Shape analysis for phenotype characterisation from high-throughput imaging

Guo, Y.; Guo Y.

Summary

In this thesis we have studied shape with a particular focus on the zebrafish model system. The *shape* is an essential appearance of the phenotype of a biological specimen and it can be used to read out a current state or response or to study gene expression. Therefore, accurate shape analysis requires a precise shape description of a model system such as the zebrafish. Moreover, a sufficiently large sampling size of the specimens is necessary to ensure a justified and unbiased shape analysis. The latter is, for instance, very important for high-throughput in compound screening. All in all, top performance in zebrafish analysis requires high-throughput imaging (HTI).

In order to deal with high-throughput imaging, we aim to design an elaborate and well-performing HTI architecture. For the essential operations we need computational approaches to obtain the 2D/3D shape representations that are precise and yet can be acquired fast. The quality of the obtained shape descriptions will be validated in a straightforward manner with scalar primitives, i.e., the volume and surface area of a 3D shape. These primitives serve as 3D measurements for a robust primary shape assessment in the phenotype characterisation. Using only shape description is not sufficient, e.g., for high-resolution imaging on tissue and cellular level, so texture should be considered to complement and enhance the shape analysis.

The work in this thesis is divided in 5 research chapters that each have their own research question. The overall problem we are addressing is:

> *To what extent can we develop a stable HTI architecture and produce a robust and accurate shape analysis for the phenotype characterisation from the HTI architecture?*

In Chapter 2, we focus on methods to obtain accurate 2D shape information from microscope images. For our particular case these images result from high-

throughput imaging. So we need to extract the object, i.e. the zebrafish, from these images and this must be done as precise as possible, as the shape from the 2D images is required to construct the 3D image. In this manner we can provide the basis for a fast and accurate 3D measurement. The method will be embedded in the high-throughput axial-view imaging (HTAI) architecture that we propose. So, a hybrid segmentation method is developed which integrates (a) the mean shift algorithm and (b) the improved level set method. This method enables us to achieve an accurate 2D shape description of the zebrafish larvae.

In Chapter 3 we elaborate the 3D reconstruction of shape from axial views of the object. We investigate the architecture for axial-view imaging and question if accurate 3D measurements can be obtained from the imaging architecture. The imaging architecture, the VAST-BioImager, is the basis for the development of a new shape-based 3D reconstruction method. We demonstrated that with this method we can obtain accurate 3D shape descriptions in an efficient manner. From the 3D models, we obtain the volume and surface area. This is applied in an experiment with a large collection of zebrafish larvae of different developmental stages. For three larval stages we have produced a statistical representation of shape from the 3D measurements of the zebrafish.

In Chapter 4, the same input as in Chapter 3 is considered. However, we specifically focus on some characteristics with the objects which complicate 3D reconstruction in a direct manner. In our specific case the objects, i.e. zebrafish larvae, are partially transparent and translucent. To that end we elaborate a probabilistic approach with probabilistic models from the image textures. In this manner we are less limited by the accuracy of the segmentation; this is especially true for some sub-optimal illumination conditions. This new approach for 3D reconstruction from axial views is referred to as the two-phase 3D reconstruction approach (2-3DLA). The evaluations demonstrate a good performance at the cost of a higher computation time.

Initially, the 3D reconstructions were built from bright-field microscopy images. In high-throughput imaging other modalities are equally important. In particular fluorescence microscopy, as it allows to specifically visualise parts of the object. Therefore, in Chapter 5, we further develop the imaging architecture to be able to obtain 3D reconstructions from different imaging modalities and fuse these modalities in one model. We demonstrate this with an application of modelling of

the zebrafish liver using fluorescence while the shape is reconstructed from bright-field microscopy. Our results demonstrate a multi-modal 3D reconstruction from the fusion of 3D models on the organism- and organ-level.

Besides pure shape analysis, in Chapter 6, we investigate the application of classification models (or regression models) with the help of image features in annotated datasets. We question if in this manner we can be able to validate the performance of the features in shape analysis. We use completely different material compared to the previous chapters. We use four annotated datasets including human faces, butterflies, orchids and woods. From the human faces we develop graphical model for the kinship recognition of a group of faces in images. For the butterfly, orchid and wood datasets we have adapted a convolutional neural networks (CNN) architecture, a.k.a. deep learning, for learning representative features and developing a classification for prediction of the taxonomy of the species in the datasets. For all datasets we have demonstrated very good results.

Finally, in Chapter 7 we enumerate the conclusions of the research presented in this thesis by summarising the answers of the research questions that we have introduced in Chapter 1. We then provide a balanced discussion on the proposed approaches. Finally, we offer recommendations for further research.