



Doctoral Thesis

Unsupervised modeling of cell morphology dynamics for time-lapse microscopy and phenotypic profiling

Author(s):

Zhong, Qing

Publication Date:

2013

Permanent Link:

<https://doi.org/10.3929/ethz-a-007625013> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH NO. 20993

Unsupervised modeling of cell morphology dynamics for time-lapse microscopy and phenotypic profiling

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
QING ZHONG
MSc in Informatik, ETH ZURICH

born July 3 1979
citizen of Switzerland

accepted on the recommendation of

Prof. Dr. Matthias Peter, examiner
Dr. Daniel Gerlich, co-examiner
Dr. Wolfgang Huber, co-examiner

2013

Abstract

Automated fluorescence live-cell microscopy empowers high-throughput screening with ultra high-content assays. The analysis of such image datasets, typically several terabytes large, requires automated computational procedures for the identification of cell phenotypes. Conventional supervised learning methods have yielded accurate results in diverse biological assays for classification of cellular morphologies. However, their applications require domain knowledge about the phenotypes and user-annotated training data for different fluorescent markers and experimental conditions. This thesis introduces an unsupervised computational framework that uses hidden Markov models to predict cell morphology classes and model cellular dynamics, based on temporally constrained combinatorial clustering initialized Gaussian mixture models. The generality of the unsupervised framework is demonstrated by its application to a great variety of biological assays and perturbation experiments.

The thesis is organized as follows. First, a brief introduction is given to bioimage informatics, cell cycle, computational image analysis, and machine learning. Then, a concise method development of the unsupervised framework is presented. Next, the versatility of the framework is explored by applying a variety of methods to diverse fluorescent markers. Additionally, the unsupervised framework is compared with the state-of-the-art supervised method in terms of classification performance on cell morphology. Moreover, the framework's ability to derive spatiotemporal models of cellular phenotypes from experimentally perturbed RNAi screening data is demonstrated. Furthermore, the mathematical expression of the temporally constrained combinatorial clustering is formally defined and its useful properties are proved. Last, the thesis is completed by extensive discussions about the unsupervised framework's general impacts on basic biology research, its limitation, and future method developments.

In conclusion, I developed an unsupervised framework that requires no user annotation, achieves high classification accuracies, generates statistical models for cellular dynamics, and provides an objective computational strategy for phenotypic profiling in image-based systems biology.

Zusammenfassung

Automatisierte Fluoreszenz-Live-Mikroskopie ermöglicht High-Throughput Screening mit ultra High-Content Assays. Die Analyse solcher Bilddatensätze, typischerweise mehrere Terabyte groß, erfordert automatische Berechnungsverfahren zur Identifizierung von Zell-Phänotypen. Konventionell überwachte Lernmethoden haben akkurate Ergebnisse in vielfältigen biologischen Assays für die Klassifizierung der zellulären Morphologie ergeben. Allerdings erfordern ihre Anwendungen Domain-Wissen über die Phänotypen und mit Benutzeranmerkungen versehene Trainingsdaten für verschiedene fluoreszierende Marker und experimentelle Bedingungen. Diese Arbeit stellt ein unüberwachtes computational Framework dar, die Hidden-Markov-Modellen verwendet, um Zellmorphologie Klassen vorherzusagen und zellulären Dynamik zu modellieren, basierend auf temporally constrained combinatorial Clustering initialisierte Gauss-Mischung Modelle. Die Allgemeinheit des unüberwachten Rahmen wird durch ihre Anwendung auf einer Vielzahl von biologischen Assays und Störung Experimenten demonstriert.

Die Arbeit ist wie folgt gegliedert: Zunächst wird eine kurze Einführung in Bioimage Informatik, Zellzyklus, Computational Bildanalyse und Maschinelles Lernen gegeben. Infolge wird eine prägnante Methodenentwicklung des unüberwachten Frameworks präsentiert. Anschliessend wird die Vielseitigkeit des Frameworks durch eine Vielzahl von Verfahrensanwendungen zu verschiedenen fluoreszierenden Markern untersucht. Zusätzlich wird das unüberwachte Framework mit dem modernsten überwachten Verfahren in Bezug auf die Klassifizierungsleistung der Zellmorphologie verglichen. Ferner wird die Fähigkeit des Frameworks zur Ableitung der raumzeitlichen Modellen der zellulären Phänotypen von experimentell gestörten RNAi Screening-Daten nachgewiesen. Darüber hinaus wird der mathematische Ausdruck des temporally constrained combinatorial Clustering formal definiert und dessen nützliche Eigenschaften bewiesen. Die Arbeit wird durch umfangreiche Diskussionen über die allgemeinen Auswirkungen des unüberwachten Frameworks auf die Grundlagenforschung der Biologie, die Begrenzung dieses Frameworks und zukünftige methodische Entwicklungen abgeschlossen.

Zusammenfassend wird von mir ein unüberwachtes Framework entwickelt, das keine Benutzerannotation erfordert, hohe Klassifizierungsgenauigkeit erreicht, statistische Modelle für zelluläre Dynamik erzeugt und eine objektive rechnerische Strategie für phänotypisches Profiling in der bildbasierten Systembiologie ermöglicht.