# Image Analysis and Platform Development for Automated Phenotyping in Cytomics

Yan, K.

# English Summary

Cytomics is the study to understand cellular or subcellular behavior. In order to do research in cytomics on a large scale, automation is needed. The goal of a cytomics experiment is to study the cell behavior from images captured with an automated microscope setup; for example measuring migration of a cancer cell. The goal is achieved by using an image analysis pipeline. The aim of image analysis is to extract numerical descriptors from image sequences using digital image processing. When applied on a large-scale to images from bio-experiments, this is often referred as a high-content screening (HCS) experiment. Compared to other image analysis problems, HCS experiments as employed in cytomics studies often produce image sets containing objects significantly varied in both intensity and shape. Generic image analysis solutions often overlook the between-objects variation, thereby producing false results. To that end, our research focuses on designing dedicated image analysis solutions to cope with the between-object variation in HCS images using a strategy through which the computation adapts to the variation in the data. In this thesis, dedicated solutions for two procedures, namely image segmentation and object tracking, in the image analysis of HCS experiment are illustrated:

(1) Image segmentation is the procedure to extract objects from an image, i.e. cells. Often it is difficult to tune the parameters of a generic segmentation solution to the variation between individual images.

In the research described in this thesis, a self-adaptive segmentation algorithm, namely the Watershed Masked Clustering (WMC) algorithm has been developed. Compared to other algorithms, the WMC that we developed has demonstrated a robust performance in HCS studies.

(2) Object tracking is a procedure that will extract dynamic information from the objects, i.e. speed of migration. An object tracking algorithm builds linkages between objects from consecutive frames using information such as object shape and/or position. In cell biology studies, such an object can be a cell or a subcellular structure. Objects that are in proximity over frames and/or of similar shape will be related. From the trajectory, measurements such as a velocity or motion pattern can be extracted.

In the research described in this thesis, two tracking algorithms, namely the kernel density estimation with mean shift algorithm and the energy driven linear modeling algorithm, are introduced. From our testing, these two algorithms produce accurate tracking results in both case studies illustrated in this thesis.

The following two case studies demonstrate the applicability of our dedicated image analysis solutions:

Case Study 1: Our image analysis pipeline was capable of extracting both motility and morphology measurements from growth factor treated cancer cell. These measurements can be used to distinguish the changes in cell behavior induced by different growth factors. These measurements can further contribute to an objective understanding of drug-induced responses in terms of cell motility and morphology.

Case Study 2: In this case study, our image analysis pipeline demonstrated the possibility to extract measurements for subcellular structures in a study on the dynamics of matrix adhesions. The analysis has numerically confirmed that the turnover of matrix adhesions, in terms of assembly and disassembly duration, is strongly associated with the regulation of cell motility.

With a good data management system available, the image analysis can be performed in an efficient manner. The data management system shields end-users from the underlying complexity of the computational approaches in the image analysis by providing an integrated high-end graphic user interface (GUI). The automation underlying the GUI will enable to scale to a higher volume of image data. It further increases the accessibility of interdisciplinary data by standardizing different data formats and measurement structures between experiments. The standardization of data formats is an important feature since the ultimate goal of cytomics is to be integrated with other –oimcs data in order to study the genotype-to-phenotype correlation.

In conclusion, the research described in this thesis aims to design efficient and robust HCS analysis solutions. By studying existing algorithms, several dedicated image analysis algorithms are designed to fit the unique image characteristics of HCS experiment. From the case studies, it shows that analysis pipelines using these dedicated algorithms can well provide a platform of extracting objective understanding of biological questions. Compared to manual analysis, an automated solution will objectivize HCS analysis. It further paves the way in the understanding of control mechanisms of cell behavior.