle.

5.2.2 Imaging

72 hpf control and arsenite-treated Tg(kdrl:EGFP)*mitfa*^{b692} embryos are manually dechorionated, if necessary, and anesthetized with 0.04% MS-222 (Pentair Aquatic Eco-systems, Apopka, FL). Embryos are then manually oriented and mounted in 0.8% low-melt agarose (LMA; Sigma), and imaged using an Olympus Fluoview 1000 confocal fluorescence microscope with a 4X or 20X objective and 50 or 30 z-plane optical slices, respectively. Images are then rendered by Olympus Fluoview software and projection are generated using ImageJ. Brightfield images are captured with a Nikon DS-Fi1 color camera attached to a Nikon AZ100M microscope with a 4X objective and 25 z-plane optical slices, and then rendered by Nikon NIS-Elements software.2.7.

5.2.3 Time-lapse imaging

Tg(kdrl:EGFP)*mitfa*^{b692} embryos are treated with arsenite, as described above. At approximately 18 hpf, control and treated embryos are manually dechorionated, anesthetized with 0.04%MS-222 and mounted in 0.3% LMA on MatTek glass bottom dishes (MatTek Corporation, Ashland, MA). The 0.3% LMA mounting media is supplemented with 0.04% MS-222 and arsenite at concentrations matching to that of the exposure solutions. E3 with MS-222 and arsenite is added following LMA solidification and the dishes are sealed with parafilm. 100 z-plane optical slices are acquired on an Olympus Fluoview 1000 confocal fluorescence microscope over a span of 15 h with 15 min time intervals. Images are then processed by Olympus Fluoview software and projections are animated into a time-lapse movie using ImageJ.

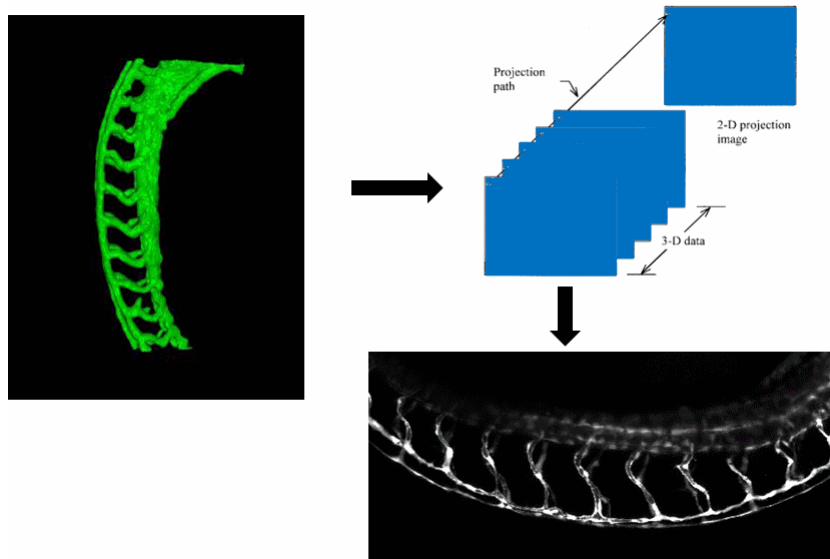


Figure 5.2: Maximum intensity projection is computed by selecting the brightest voxel along z-axis.

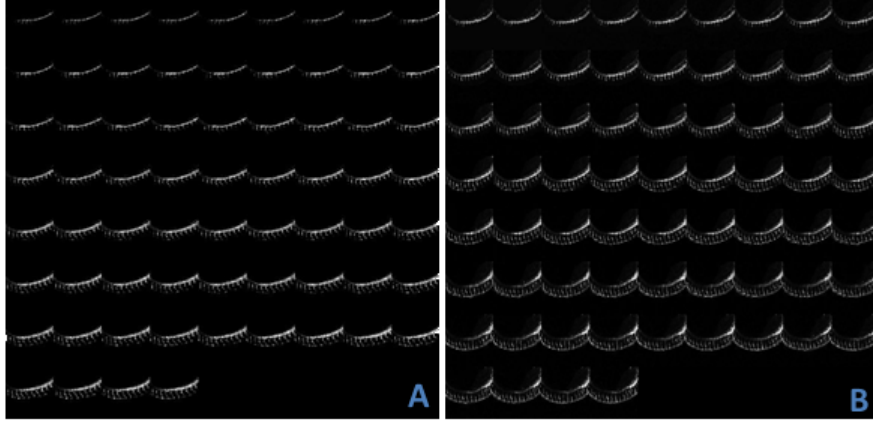


Figure 5.3: (A) Vascular development for unexposed embryo. (B) Vascular development for arsenic exposed embryo.

Images are generated in TIFF format with 8-bit intensity depth. Image size is 512 pixels by 512 pixels and 640 pixels by 480 pixels. Maximum Intensity projection (MIP) (fig. 5.2) is computed for stack of images. MIP algorithm selects the brightest voxel along the z-axis and projects it on the orthogonal image plane. Figure 5.3 shows MIPs for unexposed and exposed embryo.

Chapter 6

Zebrafish Analysis

This chapter is divided into three sections. First section presents result related to ISV analysis. We will compare proposed automated segmentation of ISV against manual segmentation. Further to identify whether chemicals from ToxCast Phase I library are vascular disruptor compounds, we present how ISV features can be used to discriminate healthy embryo from abnormal embryos. ISV features are used to train a linear classifier for depicting embryos impacted with toxicity.

Second section will focus on CVP analysis, presenting classification results using GLCM, HOG, Co-HOG, and proposed gCo-HOG. In third section, we will explore the relationship between increasing dosage of chemicals with impact on ISV, and CVP separately. We will discuss how many chemicals impacted ISV and CVP or both.

Last section will present results showing variations in temporal growth of embryo

for healthy against treated.

6.1 Intersegmental Vessel Analysis

ISV results are presented for entire zebrafish vasculature recorded from the fluorescence microscope.

6.1.1 Segmentation Analysis

Segmentation of ISV can be a time consuming process. ISVs occur in various size and shows huge variation in intensity. Method in section 3.3 to segment the ISV from zebrafish images. The accuracy of the automatic segmentation of ISVs is determined by the comparisons between manual segmentation and the automated. 30 randomly chosen ISVs are manually segmented in the RGB color space, and compared. Vessels are visually inspected by two users (fig. 6.1).

Accuracy, Precision and recall measures and F-score are used to define error [41]. The error measures used are defined as follows:

$$Accuracy = \frac{tp + tn}{(tp + tn + fp + fn)} \quad (6.1)$$

$$Precision = \frac{tp}{(tp + fp)} \quad (6.2)$$

$$Recall = \frac{tp}{(tp + fn)} \quad (6.3)$$

$$F - score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (6.4)$$

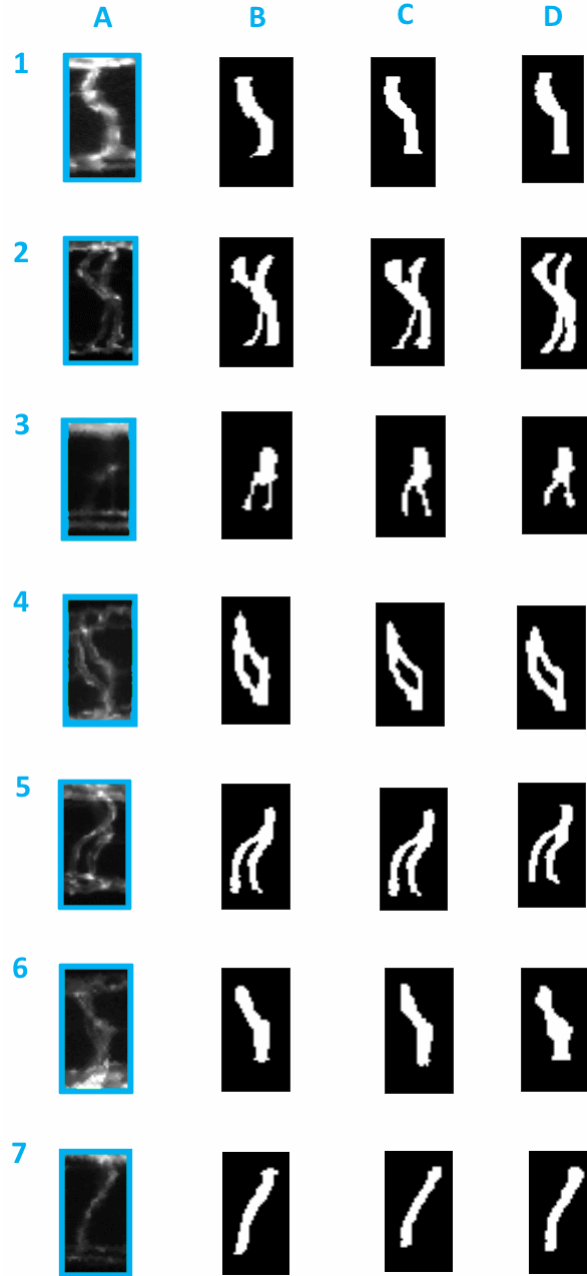


Figure 6.1: Comparison between automated segmentation and manual segmentation by user1 and user 2 of zebrafish embryos. (A) Analyzed ISV (B) Automated segmentation (C) User 1 segmentation (D) User 2 segmentation

where tp is the number of true positives, tn is the true negatives count, fp is the number of false positives and fn is the number of false negatives. The output of the segmentation is a binary vector with the same size as the image. A true positive is when output of our segmentation is 1 when manual users marked it as 1, a true negative is when the output of the segmentation is 0 while the manual users labeled it as 0, a false positive is when the output of the segmentation is 1 when users label is 0, and a false negative is when the output of the segmentation is 0 while the users label is 1. We used the above mentioned error to quantitate how close the automated method is to manual segmentation. We have shown the results for 30 randomly chosen zebrafish ISVs by two users in fig. 6.2.

Table 6.1: Segmentation results for ISV

Analysis	User 1	User 2
Precision	0.847	0.839
Recall	0.826	0.828
Accuracy	0.952	0.951
F - score	0.834	0.830

The average accuracy for the ISVs with manual users are 0.952 and 0.951. Average f-score for both users are 0.830 and 0.832. More detailed results are presented in table 6.1. These results indicate that the automated segmentation is comparable to that by manual segmentation.

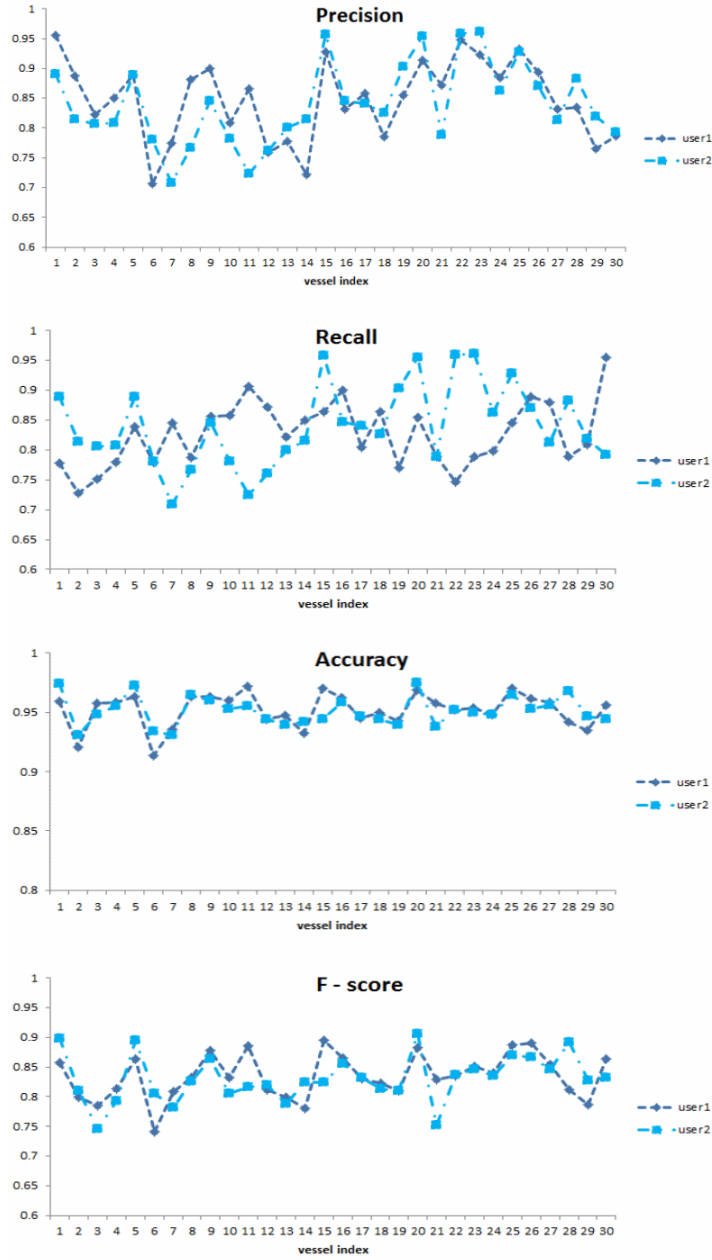


Figure 6.2: (a) Our approach have an average accuracy of 0.95 with user 1 and user 2. (b) Our approach have an average precision of 0.83 and 0.84 with with user 1 and user 2. (c) Average recall for both users is 0.83. (d) Lastly, average f - score for user are 0.83.

6.1.2 Toxicity impacts Intersegmental Vessels

To identify whether chemicals from ToxCast Phase I library are vascular disruptor compounds (VDCs), 38 chemical are tested in a range of concentrations on zebrafish embryos from 3 hpf to 72 hpf in a single static exposure. Typically, the chemicals are tested from 100 nM to 20 μ M. If the chemical exposures are lethal, lower and narrower dosages are tested. Visual analysis showed changes in ISV morphology due to chemical treatment. ISV abnormalities included the absence of ISV (fig. 6.3), or thin and underdeveloped ISV (fig. 6.3C). Quantitative image analysis is performed on the vascular disruption in the ISVs.

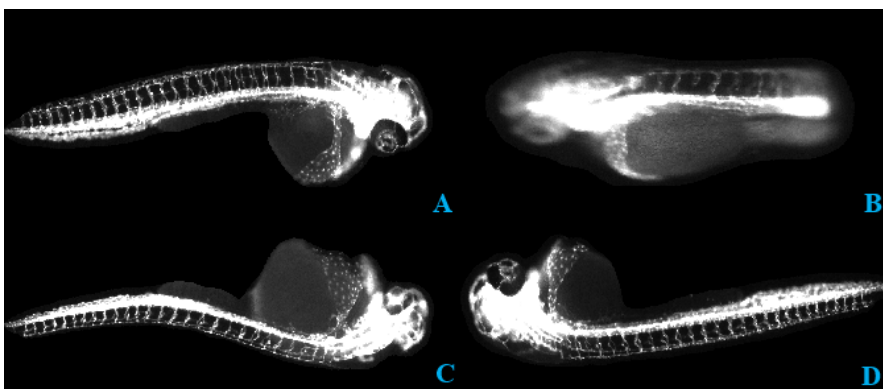


Figure 6.3: Severity of toxins effect on zebrafish ISVs. High dosage can act as a ISV disruptor, and these effects can be quantified in terms of ISV count; average distance between ISV; total area of ISV; and an average ISV length. Figure 6.3A shows an increment in ISV count, whereas 6.3B shows decrement in ISV count.

Figure 6.3C, 6.3D embryos have small size ISV and occupy less area.

ISVs are first isolated from images using algorithm presented in section 3.3 from

the whole embryo and measured for ISV length, ISV distance, number of ISVs and ISV area. All terms are defined in pixels. We reported quantification results for 38 toxins. Previous research have shown the quantification terms used in this work can be used to analyze toxicity effect in ISV [20], [54], [55]. Figure 6.4 shows with increasing toxins dosage, the ISV disruption becomes more prominent. Average ISV count, average distance between ISV, average ISV area and an average ISV length gradually decreases with increasing chemical dosage.

To aid HTS of ISV for toxicity, we utilized features quantification to train a linear SVM classifier to identify zebrafish embryo with morphological changes. Dataset consists of 380 images of ISV. 190 images in the dataset did not show any abnormalities, while the remaining 190 images shows morphological changes to the ISV due to toxicity. Labels are visually obtained. We split the data into two parts. 1/3 of data is used for parameter estimation for SVM and rest 2/3 is used for training and testing. Parameters (C, γ) are determined based on a grid-search that is conducted among $C \in 2^{-10}, 2^{-4}, \dots, 2^{10}$ and $\gamma \in 2^{-10}, 2^{-4}, \dots, 2^{10}$ with 3-fold cross-validation. Using the parameter values that achieved the best cross-validation accuracy, we then built a new model on the remaining data set, using 3-fold cross-validation. The optimal C and γ parameters are found to be 0.5 and 16, respectively, resulting in SVM model classification accuracy of 93.03%.

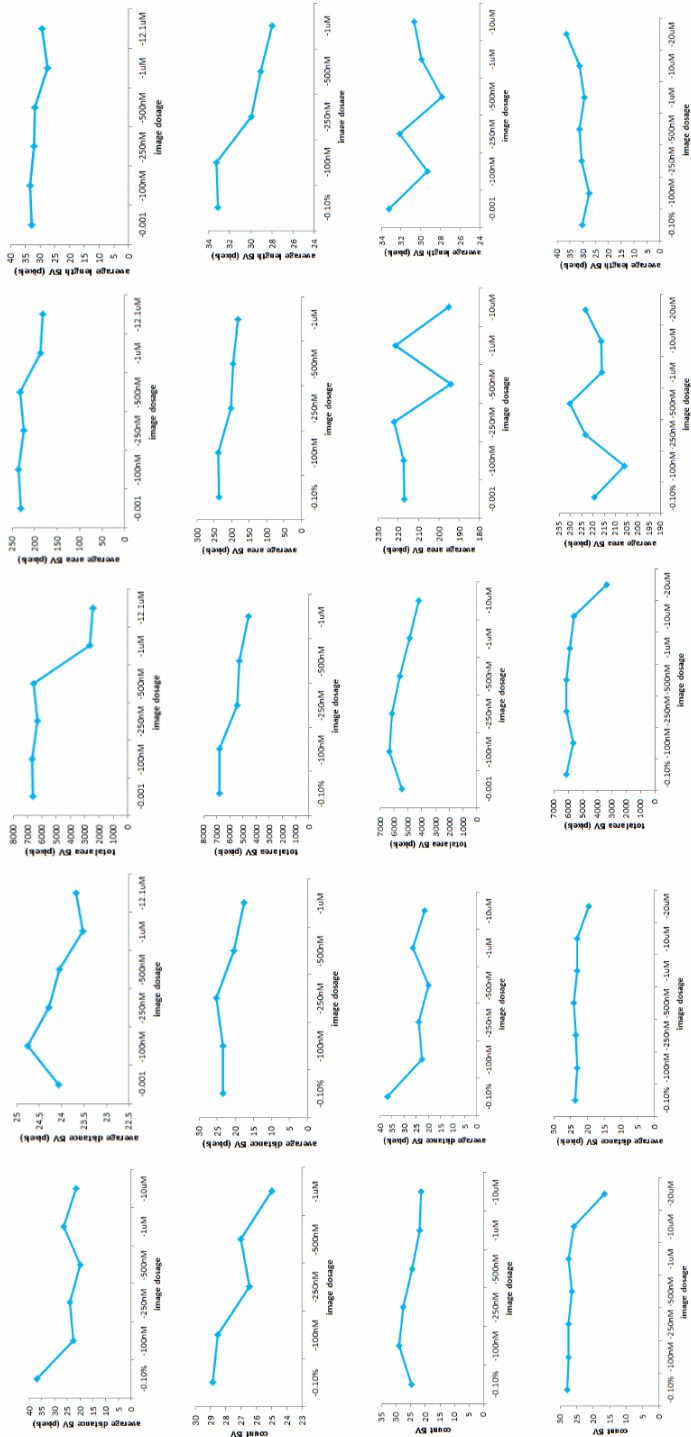


Figure 6.4: Plots for various chemicals with variable dosage in comparison too untreated embryo (0.1%). Each row corresponds to different chemicals. Within each row first plot shows variation in ISV count; Average ISV distance; Average ISV Area; Total SIV area;

6.2 Toxicity impacts Caudal Vein Plexus

CVP analysis results are presented for entire zebrafish vasculature recorded from the fluorescence microscope. We will present CVP analysis results using features explained in section 4.2. We will empirically validate choice of parameters used for analysis. Dataset consists of 180 images of various size ($163 \times 55 - 331 \times 150$ pixels). 90 images in the dataset do not show any abnormalities, while the remaining 90 images show structural changes to the CVP due to toxicity effects (fig. 6.5 shows the images of healthy and treated zebrafish embryo). Labels are visually obtained. On similar lines ad CVP toxicity analysis We split the data into two parts. 1/3 of data is used for parameter estimation and rest 2/3 is used for training and testing. Parameters (C, γ) are determined based on a grid-search that is conducted among $C \in 2^{-10}, 2^{-4}, \cdot, 2^{10}$ and $\gamma \in 2^{-10}, 2^{-4}, \cdot, 2^{10}$ with 3-fold cross-validation. Using the parameter values that achieved the best cross-validation accuracy, we then built a new model on the remaining data set, using 3-fold cross-validation.

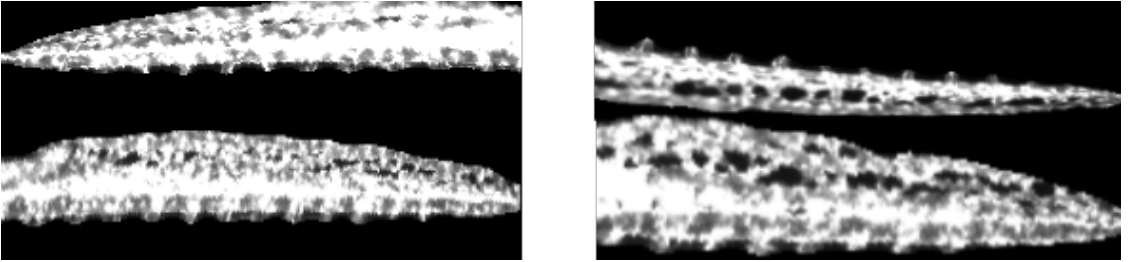


Figure 6.5: Image on left shows healthy CVP and on right is the CVP observed for an embryo exposed to toxin. Loopy structure develops in CVP region for the treated embryo.

6.2.1 Gray Level Co-occurrence Matrix Analysis

Features energy, homogeneity, correlation, and contrast are calculated from each image using GLCM representing our image. We computed results with varying d ranging from 1-15. We obtain best results using $d = 9$. Table 6.2 shows comparison between accuracies with varying d .

Table 6.2: Recognition results (%) for GLCM with varying d

d	accuracy
1	73.02
2	71.42
3	80.15
4	76.19
5	76.91
6	79.36
7	80.95
8	76.98
9	80.95
11	75.39
12	80.74
13	80.95
14	78.57
15	80.57

The optimal C and γ parameters are found to be 65536 and 0.03, respectively, resulting in SVM model classification accuracy of 80.95%.

6.2.2 Histogram of Gradient Analysis

8 rectangular regions are tiled $M/4 \times N/2$ with no overlap. More tiles are made along the image width as CVP occupies more region along width as compared to height. To compare the effect of orientations, we generate histograms with orientation range from $[0, 360]$ and $[0, 180]$. Histograms with bin size K ranging from 5 to 10 are evaluated. We got a best performance with bin size 6, 8, 9, 10 for $[0, 360]$ orientation (table 6.3 and 6.4). We choose bin size of 6 and orientation range as $[0, 360]$. Thus the dimension of our feature is $(6 \times (4 \times 2)) = 1$. The optimal C and γ parameters are found to be 16 and 0.5, respectively.

Table 6.3: Recognition results (%) for Co - HOG with varying bin size K for $[0, 360]$ orientation

d	5	6	7	8	9	10
accuracy	90.47	92.06	92.00	92.06	92.06	92.06

Table 6.4: Recognition results (%) for Co - HOG with varying bin size K for $[0, 180]$ orientation

d	5	6	7	8	9	10
accuracy	79.36	81.74	84.12	83.34	85.12	85.71

6.2.3 Co - occurrence of Histogram of Gradient Analysis

Similar to HOG, 8 rectangular regions are tiled $M/4 \times N/2$ with no overlap. As mentioned in 4.2.3 we obtain occurrence matrix for 31 offset. To compare the effect of

orientations, we generate histograms with orientation range from $[0, 360]$ and $[0, 180]$. Histograms with bin size K ranging from 5 to 10 are evaluated. We got a best performance with bin size 8 for $[0, 360]$ orientation (table 6.5 and 6.6). Final, feature vector size is $((8 \times 8) \times 30 + 8) \times (4 \times 8) = 15424$. The optimal C and γ parameters are found to be 0.5 and 256, respectively.

Table 6.5: Recognition results (%) for Co - HOG with varying bin size K for $[0, 360]$ orientation

d	5	6	7	8	9	10
accuracy	92.44	92.44	92.40	92.86	92.85	92.85

Table 6.6: Recognition results (%) for Co - HOG with varying bin size K for $[0, 180]$ orientation

d	5	6	7	8	9	10
accuracy	83.33	85.71	83.33	85.71	86.5	82.53

6.2.4 Gradient Co - occurrence of Histogram of Gradient Analysis

Similar to HOG and Co - HOG, 8 rectangular regions are tiled $M/4 \times N/2$ with no overlap. As mentioned in 4.2.3 we obtain occurrence matrix for 31 offset. To compare the effect of orientations, we generate histograms with orientation range from $[0, 360]$ and $[0, 180]$. Histograms with bin size K ranging from 5 to 10 are evaluated. We got a best performance with bin size 8 for $[0, 360]$ orientation (table 6.7 and 6.8). Final,

feature vector size is $((8 \times 8) \times 30 + 8) \times (4 \times 8) = 15424$. The optimal C and γ parameters are found to be 0.5 and 128, respectively.

Table 6.7: Recognition results (%) for HOG with varying bin size K for $[0, 360]$ orientation

d	5	6	7	8	9	10
accuracy	93.65	91.26	93.66	94.44	92.85	92.85

Table 6.8: Recognition results (%) for HOG with varying bin size K for $[0, 180]$ orientation

d	5	6	7	8	9	10
accuracy	86.5	85.71	85.71	87.3	85.71	86.5

Experimental results in table 6.9 shows that our proposed descriptor gCo-HOG outperforms the other two (HOG, Co-HOG) and has an accuracy of 94.44%.

Table 6.9: Recognition results (%) for descriptors with 360 orientation

gCo-HOG	Co-HOG	HOG
94.44	92.85	92.06

6.3 Toxicity Screening with ISV, CVP

Among so many chemicals that exist in the market, it is crucial to be able to identify which chemicals can target the vasculature blood vessel development. Further,

its imperative to determine safe dosage of chemicals, which does not impose harmful effects on vasculature growth. In this section we will discuss, how number of zebrafish impacted by chemical treatment increases with high dosages. More specifically, we will talk about how number of zebrafish showing abnormalities in ISV and CVP region increases with increasing dosage. Our screening method analyzed 38 chemicals.

Toxicity screening uses the model derived from section 6.1 and 6.2 to study the impact of different dosages on zebrafish. We retrain the model on all the 380 images for ISV (using morphological properties of ISV) and 180 images for CVP (using proposed gCo-HOG). We tested both the models on 2039 zebrafish images. To study the effects of malformations with increasing dosage and to narrow down on safe dosage for chemicals, we coined the term lethality rate. Lethality rate is the ratio of number of zebrafish embryos with abnormal growth based on proposed model to the total number of embryos.

$$\text{lethality rate} = \text{no. of abnormal zebrafish} / \text{total no. of zebrafish}$$

This term is calculated for each dosage for all chemicals. We calculated the lethality rate with both ISV, and CVP models.

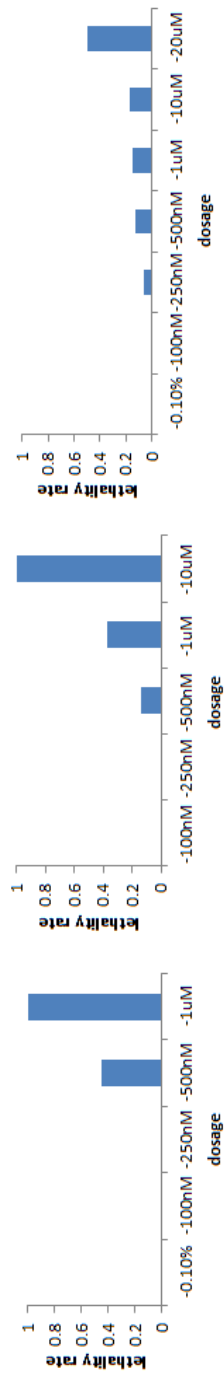


Figure 6.6: Lethality rate for three chemicals on ISV. There is an increase in lethality rate with increase in dosage. 0.1% represents the untreated (control) zebrafish image.

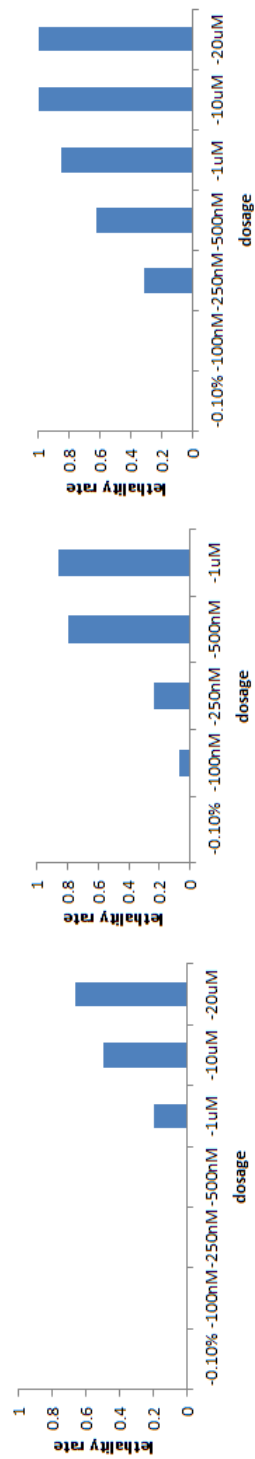


Figure 6.7: Lethality rate for three chemicals on CVP. There is an increase in lethality rate with increase in dosage. 0.1% represents the untreated (control) zebrafish image.

Based on lethality rate, we deduced out of 38 screened chemicals, our ISV model identified 28 chemicals that severely ($\geq 10\%$ of embryos are impacted for each dosage within each chemical) impacted ISV growth. Out of 28 chemicals impacting ISV growth, 22 of them showed increasing lethality rate with increasing dosage. Remaining 6 chemicals did not show an increment in lethality rate with increasing dosage, but the effects are more severe at high dosages. Figure 6.6 shows lethality rate for three chemicals with increasing dosage. It's worth pointing out that in some cases our model identified untreated zebrafish as unhealthy. Reasons for the miss-classification is due to the imaging artifact. There are 20 images out of 2039, in which zebrafish turned on its back during imaging. We have isolated those images while calculating lethality rate. We can calculate safe dosage as maximum dosage within each chemical with 0 lethality rate.

On similar line as above, we computed lethality rate for CVP analysis. 30 chemicals out of 38 severely ($\geq 10\%$ of embryos are impacted for each dosage within each chemical) impacted development in CVP region. Out of 30 chemicals impacting CVP, 15 showed increasing impacts on CVP with increasing dosage (in terms of lethality rate). Figure 6.7 shows lethality rate for three chemicals with increasing dosage.

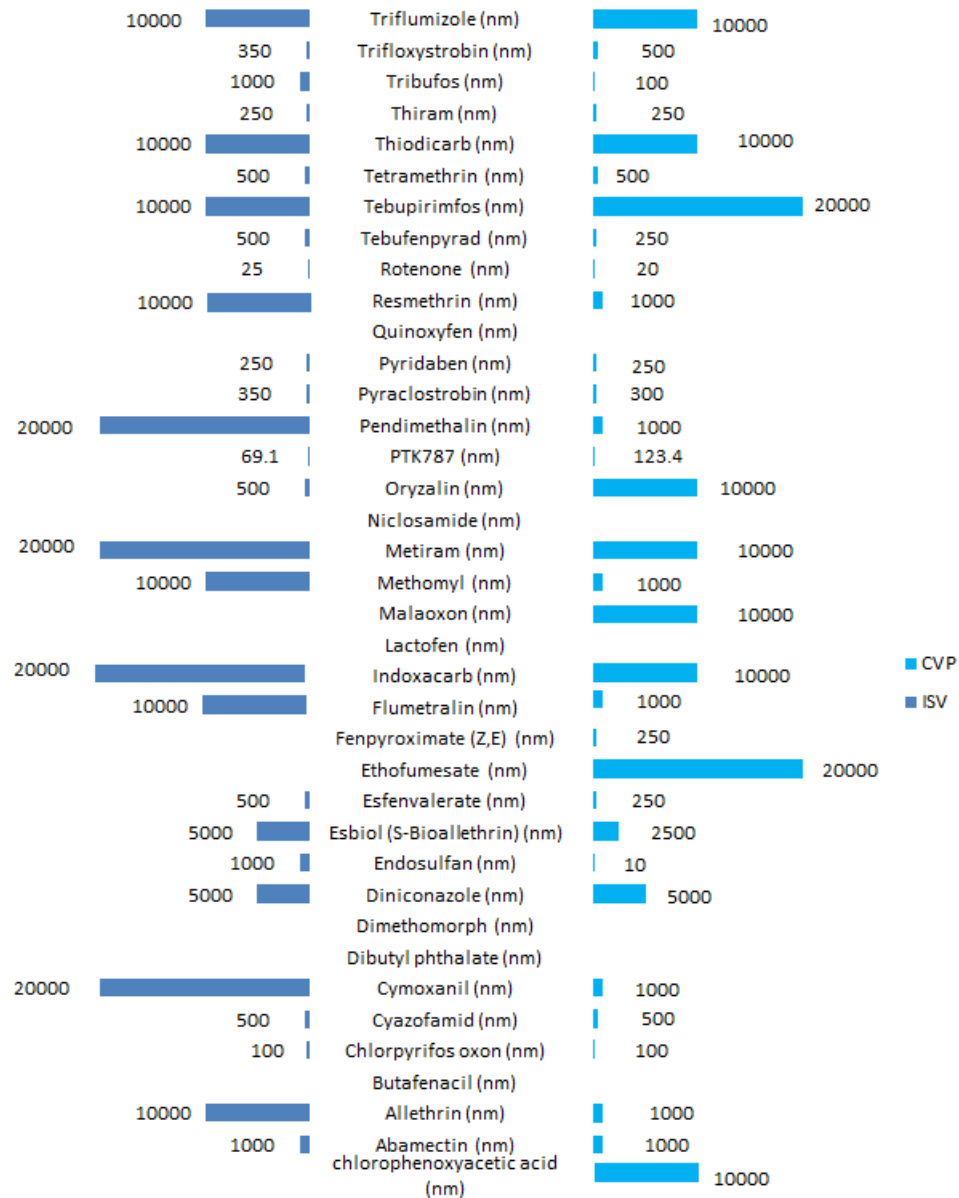


Figure 6.8: Figure tells us chemicals, which impact ISV only, CVP only, both ISV and CVP. Safe dosage for each chemical for ISV and CVP region is also illustrated in this plot. Each dosage is given in nanometer (nm). Chemical with the blank value, shows there is no impact due to toxicity for tested dosages.

A representation is illustrated to show what is the safe dosage for chemical based on ISV and CVP based screening (fig.6.8) . This illustration also tells us which chemicals impacted ISV only, CVP only or both ISV and CVP. Moreover it tells us the safe dosage for various chemicals used to treat zebrafish embryo. For 15 chemicals, we found exactly same safe dosage for both ISV and CVP. In 15, of them difference is one step of dosage. By, one step of dosage we mean, lets say dosages tested are 10nm, 100nm, 250nm for a chemical. ISV and CVP shows 100nm and 250nm as safe dosage respectively. Hence we say dosage difference is one step between ISV and CVP. 4 chemicals had a difference of two steps for safe dosages for ISV and CVP. In remaining 4 the difference is more than two steps.

6.4 Vasculature Time - Series Analysis

The computational framework presented in section 4.3 is applied to quantify vascular changes in both toxin treated and untreated zebrafish embryos. A uniform B-splines control grid with spacing [20, 20], and 3 grid refinement levels are used. Totally, 7 embryos (5 untreated and 2 treated) are analyzed. Vascular growth is computed using method described in section 4.3. Figure 6.9 shows the dynamics of blood vessel growth in terms of average pixel change for untreated and arsenic treated embryos. As seen in the figure, although the overall rate of growth is constant for both treated and untreated embryos, the growth rate for untreated embryos is approximately 5 times higher.

Figure 6.10 shows the dynamics of blood vessel growth in terms of average pixel

change for each of the seven embryos analyzed. Pixel change measured in untreated fish is represented by $uexposed1 - uexposed5$ and $exposed1 - exposed2$ shows pixel change measured for arsenic exposed embryos.

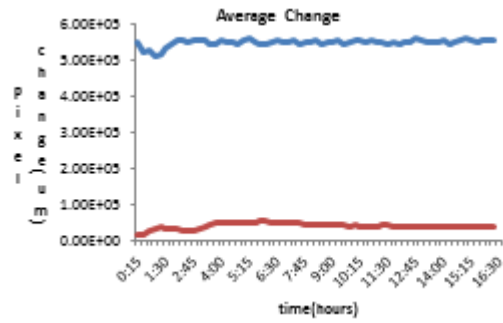


Figure 6.9: Blue curve shows average change in number of pixels over all the unexposed embryos whereas red curve shows average pixel change for arsenic treated embryo.

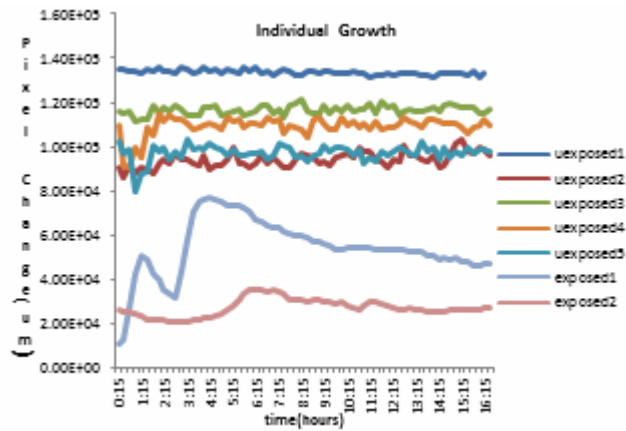


Figure 6.10: uexposed1 - uexposed5 represents change in number of pixels for unexposed embryos growth, whereas exposed1 and exposed2 represents change in number of pixels for arsenic treated embryos.

Chapter 7

Conclusions and Perspectives

As the application of zebrafish in research is becoming more sought after, manual analysis of images has become a bottleneck. This thesis has described a complete framework for the segmentation, quantification, analysis and classification of zebrafish images for ISV and CVP. We also presented methods for ISV and CVP quantification for applications not limited to toxicology analysis. Our results show that ISV and CVP analysis can be used to capture the effect of vascular disruptors.

Roughly 80,000 industrial chemicals are registered on the US market, and very few of them have been screened for disrupting properties. Thus, development of high throughput screening models that can recognize impacts of toxicity can prove to be very vital. The work presented in this thesis, paves the way for all existing research which has been restricted by analyzing huge amount of data. In conclusion, this work will directly impact the transition from qualitative observation to quantitative measurement of zebrafish by not only providing effective segmentation, but also

capturing effective dynamics.

7.1 Discussion and Limitations

Zebrafish being a versatile model, is used in analysis of many biological processes, toxic analysis, and drug screening. Despite the potential of the zebrafish as a model, the actual number of analysis reported is small, involving limited numbers of compounds, with analysis performed manually. Many of the studies reported on zebrafish, focuses on ISV, and yet the research to segment and extract the features from ISV has not taken a lift. The ISVs are particularly interesting because their pattern appears to be established primarily, unlike other vessels such as those of the yolk, brain or retina. Thus, ISVs can give us an initial pathway guidance for capturing dynamics related to growth or environmental cues.

The key bottleneck restricting these analysis is quantification of the high-throughput experiments, because there are few image analysis methods capable of capturing the complexity ISV. The huge amount of data generated for analysis makes manual analysis a time-inhibiting and error prone process subject to inter-observer variations. Computer-based image processing methods are the only viable way to analyze and quantify the data. Current methods analyze images based on pixel information and thresholding. These methods perform well on clearly resolved objects against a uniform background but often struggle with images that possess information with variable sizes, noise, and heterogeneity across images. This is exemplified in many studies where zebrafish vascular became quantifiable only after manual identification

or manual analysis.

In this work, we focus on developing an integrated image processing methodology to analyze zebrafish embryos post chemical treatment. The pipeline consists of Segmentation of the zebrafish embryo, Region Detection, ISV Extraction, and ISV refinement. We quantify extracted ISV to capture the effect of chemical treatment. A critical issue with the quantitative analysis of zebrafish is that it should be invariant to orientation of zebrafish, and it is highly desirable to have a description that is invariant to noise, intensity and scales. Our algorithm takes care of this critical issue.

ISV segmentation method perform reasonably well for noisy and blur data (fig. 7.1A and 7.1B). Our method does not perform well for very noisy and blurred ISV (fig. 7.1A and 7.1B near the tail region). Also, ISVs shows immense variation in size. They grow small and thin near the end points, as opposed to center as depicted in figure 7.1D and thin again near head region. Our method is limited by very thin and faded vessels. In figure 7.1D faded vessel does not show in our segmentation. Lastly for figure 7.1C, since ISV curled around zebrafish head, we missed many vessels which does not respond within direction parameters of our method.

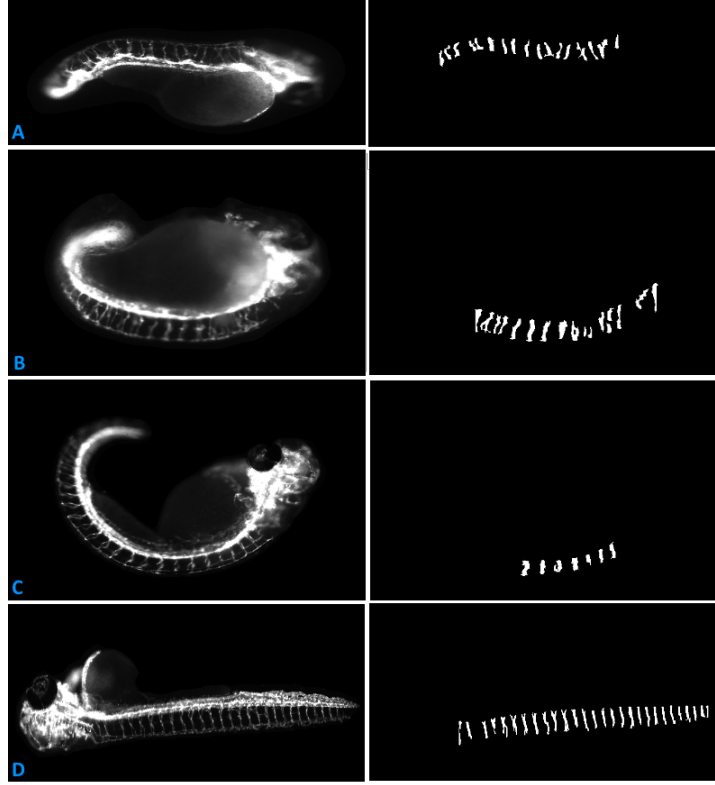


Figure 7.1: (A), (B) The vessels are more noisy and blurry (C) We lost ISV, due to direction limitation opposed to in tail region. (D) ISVs occur in various sizes, with weak boundary response. Vessels that does not produce a strong response during eigenanalysis also gets excluded during segmentation. Overall, Our algorithm did segmentation of most of ISV. Parameter selection for ISV response and merging causes some loss in data.

Choice of our parameter effects ISV merging. ISVs remain disjoint is they do not lie within parameter range. In order to negate the effect of direction restriction, we also get response from ISV image in all direction as the probable ISV pixels. Morphological closing operation is applied on direction based ISV and later merged

with direction free ISV response. This helps us merge vessels, but if there is no response from direction free ISV, they remain unlinked.

The developed segmentation presents several advantages. Firstly, the parameters are established directly from images so that user interaction is minimal. Secondly, it is not iterative and it can be performed rapidly. Thirdly, this method performs well for images with varied intensity profile and varied texture, orientation and size. Hence, this method possesses the capacity to capture highly dynamic vessels, regardless of phenotype types that exhibit a wide range of morphologies.

We have presented method for segmentation and analysis for CVP. CVP segmentation is based on prior knowledge about change in curvature around tail region. If, our algorithm is not able to find point with the steepest change in curvature, we assume the point where tail starts is the mid point of the curve tracing tail + yolk region. High toxicity impacts the structure of zebrafish, hence perturbing tail region. In those, cases we resort to mid point of curve, which does the good job capturing CVP structure properties. But, it captures more then CVP region (fig. 7.1A, 7.1B, 7.1C).

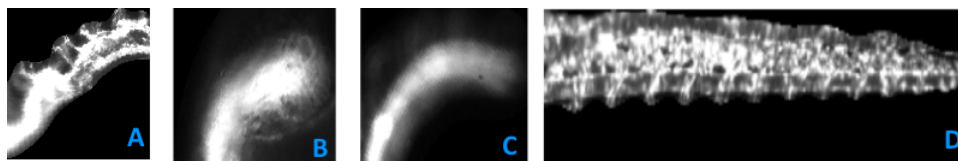


Figure 7.2: (A), (B), (C) The tail vessel is structurally damaged due to high dosage. We do get a good estimate of CVP region, but not exact. (D) We are able to segment CVP region well.

7.2 Future Work

Segmentation

In terms of related work, there is a definitely a place for improvement of segmentation of ISV and CVP. Some zebrafish curve around its head under the influence of chemicals. Peng et al. [48] proposed method to straighten curved worm based on formulating the backbone of a worm as a parametric cubic spline defined by a series of control points. This method, can be applied to backbone of zebrafish. This method will improve both the ISV and CVP segmentation. **Features**

Improvement in segmentation, will concurrently improve extracted features. For CVP analysis fusion of morphological properties, with shape properties may improve the classification accuracy. It will be interesting to be able to fuse model for both ISV and CVP to study the toxicology effects and determine safe dosages from chemicals. There is also a scope to be able to extend ISV and CVP analysis algorithm for various other dimensions including $2D + t$, $3D$, $3D + t$. **Applications**

We have tested our algorithms for application in toxicology. It can be easily extended for drug screening, gene expression modification, or disease modeling.

Bibliography

- [1] ALSHUT, R., LEGRADI, J., LIEBEL, U., YANG, L., VAN WEZEL, J., STRÄHLE, U., MIKUT, R., AND REISCHL, M. Methods for automated high-throughput toxicity testing using zebrafish embryos. In *KI 2010: Advances in Artificial Intelligence*. Springer, 2010, pp. 219–226. 3, 4, 27
- [2] AMATRUDA, J. F., AND PATTON, E. E. In *Chapter 1 Genetic Models of Cancer in Zebrafish*, K. W. Jeon, Ed., vol. 271 of *International Review of Cell and Molecular Biology*. Academic Press, 2008, pp. 1 – 34. 2
- [3] ANNILA, T., LIHAVAINEN, E., MARQUES, I. J., WILLIAMS, D. R., YLIHARJA, O., AND RIBEIRO, A. Zebiat, an image analysis tool for registering zebrafish embryos and quantifying cancer metastasis. *BMC bioinformatics* 14, Suppl 10 (2013), S5. 23
- [4] ARSLANOVA, D., YANG, T., XU, X., WONG, S., AUGELLI-SZAFRAN, C., AND XIA, W. Phenotypic analysis of images of zebrafish treated with alzheimer’s -secretase inhibitors. *BMC Biotechnology* 10, 1 (2010). 19
- [5] BANG, P. I., YELICK, P. C., MALICKI, J. J., AND SEWELL, W. F. High-throughput behavioral screening method for detecting auditory response defects in zebrafish. *Journal of Neuroscience Methods* 118, 2 (2002), 177 – 187. 3, 4
- [6] BANG, P. I., YELICK, P. C., MALICKI, J. J., AND SEWELL, W. F. High-throughput behavioral screening method for detecting auditory response defects in zebrafish. *Journal of neuroscience methods* 118, 2 (2002), 177–187. 30
- [7] BETKE, M., HONG, H., AND KO, J. Automatic 3d registration of lung surfaces in computed tomography scans. Tech. rep., Boston, MA, USA, 2001. 12
- [8] BOEHMLER, W., PETKO, J., WOLL, M., FREY, C., THISSE, B., THISSE, C., CANFIELD, V. A., AND LEVENSON, R. Identification of zebrafish {A2}

- adenosine receptors and expression in developing embryos. *Gene Expression Patterns* 9, 3 (2009), 144 – 151. 2
- [9] BOSCH, A., ZISSERMAN, A., AND MUNOZ, X. Representing shape with a spatial pyramid kernel. *CIVR '07*, pp. 401–408. 55
 - [10] BRO-NIELSEN, M., GRAMKOW, C., AND KREIBORG, S. Non-rigid image registration using bone growth model. In *CVRMed-MRCAS'97*, J. Troccaz, E. Grimson, and R. Mages, Eds., vol. 1205 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 1997, pp. 1–12. 12
 - [11] CAMPBELL, W. A., YANG, H., ZETTERBERG, H., BAULAC, S., SEARS, J. A., LIU, T., WONG, S. T., ZHONG, T. P., AND XIA, W. Zebrafish lacking alzheimer presenilin enhancer 2 (pen-2) demonstrate excessive p53-dependent apoptosis and neuronal loss. *Journal of neurochemistry* 96, 5 (2006), 1423–1440. 2
 - [12] CHEN, E., STRINGER, S. E., RUSCH, M. A., SELLECK, S. B., AND EKKER, S. C. A unique role for 6-o sulfation modification in zebrafish vascular development. *Developmental Biology* 284, 2 (2005), 364 – 376. 11, 21
 - [13] CHEN, S., ZHU, Y., XIA, W., XIA, S., AND XU, X. Automated analysis of zebrafish images for phenotypic changes in drug discovery. *Journal of Neuroscience Methods* 200, 2 (2011), 229 – 236. 4, 26
 - [14] CHENG, S. H., CHAN, P. K., AND WU, R. S. S. The use of microangiography in detecting aberrant vasculature in zebrafish embryos exposed to cadmium. *Aquatic Toxicology* 52, 1 (2001), 61–71. 22
 - [15] CHOI, J., MOUILLESSEUX, K., WANG, Z., FIJI, H. D., KINDERMAN, S. S., OTTO, G. W., GEISLER, R., KWON, O., AND CHEN, J.-N. Aplexone targets the hmg-coa reductase pathway and differentially regulates arteriovenous angiogenesis. *Development* 138, 6 (2011), 1173–1181. 7
 - [16] DALAL, N., AND TRIGGS, B. Histograms of oriented gradients for human detection. *CVPR '05*, pp. 886–893. 55
 - [17] DESCOTEAUX, M., AUDETTE, M., CHINZEI, K., AND SIDDIQI, K. Bone enhancement filtering: Application to sinus bone segmentation and simulation of pituitary surgery. In *Medical Image Computing and Computer-Assisted Intervention MICCAI 2005*, J. Duncan and G. Gerig, Eds., vol. 3749 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, pp. 9–16. 45

- [18] DNIZ, O., BUENO, G., SALIDO, J., AND LA TORRE, F. Face recognition using histograms of oriented gradients. *Pattern Recognition Letters* 32, 12 (2011), 1598–1603. 55
- [19] FAN, C.-Y., COWDEN, J., SIMMONS, S. O., PADILLA, S., AND RAMABHADRAN, R. Gene expression changes in developing zebrafish as potential markers for rapid developmental neurotoxicity screening. *Neurotoxicology and Teratology* 32, 1 (2010), 91 – 98. 2
- [20] FENG, J., CHENG, S. H., CHAN, P. K., AND IP, H. H. Reconstruction and representation of caudal vasculature of zebrafish embryo from confocal scanning laser fluorescence microscopic images. *Computers in Biology and Medicine* 35, 10 (2005), 915 – 931. 3, 5, 6, 11, 30, 31, 32, 52, 54, 77
- [21] FOLKMAN, J. Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nature medicine* 1, 1 (1995), 27–30. 5
- [22] FORSEY, D. R., AND BARTELS, R. H. Hierarchical b-spline refinement. *SIGGRAPH Comput. Graph.* 22, 4 (June 1988), 205–212. 61
- [23] FOUQUET, B., WEINSTEIN, B. M., SERLUCA, F. C., AND FISHMAN, M. C. Vessel patterning in the embryo of the zebrafish: guidance by notochord. *Developmental biology* 183, 1 (1997), 37–48. 2
- [24] FRANGI, R. F., NIESSEN, W. J., VINCKEN, K. L., AND VIERGEVER, M. A. Multiscale vessel enhancement filtering. Springer-Verlag, pp. 130–137. 45
- [25] HARALICK, R. M., SHANMUGAM, K., AND DINSTEN, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 6 (1973), 610–621. 54
- [26] HARALOCK, R. M., AND SHAPIRO, L. G. *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc., 1991. 54
- [27] ISOGAI, S., HORIGUCHI, M., AND WEINSTEIN, B. M. The vascular anatomy of the developing zebrafish: an atlas of embryonic and early larval development. *Developmental biology* 230, 2 (2001), 278–301. 6, 7
- [28] KAMALI, M., DAY, L., BROOKS, D., ZHOU, X., AND O’MALLEY, D. Automated identification of neurons in 3d confocal datasets from zebrafish brainstem. *Journal of microscopy* 233, 1 (2009), 114–131. 26

- [29] KATO, S., NAKAGAWA, T., OHKAWA, M., MURAMOTO, K., OYAMA, O., WATANABE, A., NAKASHIMA, H., NEMOTO, T., AND SUGITANI, K. A computer image processing system for quantification of zebrafish behavior. *Journal of Neuroscience Methods* 134, 1 (2004), 1 – 7. 4, 31
- [30] KAUFMAN, C. K., WHITE, R. M., AND ZON, L. Chemical genetic screening in the zebrafish embryo. *Nature protocols* 4, 10 (Sep 2009). 2
- [31] KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B., AND SCHILLING, T. F. Stages of embryonic development of the zebrafish. *Developmental Dynamics* 203, 3 (1995), 253–310. 2, 3, 64
- [32] KLEINSTREUER, N. C., JUDSON, R. S., REIF, D. M., SIPES, N. S., SINGH, A. V., CHANDLER, K. J., DEWOSKIN, R., DIX, D. J., KAVLOCK, R. J., AND KNUDSEN, T. B. Environmental impact on vascular development predicted by high-throughput screening. *Environmental health perspectives* 119, 11 (2011), 1596. 5
- [33] KOENDERINK, J. The structure of images. *Biological Cybernetics* 50, 5 (1984), 363–370. 45
- [34] KOSTELEK, P. J., AND PERIASWAMY, S. Image registration for mri, 2003. 12
- [35] LEE, S., WOLBERG, G., AND SHIN, S.-Y. Scattered data interpolation with multilevel b-splines. *Visualization and Computer Graphics, IEEE Transactions on* 3, 3 (Jul 1997), 228–244. 60, 61
- [36] LEE, T.-C., KASHYAP, R. L., AND CHU, C.-N. Building skeleton models via 3-d medial surface/axis thinning algorithms. *CVGIP: Graph. Models Image Process.* 56, 6 (Nov. 1994), 462–478. 53
- [37] LEONARD I. ZON, R. T. P. In vivo drug discovery in the zebrafish, 2005. 2
- [38] LINDBERG, T. Edge detection and ridge detection with automatic scale selection. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on* (Jun 1996), pp. 465–470. 46
- [39] LIU, R., LIN, S., RALLO, R., ZHAO, Y., DAMOISEAUX, R., XIA, T., LIN, S., NEL, A., AND COHEN, Y. Automated phenotype recognition for zebrafish embryo based in vivo high throughput toxicity screening of engineered nanomaterials, 2012. 5, 27

- [40] LIU, T., LU, J., WANG, Y., CAMPBELL, W. A., HUANG, L., ZHU, J., XIA, W., AND WONG, S. T. Computerized image analysis for quantitative neuronal phenotyping in zebrafish. *Journal of Neuroscience Methods* 153, 2 (2006), 190 – 202. 3, 4, 29
- [41] MARTIN, D., FOWLKES, C., TAL, D., AND MALIK, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (2001), vol. 2, pp. 416–423 vol.2. 71
- [42] MCCOLLUM, C. W., DUCHARME, N. A., BONDESSON, M., AND GUSTAFSSON, J.-A. Developmental toxicity screening in zebrafish. *Birth Defects Research Part C: Embryo Today: Reviews* 93, 2 (2011), 67–114. 22
- [43] MCCOLLUM, C. W., HANS, C., SHAH, S., MERCHANT, F. A., GUSTAFSSON, J.-Å., AND BONDESSON, M. Embryonic exposure to sodium arsenite perturbs vascular development in zebrafish. *Aquatic Toxicology* 152 (2014), 152–163. 3
- [44] MIKUT, R., DICKMEIS, T., DRIEVER, W., GEURTS, P., HAMPRECHT, F., KAUSLER, B. X., LEDESMA-CARBAYO, M. J., MARÉE, R., MIKULA, K., PANTAZIS, P., RONNEBERGER, O., SANTOS, A., STOTZKA, R., STRÄHLE, U., AND PEYRIÉRAS, N. Automated processing of zebrafish imaging data - a survey. *Zebrafish* 10, 3 (Aug 2013), 401–421. 8
- [45] NEWMAN, M., VERDILE, G., MARTINS, R. N., AND LARDELLI, M. Zebrafish as a tool in alzheimer’s disease research. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1812, 3 (2011), 346 – 352. Including the Special Section: Zebrafish Models of Neurological Diseases. 2
- [46] NIBLACK, W. *An introduction to digital image processing*. Strandberg Publishing Company, 1985. 47
- [47] PANG, Y., YAN, H., YUAN, Y., AND WANG, K. Robust cohog feature extraction in human-centered image/video management system. *Systems, Man, and Cybernetics* 42, 2 (2012), 458–468. 57
- [48] PENG, H., LONG, F., LIU, X., KIM, S. K., AND MYERS, E. W. Straightening caenorhabditis elegans images. *Bioinformatics* 24, 2 (2008), 234–242. 96
- [49] PERAVALI, R., GEHRIG, J., GISELBRECHT, S., LÜTJOHANN, D. S., HADZHIEV, Y., MÜLLER, F., AND LIEBEL, U. Automated feature detection

- and imaging for high-resolution screening of zebrafish embryos. *Biotechniques* 50, 5 (2011), 319. 24
- [50] SHALABY, F., ROSSANT, J., YAMAGUCHI, T. P., GERTSENSTEIN, M., WU, X.-F., BREITMAN, M. L., AND SCHUH, A. C. Failure of blood-island formation and vasculogenesis in flk-1-deficient mice. *Nature* 376, 6535 (1995), 62–66. 5
 - [51] SHIRINIFARD, A., MCCOLLUM, C. W., BOLIN, M. B., GUSTAFSSON, J.-K., GLAZIER, J. A., AND CLENDENON, S. G. 3d quantitative analyses of angiogenic sprout growth dynamics. *Developmental Dynamics* 242, 5 (2013), 518–526. 21
 - [52] STERN, O., MARE, R., ACETO, J., JEANRAY, N., MULLER, M., WEHENKEL, L., AND GEURTS, P. Automatic localization of interest points in zebrafish images with tree-based methods. In *Pattern Recognition in Bioinformatics*, M. Loog, L. Wessels, M. Reinders, and D. de Ridder, Eds., vol. 7036 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011, pp. 179–190. 4, 28
 - [53] TAL, T., MCCOLLUM, C., HARRIS, P., OLIN, J., KLEINSTREUER, N., WOOD, C., HANS, C., SHAH, S., MERCHANT, F., BONDESSON, M., KNUDSEN, T., PADILLA, S., AND HEMMER, M. Immediate and long-term consequences of vascular toxicity during zebrafish development. *Reproductive Toxicology* 48, 0 (2014), 51 – 61. 42nd Annual Conference of the European Teratology Society. 23
 - [54] TRAN, T. C., SNEED, B., HAIDER, J., BLAVO, D., WHITE, A., AIYEJORUN, T., BARANOWSKI, T. C., RUBINSTEIN, A. L., DOAN, T. N., DINGLEDINE, R., AND SANDBERG, E. M. Automated, quantitative screening assay for antiangiogenic compounds using transgenic zebrafish. *Cancer Research* 67, 23 (2007), 11386–11392. 2, 5, 6, 8, 10, 11, 22, 52, 54, 77
 - [55] VOGT, A., CHOLEWINSKI, A., SHEN, X., NELSON, S. G., LAZO, J. S., TSANG, M., AND HUKRIEDE, N. A. Automated image-based phenotypic analysis in zebrafish embryos. *Developmental Dynamics* 238, 3 (2009), 656–663. 5, 6, 28, 32, 52, 77
 - [56] WADA, T., HUANG, F., AND LIN, S. *Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection*, vol. 5414. 2009. 55
 - [57] WANG, K., AND BELONGIE, S. Word spotting in the wild. In *Computer Vision ECCV 2010*, vol. 6311. 2010, pp. 591–604. 55

- [58] WESTERFIELD, M. *The zebrafish book: a guide for the laboratory use of zebrafish (Danio rerio)*. University of Oregon Press, 2000. 64
- [59] XU, X., XU, X., HUANG, X., XIA, W., AND XIA, S. A high-throughput analysis method to detect regions of interest and quantify zebrafish embryo images. *Journal of Biomolecular Screening* 15, 9 (2010), 1152–1159. 4, 29
- [60] YANG, L., HO, N. Y., ALSHUT, R., LEGRADI, J., WEISS, C., REISCHL, M., MIKUT, R., LIEBEL, U., MLLER, F., AND STRHLE, U. Zebrafish embryos as models for embryotoxic and teratological effects of chemicals. *Reproductive Toxicology* 28, 2 (2009), 245 – 253. 37th Annual Conference of the European Teratology Society. 2
- [61] YANIV, K., ISOGAI, S., CASTRANOVA, D., DYE, L., HITOMI, J., AND WEINSTEIN, B. M. Live imaging of lymphatic development in the zebrafish. *Nature medicine* 12, 6 (2006), 711–716. 7
- [62] YOZZO, K. L., ISALES, G. M., RAFTERY, T. D., AND VOLZ, D. C. High-content screening assay for identification of chemicals impacting cardiovascular function in zebrafish embryos. *Environmental Science & Technology* 47, 19 (2013), 11302–11310. PMID: 24015875. 22
- [63] ZACK, G., ROGERS, W., AND LATT, S. Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry & Cytochemistry* 25, 7 (1977), 741–753. 35