

# Zusammenfassung

Die Mikroskopie vollzog vor einigen Jahren den Schritt von einer eher qualitativen biologischen Technik zu einer quantitativen Technik. Grund hierfür sind technologische Entwicklungen im Bereich der Robotik, bei Fluoreszenzfarbstoffen und in der digitalen Bildgebung. Die großen Datenmengen, die durch neuartige Mikroskope produziert werden, lassen nicht mehr die eigentliche Bildgebung, sondern die Analyse der Daten zu einem technischen Engpass werden.

In meiner Doktorarbeit präsentiere ich vier Methoden für die Analyse biologischer Bilder. Zwei Methoden behandeln die Überführung von Bildern in quantitative Daten. Die erste Methode ist ein Algorithmus, der die Blätter der Pflanze *Cardamine Hirsuta* in phänotypisch relevante Untereinheiten aufteilt.

Die zweite Methode quantifiziert die Musterbildung von Trichomen auf Blättern der Pflanze *Arabidopsis Thaliana*. Dazu habe ich eine Software namens TrichEratops erstellt, die es dem Biologen ermöglicht, Trichompositionen auf dem Blatt festzulegen und Statistiken zur Musterbildung berechnen zu lassen. Um genaue Distanzen zwischen den Trichomen zu ermitteln, modelliert die Software die 3D Oberfläche des Blattes.

Mit den zwei anderen Methoden zeige ich, wie man Zellen in Zeitraffer-Mikroskopie in diskrete, phänotypische Klassen überführt. Konkret werden die Zellen nach einem Gen Knockdown in morphologisch ähnliche Klassen eingeteilt. Dafür verwende ich zum einen Hidden Markov Modelle in Kombination mit multivariaten Gaußverteilungen. Diese Klassen benutze ich, um in genetischen Perturbationsscreens nach knockdown-spezifischen Phänotypen zu suchen. Wird die Aktivität von Zellen über eine bestimmte Zeit verfolgt, so ist die Datenstruktur typischerweise in einer Baumstruktur angeordnet. Allerdings können HMMs nur sequentielle Daten prozessieren. Darum führe ich in meinem vierten Projekt ein neues Model namens treeHFM ein, welches eine Generalisierung von Hidden Markov Modellen auf in Baumstrukturen angeordneten Daten darstellt. Ich zeige, wie dieses neue Model dazu benutzt werden kann, nach Gen-Knockdowns zu suchen, die die zelluläre Heterogenität beeinflussen.

## Summary

Optical microscopy is a central technique of biological research. Recently, microscopy made the step from a qualitative to a quantitative biological research tool. The reasons for this is technological progress in the robotics for sample handling, in fluorescence probes and in digital image capturing. Nowadays, the technical bottleneck is not anymore the generation of the data but its processing and analysis.

Within the scope of my PHD thesis, I have developed four methods for the analysis of biological images. In the first two projects, I show how to convert biological images of plant leaves into quantitative measurements. These measurements are then used to compare the phenotypic variation of morphological patterns in different ecotypes. One method partitions leaves of *Cardamine Hirsuta* into phenotypically relevant units and generates morphological and spatial measurements from these units. The second method quantifies the patterning of hairs on the leaf surface of *Arabidopsis thaliana*. I created TrichEratops, a software that enables the biologist the specification of trichome positions on leaf surfaces and the calculation of patterning statistics. In order to obtain accurate distances between the trichomes, the software also models the 3D surface of the leaf.

In the third method, I present a statistical method for assigning biologically relevant phenotype classes to cells in time-lapse microscopy. I use hidden Markov models (HMMs) in combination with multivariate Gaussian distributions in order to cluster morphologically similar cells into phenotype classes. Afterwards, these classes can be used to search for knockdown-specific phenotypes in large genetic perturbation screens. Since cells can divide, the trajectories of tracked cells typically are tree-like structures. However, HMMs can only process sequential data. To overcome this limitation, I introduce a new model called treeHFM, which is a generalization of HMMs to tree structured data. I demonstrate how this

model can be used to screen for gene knockdowns that influence the variability of cell development.