Shape analysis for phenotype characterisation from high-throughput imaging

Guo, Y.; Guo Y.

Chapter 4

A Novel 3D Reconstruction Approach

Based on:

- Y. Guo, Y. Zhang & F.J. Verbeek, "A two-phase 3D reconstruction approach for light microscopy axial-view imaging," in IEEE Journal of Selected Topics in Signal Processing, 2017.

This chapter addresses RQ 4.

> **RQ 4**: *How can we efficiently deal with the translucency and transparency of specimen in light microscopy and still obtain a good 3D shape description from the MM-HTAI architecture*?

Abstract – Three-dimensional representations in light microscopy are important for accurate shape assessment of model systems in biosciences (see Section 4.1). The computational multi-view 3D reconstruction seems feasible in obtaining the 3D representations, in particular for high-throughput. The specimen for imaging can have properties, i.e., transparency and translucency, that impede the detection of well-defined boundaries (see Section 4.2). Consequently, 3D reconstruction and measurements, i.e., volume and surface area, will be inaccurate. The motivation in this chapter is to develop a two-phase 3D reconstruction approach for light microscopy axis-view imaging that can deal with these properties (see Section 4.4). In phase I of this approach, we develop an improved 3D volumetric representation defined as the confidence map. It is derived from texture-augmented axial-view images of the specimen. In phase II, the 3D reconstruction is accomplished by searching the optimal surface for the specimen over the confidence map. Subsequently, from the obtained 3D reconstruction, 3D measurements can be extracted. We apply our MM-HTAI architecture presented in Chapter 3 and propose three typical datasets with different imaging modalities, including (1) standard RGB images, (2) the bright-field images of zebrafish larvae, and (3) zebrafish liver in fluorescence (see Section 4.3). In the experiments, we have applied our approach on these datasets. We find that our approach yields a precise 3D shape representation and a natural visualisation (see Section 4.5). In comparison with a groundtruth setup, we have obtained accurate 3D measurements both for the organism and the organ, which holds a promising shape assessment for model systems in biosciences (see Section 4.6).

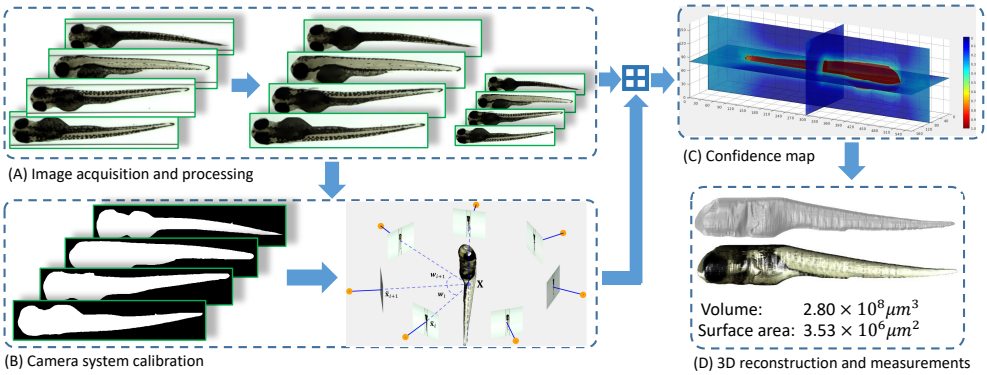

Figure 4.1: A schematic of the 2-3DLA approach. Phase I: (A) A series of axial-view images of a specimen are acquired. The textures are augmented by the mean shift filtering, and the multi-scale images are used as input of the system. (B) 2D shape approximations of the specimen are obtained from the augmented textures, by which the camera calibration can be accomplished. (C) An improved 3D volumetric representation in the form of a confidence map is derived from the textures through the volume intersection model. Phase II: (D) Region based level set method is adapted and applied to the confidence map to estimate the optimal enclosed surface retaining a natural shape and smooth appearance of the specimen, from which the volume and surface area are derived.

4.1 Improved 3D reconstruction

In modern life-science research, large volumes of microscopy data, i.e., representative images of a specimen from cellular to whole-mount scale are used. Robust and reproducible methods for data acquisition, image processing, and analysis are essential for further handling of the data. As data volumes get larger, comprehensive visualisations summarising the data content also become important.

Images from samples, i.e., specimens, acquired from light microscopy carry a great deal of information that can be expressed by features such as shape and texture. Variations in these features provide information for a classification according to a specific condition. Such an approach is typically employed in imaging applications in the fields of cytology, toxicology, oncology and others. At a glance, experts are able to observe variations in colour, though subtle differences of size and shape are more difficult to capture. Moreover, as the volume of the data increases, it is not possible to classify the differences objectively. Therefore, these features need to be extracted in an automated manner. So, more advanced systems need to be developed [38]. For high-throughput systems this is especially true. Manual

inspection is not feasible and in order to make inferences from the data, robust methods are required that produce accurate and reproducible measurements for shape analysis.

For the study of whole specimens, we should acknowledge that the nature of the shape is, de facto, in 3D. Thus, from a 3D shape representation of the specimen one will be able to produce an informative shape description, e.g., volume, surface area and 3D shape factors. With such features, we can accurately assess and compare shape variations in applications. These descriptors are also important for phenotypical systems evaluation that requires volume normalisation into metric measurements [39].

3D images can be generated by different techniques. With a confocal laser scanning microscope (CLSM) 3D images can be acquired from which 3D representations can be derived. For high-throughput imaging, CLSM is less efficient and also larger objects are not feasible for CLSM imaging. The post-processing of the images is not straightforward [94]. Alternative to CLSM is the optical projection tomography (OPT) [95]. However, a serious limitation of this technique is that it does not work for live specimens. Moreover, the specimen preparation is rather time consuming. In light field microscopy [96], microlenses are configured between the main lens and image sensor. Through 3D deconvolution a larger focal range is addressed and in this manner focal images are combined into a 3D image [97]. This technique is suitable for semi-transparent objects. In our research, we deal with specimens that consist of opaque, specular and transparent regions. A second limitation of light-field microscopy is the trade-off between spatial and angular resolution.

In this chapter, we present an axial-view imaging architecture based on light microscopy using the Vertebrate Automated Screening Technology (VAST BioImager) [26]. From this imaging architecture, we can generate a series of 2D axial-view images for the specimen. In fact, this imaging belongs to the class of multi-view imaging techniques, which is also referred to as turn-table sequence [31]. In the field of computer vision, multi-view stereo (MVS) approaches have been developed to recover a 3D scene, or an object, from a range of 2D multi-view images. In these approaches, surface points, or a depth map, of an object can be estimated through the matching of correlated images [30]. However, for light microscopy imaging, objects of interest can be (partially) translucent and/or transparent; this holds for the microscope modalities that we are using, i.e., bright-field and

fluorescence. These qualities are in favour of the observation of internal structures of the specimen but prevent the feasibility of surface points matching which is required for a MVS approach.

In Chapter 3, we have obtained a binary 3D volumetric representation from the shape-based 3D reconstruction which is intuitive and suitable for the goal of shape analysis through 3D measurements [98]. This method requires precise 2D shape segmentation. However, the accurate shape segmentation is sometimes difficult to obtain from images depicting specimens with poor-defined boundaries as a consequence of translucent and transparent properties of the specimen. Instead, a probabilistic framework for the 3D volumetric representation generates more flexible but not always accurate 3D shape description. To address the problems, we propose a **two**-phase approach for the **3D** reconstruction and measurements from **light** microscopy **axial**-view imaging; this is abbreviated as 2-3DLA. A schematic representation of this approach is shown in Fig. 4.1. The system takes a series of axial-view images as input and reconstructs a precise 3D model of the object from which accurate 3D measurements, i.e., volume and surface area, can be derived.

Specifically, in **Phase I** of 2-3DLA, an improved 3D volumetric representation in the form of a confidence map is constructed. First, the mean shift algorithm [29] is applied to improve the texture representation of the original images. Thus, the translucent and transparent regions of the specimen are enhanced so that they become more separable. Subsequently, approximations of the 2D shapes of the specimen are obtained and used for camera system calibration. Together with all the data, based on the camera projection intersection model, we obtain the confidence map by imposing a score, instead of a binary value or a probability, to each voxel element in the 3D space. The score is obtained by integrating two probabilistic models. These probabilistic models are jointly estimated from all the axial-view images in a multi-scale fashion and aims to generalise the texture distributions of the object and background. The confidence map indicates the likelihood of each voxel in 3D space to be part of the object. This is a more flexible 3D representation for the optimisation of the surface. In **phase II**, we present the assumption that the optimal surface which includes the specimen will be able to (1) maximally separate the voxels from the object to the background as well as (2) retain a smooth appearance of the object. The 3D reconstruction and measurements are accomplished by searching for such an enclosed surface over the confidence map. Consequently, we formulate the 3D reconstruction as a

3D segmentation problem which can be solved by employing a region based level set method.

In modern life-sciences, e.g., developmental biology and pharmacokinetics, zebrafish are widely used as model systems in various experimental settings [18]. Zebrafish are small in embryonic stages and can be easily studied with different types of microscopes; zebrafish are transparent in the early stages. In recent years, zebrafish are augmented with a large amount of reporter lines and these lines are extensively used in disease studies. We use our light microscopy axial-view imaging architecture to acquire three representative datasets including bright-field images of zebrafish and fluorescence images of zebrafish liver. We use the zebrafish and the corresponding axial-view images as a case study for the evaluation of 2-3DLA. We have found that the 2-3DLA can be successfully applied in this research field; we have achieved promising results for 3D reconstruction and measurements of zebrafish and its liver.

We summarise our major contributions as follows.

I We present a multi-modal axial-view imaging architecture using light microscopy.

II We present three representative datasets of light microscopy axial-view imaging including the zebrafish in bright-field and the zebrafish liver in fluorescence.

III We propose a computational and automated system named 2-3DLA to solve the problem of accurate 3D reconstruction and measurements.

IV The proposed 2-3DLA is applicable in the datasets to obtain detailed 3D shape description both for the zebrafish and its liver. This makes the 2-3DLA generic for shape analysis on the level of the organism as well as on the level of the organ.

The remainder of the chapter is structured as follows. In Section 4.2, the background and related topics of our approach is introduced and the proposed system is motivated. In Section 4.3, we introduce three different complementary datasets using the MM-HTAI architecture. In Section 4.4, we describe the prior knowledge of the light microscope camera calibration and the binary 3D volumetric representation. This is followed by a detailed description of the proposed system. In Section 4.5 we present experiments and discuss the results. Finally, in Section

4.6 we summarise the answer to RQ 4, present conclusions and describe future work.

4.2 Background and related work

Given a sequence of calibrated multi-view images of an unknown scene, one category of the multi-view stereo (MVS) approaches aims to estimate a depth map for each view by matching each pixel of a binocular image pair [99, 100, 101]. The matching criterion is usually defined within a support window, i.e., a local neighbourhood, centred around a target pixel [102, 103]. Another approach formulates the depth map estimation as a continuous optimisation problem of an energy functional integrating colour, spatial consistency of neighbouring views and a global smoothness constraint [104]. A merging strategy is developed to integrate the multi-view depth maps. A patchmatch method [105] takes the slanted planar surface into account for a better depth map estimation in binocular stereo. An extension of the patchmatch method is developed for massive parallelisation and integration of multi-view depth maps [106].

Alternatively, a conventional MVS approach directly estimates the surface points of an object using a sparse-to-dense strategy [107]. A novel approach [108] extends the patch-based method into multi-view 4D reconstruction, solving the problem of temporally consistent 3D modelling in videos.

Regarding MVS, an accurate estimation of epipolar geometry is essential, which could significantly reduce the search space on the epipolar line and thereby increase the matching quality [67]. Structure from motion (SFM) enables the estimation of camera poses from the cooperation of salient point detection [109, 110] and bundle adjustment [89, 111]. However, both SFM and MVS rely on the quantity and quality of the salient point detection and matching from correlated images. In our light-microscopy imaging, the challenge is that the object surface and boundaries are not always well-defined. In most cases, volume instead of surface for an object is observable.

For image based 3D reconstruction, the volumetric representation approach is an important category. The goal is to estimate a convex hull in the 3D space represented as discrete voxels according to their visibility to each view [31, 73, 74, 112]. Accordingly, the shape-based method defines the 3D object through the

intersection of a set of projections exposed from the 2D shapes of the objects in the images. A space carving algorithm aims to recover the 3D object by wiping out the voxels which are consistently invisible to the views [75, 76]. These methods, however, require accurate image segmentations which are not always available in light microscopy imaging.

In addition, the textures from the multi-view images can be used to optimise the 3D volumetric representation based 3D reconstruction [79, 80, 81]. Instead of using a binary representation, the probabilities indicating the membership of each voxel are estimated by applying multivariate Gaussian kernels on the textures of foreground and background, respectively. However, this requires user specification. Consequently, the quality of user input will, to a certain extent, determine the quality of the 3D reconstruction. Other methods attempt to fuse the shape and texture consistency in an integral deformable framework [113].

There are also other well-designed 3D reconstruction methods. Some in particular solve the problem of specular and transparent object reconstruction [114]. These methods work well on a macroscopic scale as they position special patterns behind the object to enable the shape-from-distortion or set up various lighting to collect surface reflective-highlights [115, 116, 117]. A recent study [118] presents a semantic reconstruction as a convex-relaxation formulation which combines a data term and a regularisation constraint, achieving elaborate results on public datasets [100]. Nowadays, deep learning is used in the MVS to improve the matching quality [119, 120]. This type of method requires a large volume of training data, which, in our case, is not available.

From the study of related work, we may conclude that the volumetric representation based approaches, such as shape-based 3D reconstruction with binary volumetric representation are most promising in addressing the challenge of 3D reconstruction and measurements for light microscopy imaging. However, this conclusion is not sufficient in itself, as in some cases accurate 2D shapes cannot be obtained. Therefore, we have developed the 2-3DLA method to solve the problem. In our approach, we first estimate a confidence map using the augmented textures from the axial-view images. The confidence map is actually an improved 3D representation. Subsequently, we accomplish the 3D reconstruction using the region based level set method on the confidence map. A validation has shown that our method yields accurate 3D measurements.

4.3 Dataset collection

In Chapter 3, we have developed our MM-HTAI architecture. In this section, we provide a comprehensive depiction of the dataset collection and show how the multi-modal images are produced.

The VAST BioImager is specifically set up to work for high-throughput imaging of zebrafish. For the remainder of the chapter, the concept of specimen and zebrafish are both used. Specimen is used in a generic context while zebrafish is used in the particular application of the zebrafish high-throughput imaging. In the VAST BioImager, the specimens, i.e., zebrafish larvae, are positioned along their longitudinal axis as in this manner the most important features can be readily observed; it is also related to the manipulation of the specimen in the capillary. In this orientation we obtain images from all axial-views along the profile axis from one full revolution. A schema of the imaging architecture is depicted in Subsection 3.3.1 (Fig. 3.2). A positioning module consists of a capillary (the holder of specimen) and a set of stepper motors that accomplish the rotation manipulation of the specimen.

A VAST camera (#1) is mounted with the device and used to detect the location and orientation of the specimen so as to keep it in the field of view for image acquisition. An object is revolved over 360 degrees by the stepper motors so that bright-field images for the specimen in the axial-views can be acquired. These image always depict the whole specimen. The VAST camera (#1) is an Allied Vision Systems, Pro Silica GE 1050 CCD (pixel size 5.5 $\mu m \times 5.5$ μm.). This camera acquires images of 1024×1024 pixels. The resolving power for this camera system is about 13.4 μm.

The VAST unit is mounted on a microscope, of which the microscope-camera (#2) is able to acquire detailed microscopic images both in bright-field and fluorescence from arbitrary axial positions; this setup uses the Leica DFC450C CCD (pixel size 3.4 $\mu m \times 3.4$ μm) and it acquires images of 1920×2560 pixels (5 Megapixel). We use two objectives in this case. The resolving power for a 2.5× objective is about 4.8 μm and for a 4× lens (red fluorescence) it is about 3.1 μm.

The image acquisition protocol is as follows. In a full revolution of the capillary, 84 axial-view images are acquired for the specimen. The step size between adjacent axial-views is about 4.3° ($360° \div 84 \simeq 4.3°$). We have shown that 21 evenly sampled axial-views ($N = 21$) are sufficient to obtain accurate results [98]. We

apply this acquisition protocol to obtain 3 different datasets that we will use as case-studies for the evaluation of our 2-3DLA approach under different imaging conditions.

Dataset A is obtained by the VAST camera. A number of zebrafish larvae during different development stages, i.e., 3, 4 and 5 days post fertilisation (dpf) are acquired. The dataset finally contains 12 examples for 3 dpf, 24 examples for 4 dpf and 24 examples for 5 dpf, respectively, and 60 examples in total. We will use the images from this dataset as example to interpret our 2-3DLA system. Examples of the images in this dataset are shown in Fig. 4.3. This dataset is also used in Chapter 3.

Dataset B is obtained using the microscope-camera; the VAST unit manipulates the position of the specimen while, in this case, the microscope-camera acquires bright-field images. For the acquisition the objective $2.5 \times /0.07NA$ is used. *Dataset B* consists of a set of representative examples, i.e., 3 specimens of 5 dpf. Examples of the images are depicted in Fig. 4.6.

Dataset C represents a collection of images of an internal structure of the zebrafish, i.e., an organ. These images are obtained using the fluorescence imaging modality as the organ is specifically visible with fluorescence. For fluorescence we use the objective $4 \times /0.12NA$ to collect the dataset for the zebrafish liver. We used 7 zebrafish samples of 3 dpf. The microscope is equipped for fluorescence and the images are acquired with the microscope-camera. Examples of the images are depicted in Fig. 4.7; in Appendix A the preparation of these samples is given.

4.4 Two-phase 3D reconstruction from axial-views

In this section we elaborate on the 2-3DLA approach. For the 3D reconstruction using axial-views, a microscope camera calibration is necessary. This can be solved by the algorithm of voxel residual volume maximization. As a baseline method, the shape-based 3D reconstruction is used for comparison with our method. For the details of the microscope camera calibration and the shape-based 3D reconstruction method, we refer to Section 3.3. Below we discuss: the improved 3D volumetric representation as confidence map in Subsection 4.4.1 and 3D reconstruction as objective function optimisation in Subsection 4.4.2.

4.4.1 Improved 3D volumetric representation as confidence map

In shape-based 3D reconstruction we need accurate segmentations of the original images. These are, however, not always satisfactory. One of the important complications is the translucency and transparency of the specimen in the images. We, therefore, propose the improved 3D volumetric representation in the form of a confidence map which is derived from the probabilistic models as estimated from the object and background presented in the axial-view images.

We observe that the partial transparency of the specimen is difficult to recover. Severe texture variation within the object challenges the generality of the probabilistic models. We therefore need to collect prior knowledge on the approximate locations of the objects of interest and background. It can be implemented in an interactive manner. In this case, the results of the 3D reconstruction will rely on the quality of the user input; insufficient user input tends to produce an underestimated model.

In following, we propose to apply the MS algorithm to obtain the approximations of the 2D shapes for the specimen. The MS algorithm improves the texture representation of the object so that the translucent and transparent regions of the specimen become more separable from the background. We threshold the texture-augmented images to obtain the 2D shape approximations. This also results in an augmentation of the texture representation in the transparent parts of the specimen. The approximations of the 2D shapes can roughly separate the object and background. Although such shape approximations are inaccurate, morphological operations to the envelope of the shape (dilation and erosion) enforce the discrimination of the object and background. In this manner, we have solved the problem of texture augmentation for the partial transparent specimens. In addition, the obtained 2D shape approximations can serve as the shape constraints, such that almost all texture information included by the object can be sampled. This improves the generalisation of the probabilistic models and contributes to the successful automation of the whole system.

We combine the multi-scale textures interpreted as image pyramids to further enhance the probabilistic models. Now, let $\mathcal{I}_{sc} = \{\mathbf{I}_{(1,sc)}, \mathbf{I}_{(2,sc)}, ..., \mathbf{I}_{(N,sc)}\}$ and $\mathcal{S} = \{\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_N\}$ denote the images (after mean shift filtering) in scale of sc and the approximations of 2D shapes in original scale, where N represents the number of views. In any of the views, the foreground and background can be approximately discriminated and indexed by $\mathbf{x}_f$ and $\mathbf{x}_b$, where $\mathbf{S}(\Omega) \approx \{\mathbf{x}_f, \mathbf{x}_b\}$

and Ω represents the whole image domain. Then, the multi-scale textures are represented as $\mathbf{I}_{sc}(\mathbf{x}_f)$ and $\mathbf{I}_{sc}(\mathbf{x}_b)$, where $\mathbf{I}_{sc}(\mathbf{x}) = (r,\ g,\ b)^T$ denote the RGB values in colour space for any pixel in scale sc. Next, a multivariate Gaussian is applied to estimate the probabilistic models of the range images for each specific scale. Thus:

$$\begin{aligned} p_{(f,sc)}(\mathbf{x}_f) &= \frac{1}{\sqrt{(2\pi)^3|\mathbf{\Sigma}_f|}} e^{-\frac{1}{2}(\mathbf{I}_{sc}(\mathbf{x}_f)-\mu_f)^T \mathbf{\Sigma_f^{-1}} (\mathbf{I}_{sc}(\mathbf{x}_f)-\mu_f)} \\ p_{(b,sc)}(\mathbf{x}_b) &= \frac{1}{\sqrt{(2\pi)^3|\mathbf{\Sigma}_b|}} e^{-\frac{1}{2}(\mathbf{I}_{sc}(\mathbf{x}_b)-\mu_b)^T \mathbf{\Sigma_b^{-1}} (\mathbf{I}_{sc}(\mathbf{x}_b)-\mu_b)} \end{aligned} \tag{4.1}$$

where $p_{(\bullet,sc)}$ represents the probabilistic model estimated for the textures in scale sc; and $(\mu_f,\ \mathbf{\Sigma}_f)$ and $(\mu_b,\ \mathbf{\Sigma}_b)$ denote the mean vectors and the covariance matrices estimated for the foreground and background, respectively. To avoid overestimation of the probabilistic models, we randomly select a subset of the axial-views for computation.

A voxel $\mathbf{X} \in \mathcal{X}$ in 3D space corresponds to a pixel location in $\mathbf{I}_{(i,sc)}$ (the i^{th} image in scale sc) through the pinhole camera model $\mathbf{x} = \mathbf{P}_i\mathbf{X}$, where $\mathbf{P}_i \in \mathcal{P}$ is the camera projection matrix estimated for each view. We can obtain the probabilities of this voxel being the foreground and background by $p_f(\mathbf{x})$ and $p_b(\mathbf{x})$. Suppose that all the N axial-view images $\mathcal{I}_{sc}$ in scale sc are independent, the joint probabilities of a voxel $\mathbf{X}$ indicating its visibility to the foreground and background can then be modelled as follows:

$$\begin{aligned} P_{(f,sc)}(\mathbf{X}) &= \left(\prod_{i=1:N} p_{(f,sc)}(\mathbf{x}) \right)^{\frac{1}{N}} \\ P_{(b,sc)}(\mathbf{X}) &= 1 - \left(\prod_{i=1:N} (1 - p_{(b,sc)}(\mathbf{x}) \right)^{\frac{1}{N}} \end{aligned} \tag{4.2}$$

We can interpret Eq. (4.2) as the intersection of the N camera projections for each voxel. We use the logarithm to re-scale the two probabilities. The multi-scale probabilistic models are then fused to obtain the confidence map defined as:

$$\mathbf{I}^* = \sum_{sc=1}^{M} \beta_{sc} \left[log(P_{(f,sc)}) - log(P_{(b,sc)}) \right] \tag{4.3}$$

where $\beta = \{\beta_{sc}\}_{sc=1:M}$ specifies the weight for each scale and $\mathbf{1}^T \cdot \beta = 1$. In our specific implementation, a uniform distribution of β is used, which simply averages the multi-scale probabilities. We use three-scale textures ($M = 3$) in a hierarchical fashion, such that the higher scales can be generated by halving the images in the respective lower scales. The confidence map takes a confidence score as the entry for each voxel to indicate its likelihood to be the object (or background). The larger the confidence score of a voxel is, the more possible that voxel is classified as the object.

4.4.2 3D reconstruction as objective function optimisation

With the confidence map, we can accomplish the 3D reconstruction by searching for an enclosed surface which is able to correctly classify all the voxels into the object and background. The marching cubes algorithm [121] straightforwardly estimates a specific surface. However, it is difficult to determine which is the optimal one; and serious fractal behaviour of the estimated surface will occur. This will result in inaccurate 3D measurements of volume and surface area. An empirical solution can be applied to cascade a surface refinement module which needs to be carefully validated. In this work, we apply the region based level set method (CV model) [41] to accomplish this task. The employment of the CV model transfers the 3D reconstruction problem into a 3D segmentation. The optimal enclosed surface retaining a smooth appearance is achieved.

According to the CV model, an enclosed 3D surface $\mathcal{C}$ is embedded in a distance regularised level set function (LSF) $\mathcal{C} := \{\mathbf{X} \in \Omega; \phi(\mathbf{X}) = c\}$, where Ω and $\mathbf{X}$ are consistently defined as the image domain (3D in this case) and the spatial locations of the voxels in world frame. To separate the object and background according to the similarity in the confidence map we define the objective function Eq. (4.4). This includes the external force as the confidence map and the internal force as the smooth appearance of the surface. Thus:

$$
\begin{aligned}
E(\phi) \;=\; & \mu \underbrace{\int_{\Omega} |\nabla H(\phi)| d\mathbf{X}}_{\text{internal force}} \\
& + \upsilon \underbrace{\int_{\Omega} \left[|\mathbf{I}^* - u_{in}|^2 H(\phi) \;+\; |\mathbf{I}^* - u_{out}|^2 (1 - H(\phi)) \right] d\mathbf{X}}_{\text{external forece}}
\end{aligned}
\tag{4.4}
$$

where, u_{in} and u_{out} represent the mean intensity of the confidence map in and outside of the surface; μ and υ are constants used to balance different terms. H is the Heaviside function defined as follows:

$$
H_{\epsilon}(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \tag{4.5}
$$

Using the Euler-Lagrange equation, as:

$$
\frac{\partial \phi}{\partial t} = -\frac{\partial E}{\partial \phi}, \tag{4.6}
$$

one can implement the surface evolution by computing the gradient flow for the objective function defined as:

$$
\frac{\partial \phi}{\partial t} = \delta(\phi) \Big\{ \mu \; div \left(\frac{\Delta \phi}{|\Delta \phi|} \right) + \upsilon \left[\left(|\mathbf{I}^* - u_{out}|^2 - |\mathbf{I}^* - u_{in}|^2 \right) \right] \Big\}, \tag{4.7}
$$

where $\delta(\phi)$ is the derivative of $H(\phi)$. By setting a proper step size Δt, we can complete the 3D reconstruction via gradient descent, defined as:.

$$
\phi_{t+1} = \phi_t + \Delta t \frac{\partial \phi}{\partial t} \tag{4.8}
$$

3D measurements inference Given the 3D reconstruction represented as the optimal surface embedded in ϕ, we can derive the 3D measurements. The volume is obtained by the integration over all the voxels which are included in the object ϕ: $V = \sum_{\phi} \mathbf{1}[\phi > c^*] \times v$, where $\mathbf{1}[\bullet]$ keeps its representation as an indicator function; c^* is the optimal c-level set which we will investigate in following sections; v is the unit volume for the voxels which is pre-defined. A set of surface points are generated from the c^*-level set of ϕ, which will be used to produce a triangulated mesh. The surface area is obtained by the integration over all the facets in the

triangulated meshes using Heron's formula $A = (s(s-a)(s-b)(s-c))^{1/2}$, where $s = (a+b+c)/2$ and $a,\ b,\ c$ represent the edges of a triangle.

4.5 Experiments

In this section, we describe our experiments to evaluate the performance of the proposed approach for our datasets. In Subsection 4.5.1, we perform a visual inspection of the results on *Dataset A*. In Subsection 4.5.2, we discuss how to obtain accurate 3D measurements for *Dataset A*. In Subsection 4.5.3, we perform visualisation and 3D measurements on *Dataset B & C*. In Subsection 3.5.4, we evaluate the proposed method by runtime.

4.5.1 Visual inspection of the results on *Dataset A*

We first apply (A) different methods on the *Dataset A* to compare the obtained 3D models through visualisation. We next (B) apply our 2-3DLA approach on the whole *Dataset A* to demonstrate its performance on the zebrafish in various developmental stages.

(A) Comparison with various methods using visualisation

We implemented an interactive method [81] which requires user input to indicate the approximate locations of the object and background. So, the quality of the 3D models will rely on the user input. The shape-based 3D reconstruction is also used in this comparison.

In Fig. 4.2 (A), the results obtained by the interactive method are shown. Two types of user input are employed to estimate the probabilistic models for each specimen. One input collects less textures by drawing shorter scribbles on the object. The other input covers a larger region of the object, thus collecting more texture information. In Fig 4.2 (A), the corresponding 3D reconstructions are shown next to the input drawings. Our assumption is confirmed that variations in user input impact the quality of the 3D reconstructions. Sufficient texture sampling usually results in better 3D reconstructions preserving most of the shape of the specimen. However, even with "good" user input, parts of the zebrafish, mostly at the tail area, still fail to be faithfully reconstructed. The main reason is the considerable texture inhomogeneity of the specimen which hampers the

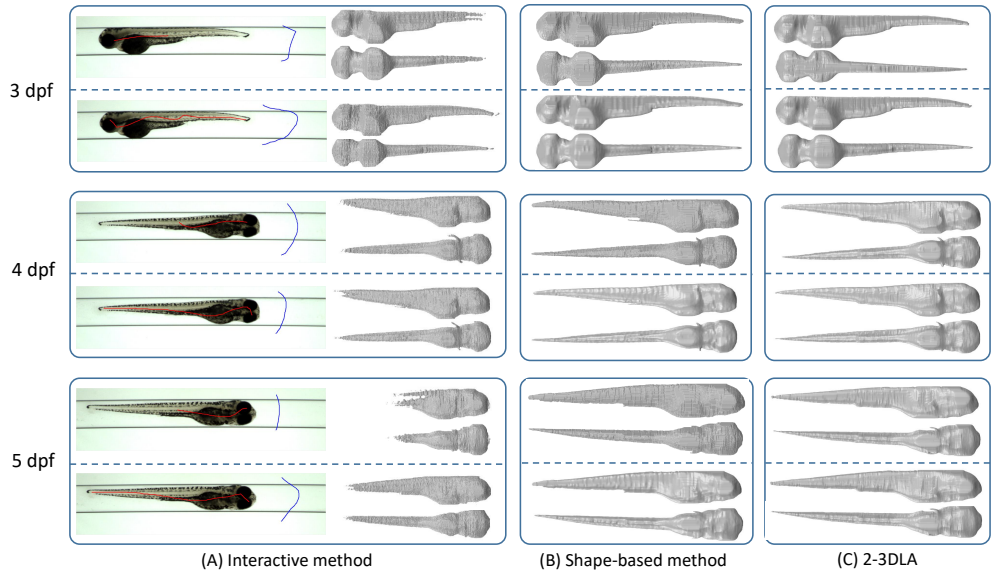

Figure 4.2: Comparison of visualised results of different methods. We select one example from each larval stage, i.e., 3, 4 and 5 dpf. For each 3D model, we visualise the same perspectives, i.e., lateral and ventral. (A) The 3D reconstruction of the interactive method. The users randomly draw on the object of interest (red scribble) and background (blue scribble), respectively. Two types of user input are considered, as shown on the left. Corresponding 3D reconstructed models are shown on the right. (B) The 3D reconstruction of the shape-based 3D reconstruction. The raw 3D models are shown in the top two rows. The 3D models with extra surface refinement are shown below. (C) The 3D reconstructions of the 2-3DLA using only the original-scale textures and multi-scale textures are separately shown in each box above and below.

generality of the probabilistic models and thereby results in a misclassification of the translucent part of the specimen as background. From the results of the interactive method, we can also observe the fractal behaviour of the surface. It is apparent that inaccurate 3D reconstruction will result in inaccurate 3D measurements.

In Fig. 4.2 (B) we show results of the shape-based method with a set of accurate 2D shapes, which generates very accurate 3D shape representation of the specimen. Each shape shows a natural and detailed surface, but fractal behaviour also occurs due to the carving effects of the method. This is depicted in the top two rows of Fig. 4.2 (B). After extra surface refinement [93], accurate 3D measurements can be available. The refined 3D models are shown in the bottom two rows of Fig. 4.2 (B). We need to accept that the shape-based method requires

accurate 2D segmentations from the original images which are sometimes difficult to obtain in light microscopy imaging.

In Fig. 4.2 (C) we show the 3D models obtained by the 2-3DLA system. In the top two rows, the 3D reconstructed models are obtained by the confidence map derived with one-scale texture. In the bottom two rows, the results obtained with three-scales texture are shown. One can observe a better 3D visualisation for both of the results. In the 2-3DLA, the employment of the MS algorithm ensures improved discrimination between the object and background. Furthermore, the shape constrained texture collection augments the generality of the confidence map. These improvements cooperate in obtaining a better volumetric representation. The CV model aims to search for the optimal surface which can largely separate the object and background in the confidence map and preserves a smooth appearance of the specimen, which subsequently results in accurate 3D measurements. However, a little fractal behaviour is still showing in the results with one-scale texture, resulting in a slightly deficient tail region of the zebrafish.

(B) 3D visualisation for *Dataset A* using 2-3DLA

Here we apply the 2-3DLA on the whole *Dataset A*. In Fig. 4.3 some results are depicted. For each developmental stage of the zebrafish, we randomly select one specimen in the dataset. To create additional value for 3D models, we map the textures from the original images onto the surface of the 3D models to produce texturised 3D shapes. For practical reasons, the texturing of the partial transparent regions is artificial. We separately show the binary and the texturised 3D models in the middle and bottom rows for each specimen. We select three typical axial-views, i.e., lateral, tilted, and ventral.

From the visualisation, we can assess that the 2-3DLA obtains quite natural and vivid 3D shapes for the specimens. The problem arising from the translucency and transparency seems to be solved. From the 3D shapes it is obvious that older zebrafish are larger compared to the younger ones. The size of yolk in the various stages is different. During the development, the yolk of the zebrafish provides nutrient for its growth. Consequently, older zebrafish have smaller yolk, simply as a result of their growth. Furthermore, the 3D shape obtained by our approach offers a good framework for the visualisation of the biological system by introducing organ-level 3D systems, the circulatory system, the liver, the brain, etc.. Using specific staining, the 3D systems can be incorporated and visualised. The whole

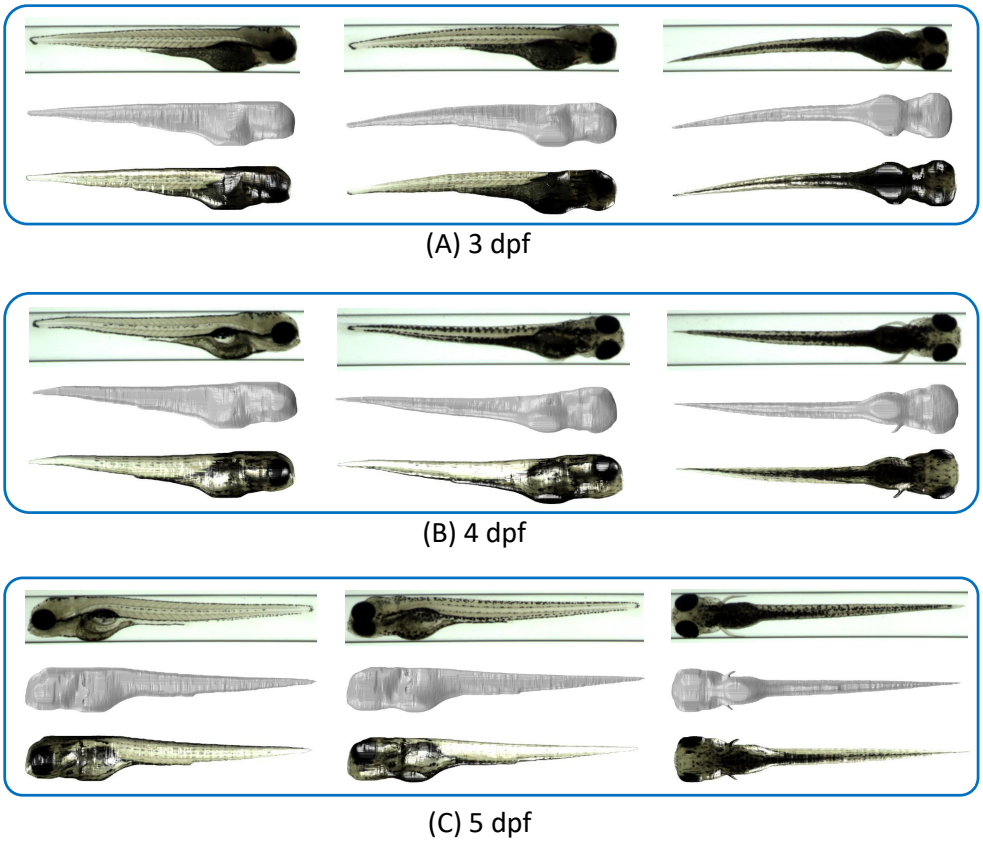

Figure 4.3: 3D models visualisation of the zebrafish in various larval stages from *Dataset A*. Each bounding box denotes a staged zebrafish. For each specimen, three typical perspectives are shown in different columns. The original axial-view images are shown in the top row; the pure 3D models are shown in the middle row; the texturised models are shown in the bottom row.

process is reviewed in a 2-3DLA spotlight including animated visualisations of the 3D models at `http://bio-imaging.liacs.nl/galleries/VAST-2-3DLA/`.

4.5.2 3D measurements for *Dataset A*

We investigate how to obtain accurate 3D measurements from the 3D models obtained by the 2-3DLA. To this end, the hyperparameter of c-level set needs to be tuned, because an unvalidated c-level set may lead to inaccurate 3D measurements. We present two interpretations, (A) the 2D shape coherence and (B) the 3D shape coherence, for the inference of accurate 3D measurements. We compute the (C) 3D measurements based on the inference for *Dataset A*.

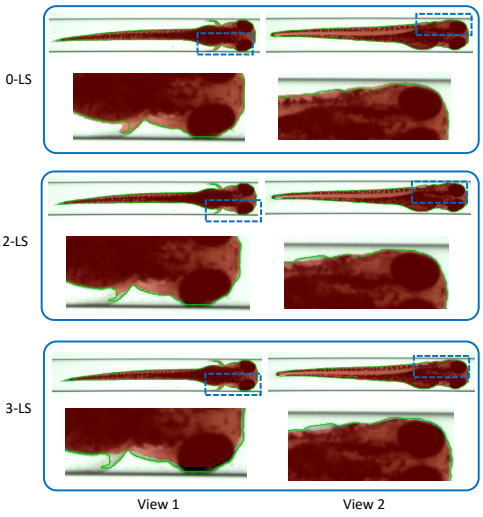

Figure 4.4: 2D shape coherence (red) between the projections from 3D models to axial-view images and the groundtruth 2D shapes (green contours) with various c-level sets. A subsection of the original 2D shape coherence (dashed bounding box) is zoomed in and shown at the bottom of each row.

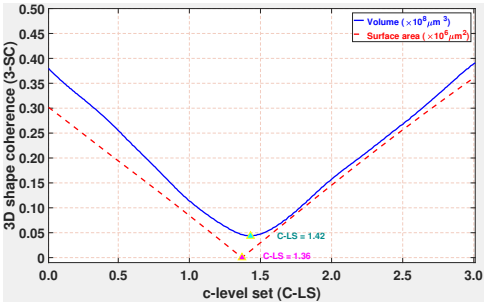

Figure 4.5: 3D shape coherence with various c-level sets. The triangles indicate the minimum of the measurements, which are found around 1.4-level set.

(A) 2D shape coherence

A good 3D reconstruction must approximate the real shape of the object. We evaluate and interpret this as shape coherence. We define the 2D shape coherence as a disparity map from the projected shape of a 3D model to a "groundtruth 2D shape" in each original axial-view. We have developed a robust method for the segmentation of zebrafish larvae. Using manual segmentations, the method is validated and turned out to be accurate. We thereby use the segmentation results as the approximation of the groundtruth 2D shape for the *Dataset A*. Accordingly, we interpret the 2D shape coherence as the overlap from the projected shape to the groundtruth 2D shape. We chose three different c-level sets (c-LS), i.e., 0-LS, 2-LS, and 3-LS, to extract the corresponding 3D models. The 2D shape coherence that are obtained in this manner are shown in Fig. 4.4. We depict two typical axial-views of the object, i.e., ventral and lateral. We represent the groundtruth 2D shapes as green contours. In order to appreciate the results, a subsection of the 2D shape coherence is zoomed in and its visualisation is shown in the bottom

row for each view.

We observe that a small c-LS, i.e., 0-LS, produces an overestimation of the 3D model, which presents a large 2D shape coherence. This is illustrated as the red area exceeds the green contours in the top row of Fig. 4.4. A large c-LS, i.e., 3-LS, results in a more compact 3D shape of the object, which obtains a smaller 2D shape coherence. It is illustrated as the red area separating from the green contours in the bottom row of Fig. 4.4. In contrast, a reasonable c-LS, i.e., 2-LS, obtains highly correlated 2D shape coherence producing a better 3D reconstruction.

(B) 3D shape coherence

The shape-based 3D reconstruction method results in a binary volumetric representation. If accurate 2D shapes are used in this method, the obtained 3D model can serve as the "groundtruth". The 3D shape coherence is then defined as the disparity from the 3D models obtained by the 2-3DLA to the "groundtruth 3D model". We propose two measurements to evaluate the 3D shape coherence: volume disparity ($\overline{V}$) and surface area disparity ($\overline{A}$). They are separately defined as $\overline{V} = [(1/M)\sum_i (V_i^* - V_i^m)^2]^{(1/2)}$ and $\overline{A} = [(1/M)\sum_i (A_i^* - A_i^m)^2]^{(1/2)}$, where V^* and A^* are the volume and surface area measured from the 3D model groundtruth; V^m and A^m are the volume and surface measured from the 3D models obtained using a specific c-LS; and M is the number of specimens. We use half of the *Dataset A* for the validation of this experiment. In Fig. 4.5, we show a graph of the 3D shape coherence for various c-LS. We illustrate that the choice of a small c-LS, e.g., 0-LS, tends to produce an overestimation of the measurements. The choice of a large c-LS, e.g., 2-LS, will lead to an underestimation. This corresponds nicely with the findings in the 2D shape coherence as shown in Fig. 4.4. Finally, we find that the 1.4-LS is a good compromise and gives the best estimation of the volume and surface area. For the next paragraph (4.5.2 (C)), we use the 1.4-LS to obtain the deterministic 3D models from which we can derive the 3D measurements.

(C) 3D measurements for *Dataset A*

We use the remaining half of the *Dataset A* to compute the 3D measurements of the zebrafish in various developmental stages, i.e., 3, 4, and 5 dpf. In this experiment, we compute the statistics for the volume and surface area of the zebrafish and show the results in Table 4.1 and 4.2. The 3D measurements obtained from

Table 4.1: Volume statistics ($\times 10^8 \mu m^3$) of the 3D models for the zebrafish in various developmental stages

	Shape	2-3DLA
3 dpf	2.53 ± 0.11	2.56 ± 0.12
4 dpf	2.63 ± 0.19	2.66 ± 0.19
5 dpf	3.00 ± 0.18	2.96 ± 0.18

Table 4.2: Surface area statistics ($\times 10^6 \mu m^2$) of the 3D models for the zebrafish in various developmental stages

	Shape	2-3DLA
3 dpf	3.20 ± 0.15	3.21 ± 0.14
4 dpf	3.34 ± 0.17	3.36 ± 0.17
5 dpf	3.63 ± 0.14	3.61 ± 0.14

the shape-based 3D reconstruction method are used to assess the performance of the 2-3DLA.

From the tables we may conclude that the 2-3DLA can obtain accurate 3D measurements of volume and surface area comparable to the groundtruth method. This holds for the zebrafish in the three developmental stages. The results of the 2-3DLA are slightly larger, but the difference is small and acceptable. One should note that the 2-3DLA will be advantageous for the cases in which accurate segmentations are not available. Furthermore, the increasing trends of the volume and surface area for the zebrafish staged from 3 dpf to 5 dpf is consistent with the expected physical growth of the zebrafish. Importantly, Table 4.1 and 4.2 offer baseline metrics for volume and surface area of the zebrafish in various stages; this facilitates quantitative and qualitative analysis using the zebrafish as model system.

4.5.3 3D reconstruction and measurements of *Dataset B & C*

We have shown that the 2-3DLA can obtain an accurate 3D shape description of the whole specimen offering the baseline reference for the 3D measurements. This is, however, not sufficient to evaluate a specimen on the level of organs. From the optics used, the resolving power of *Dataset B & C* (see Section 4.3) is much higher. Moreover, different microscope modalities are used for these datasets, i.e., bright-field and fluorescence. As a case study, in addition to the whole specimen, we use the zebrafish liver as a model organ system. The liver size is important to pharmacokinetics, as it is an crucial organ for drug metabolism. For *Dataset B & C*, 3D measurements of volume and surface area are separately reported in Table 4.3 and 4.4. As a comparison, we still use the results obtained from the shape-based 3D reconstruction approach.

Table 4.3: Volume ($\times 10^8 \mu m^3$) and surface area ($\times 10^6 \mu m^2$) of the 3D models for the zebrafish in *Dataset B*

	Volume		Surface area	
	Shape	2-3DLA	Shape	2-3DLA
#1	3.00	3.05	3.61	3.65
#2	2.95	3.04	3.52	3.60
#3	3.06	3.13	3.69	3.74

Table 4.4: Volume ($\times 10^5 \mu m^3$) and surface area ($\times 10^4 \mu m^2$) of the 3D models for zebrafish liver in *Dataset C*

	Volume		Surface area	
	Shape	2-3DLA	Shape	2-3DLA
#1	7.70	7.56	4.67	4.57
#2	5.38	4.15	3.61	3.15
#3	8.10	8.47	4.91	5.04
#4	9.06	9.28	5.20	5.18
#5	11.6	9.77	6.07	5.49
#6	15.1	16.6	6.89	7.69
#7	6.55	6.79	4.28	4.31

In Table 4.3 we observe only a small difference of the volume and surface area produced by the 2-3DLA compared to baseline method. This difference is acceptable. It is apparent that both of the measurements are within the distributions given in Table 4.1 and 4.2. This suggests that the 2-3DLA system shows stable performance for bright-field microscopy axial-view imaging. Subsequently, it can obtain accurate 3D measurements for the specimen as a whole. In Fig. 4.6, we show the 3D models for two specimens. The results are presented in the similar fashion as used in Fig. 4.3. An interesting phenomenon shown in the Example #1 is that the tail of the 3D model is thinner than the original specimen. The reason is that, in this experiment, the specimen were alive and anesthetised. During the imaging, the motion of the positioning capillary introduces slight shape deformations of the specimen. This impedes accurate 3D reconstruction which requires static objects.

In Table 4.4 we can find that most of the zebrafish livers are reconstructed well and comparable 3D measurements are obtained. Significant difference for the 3D measurements are found in example #2. Both the volume and surface obtained by the 2-3DLA are smaller than those obtained by the baseline method. This

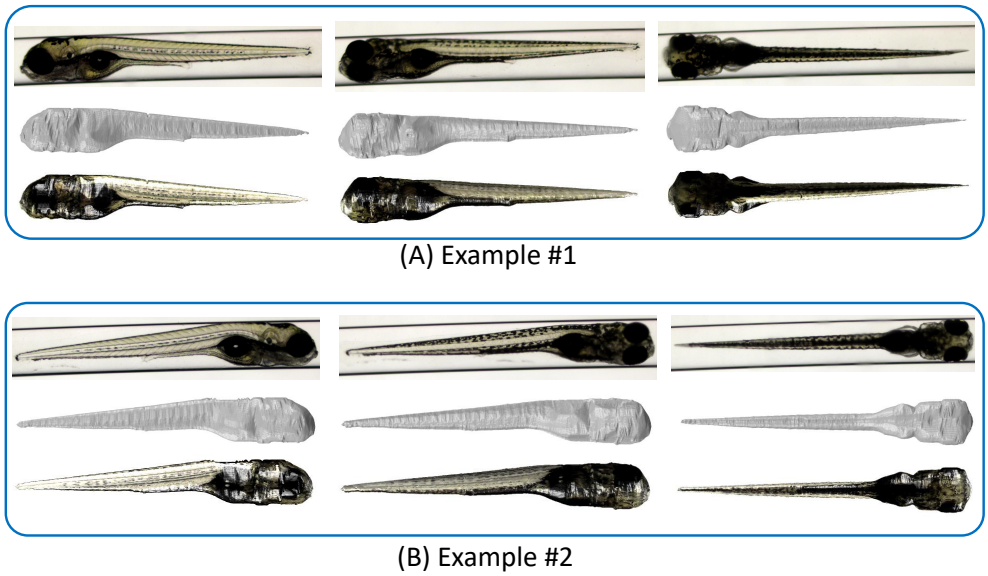

Figure 4.6: Visualisation of 3D models of the zebrafish in bright-field microscopy from *Dataset B*. For each specimen, three typical axial-view images are shown on the top; the corresponding axial-views of the pure and texturised 3D models are shown in the middle and bottom, respectively.

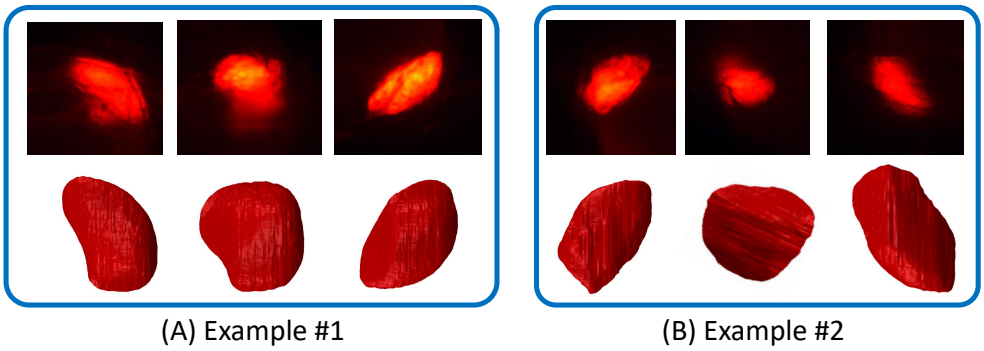

Figure 4.7: Visualisation of 3D models of the zebrafish liver in fluorescent microscopy from *Dataset C*. We show three axial-view images and the corresponding 3D perspectives in the top and bottom row for each specimen.

is caused by the image quality for the example #2. We visualise the 3D model for this example in Fig. 4.7 (B). We observe that the original axial-view images do not clearly represent the specimen. A reason for the obfuscation may be the relatively small external force in Eq. (4.7) for all the specimens, which produces compact 3D models. In fact, using a different configuration for the parameter in example #2 will result in accurate 3D measurements. However, in order to ensure

Table 4.5: Performance evaluation on runtime (s=second,min=minute)

	Step A	Step B	Step C	Step D	Total
Dataset A	$36 \pm 1.4(s)$	$12.6 \pm 0.8(s)$	$35.9 \pm 0.9(s)$	$2.2 \pm 0.5(min)$	$\sim 3.6(min)$
Dataset B	$4.3 \pm 0.2(min)$	$56.4 \pm 0.9(s)$	$37.2 \pm 0.4(min)$	$7.9 \pm 0.5(min)$	$\sim 50.3(min)$
Dataset C	$5.5 \pm 0.6(min)$	$39.0 \pm 7.4(s)$	$8.8 \pm 1.2(min)$	$12.1 \pm 1.1(min)$	$\sim 27.1(min)$

- Step A: Image capture
- Step B: 2D shape approximation
- Step C: Camera system calibration
- Step D: 3D reconstruction

a justified evaluation of this experiment, we reported the results obtained with the same parameters. In future work, we can validate the parameter specifications in relation to image quality.

4.5.4 Evaluation on efficiency

For the evaluation of the computational efficiency of our approach, we first specify the setup. We represent the process of (1) image capture, (2) 2D shape approximation, (3) camera system calibration, and (4) the 3D reconstruction as Step A, B, C, and D, respectively. In Step A, we use the VAST BioImager to automatically obtain *Dataset A*. A well-trained biologist used the microscope setup (Leica DMRB) to acquire *Dataset B* & *C*. In this data acquisition, the object is manipulated by the VAST BioImager to present in the view of the microscope. The remaining steps are implemented with the Matlab platform (and partial C/C++ implementation) on a desktop equipped with a CPU i7 and 16G RAM. In Step B, we heuristically configure the filter width in spatial, colour feature space as $(h_s, h_r) = (10, 20)$ for *Dataset A*; and $(h_s, h_r) = (20, 30)$ for *Dataset B* & *C*. In Step C, we find that the Nelder-Mead simplex method works much faster for *Dataset A* & *C*. The evolution strategy achieves robust performance for *Dataset B*, though it is not efficient. We take a compromise on the optimisers and we apply the Nelder-Mead simplex method for *Dataset A* & *C* and the evolution strategy for *Dataset B*. In phase II of our 2-3DLA, we set the parameters $\mu = 10$, $\upsilon = 1$, and $\Delta t = 0.5$ for all cases.

In Table 4.5, for each step from A to D, we report the performance evaluation in terms of runtime in minutes or seconds for the three datasets. We compute the average and standard deviation of the runtime for each specimen in the datasets. First, from the results, we may conclude that the performance of our approach

on *Dataset A* is most efficient. In contrast, the larger size of the images in the *Dataset B* and *C* results in an expensive computation for our approach. Second, Step B performs very efficiently in the three datasets due to our fast implementation of the MS algorithm. Third, Step C in *Dataset B* and *C* requires a large amount of computations. The reason is that during the camera system optimisation, massive evaluations for the cost function are executed. However, a better optimiser and a high-performance scheme such as parallelisation or GPU programming can be employed to improve the efficiency of this step. Fourth, Step D in the three datasets seems be computationally expensive. One reason is the iterative 3D surface evolution. We set up a restricted termination criterion that requires the change of the cost function to be close to zero. So, this results in more iterations. In addition, we set a small grid size for the voxel space. This leads to massive projections and demands a large amount of memory to fulfil a precision guarantee. Again, this can, to a certain extent, be solved by the introduction of high-performance computing strategies.

4.6 Chapter conclusions and future work

Imaging of partially transparent objects impedes the visualisation of a continuous surface which renders it difficult to estimate a depth map of the object. To handle this problem, we first present three typical datasets using our MM-HTAI architecture. We then have developed the 2-3DLA approach for 3D reconstruction and measurement in light microscopy axial-view imaging. It answers RQ 4: *How can we efficiently deal with the translucency and transparency of specimen in light microscopy and still obtain a good 3D shape description from the MM-HTAI architecture*?

More concrete, in phase I of our 2-3DLA approach, we estimate an improved 3D volumetric representation as a confidence map from a range of texture-augmented images by applying the MS algorithm on the original images. In phase II, we adapt the region based level set method to estimate the optimal enclosed surface for the object which balances the 3D shape integrity and the smoothness of appearance for the 3D model. Compared with the shape-based method, the proposed 2-3DLA approach obtains comparable 3D reconstructions and measurements, but it does not require accurate 2D segmentations of the original images. As segmentation of partial transparent objects can be difficult, this system provides a good solution

to this shortcoming. Unlike the shape-based 3D reconstruction method, the 2-3DLA does not need extra refinement to obtain a natural and smooth 3D surface. We have successfully applied our approach for zebrafish analysis in multi-modal light microscopy axial-view imaging including bright-field and fluorescence. The obtained results can be directly used for the model system evaluation in the fields of toxicology, infectious diseases and oncology.

Still, we admit that the efficiency of our approach needs improvement with respect to larger images. A possible solution can be a dynamic programming scheme for high-performance parallelisation. As is well known, the zebrafish model is intensively used in the field of life-sciences. Therefore, in our 2-3DLA we can learn a good probabilistic model for the zebrafish either in supervised or semi-supervised way [122]. Yet, our approach needs the support from 2D shape approximation in the process of camera system calibration. In future research, we thus aim for a method which is independent on image segmentation. In addition, other multi-view imaging modes are investigated by our method. However, the axial-view imaging is the most commonly used modality, so the 2-3DLA can be considered generic for other specimen observations.

Appendix A

The zebrafish used in *Dataset C* are from a transgenic line Tg(lfabp:dsRed; elaA:EGFP) with 2 fluorescent colours for liver and pancreas (2CLIP). The eggs were kept in 60 $\mu g/mL$ Instant Ocean Sea Salts (Sera Marin, Heinsberg Germany) in demineralised water and treated with 0.003% 1-phenyl-2-thiourea (PTU, Sigma-Aldrich, Zwijndrecht, The Netherlands) to prevent pigmentation.