

Aus dem Institut für Neuro- und Bioinformatik  
der Universität zu Lübeck  
Direktor: Prof. Dr. T. Martinetz

Algorithmen zur automatisierten Analyse  
von *in vitro*-Stammzellpopulationen

Inauguraldissertation  
zur  
Erlangung der Doktorwürde  
der Universität zu Lübeck

Aus der Sektion Informatik/Technik

vorgelegt von  
Tim Becker, M.Sc.  
aus Wesel  
Lübeck, Juni 2013





1. Berichterstatter: Prof. A. Madany Mamlouk
2. Berichterstatter: Prof. H. Handels

Tag der mündlichen Prüfung: 29. November 2013

Zum Druck genehmigt. Lübeck, den 2. Dezember 2013



Für Elisa Sophie und Tammo Noah



## Zusammenfassung

In der vorgelegten Dissertation wurden neue Methoden zur Analyse von adhärent wachsenden, *in vitro*-kultivierten Zellpopulationen mittels Langzeitmikroskopie entwickelt. Diese interdisziplinäre Arbeit umfasst einen experimentellen Teil, in dem die Durchführung der Langzeitbeobachtungen etabliert und standardisiert wurde. Es wurden Bildserien verschiedener Zellpopulationen aufgenommen, die u.a. dafür genutzt wurden, um den ersten, frei zugänglichen Referenzdatensatz für Zelltracking-Algorithmen zu erstellen und zu veröffentlichen.

Im methodischen Teil der Dissertation lag der Schwerpunkt auf der Validierung und der Evaluation. Als Validierung wird die unabhängige Analyse bzw. Identifizierung von möglichen Tracking-Fehlern verstanden. Diese erlaubt es, die Genauigkeit der erzielten Ergebnisse deutlich zu verbessern und Fehler zweiter Art (falsch positive) zu vermeiden. Die Evaluation bezeichnet dagegen die systematische Analyse und Fehlerquantifizierung, mit der eine Zelldetektion oder ein Zelltracking-Algorithmus ausgeführt wird.

Weiterhin wurde das Anwendungsspektrum des entwickelten, validierten Zelltracking untersucht und sukzessive ausgebaut. Mit der funktionalen Fluoreszenzanalyse wurde eine neue Methode vorgestellt, um das nicht-invasive Zelltracking mit einer Immunfluoreszenzanalyse zu kombinieren. So kann die genealogische Information automatisiert in einem biologischen Kontext interpretiert werden. Diese Auswertung erlaubt es z.B., das Differenzierungspotential einer Zellpopulation auf Einzelzell-Ebene zu quantifizieren. Eine anschließende Auswertung liegt in dem weitgefassten Bereich der Systembiologie. Sie ermöglicht eine holistische Analyse, in der die Ergebnisse der verschiedenen Auswertungen in einem gemeinsamen, interdisziplinären Kontext interpretiert werden.



# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
<b>2. Grundlagen</b>	<b>11</b>
2.1. <i>in vitro</i> -Kultivierung adulter Stammzellen . . . . .	12
2.1.1. Kultivierung und Passagieren von adulten Stammzellen . . . . .	12
2.1.2. Immunfluoreszenzanalyse von Zellen . . . . .	13
2.2. Langzeitmikroskopie . . . . .	14
2.2.1. Langzeitbeobachtung mit dem Mikroskop Olympus ix81 . . . . .	14
2.2.2. Durchführung der Langzeitexperimente . . . . .	16
2.3. Bildserien von adulten Stammzellen und Zelllinien . . . . .	17
2.3.1. Bildserien pankreatischer Stammzellen (rPSC) . . . . .	18
2.3.2. Mosaikaufnahmen pankreatischer Stammzellen (rPSC) . . . . .	18
2.3.3. Die größte manuell rekonstruierte Zellfamilie . . . . .	23
2.3.4. Bildserien Humaner Zellen . . . . .	23
2.3.5. Bildserien HeLa-Zellen . . . . .	25
2.4. Das Zelltracking-Framework . . . . .	28
2.4.1. Automatisierung der Auswertung . . . . .	28
2.4.2. Zelldetektion . . . . .	30
2.4.3. Zelltracking . . . . .	34
<b>3. Validiertes Zelltracking</b>	<b>39</b>
3.1. Fehleranalyse und Validierung von Zelltracking-Daten . . . . .	39
3.1.1. Fehler und Fehlerfortpflanzung in zellulären Stammbäumen . . . . .	41
3.1.2. Das Genealogische Modell . . . . .	44
3.1.3. Validiertes Zelltracking als Integritätsprüfung . . . . .	45
3.2. CeTReS: Ein Benchmark für Zelltracking-Algorithmen . . . . .	47
3.2.1. <b>Cell Tracking Reference Set (CeTReS)</b> . . . . .	48
3.2.2. Einheitliche Fehlermaße für Zelltracking-Algorithmen . . . . .	51
3.3. Der Referenzdatensatz CeTReS.A . . . . .	57
3.3.1. Proliferation und Morphologie . . . . .	57
3.3.2. Migration . . . . .	59
3.3.3. Genealogien . . . . .	62
3.3.4. Zusammenfassung Zelltracking . . . . .	64
<b>4. Mitosedetektion</b>	<b>65</b>
4.1. Analyse von Detektions- und Klassifikationsproblemen . . . . .	66
4.1.1. Fehlermaße zur Beurteilung von Klassifikatoren . . . . .	66
4.1.2. Klassifikation mit einer Support Vector Machine (SVM) . . . . .	68
4.2. Mitosedetektion in den Schräglichtbildern . . . . .	68
4.2.1. Testdaten der Schräglichtbilder . . . . .	69
4.2.2. Morphologische Mitosedetektion . . . . .	71
4.2.3. Evaluation der morphologischen Mitosedetektionen . . . . .	72

4.2.4. Grenzen der morphologischen Klassifikatoren . . . . .	75
4.3. Mitosedetektion in Phasenkontrastbildern . . . . .	78
4.3.1. Referenzmitosen der Phasenkontrastbilder (CeTReS.B) . . . . .	78
4.3.2. Evaluation der Mitosedetektion auf den Phasenkontrastbildern . . . . .	80
4.4. Ergebnisse der Mitosedetektion und des Zelltracking . . . . .	85
4.4.1. Zusammenfassung der Mitosedetektion . . . . .	89
<b>5. Anwendungen</b>	<b>91</b>
5.1. Zelltracking zur Vitalitätsanalyse von Zellpopulationen . . . . .	92
5.1.1. Induktion und Nachweis des Zelltodes (Apoptose und Nekrose) . . . . .	92
5.1.2. Vergleich Fluoreszenzfärbung vs. Morphologieänderung . . . . .	95
5.2. Migrationsstudie mitotischer Zellen . . . . .	95
5.2.1. Verwendete Datensätze . . . . .	96
5.2.2. Gerichtete Migration mitotischer Zellen . . . . .	98
5.3. Oberflächenkontrastmikroskopie zur Zelldetektion . . . . .	101
5.3.1. Zelldetektion in SCM Bildern . . . . .	102
5.3.2. Auswertung und Vergleich der Zelldetektionen . . . . .	103
5.4. Fluoreszenzbilder zur Quantifizierung von Wundregeneration . . . . .	106
5.4.1. Quantitative Auswertung . . . . .	106
<b>6. Kombination von Genealogien und Fluoreszenzanalysen</b>	<b>109</b>
6.1. Geometrisches Hashing . . . . .	112
6.1.1. Geometrisches Hashing zur Positionssuche von Zellbildern . . . . .	112
6.1.2. Hashfunktionen für Drei-, Vier-, und Fünfecke . . . . .	116
6.1.3. Validieren der Transformationen . . . . .	120
6.1.4. Implementierung und Suchstrategie . . . . .	126
6.2. Das Petrischalen-Modell . . . . .	128
6.2.1. Das <i>in vitro</i> -Rauschmodell . . . . .	129
6.2.2. Eigenschaften der Hashfunktionen . . . . .	131
6.2.3. Eigenschaften des Bayes-Tests . . . . .	135
6.2.4. Vergleich der geometrischen Figuren . . . . .	135
6.3. Ergebnisse . . . . .	138
6.3.1. Suche in den Referenzbildern . . . . .	138
6.3.2. Die erste funktionale Immunfluoreszenzfärbung . . . . .	141
6.3.3. Zusammenfassung . . . . .	144
<b>7. Diskussion und Ausblick</b>	<b>149</b>
<b>Abbildungsverzeichnis</b>	<b>155</b>
<b>Tabellenverzeichnis</b>	<b>157</b>
<b>Abkürzungsverzeichnis</b>	<b>159</b>
<b>Eigene Veröffentlichungen</b>	<b>161</b>
<b>Literaturverzeichnis</b>	<b>163</b>

# 1

## Einleitung

Das Thema dieser Dissertation ist die Entwicklung neuer Methoden zur Analyse von adhären wachsenden, *in vitro*-kultivierten Zellpopulationen mittels Langzeitmikroskopie. Mit dieser Technik wird eine Zellpopulation kontinuierlich über Tage bis Wochen beobachtet, indem die Zellen in einem konstanten Zeitintervall abfotografiert werden. Es wurden von mir verschiedene neue Algorithmen, Methoden und Ansätze implementiert, um die erstellten Bildserien zu analysieren. Diese ermöglichen es, eine Zellpopulation mit auswertbaren, verlässlichen Daten und Parametern zu beschreiben. Die Grundlage liefert eine Zelldetektion in den einzelnen Bildern. Hierauf aufbauend können verschiedene Eigenschaften extrahiert werden und es ist z.B. möglich, den Weg einzelner Zellen über die Zeit zu rekonstruieren.

Eine Anwendung mit einem sehr großen wissenschaftlichen Interesse ist die Analyse (adulter) Stammzellen, die *in vitro*-kultiviert und vermehrt werden können. Diese Zellen haben ein großes Potential für eine klinische Anwendung in der regenerativen (Danner et al. [6]) oder der personalisierten Medizin (Hankowski et al. [44]) und sind Gegenstand vieler aktueller Forschungsprojekte.

### Adulte Stammzellen

In einem Organismus erfüllt jede Zelle eine spezifische, genau definierte Aufgabe. Durch die Funktion, die Stammzellen ausführen, heben sich diese Zellen deutlich von den anderen Zellen ab: Sie sind dafür verantwortlich, neue Zellen und Zelltypen zu generieren. Abhängig von dem betrachteten Kontext kann diese Aufgabe verschiedenen Prozessen zugeordnet werden: So sind Stammzellen für die Entwicklung des gesamten Organismus von der einzelnen Eizelle bis zum vollständigen Lebewesen (Ontogenese) verantwortlich. Im ungestörten Organismus sind sie maßgeblich an der Aufrechterhaltung und dem Funktionieren des Organismus selbst beteiligt (Homöostase). Die dritte, wichtige Aufgabe besteht in der Regeneration, in dessen Rahmen diese Zelle neues Gewebe, Blutgefäße oder Haut produzieren. Um diese Aufgabe zu erfüllen, müssen die Stammzellen funktionelle Zellen generieren, die z.B. die Aufgabe einer Hautzelle oder einer Nervenzelle übernehmen (Zell-Differenzierung). Gleichzeitig muss aber die Stammzellpopulation selbst, d.h. die Anzahl der Stammzellen, konstant bleiben.

Die meisten Modelle gehen davon aus, dass dieser Prozess durch eine **asymmetrische Zellteilung** ausgeführt wird. Hierbei teilt sich eine Stammzelle in zwei Tochterzellen auf, d.h. es werden zwei neue Zellen generiert (Mitose). Die asymmetrische Teilung bezeichnet den Fall, dass eine der Tochterzellen eine Stammzelle ist, während mit der zweiten Tochter eine ausdifferenziertere Zelle generiert wird (Clarke et al. [21], Krau-

se et al. [60]). Werden zwei Stammzellen generiert, wird die Teilung als symmetrisch beschrieben.

Im adulten Körper finden sich in fast allen Organen verschiedene Stammzellen, die für die Aufrechterhaltung der Körperfunktionen und Erneuerung von Gewebe benötigt werden (Zuba-Surma et al. [126]). Dieser Prozess der Homöostase (Fadell and Orrenius [35]) ist im Körper selbst, d.h. in der *in vivo*-Umgebung, streng reguliert und kontrolliert. Werden diese Zellen allerdings außerhalb des Körpers in einer *in vitro*-Umgebung kultiviert, zeigen sie einen anderen Charakter: Sie vermehren sich schnell, d.h. sie zeigen eine hohe Proliferationsrate, behalten aber die Fähigkeit der asymmetrischen Teilung bzw. Differenzierung (Gorjup et al. [38], Krause et al. [60], Petschnik et al. [89]).

Die Langzeitmikroskopie liefert einzigartige Möglichkeiten, das Verhalten von *in vitro*-kultivierten Stammzellen über mehrere Generationen hinweg zu beobachten. Eine entsprechende Bild- und Datenverarbeitung vorausgesetzt, ist es möglich, die Wege und die Zellteilungen der einzelnen Zellen zu extrahieren. Somit kann der gesamte Stammbaum von einer beobachteten Zellfamilie rekonstruiert werden.

### **Zelltracking**

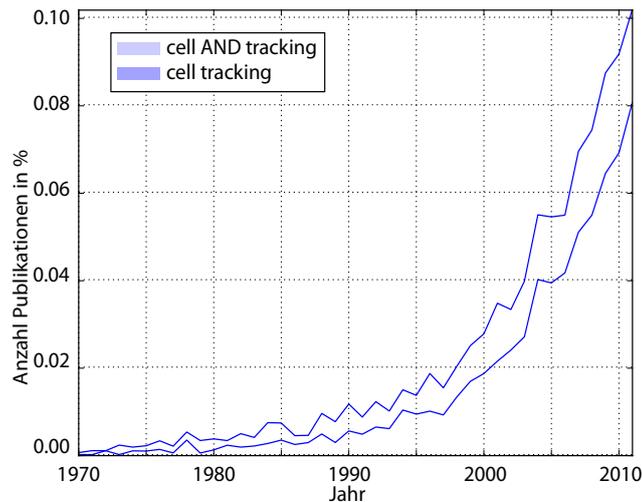
Mit dem Begriff des Zelltracking werden alle Techniken zusammengefasst, mit denen das Verhalten von einzelnen Zellen in den aufgenommenen Bildserien analysiert wird. So können z.B. die Wachstumskurven der Zellpopulationen, das Migrationsverhalten, die Zellteilungen und die Verwandtschaften der Zellen extrahiert werden.

Ein bekanntes Beispiel für eine Anwendung des Zelltracking ist in der Arbeit von Sulston et al. [111] zu finden, in der die Embryogenese und Entwicklung des Fadenwurms *C. Elegans* analysiert wurde. Mit einer Langzeitbeobachtung und Videoauswertung wurde der gesamte Stammbaum in mühsamer Kleinarbeit manuell rekonstruiert, der die Entwicklung einer einzelnen Eizelle bis zu dem ausgewachsenen Wurm zeigt. Diese aufwendige Arbeit war möglich, da die ausgewachsenen männlichen bzw. weiblichen Würmer aus exakt 1031 bzw. 959 Zellen bestehen. Durch die detaillierte Kenntnis des Ablaufs der Embryogenese hat sich *C. Elegans* als ein weit verbreiteter Modellorganismus etabliert. Für die Einführung von *C. Elegans* als Modellorganismus wurde im Jahr 2002 der Nobelpreis für Medizin vergeben.

### **Automatisierte Auswertung**

Mit der zunehmenden Automatisierung der Mikroskope und leistungsfähigeren Computern ist es machbar geworden, die erstellten Bildserien automatisiert auszuwerten (Meijering et al. [78, Kap. 9]). Zusammen mit der Möglichkeit, Zelllinien und (adulte) Stammzellen *in vitro* zu kultivieren, hat das zu einem gesteigerten Interesse für die Anwendung des Zelltracking geführt (vgl. 1.1).

Die automatisierten Zelltracking-Systeme folgen dabei einem Ansatz, der aus zwei Teilen besteht. Zuerst wird eine Bildsegmentierung genutzt, um die Zellen in den gegebenen Mikroskopaufnahmen zu identifizieren. Dieser Schritt wird als Zelldetektion bezeichnet und wird mit vielen unterschiedlichen Methoden durchgeführt, wie z.B. einer Level-Set- (Degerman et al. [28]), Wavelet- (Padfield et al. [85]), schwellwertbasierte-Segmentierung (Chen et al. [20]) oder aktive Konturen-Algorithmen (Pecreaux et al.



**Abbildung 1.1.:** Entwicklung der jährlichen Publikationen zu den Suchbegriffen *Cell Tracking* in der PubMd Datenbank. Die Anzahl ist als Prozent der Gesamtzahl an Publikationen angegeben und verdeutlicht das wachsende Interesse über die letzten Jahre. Die Daten wurden mit einem Python Programm erstellt, dessen Quellcode im Anhang unter A.1 zu finden ist.

[88], Yang et al. [122]). Das gemeinsame Ziel dieser Methoden ist es, die Form jeder Zelle zu extrahieren, d.h. es wird letztlich für jeden Pixel im Bild entschieden, ob er eine Zelle oder den Hintergrund zeigt.

Nach der Zelldetektion folgt der zweite Schritt, in dem der Weg der einzelnen Zellen über die Zeit rekonstruiert wird. Hierzu gibt es verschiedene Ansätze, die nach den naheliegenden Zellen im folgenden Bild suchen (Debeir et al. [27], Zhou et al. [125]) oder komplexere Ansätze, die eine Bildregistrierung (Hand et al. [43]), die Änderung der Zellkontur (aktive-Kontur, Degerman et al. [28], Li et al. [66], Li and Kanade [67], Li et al. [68, 69], Shen et al. [106]) oder eine Markov-Modellierung nutzen (Zhou et al. [125]). Die verschiedenen, oben genannten Methoden erlauben es, die Bewegung, Migration, das Teilungsverhalten und die Stammbäume der Zellen zu untersuchen (Eilken et al. [33], Rieger et al. [94]) oder das Schicksal (*single cell fate*), d.h. das Differenzierungsmuster, einer Zelle unter speziellen Bedingungen vorherzusagen (Cohen et al. [22]).

Ein weiteres Feld, in dem Zelltracking-Methoden eingesetzt werden, ist die automatisierte Charakterisierung und Überwachung von Stammzellpopulationen. Ein Beispiel liefert das Large Scale Digital Cell Analysis System (LSDCAS), das von Davis et al. [25] vorgestellt wurde und eine Zellpopulation kontinuierlich fotografiert bzw. überwacht. Dieses war eines der ersten Systeme, das den gesamten Aufbau für ein Zelltracking-Experiment bereitstellte, welches ein automatisiertes Langzeitmikroskop, eine automatisierte Datenverwaltung und Auswertung beinhaltet.

Diese Bildserien, die mit diesem System erstellt wurden, waren der Ausgangspunkt für weitere Studien, in denen neue Tracking-Algorithmen implementiert wurden, um das dynamische Verhalten der Zellen noch detaillierter zu untersuchen (Mosig et al. [81], Yang et al. [122]).

Die Schwierigkeit bei der Auswertung liegt hierbei nicht in der Analyse der Bilddaten allein, sondern beginnt bei dem experimentellen Aufbau und den verwendeten Mikroskopen. Zwar sind angepasste Langzeitmikroskope mit einer integrierten Klimakammer und einer automatisierten Steuerung verfügbar. Doch um eine verlässliche und automatisierte Auswertung durchführen zu können, muss eine gesamte Labor-Infrastruktur bereitgestellt werden, d.h. es werden Datenbanken und Programme für die Speicherung und Verarbeitung der Bildserien benötigt. Für die Auswertung der großen Datenmengen ist eine angepasste Bildanalyse und eine Tracking-Software erforderlich, deren Ergebnisse einheitlich abgespeichert und ausgewertet bzw. visualisiert werden müssen. Das führt dazu, dass selbst die technische Durchführung der Langzeitexperimente herausfordernd ist:

„The current lack of usable off-the-shelf hardware and software for such experiments makes their implementation technically challenging.“<sup>1</sup>

„It’s technically challenging. It requires a renaissance-type researcher to be able to span robotic imaging, database and data mining and stem cell biology. . .“<sup>2</sup>

Diese Zitate verdeutlichen, welche Anforderungen an ein gesamtes Zelltracking-System gestellt werden. Daher liegt ein Fokus der Forschungsarbeit in der Etablierung der Langzeitexperimente und der Standardisierung der Auswertung. Hierzu ist ein Zelltracking-Framework entstanden, mit dem die gesamte Verarbeitung der Bildserien vom Import der Rohdaten bis zu der Visualisierung der rekonstruierten Stammbäume realisiert ist. In dieser Software wurden verschiedene Verbesserungen und neue Methoden konzipiert und umgesetzt, die eine nahezu fehlerfreie Analyse der Zelltracking-Ergebnisse erlauben. Mit Hilfe dieser Auswertungen ist es möglich, eine beobachtete Zellkultur zu parametrisieren, d.h. verschiedene Eigenschaften der Zellen wie die Größe, Wachstumsrate oder Bewegung verlässlich und mit einer hohen Genauigkeit zu beschreiben. Weiterhin kann mit der Software erstmals die Verteilung der Zellzyklusdauer mit einer nicht-invasiven Methode sehr genau erfasst werden.

### **Das zentrale Konzept: Validierung und Evaluierung**

Trotz des breiten Anwendungsspektrums der Zelltracking-Algorithmen, die bereits in einer Vielzahl an Publikationen vorgestellt wurden, werden die interessantesten und komplexesten Auswertungen nach wie vor mit einem manuellen Zelltracking durchgeführt (vgl. Rekonstruktion des Stammbaums von *C. Elegans* in Kimble and Hirsh [58], Sulston and Horvitz [111]). Ein weiteres Beispiel liefert die Arbeit von Eilken et al. [33], in der die Entwicklung von Blutzellen untersucht wurden. Die Auswertung der Langzeitexperimente wurde auch hier manuell durchgeführt, indem jede Zelle in jedem Bild einzeln annotiert wurde.

---

<sup>1</sup>Long-term single-cell imaging of mammalian stem cells, Timm Schroeder, Nature Methods, März 2011

<sup>2</sup>Zitat von Eric Jervis, gefunden in [104], A closer look at the single cell, Megan Scudellari, Nature Reports Stem Cells, Mai 2009

---

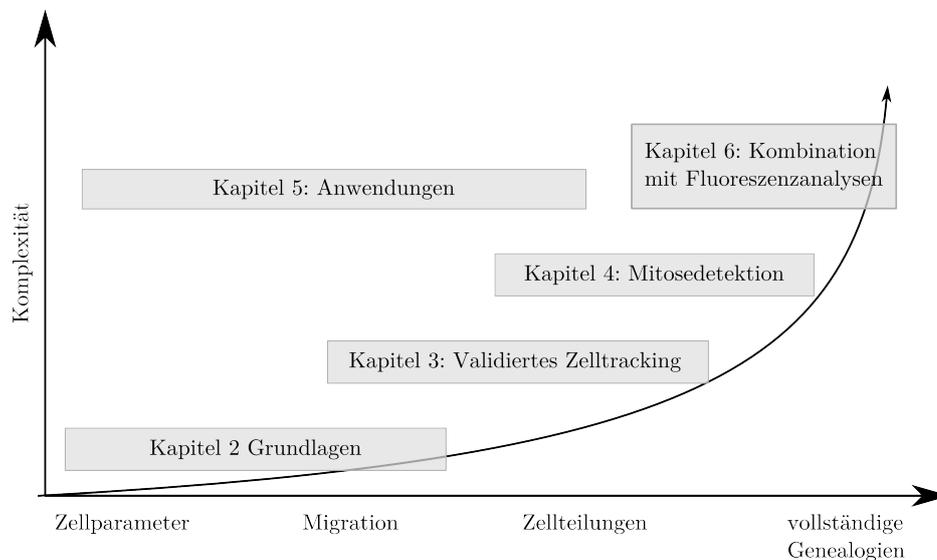
Der große Unterschied zwischen einer manuellen und einer automatisierten Auswertung liegt in der Konsistenz und der Verlässlichkeit der Tracking-Ergebnisse. Während ein „menschlicher Zelltracker“ automatisch die Korrektheit der Daten überprüft, gab es bisher keinen automatisierten Tracking-Ansatz, der die Ergebnisse validiert und somit verlässliche Ergebnisse produziert. Die Notwendigkeit dieser Validierung wird ersichtlich, sobald die starke Fehlerfortpflanzung in den Stammbäumen betrachtet wird: Ein kleiner, lokaler Fehler in den ersten Bildern wie z.B. eine übersehene Mitose oder Zelle kann dazu führen, dass der gesamte Stammbaum falsch rekonstruiert wird. So ist es zum Beispiel möglich, dass ein gesamter Ast des Stammbaums übersehen wird.

Aus diesem Grund wird in dieser Arbeit das Zelltracking um eine **Validierung** der Trackingergebnisse erweitert. Diese Validierung bezeichnet die automatisierte Identifizierung von fehlerhaften Ergebnissen, die anschließend aus der Auswertung entfernt werden. Als Basis für die Validierung wird ein genealogisches Modell genutzt, welches das Verhalten der Stammzellen wie z.B. eine Zellteilung beschreibt. Jede Abweichung von dem Modell kann direkt als Fehler entweder in der Zelldetektion oder in dem Trackingschritt interpretiert werden.

Ein wichtiger Bestandteil der Validierung ist die verlässliche Identifizierung der Zellteilungen, mit der die Mutter-Tochter-Beziehungen der Zellen und somit die Verwandtschaftsgrade festgelegt werden. Hierbei wird ein experimenteller Aufbau verwendet, bei dem jeweils zwei Aufnahmen einer Zellpopulation mit unterschiedlichen Kontrasteigenschaften erstellt werden. Die erste Aufnahme wird für die Zelldetektion und die zeitliche Rekonstruktion ihrer Wege genutzt, während der zweite Kontrast verwendet wird, um die Mitosen unabhängig von der Bildsegmentierung der Zellen zu detektieren. Das gesamte Konzept der Validierung, das die Mitosedetektion beinhaltet, ist unabhängig von dem Zelltracking und lässt sich auf andere Algorithmen übertragen, was in Abschnitt 4.4 gezeigt wird.

Ein weiteres zentrales Thema der vorgestellten Arbeit ist die systematische Evaluierung von Zelltracking-Algorithmen. Um diese durchzuführen, wird eine ausführliche Fehleranalyse gegeben und es wird der erste (validierte) Referenzdatensatz für Zelltracking-Algorithmen vorgestellt. In vielen Bereichen wie z.B. der Neuroinformatik, der Bildverarbeitung und des Maschinellen Lernens liefern solche Referenzdatensätze die Grundlage für die Evaluation und den Vergleich verschiedener Methoden (vgl. z.B. die Daten in Deng [30]). Im Bereich des Zelltracking gab es aber noch keinen korrigierten bzw. gelabelten Referenzdatensatz. Da weiterhin keine einheitlichen Fehlermaße genutzt wurden, war es bislang nicht einmal möglich, die verschiedenen Zelltracking-Ansätze zu vergleichen. Das hatte zur Folge, dass verschiedene Zelltracking Algorithmen zum Teil auf künstlich erstellten Bildern evaluiert (Degerman et al. [29], Li et al. [66]) oder das nur wenige Zellpfade bzw. Bilder genutzt wurden (Hand et al. [43], Yan et al. [121]).

Der im Rahmen dieser Dissertation erstellte Datensatz (*Cell Tracking Reference Set*, CeTReS) liefert somit erstmals die Grundlage für einen direkten Vergleich verschiedener Zelltracking-Ansätze und ermöglicht es, einen Zelltracking-Algorithmus auf systematische Fehler zu überprüfen.



**Abbildung 1.2.:** Übersicht der Kapitel. Die Themen der einzelnen Kapitel sind nach der Komplexität angeordnet. Diese gibt die Anforderung an, die an die entsprechende Aufgabe gestellt wird. Als Zellparameter werden alle Parameter zusammengefasst, die aus einer Zelldetektion extrahiert werden können und umfasst die Anzahl der Zellen, die Zelldichte, die durchschnittliche Fläche der Zellen etc.. Die komplexeste Aufgabe ist die Rekonstruktion der Stammbäume bzw. Genealogien. Hierfür ist es erforderlich, alle Zellen korrekt zu detektieren, ihren Weg über die Zeit zu rekonstruieren und korrekt miteinander zu verknüpfen. Für den letzten Punkt wird eine zuverlässige Mitosedetektion benötigt.

## Aufbau der Arbeit

Die vorgelegte Dissertation hat einen interdisziplinären Charakter und hat die Durchführung und Auswertung von Langzeitbeobachtungen adulter *in vitro*-Stammzellpopulationen im Fokus. Sie umfasst einen experimentellen Teil, in dem die Durchführung der Langzeitbeobachtungen etabliert und standardisiert werden. (vgl. 2.2). Es wurden Bildserien verschiedener Zellpopulationen aufgenommen, die die Grundlage für die Referenzdaten aus Kapitel 3, die Mitosedetektion aus Kapitel 4 und die Vitalitätsuntersuchung und Migrationsanalyse aus Kapitel 5 liefern.

Um die Bildserien und Daten automatisiert auszuwerten, wurde die gesamte benötigte Datenverarbeitung und Auswertung in dem Zelltracking-Framework implementiert, das in Abschnitt 2.4 vorgestellt wird. Es stellt die Schnittstelle zwischen dem experimentellen und dem technischen bzw. informatischen Teil der Arbeit dar. In den Kapiteln 3 und 4 wird das Framework um eine Validierung und eine Mitosedetektion erweitert, wobei verschiedene Techniken der Bildverarbeitung und des Maschinellen Lernens genutzt wurden, z.B. um die Zellen in den Mikroskopbildern zu detektieren oder das Muster einer Mitose zu lernen. Das gesamte Framework wurde in MATLAB implementiert und besteht aus ca. 32 Tausend Zeilen Quellcode, wobei Kommentare, Beispiele und Auswertungen mitgezählt wurden. Um das gesamte Projekt und die Software zu verwalten,

---

wurde das Revisionskontrollsystem git<sup>3</sup> eingesetzt, d.h. das gesamte Framework ist als Software-Repository verfügbar.

Das letzte Kapitel stellt einen weiteren Algorithmus vor, mit dem automatisiert Zelltracking-Daten mit einer fluoreszenzmikroskopischen Analyse kombiniert werden. Diese Methode basiert auf einem geometrischen Hashverfahren und wurde als eigenständiges Projekt bzw. Repository unabhängig von dem Zelltracking-Framework entwickelt. Die Hashverfahren und die darauf basierenden Suchalgorithmen wurden ebenfalls in MATLAB implementiert und umfassen ca. 23 Tausend Zeilen Quellcode.

Der Aufbau und Inhalt der Dissertation wird anhand der einzelnen Kapitel vorgestellt, die die Forschungsarbeit in chronologischer Reihenfolge widerspiegeln. Zusammenfassend ist in Abb. 1.2 eine Übersicht über die verschiedenen Themengebiete gegeben, die abhängig von der Komplexität angeordnet sind. Hier wird ersichtlich, welche Herausforderung gemeistert werden mussten, um eine gesamte Genealogie bzw. einen vollständigen Stammbaum zu extrahieren: Die Zelldetektion und das Zelltracking müssen fehlerfrei ausgeführt werden, um den Weg der einzelnen Zellen zu rekonstruieren. Diese Wege liefern die Bausteine, aus denen die Stammbäume zusammengesetzt werden. Um sie zu verknüpfen, müssen wiederum alle Verwandtschaften bekannt sein, d.h. dass alle Mitosen korrekt identifiziert werden müssen. Wenn nur eine dieser Voraussetzungen nicht oder nur zum Teil erfüllt ist, wird ein fehlerbehafteter Stammbaum erstellt.

Die folgende Auflistung liefert eine ausführliche Zusammenfassung der einzelnen Kapitel und der Publikationen, in denen jeweils Teile der Arbeit vorgestellt wurden. Hierbei sind nur die Veröffentlichungen aufgelistet, die in einer Fachzeitschrift (journal) oder in einem Tagungsband einer Konferenz (proceedings) erschienen sind.

- In dem **zweiten Kapitel** werden neue, standardisierte Langzeitexperimente etabliert, die mit einem automatisierten Langzeitmikroskop durchgeführt wurden. Dieses Mikroskop ist mit einem motorisierten Tisch und einem automatischen Kondensator ausgerüstet. Die Ausstattung erlaubt es, während eines laufenden Experimentes verschiedene Positionen einer Petrischale anzusteuern und die Zellen mit unterschiedlichen Kontrasteinstellungen zu fotografieren. Aufbauend auf dieser Grundlage wurde eine Standardmethode für die Durchführung eines Langzeitexperimentes etabliert, bei dem jede Zellpopulation mit zwei Kontrasteinstellungen fotografiert wird. Der erste Kontrast ist darauf optimiert, die Zellen zu detektieren, um ihren Weg über die Zeit zu rekonstruieren. Die zweite Kontrasteinstellung zeigt dagegen eine Zellteilung mit einem eindeutigen, optischen Muster, welches im vierten Kapitel für eine Mitosedetektion genutzt wird. Die Durchführung und der Aufbau dieser Experimente sind zusammen mit den erstellten Bildserien in den beiden Abschnitten 2.2 und 2.3 zu finden. Diese Daten liefern die experimentellen Grundlagen für die folgenden Kapitel 3-6.

Um diese Bildserien und Daten auszuwerten, ist das Zelltracking-Framework im informatischen Bereich der Arbeit entstanden. Dieses Framework, die grundlegende Zelldetektion und das Zelltracking wird mit den verwendeten Bezeichnungen und Definitionen in dem letzten Abschnitt 2.4 des zweiten Kapitels vorgestellt.

- Das zentrale Thema des **dritten Kapitels** ist die **Validierung** und **Evaluie-**

---

<sup>3</sup><http://git-scm.com/>

**rung** von Zelltracking-Algorithmen. Hierzu wird in Abschnitt 3.1 eine ausführliche Fehleranalyse gegeben und es wird das Konzept des validierten Zelltracking eingeführt. Von zentraler Bedeutung für die gesamte Arbeit sind die erstellten Referenzdaten CeTReS (*Cell Tracking Reference Set*), die zusammen mit einheitlichen Fehlermaßen zur Beurteilung von Zelltracking-Algorithmen in Abschnitt 3.2 und 3.3 vorgestellt werden. Die CeTReS-Daten sind die ersten, frei verfügbaren Referenzdaten, die für eine Evaluierung und den Vergleich von Zelltracking-Algorithmen genutzt werden können. Insgesamt wurden zwei Bildserien manuell korrigiert, wobei die vorgestellte Validierung genutzt wurde, um fehlerfreie Daten zu erstellen. Diese beiden Datensätze liefern die Grundlage für die Evaluationen der Methoden aus den folgenden Kapiteln 4 und 5. Die erstellten Referenzdaten, die Fehlermaße und das Konzept des validierten Zelltracking aus Kapitel 3 wurden in dem freien *open access* Journal PLOS ONE publiziert (Rapoport/Becker et al. [8]). Die Referenzdaten wurden ebenfalls unter der Creative Commons-Lizenz frei verfügbar veröffentlicht.

- In dem **vierten Kapitel** wird das Zelltracking-Framework um zwei **Mitosedetektionen** erweitert. Hierzu wird in Abschnitt 4.2 eine Methode vorgestellt, die eine Mitose anhand der Änderung der Zellmorphologie identifiziert, die kurz vor einer Zellteilung beobachtet wird. Da die Morphologie stark von der Zelldichte abhängt, d.h. von der Anzahl der Zellen, ändert sich das charakteristische Muster der Mitose im Laufe eines Langzeitexperimentes. Um diese Änderung auszugleichen, wurde eine adaptive Version der Mitosedetektion entwickelt, die sich an die verschiedenen Wachstumsphasen einer proliferierenden Zellpopulation anpasst. Diese Methoden wurden auf zwei Fachtagungen als Vortrag und Poster präsentiert (Becker and Madany Mamlouk [1], Becker et al. [4]) und in einer Sonderedition einer Fachzeitschrift veröffentlicht (Becker und Madany, [2]).

Die zweite Mitosedetektion aus Abschnitt 4.3 nutzt die Aufnahmen, die mit der zweiten Kontrasteinstellung aufgenommen wurden. Diese Bilder zeigen die Zellteilungen mit einem charakteristischen, optischen Muster, das durch den verwendeten Phasenkontrast erzeugt wird. Anhand dieses Musters können die Mitosen mit einer sehr hohen Genauigkeit identifiziert werden. So ist es möglich, die Zellteilungen unabhängig von dem Zelltracking durchzuführen, da die Detektion der Zellen in den Bilddaten der ersten Kontrasteinstellung erfolgt. Diese Mitosedetektion und die Kombination mit dem unabhängigen Zelltracking wurde auf einer weiteren IEEE-Konferenz als Vortrag vorgestellt (Becker et al. [3]).

Da die Mitosedetektion ein essentieller Bestandteil für die Rekonstruktion von zellulären Stammbäumen ist, wird mit ihrer Hilfe eine ausführliche Analyse der Zelltracking-Algorithmen in dem folgenden Abschnitt 4.4 gegeben. Hier werden die erstellten Referenzdaten CeTReS genutzt, um das Zelltracking zu evaluieren. Um zu zeigen, dass sich das Konzept der Validierung und die vorgestellte unabhängige Mitosedetektion auf andere Tracking-Ansätze übertragen lässt, wird ein weiterer Zelltracking-Algorithmus ausgewertet und mit der Validierung und Mitosedetektion kombiniert.

- Das Thema des **fünften Kapitels** sind verschiedene **Anwendungsszenarien**

---

der Methoden aus den Kapiteln 2-4. In dem ersten Abschnitt 5.1 wird ein Langzeitexperiment genutzt, um eine Vitalitätsanalyse mit Hilfe des Zelltracking-Frameworks an einer Krebszelllinie durchzuführen. Die zweite Studie aus Abschnitt 5.2 stellt eine Untersuchung zu dem Migrationsverhalten von adulten Stammzellen vor. Die Auswertungen zeigen, dass die Bewegungsrichtungen einer mitotischen Zelle mit der Migration der beiden entstehenden Tochterzellen korreliert. In dem folgenden Abschnitt 5.3 werden Zellbilder analysiert, die mit einer neuen, kontrasterzeugenden Technik, der Oberflächenkontrastmikroskopie, aufgenommen wurden. Das vorgestellte Projekt wurde in Kooperation mit Dr. Y. Nazirizadeh aus der Christian-Albrechts-Universität zu Kiel (CAU) durchgeführt. Die gemeinschaftliche Arbeit wurde in einer Fachzeitschrift veröffentlicht (Nazirizadeh et al. [7]). Abschließend wird in diesem Kapitel in Abschnitt 5.4 eine Studie vorgestellt, in der die Zelldetektion genutzt wurde, um den Einfluss von adulten Stammzellen in der Wundheilung zu quantifizieren. Diese Studie wurde in Zusammenarbeit mit Dr. S. Danner an der Fraunhofer EMB durchgeführt und ebenfalls in einer Fachzeitschrift publiziert (Danner et al. [6]).

- In dem **sechsten Kapitel** wird mit der **funktionalen Fluoreszenzanalyse** eine neue Technik bzw. Methode vorgestellt, mit der ein nicht-invasives Langzeitexperiment mit einer Fluoreszenzanalyse kombiniert wird. Diese Technik erlaubt es, eine Analyse der Proteinexpression mit den Informationen des Zelltracking zu kombinieren und liefert die einzigartige Möglichkeit, sowohl den Verwandtschaftsgrad der beobachteten Zellen als auch ihren Differenzierungszustand zu extrahieren. Die Anwendung und Auswertung der funktionalen Fluoreszenzanalyse wird es daher erlauben, das Verhalten der Stammzellen auf einer neuen Ebene zu analysieren und zu verstehen.

In Abschnitt 6.1 wird hierzu ein Algorithmus vorgestellt, mit dem dieselben Zellen in der Bildserie und den Fluoreszenzaufnahmen gefunden werden. Die vorgestellte Methode wird als geometrisches Hashverfahren bezeichnet und hat ihren Ursprung in der Astronomie. Hier wird sie genutzt, um die Sterne in Aufnahmen des Nachthimmels zu identifizieren. Da sich die Anwendungsszenarien der Astronomie und der Petrischale stark unterscheiden, werden die zugrundeliegenden Modelle und Parameter in Abschnitt 6.2 an die *in vitro*-kultivierten Zellen angepasst. Abschließend wird in dem letzten Abschnitt 6.3 die erste Kombination von Zelltracking-Daten und einer Fluoreszenzanalyse gezeigt und es wird bestätigt, dass diese Methode zuverlässig angewendet werden kann. Diese Algorithmen wurden mit einer ersten Auswertung auf einer IEEE-Konferenz vorgestellt (Becker et al. [5]).

- In dem **letzten Kapitel** werden die gesamten Ergebnisse dieser Dissertation abschließend zusammengefasst und diskutiert. Es wird das gesamte validierte Zelltracking, das aus dem initialen Zelltracking aus Kapitel 2, der Validierung aus Kapitel 3 und der Mitosedetektion aus Kapitel 4 besteht, in einem einheitlichen Kontext betrachtet.

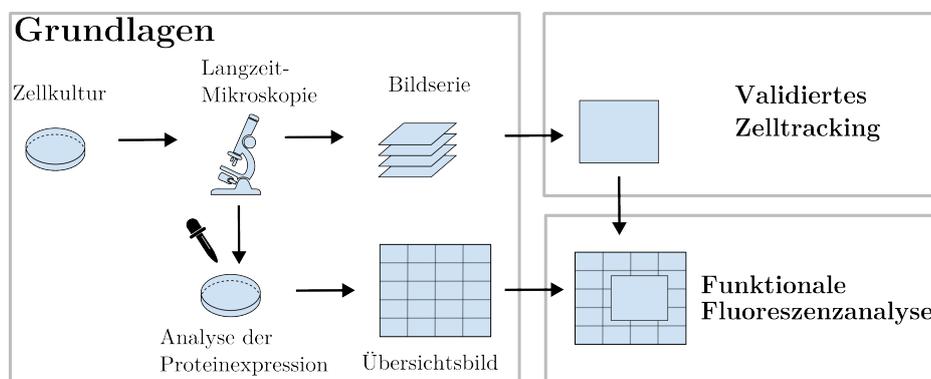


# 2

## Grundlagen

Im meiner Forschungsarbeit liegt die Entwicklung und Etablierung neuer Analysemethoden, mit denen *in vitro*-kultivierte (Stammzell-) Populationen in Langzeitexperimenten untersucht und ihr Verhalten quantifiziert werden kann. Das vorliegende Kapitel gibt einen Überblick über die benötigten biologischen Grundlagen und technischen Voraussetzungen, die für eine Langzeitbeobachtung von Stammzellpopulationen erfüllt sein müssen. Es werden die verwendeten Zelllinien und Stammzellpopulationen vorgestellt, die in verschiedenen Experimenten analysiert wurden. Anschließend werden die erstellten Bildserien beschrieben, die die Zellen über mehrere Tage zeigen sowie die verwendeten Mosaikbilder, die aus vielen Einzelbildern zusammengesetzt wurden und einen großen Bereich der beobachteten Zellpopulation zeigen. Ferner ist der Ablauf der zugrundeliegenden Experimente als schematische Übersicht in Abb. 2.1 gegeben.

In dem letzten Abschnitt des Kapitels wird das Zelltracking-Framework beschrieben, dass die Grundlage für die Zelldetektion und die Rekonstruktion der Zellpfade über die Zeit liefert. Es werden die verwendeten Definitionen, Bezeichnungen und ein grundlegendes Zelltracking eingeführt.



**Abbildung 2.1.:** Schematischer Ablauf der durchgeführten Langzeitexperimente. Die Grundlage für das Zelltracking liefern die Bildserien, die eine Zellpopulation über mehrere Tage zeigen. Nach Abschluss der Aufnahmesequenz konnte ein Übersichts- oder Mosaikbild der Zellpopulation erstellt werden, um die Tracking-Daten mit einer Fluoreszenzanalyse zu kombinieren.

## 2.1. *in vitro*-Kultivierung adulter Stammzellen

Die charakteristische Eigenschaft einer (adulten) Stammzelle ist die Fähigkeit, sich durch Zellteilungen zu vermehren (Proliferation) und durch eine asymmetrische Teilung differenziertere Zellen zu erzeugen, die ihren Stammzellcharakter verlieren und eine spezifische Aufgabe übernehmen. So bilden sie z.B. in der Wundregeneration neue Hautzellen (Differenzierung). In ihrer ursprünglichen *in vivo*-Umgebung, d.h. im Körper eines erwachsenen Lebewesens, liegt die Aufgabe dieser Zellen in der Erneuerung von Gewebe, was der Aufrechterhaltung der Körperfunktionen dient und als Homöostase bezeichnet wird.

Die große Hoffnung der Wissenschaft liegt darin, die Fähigkeiten der Proliferation und Differenzierung in den *in vitro*-kultivierten Zellpopulationen zu erhalten und z.B. in der Wundregeneration gezielt einsetzen zu können (vgl. Abschnitt 5.4). Um dies zu erreichen, ist es erforderlich, dass zwei wesentliche Punkte erfüllt sind:

- Die Zellen müssen in einer größeren Menge verfügbar sein, d.h. es werden Techniken benötigt, mit denen die Zellen reproduzierbar vermehrt und bei gleichbleibender Qualität kultiviert werden können.
- Der Status der Zellen muss vorhersagbar und kontrollierbar sein. Hierzu ist es erforderlich, die Prozesse und die Parameter, durch die die Zellen beeinflusst werden, zu untersuchen und zu verstehen.

Im Labor ist es möglich, die Zellen in einem Nährmedium außerhalb des Körpers zu analysieren und zu vermehren, was als *in vitro*-Zellkultur bezeichnet wird. Mit Hilfe dieser Kultivierung ist es möglich, die (adulten) Stammzellen in einer gleichbleibenden Umgebung zu züchten. Dies wird erreicht, indem sie in einer angepassten Umgebung gehalten werden, in der die *in vivo*-Bedingungen bestmöglich hergestellt werden.

In dieser kontrollierten Umgebung kann das Verhalten der Zellen detailliert untersucht und die Zellen können gezielt beeinflusst werden. Im Folgenden werden die verwendeten Kulturbedingungen und die verwendeten Analysemethoden zusammengefasst.

### 2.1.1. Kultivierung und Passagieren von adulten Stammzellen

Um die Zellen *in vitro* am Leben zu erhalten, werden sie in einer angepassten Umgebung unter speziellen Bedingungen gehalten: Hierzu werden sie in einem Brutschrank (CB 50, Binder, Deutschland) in einer 5% CO<sub>2</sub>-Atmosphäre bei 37°C und 70% Luftfeuchtigkeit kultiviert. Die betrachteten Zellen wachsen adhärent in den verwendeten Polystyrol-Kulturschalen (TPP, Schweiz), in denen sie mit CO<sub>2</sub> gepuffertem DMEM-Medium (Gibco, USA) mit einem Zusatz von 10% (v/v) fetalem Kälberserum (FKS) (FBS GOLD, PAA, Deutschland) und 1% (v/v) Penizillin / Streptomycin (PAA 100×, Penizillin 10.000 Einheiten/ml, Streptomycin 10 mg / ml) kultiviert werden.

Erreicht eine Zellpopulation eine kritische Zelldichte, verlangsamt sich zuerst die Proliferationsrate bevor sich die Zellen nicht mehr teilen, was als Kontaktinhibition bezeichnet wird. Die Zellen einer vollgewachsenen Kulturschale werden dann geerntet und mit einer geringeren Zelldichte wieder ausgesät. Hierzu wird das Zellmedium abgesaugt und die adhärennten Zellen mit Hilfe einer Trypsinlösung abgelöst.

Der gesamte Prozess vom Trypsinieren bis zur erneuten Aussaat wird als Passagieren bezeichnet und u.a. genutzt, um das Alter von *in vitro*-kultivierten Zellen zu bestimmen: Eine Zellpopulation in Passage sechs bedeutet, dass die Zellen bereits sechs mal passagiert wurden.

### 2.1.2. Immunfluoreszenzanalyse von Zellen

In der Zellbiologie ist die Immunfluoreszenzfärbung eine weit verbreitete Analysemethode, um intra- oder extrazelluläre Proteine und Strukturen einer Zelle sichtbar zu machen (Stephens [110, Kap. 5]). Bei diesem Verfahren werden spezifische Antikörper genutzt, die sich gezielt an das gesuchte Protein bzw. die gesuchte Struktur anheften und diese somit anfärben. Es ist möglich, die Antikörper mit einem Fluoreszenzfarbstoff zu markieren, was als Immunfluoreszenzanalyse bezeichnet wird. Hierbei beschreibt die Fluoreszenz die physikalische Eigenschaft von den Markern, spontan Licht zu emittieren, d.h. auszustrahlen, wenn es mit Licht einer anderen Wellenlänge angeregt wurde.

#### Zellkernfärbung

Eine Standardfärbung, die in der Zellanalyse durchgeführt wird um die Position von Zellen in Fluoreszenzaufnahmen sichtbar zu machen, ist die Zellkernfärbung mit dem Fluoreszenzfarbstoff 4,6-Diamidin-2-phenylindol (DAPI). Dieses Molekül bindet an die doppelsträngige DNA einer Zelle und interkaliert mit dieser, d.h. es setzt sich zwischen zwei benachbarten Basenpaaren fest. Dabei verschiebt sich das Absorptionsmaximum, so dass interkaliertes DAPI ein Licht mit einer anderen Wellenlänge aussendet als freies, nicht interkaliertes DAPI.

Um die interkalierten DAPI-Moleküle anzuregen, wird ein Licht mit einer Wellenlänge von  $358\text{nm}$  im ultravioletten Bereich gewählt, bei dem die Moleküle ein Emissionsmaximum von  $461\text{nm}$  besitzen, d.h. die Zellkerne leuchten blau. Aufgrund dieser spezifischen Wellenlänge kann der Farbstoff gut mit weiteren Fluoreszenzmarkern kombiniert werden.

#### Fluoreszenzmikroskopie

Das ZEISS Observer.Z1 Mikroskop wird für Aufnahmen von (Immun-) fluoreszenzgefärbten Zellen verwendet und ist mit einem Colibri-LED-Beleuchtungssystem ausgerüstet (ZEISS, Deutschland). Weiter ist es mit einem motorisierten Tisch und automatisiertem Kondensor- und Objektivrevolver ausgestattet, so dass verschiedene Vergrößerungen und Kontrasteinstellungen gewählt werden können. Das Mikroskop wird mit der Software AxioVision (Version 4.7.10, ZEISS, Deutschland) gesteuert.

Um große Bereiche zu fotografieren, wird das AxioVision Modul Mosaik genutzt. Dieses gestattet, mehrere teilüberlappende Einzelaufnahmen zu einem großen Gesamtbild zusammenzufügen. Hierzu werden die Positionen der einzelnen Bilder automatisch bestimmt und hintereinander angefahren und abfotografiert.

## 2.2. Langzeitmikroskopie

Das wichtigste Instrument für die Langzeitanalyse von *in vitro*-kultivierten Zellen ist ein Mikroskop mit einer integrierten Klimakammer, das über eine automatisierte Ansteuerung verfügt. In Abb. 2.2 ist das verwendete Mikroskop Olympus ix81 gezeigt, das in den nächsten Abschnitten genauer beschrieben wird. Weiterhin werden die verwendeten Aufnahme- und Kontrasteinstellungen erläutert.

### 2.2.1. Langzeitbeobachtung mit dem Mikroskop Olympus ix81

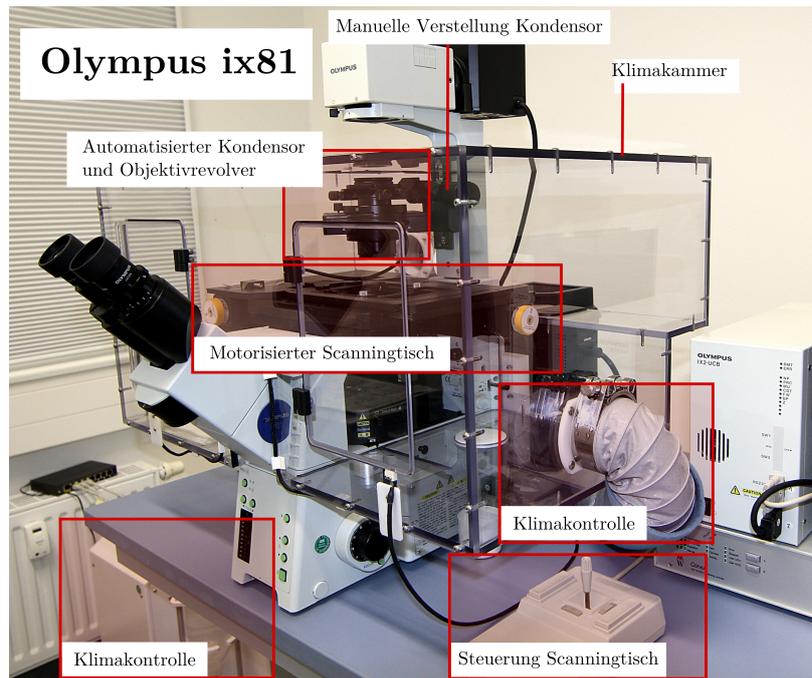
Die Langzeitexperimente werden mit einem Olympus ix81 Mikroskop durchgeführt, das mit einer integrierten Klimakammer ausgestattet ist (Evotec Air Conditioning Unit ACU04/B-0048, Deutschland, vgl. Abb. 2.2). Diese erlaubt es, die Temperatur, Luftfeuchtigkeit und den CO<sub>2</sub> Gehalt der Atmosphäre einzustellen, über einen Zeitraum von mehreren Tagen aufrechtzuerhalten und zu überwachen.

Um verschiedene Positionen anzufahren, wird ein motorisierter Tisch (Scanningtisch SCAN IM 120 × 80, Märzhäuser, Deutschland) verwendet. Eine CCD-Kamera (F-View FireWire Camera, Olympus Soft Imaging Solution) ermöglicht es, Ausschnitte der beobachteten Zellpopulation zu fotografieren. Die Aufnahmen werden mit einer Größe von 1376 × 1038 Pixel als 12-Bit Graustufenbild gespeichert, was bei einer vierfachen Vergrößerung einer Pixelauflösung von 1.6 μm entspricht. Zusätzlich konnten verschiedene Kontrasteinstellungen mit einem automatisierten Objektivrevolver und Kondensator gewählt werden. Für die Versuchsplanung und Durchführung wird die Software CELL<sup>M</sup> (Version 3.1, Olympus, Japan) genutzt, die auf einem Arbeitsplatzcomputer installiert ist, von dem das Mikroskop überwacht und gesteuert wird.

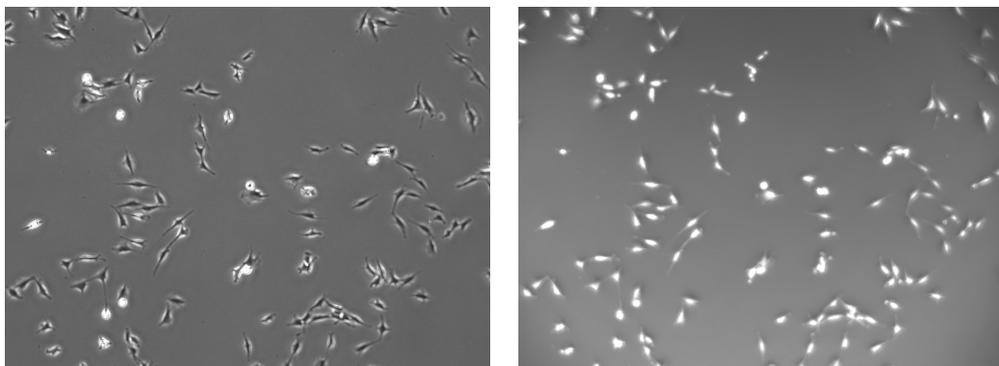
### Verwendete Kontrasteinstellungen

Für die Aufnahme der Zellen wird ein Objektiv mit einer vierfachen Vergrößerung genutzt (UPLFLN4XPH, Olympus, Japan). Es wird ein Phasenkontrastfilter mit der Bezeichnung PH1 für eine vierfache Vergrößerung und ein weiterer Phasenkontrastfilter mit der Bezeichnung PHL verwendet. Der PHL Filter ist für eine zehnfache Vergrößerung ausgelegt, wird jedoch mit dem genannten Vierfach-Objektiv verwendet.

Die Mikroskopbilder in Abb. 2.3 zeigen dieselbe Zellpopulation, die in einem kurzen Abstand mit den beiden Kontrasten fotografiert wird. Hier ist zu erkennen, dass die Aufnahmen des PHL Objektivs den typischen Charakter eines Phasenkontrastbildes zeigen: Die Zellen sind als dunkle Objekte zu erkennen, die von einem hellen Schein umgeben sind, der als Halo bezeichnet wird (Wayne [119]). Die Bilder des PH1-Objektivs ähneln dagegen Mikroskopaufnahmen, die mit einer Schräglichtbeleuchtung erzielt werden. Sie zeigen einen sehr guten Kontrast und die Zellen erscheinen als helle, fast weiße Objekte. Dafür verliert das Bild an Auflösung und es können nur wenige intrazelluläre Details erkannt werden. Um die verschiedenen Bildmodalitäten eindeutig zu unterscheiden, werden die zuletzt genannten Bilder als Schräglichtaufnahmen bezeichnet. Da der verwendete Schräglichtkontrast keiner Standardeinstellung des ix81 Mikroskops entspricht, ist eine genaue Beschreibung, wie dieser Kontrast erzielt werden kann, im Anhang in der Zusammenfassung auf Seite 175 gegeben.



**Abbildung 2.2.:** Langzeitmikroskop. Das verwendete Mikroskop Olympus ix81 verfügt über eine weitgehend automatisierte Ausstattung. Sämtliche benötigten Geräte wie der Scanningtisch, der Objektivrevolver, die Kamera und die Klimasteuerung werden von einem Computer gesteuert und überwacht.



(a) Phasenkontrastaufnahme (PHL).

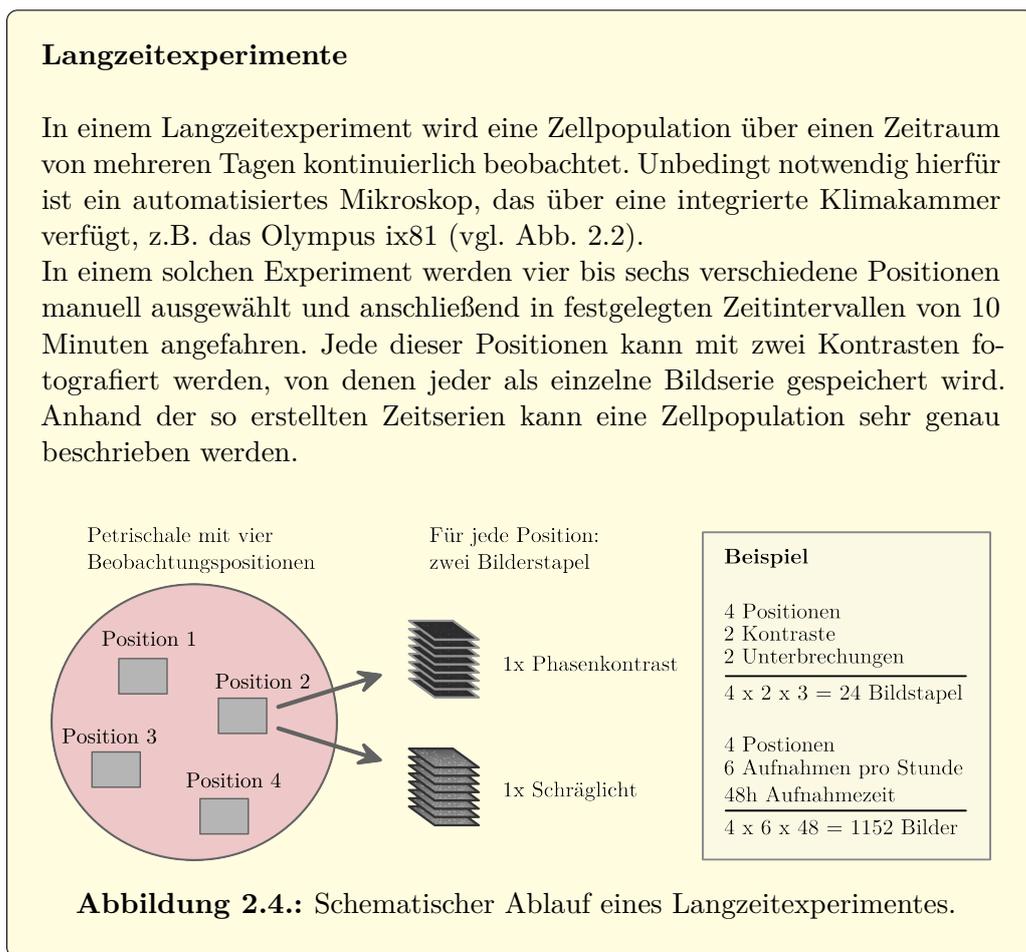
(b) Schräglichtaufnahme (PH1).

**Abbildung 2.3.:** Kontrasteinstellungen. Das verwendete Olympus ix81 Mikroskop ermöglicht es, die Zellen mit Hilfe von zwei unterschiedlichen Kontrasteinstellungen aufzunehmen. In (a) sind die Zellen als typisches Phasenkontrastbild gezeigt, während der zweite Kontrast (b) einer Schräglichtaufnahme entspricht und durch eine sorgfältige Anpassung der Einstellungen erzielt wurde. Die Aufnahmen sind durch einen sehr hohen Kontrast charakterisiert, der eine ideale Grundlage für eine Zelldetektion liefert.

### 2.2.2. Durchführung der Langzeitexperimente

Die Klimakontrolle des ix81 Mikroskops ermöglicht es, die Zellen über einen langen Zeitraum zu kultivieren, während sie kontinuierlich beobachtet werden. Der motorisierte Tisch erlaubt es, verschiedene Positionen anzufahren, die mit Hilfe des automatisierten Kondensors mit beiden Kontrasten abfotografiert werden können, die in Abb. 2.3 gezeigt sind. Prinzipiell ist die Anzahl der Positionen nur durch den zeitlichen Abstand zwischen den Aufnahmen limitiert. Es dauert durchschnittlich 30 Sekunden eine Position anzufahren, zu stoppen und die Bilder aufzunehmen. Bei einem Aufnahmeintervall von 10 Minuten ist es so maximal möglich, 20 Positionen anzufahren. Um aber die Bewegung des Tisches und Beleuchtungszeit zu minimieren, werden vier bis sechs Positionen gewählt. Der genaue Aufbau der Experimente ist in Abb. 2.4 gegeben.

Als Ausgabe wird für jede Position und jeden Kontrast eine Bildserie erstellt, die als Grundlage für weitere Auswertungen dienen. Die Bildserien werden auf einem NAS (Network Attached Storage) gespeichert, der über das Labornetzwerk erreicht werden kann, an dem das ix81 Mikroskop angeschlossen ist.



## 2.3. Bildserien von adulten Stammzellen und Zelllinien

Für die Etablierung der benötigten Methoden und die Durchführung der Experimente wird auf verschiedene Zellpopulationen und Zelllinien zurückgegriffen, die unterschiedliche Eigenschaften besitzen und unterschiedlich genutzt werden können. Im Folgenden wird ein Überblick über die verwendeten Zellen gegeben, um die Vielfalt der bereits untersuchten Zellpopulationen zu illustrieren und die Eckdaten der Experimente zusammenzufassen:

1. Zur Etablierung und Standardisierung der Langzeitexperimente wird auf eine adulte Stammzellpopulation zurückgegriffen, deren ursprüngliche Zellen aus einem Rattenpankreas extrahiert wurden (*rat pancreatic stem cell*, **rPSC**). Diese Zellen wurden aufgrund der folgenden Eigenschaften für die Etablierung und Standardisierung der Experimente genutzt: Sie lassen sich sehr gut kultivieren und sie vermehren sich sehr schnell, d.h. sie sind durch eine sehr hohe Proliferationsrate ausgezeichnet und stehen daher in großen Mengen zur Verfügung.
2. In Kapitel 5 wird die Auswirkung eines induzierten Zelltodes beschrieben und charakterisiert. Diese Versuche wurden mit der **HeLa**-Krebszelllinie durchgeführt, die 1951 als erste humane Zelllinie aus den Epithelzellen eines Zervixkarzinoms etabliert wurde (Scherer et al. [98]). Die Zellen gehören zu der am weitesten verbreiteten und verwendeten Zelllinie (Rahbari et al. [91]) und wurden bzw. werden häufig für Langzeitexperimente (Chen et al. [20], Hadjidemetriou et al. [42], Zhou et al. [125]) und Untersuchungen des Zelltodes genutzt (Shin et al. [107]). Daher liefern diese Zellen einen guten Ausgangspunkt, um die Ergebnisse des Zelltracking zu analysieren.
3. In weiteren Langzeitexperimenten wurden verschiedene humane, adulte Stammzellen untersucht, die aus der Haut oder aus der Speicheldrüse abgeleitet wurden. Viele dieser Zellpopulationen sind an der Fraunhofer EMB etabliert worden und werden direkt aus einer Primärpräparation isoliert.

Die wichtigsten Informationen und der Verwendungszweck sind für die verschiedenen Datensätze als Steckbrief zusammengefasst. Abschließend ist in Tab. 2.1 eine Übersicht der durchgeführten Experimente gegeben.

### 2.3.1. Bildserien pankreatischer Stammzellen (rPSC)

Die aus dem Pankreas eines adulten Nagers (*rat-tus norvegicus*) abgeleiteten Stammzellen (*rat pancreatic stem cell*, rPSC) wurden aufgrund ihrer guten Verfügbarkeit als Testzellen ausgewählt, um die Langzeitexperimente zu etablieren und die Auswertungen zu automatisieren bzw. zu validieren. Die rPSCs waren die ersten adulten Stammzellen, die an der Fraunhofer EMB etabliert wurden (Kruse et al. [62]) und zeichnen sich durch eine hohe Proliferationsrate und ein hohes Differenzierungspotential aus (Danner et al. [24], Gorjup et al. [38], Rapoport et al. [92]). Sie lassen sich über viele Passagen kultivieren (>150) und zeigen auch bei einer geringen Zelldichte bzw. Aussaatdichte eine starke Proliferation. Die daraus resultierende gute Verfügbarkeit der Zellen macht sie zu einem idealen Testobjekt für Langzeitexperimente. In Abb. 2.5 ist die typische Morphologie der Zellen zu erkennen. Die länglichen Zellkörper zeigen kurze Fortsätze, die als Podien bezeichnet werden und der Zelle zur Fortbewegung dienen.

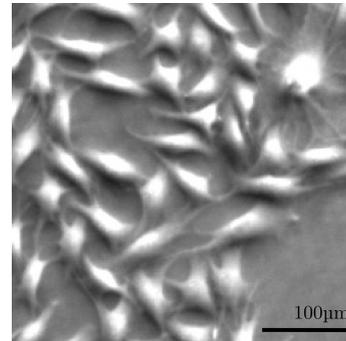


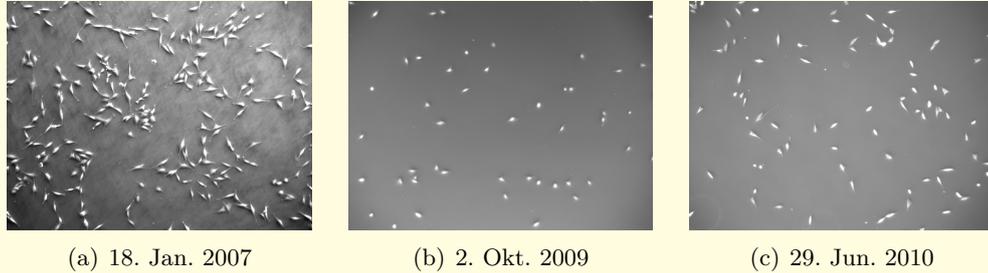
Abbildung 2.5.: rPSC.

Wie in Tab. 2.1 gezeigt, wurde eine Vielzahl an Versuchen mit diesen Zellen durchgeführt. In den ersten Langzeitexperimenten wurde nur eine einzelne Position mit einer Kontrasteinstellung aufgenommen, die späteren Versuche wurden dann erweitert und verschiedene Positionen jeweils mit unterschiedlichen Kontrasten aufgezeichnet (vgl. Abschnitt 2.2.2). Zwei dieser Bildserien wurden in der Publikation von Rapoport/Becker et al. [8] genutzt, um den Referenzdatensatz CeTReS zu erstellen. Diese Referenzdaten lieferten die Grundlage für weitere Publikationen Becker and Madany Mamlouk [1, 2], Becker et al. [3, 4, 5]. Eine Zusammenfassung der Experimente ist auf Seite 19 gegeben. Beispielbilder sind in Abb. 2.6 gezeigt.

### 2.3.2. Mosaikaufnahmen pankreatischer Stammzellen (rPSC)

Der hier beschriebene Datensatz wurde verwendet, um die in Kapitel 6 vorgestellte Methode, die automatisiert Phasenkontrastbilder und Immunfluoreszenzbilder kombiniert, zu evaluieren und zu optimieren. Hierzu wurde ein Datensatz benötigt, in dem dieselben Zellen sowohl im Phasenkontrast bzw. Schräglicht als auch im Fluoreszenzkanal mit DAPI-gefärbten Zellkernen sichtbar sind. Um das zu erreichen, wurde mit einem Skalpell ein Quadrat der Größe  $5\text{mm} \times 5\text{mm}$  an der Unterseite einer Petrischale eingritzelt (vgl. Abb. 2.7). Diese Markierung ist im Phasenkontrast- und Durchlichtmikroskop sichtbar, im Fluoreszenzkanal (DAPI) nicht. Daher wurde der Fluoreszenzkanal mit einem Durchlichtbild kombiniert (vgl. Abb. 2.7(d)), so dass die resultierenden Aufnahmen übereinander gelegt und jede Zelle ihren Zellkern zugeordnet werden konnten. In dieser präparierten Petrischale wurde eine Zellpopulation rPSC ausgesät und nach sechs Stunden der gesamte manuell markierte Bereich mit dem Langzeitmikroskop Olympus ix81 abfotografiert (vgl. Abb. 2.7(a-b)). Hierzu wurden 20 Positionen manuell ausgewählt, die den gesamten

**Steckbrief: Zeitserien rPSC**



**Abbildung 2.6.:** Erste Bilder der Zeitserien.

In vielen Langzeitexperimenten sind adulte Stammzellen aus dem Rattenpankreas (rPSC) genutzt worden. Die hier exemplarisch gezeigten Bildserien sind in der Arbeit von Rapoport/Becker et al. [8] verwendet worden.

Datum	18. Jan. 2007	2. Okt. 2009	29. Jun. 2010
Zellen	rPSC, P48	rPSC, P39	rPSC P31
Mikroskop	Olympus ix8	Olympus ix81	Olympus ix81
Kontraste	PH1	PH1,PHL	PH1,PHL
Vergrößerung	vierfach	vierfach	vierfach
Positionen	1	4	10
Zeitpunkte	209	399	428
Zeitintervall	15 Minuten	10 Minuten	10 Minuten
Aufnahmedauer	53 Stunden	67 Stunden	72 Stunden
Bilder gesamt	209	798	8560

**Besonderheiten:** Die Bildserien der Experimente vom 18. Januar 2007 und die Bilder der ersten Aufnahmeposition des Experimentes vom 2. Oktober 2009 sind verwendet worden, um den Benchmarkdatensatz CeTReS.A und CeTReS.B zu erstellen.

markierten Bereich abdeckten. Anschließend ist die Petrischale an allen Positionen im Phasenkontrast und im Schräglichtkontrast jeweils mit einer vierfachen Vergrößerung abfotografiert worden. Die gesamte Aufnahmezeit der 40 Bilder lag unter drei Minuten, in denen die Zellen noch lebten und sich bewegen konnten. Um diese Bilder zu den großen Übersichtsbildern aus Abb. 2.7(a-b) zusammzusetzen, wurde der *Image Composite Editor* (ICE Version 1.4.4.0, Microsoft Research, Redmond, USA) verwendet.

**Steckbrief: Mosaikaufnahmen**

Der Mosaikdatensatz zeigt eine sehr großen Bereich einer Zellpopulation von adulten Stammzellen aus dem Rattenpankreas (rPSC), der aus vielen Einzelbildern zusammengesetzt worden ist. Die gezeigten Bilder bestehen aus insgesamt 240 unterschiedlichen Aufnahmen derselben Zellpopulation, die mit verschiedenen Kontrastmethoden und Mikroskopen erstellt wurden.

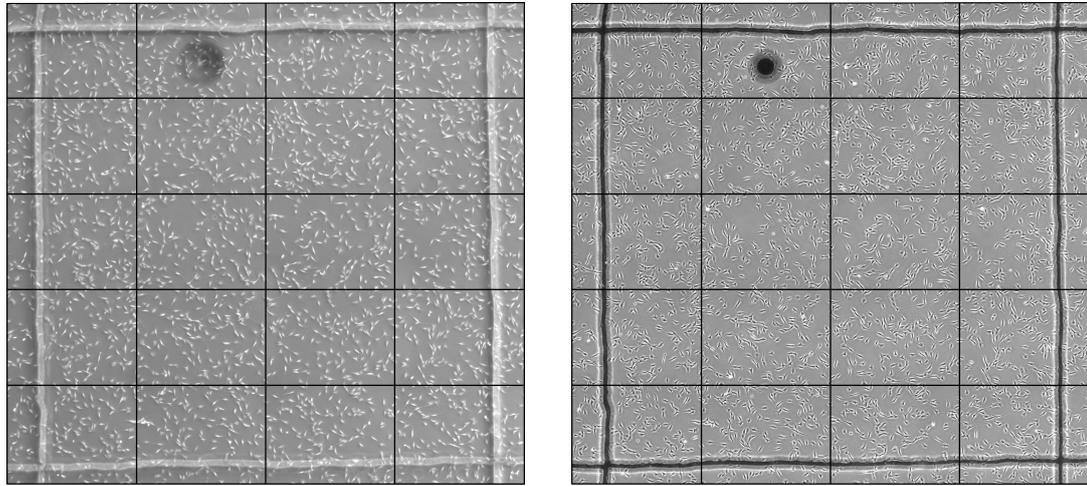
Datum	17. April 2012	17. April 2012
Zellen	rPSC, P25	rPSC, P25
Mikroskop	Olympus ix81	Zeiss Observer.Z1
Kontraste	PH1,PHL	DAPI Kanal, Durchlicht
Vergrößerung	vierfach	zehnfach
Positionen	$4 \times 5 = 20$ (Mosaik)	$8 \times 10 = 80$ (Mosaik)
Zeitpunkte	1	1
Bilder gesamt	80	160

**Besonderheiten:** Es wurde eine manuell markierte Petrischale verwendet. Die Mosaikaufnahmen in (a) und (b) zeigen jeweils den gesamten markierten Bereich der Größe  $5\text{mm} \times 5\text{mm}$  (vgl. Abb. 2.7).

Es war allerdings nicht möglich, die nötigen Mosaikbilder nur mit linearen Transformationen zu einem großen Übersichtsbild zusammenzufügen, d.h. es wurde eine Transformation für jedes Einzelbild berechnet, die aus einer Rotation, Skalierung, Translation und zusätzlich einer Scherung bestand.

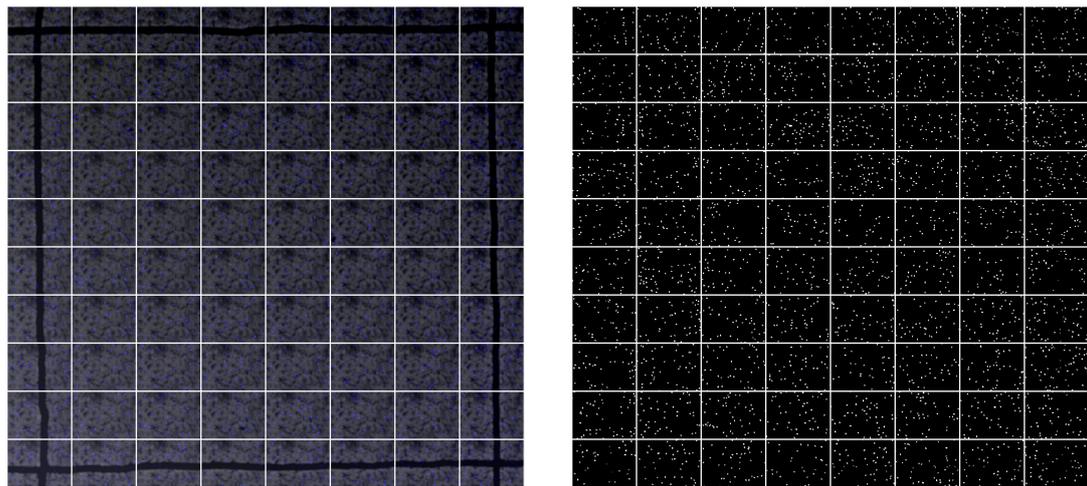
Die Algorithmen, die in Kapitel 6 vorgestellt werden, identifizieren die Zellen in einem Bild anhand der eindeutigen Zellkonstellationen der Zellen bzw. Zellkerne. Der wichtigste Parameter für diesen Algorithmus ist die exakte Verteilung der Differenzen zwischen den Zellen und den Zellkernen. Um diese zu bestimmen, wurde ein Mosaikbild benötigt, in dem die Zellbilder keine Verzerrungen enthalten, d.h. mit Transformationen die keine Scherung beschreiben, zusammengesetzt wurden. Um das zu erreichen, wurde zusätzlich mit dem GNU Image Manipulation Programm (GIMP, Version 2.6) ein manuell zusammengesetztes Bild erstellt (vgl. Abb. A.2 im Anhang).

Nach der letzten Aufnahme wurde die Zellschale aus dem Mikroskop entfernt, die Zellen mit Aceton fixiert und die Zellkerne mit DAPI gefärbt. Um die Markierung und die Zellkerne aufzunehmen, wurde das Zeiss Observer.Z1 Mikroskop genutzt. Mit Hilfe des Mosaik Moduls der AxioVision Software wurde eine Mosaikaufnahme des gesamten Bereiches erstellt. Die Aufnahmepositionen wurden automatisch ermittelt und mit einer Überschneidung von 20% aufgenommen. Alle Positionen wurden automatisch angefahren und es wurde jeweils ein Bild im DAPI Kanal sowie im Durchlichtkanal aufgenommen (vgl. Steckbrief auf Seite 20).



(a) Schräglichtaufnahme.

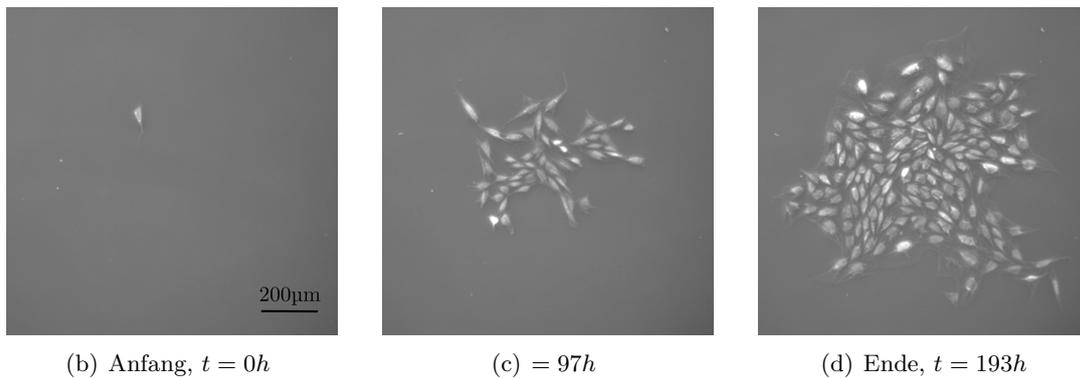
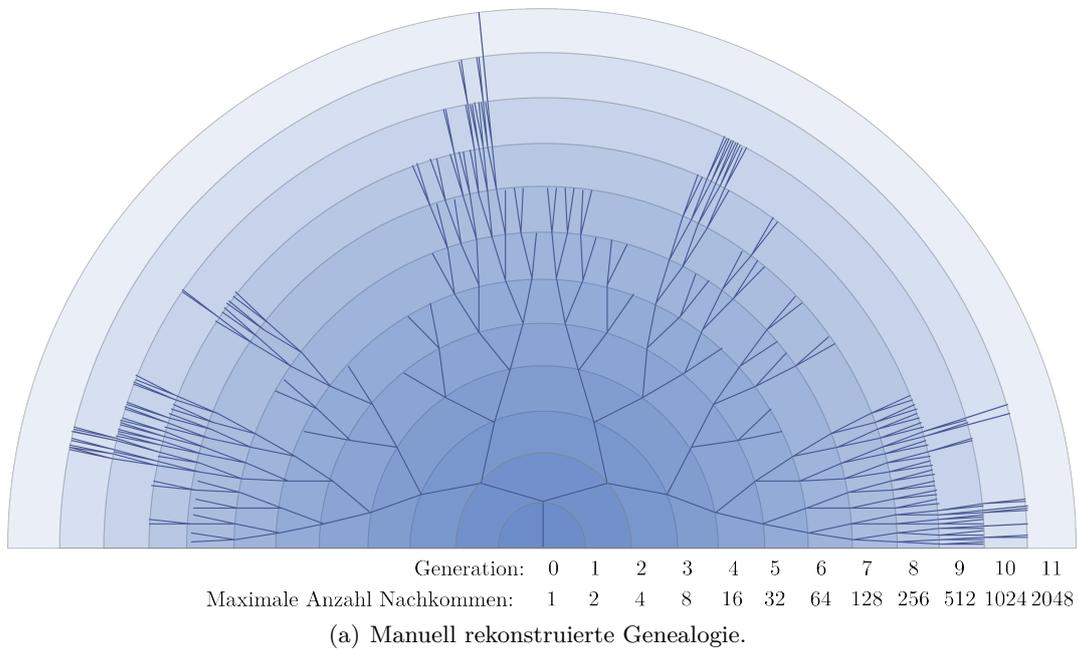
(b) Phasenkontrastaufnahme.



(c) Kombination Fluoreszenzkanal (DAPI) und Durchlicht.

(d) Fluoreszenzkanal (DAPI).

**Abbildung 2.7.:** Mosaikaufnahme. In einer markierten Petrischale wurde ein Datensatz erstellt, der eine rPSC-Population als Phasenkontrastbild und Fluoreszenzaufnahme zeigt. Die obere Reihe zeigt die zusammengesetzten Mosaikbilder, welche am Langzeitmikroskop Olympus ix81 aufgenommen wurden (vierfache Vergrößerung, 20 Bilder). Die unteren Aufnahmen zeigen dieselben Zellen als Mosaikbild im Fluoreszenzmikroskop (Zeiss Observer Z1) (zehnfache Vergrößerung, 80 Bilder). Um die Zellen und die Markierung sichtbar zu machen, wurde der Durchlichtkanal mit dem Fluoreszenzkanal aus (d) kombiniert. Das Mehrkanalbild ist in (c) gezeigt. Die Linien zeigen eine schematische Bildaufteilung, die richtigen Bildgrenzen sind nicht gezeigt.



**Abbildung 2.8.:** Größte rekonstruierte Genealogie. Das klonale Wachstum einer rPSC Population wurde über einen Zeitraum von acht Tagen beobachtet. Hierbei wurde die Genealogie einer einzelnen Zelle aufgezeichnet (b), die sich in dem Beobachtungszeitraum zu einer Population von 185 Zellen entwickelte (c-d). Anschließend wurde die entstandene Genealogie manuell rekonstruiert. Der resultierende Stammbaum in (a) zeigt die einzelnen Zellen geordnet nach ihren Generationen. Er verdeutlicht die asymmetrische Zellteilung, bei der sich einige Tochterzellen bis zu ihrer elften Generation teilen, während die Geschwister nur eine Teilung bis in die vierte Generation zeigen. Diese Asymmetrie wird bedingt durch die hohe Zelldichte in der Mitte der Wachstumsfläche (Kontaktinhibition).

### 2.3.3. Die größte manuell rekonstruierte Zellfamilie

In einem weiteren Experiment ist das klonale Wachstum von einer einzelnen rPSCs beobachtet worden (vgl. Abb. 2.8). Hierzu wurden die Zellen in einer sehr geringen Dichte ausgesät und über einen Zeitraum von acht Tagen beobachtet. In diesem Zeitraum wurde ein klonales Wachstum einer einzelnen Zelle beobachtet, die zu einer Zellfamilie mit einer Größe von 185 Zellen heranwuchs. Das Wachstum des Klons ist zu unterschiedlichen Zeitpunkten in Abb. 2.8(b-d) dokumentiert.

### 2.3.4. Bildserien Humaner Zellen

Weiterhin wurden adulte Stammzellen untersucht, die aus der humanen Haut gewonnen wurden. Diese humanen, dermalen Fibroblasten (*human dermal fibroblast*, HDF) werden aus der Dermis (Lederhaut) gewonnen und sind schnell wachsende Zellen (Takashima [113]). Die verwendeten Zellen sind größer als HeLa oder die Stammzellen aus dem Rattenpankreas (rPSC), zeigen aber eine Morphologie ähnlich der rPSC (vgl. Abb. 2.9). Die Proliferationsrate dieser Zellen ist geringer als die der rPSC und die Zellen werden seneszent, d.h. sie altern und lassen sich nicht über längere Zeiträume kultivieren (Hayflick and Moorhead [47]). Daher stehen sie nur begrenzt für Langzeitexperimente zu Verfügung. In Rapoport/ Becker et al. [8] wurde gezeigt, dass sich diese Zellen dennoch für die automatisierte Auswertung eignen. Hierzu wurde eine Population humaner dermalen Fibroblasten (HDF) in einem Langzeitexperiment über einen Zeitraum von zehn Tagen aufgenommen, in denen insgesamt 1411 Bilder erstellt wurden. Die Zellpopulation wurden an vier Positionen beobachtet, die für eine automatisierte Auswertung in Rapoport/Becker et al. [8] genutzt wurden<sup>1</sup>.

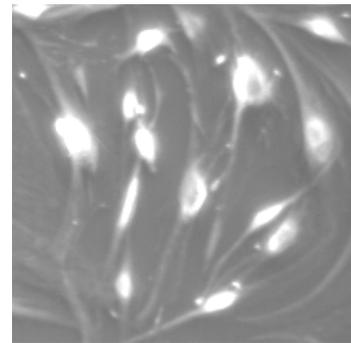


Abbildung 2.9.: HDF.

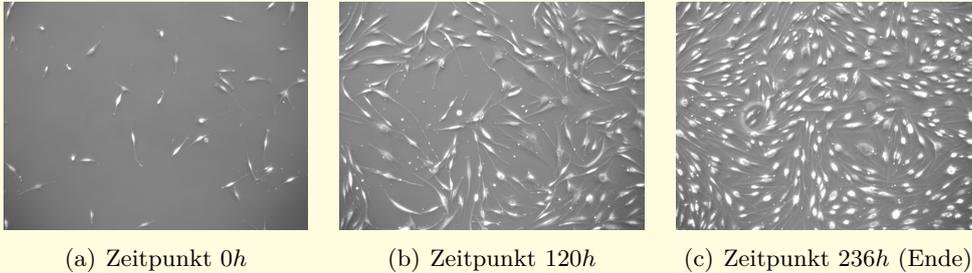
Eine der vier erstellten Bildserien wurde genutzt, um einen manuell korrigierten Datensatz zu erstellen. Dieser Datensatz enthält die korrigierten Zellmasken und Stammbäume und diente als Grundlage, um in Abschnitt 5.2 die Migration von mitotischen Zellen zu untersuchen.

### Humane, schweißdrüsenabgeleitete Stammzellen

Als weitere humane Zellen wurden adulte, schweißdrüsenabgeleitete Stammzellen verwendet. Diese Zellen wurden in der ersten funktionalen Fluoreszenzanalyse in Kapitel 6 verwendet. Zusätzlich wurden diese Zellen in einem Tierversuch verwendet, dessen Auswertung in Abschnitt 5.4 beschrieben ist.

---

<sup>1</sup>Siehe Abb. doi: 10.1371/journal.pone.0027315.s002 aus [8]

**Steckbrief: Bildserien HDF**

(a) Zeitpunkt 0h

(b) Zeitpunkt 120h

(c) Zeitpunkt 236h (Ende)

**Abbildung 2.10.:** Anfangsbilder der Zeitserien.

Mit den humanen dermalen Fibroblasten (HDF) sind Langzeitexperimente mit humanen Zellen durchgeführt worden. Da diese Zellen eine geringe Proliferation zeigen, wurden sie über einen Zeitraum von zehn Tage (236 Stunden) beobachtet. In Abb. 2.10 sind drei Bilder der beobachtete Population zu unterschiedlichen Zeitpunkten gezeigt.

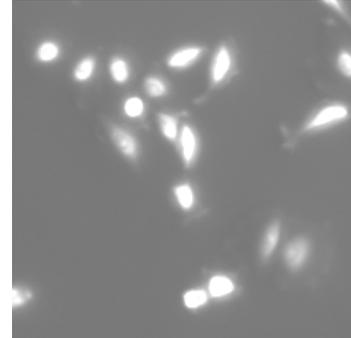
Datum	27. August 2010
Zellen	HDF, P7
Mikroskop	Olympus ix8
Kontraste	PH1,PHL
Vergrößerung	vierfach
Positionen	4
Zeitpunkte	1411
Zeitintervall	10 Minuten
Aufnahmedauer	236 Stunden ( $\approx$ 10 Tage)
Bilder gesamt	11528 Bilder

**Besonderheiten:** Die Bilder der vierten Aufnahmeposition wurden genutzt, um einen manuell korrigierten Datensatz zu erzeugen (korrigierte Zellkonturen und korrigierte Stammbäume). Um die Anzahl der Bilder und den Arbeitsaufwand zu reduzieren, wurde nur jedes dritte Bild betrachtet.

### 2.3.5. Bildserien HeLa-Zellen

Mit der HeLa-Krebszelllinie wurde eine weit verbreitete, humane Zelllinie verwendet (Rahbari et al. [91]). Diese Zellen wurden aus den Epithelzellen eines Zervixkarzinoms isoliert und sind die erste humane Zelllinie, die seit 1951 durchgehend kultiviert wurde bzw. wird (Scherer et al. [98]). In Kapitel fünf wird die HeLa-Zelllinie als Referenz für Experimente verwendet, in denen ein induzierter Zelltod untersucht wird.

Die Zellen zeigen eine sehr gute Proliferation und sind leicht zu kultivieren. Ihre Morphologie ähnelt stark den rPSC-Zellen (vgl. Abb. 2.5 und Abb. 2.11). Aufgrund der weiten Verbreitung und guten Verfügbarkeit wurden diese Zellen auch vielfach in Zelltracking-Experimenten genutzt (Chen et al. [20], Hadjidemetriou et al. [42], Zhou et al. [125]).



**Abbildung 2.11.:** HeLa.

Mit dieser HeLa-Krebszelllinie wurden verschiedene Langzeitexperimente durchgeführt, in denen die Charakteristik eines induzierten Zelltodes untersucht wurde (vgl. Abschnitt 5.1). Um eine sogenannte Apoptose einzuleiten, wurde Staurosporin (STS) verwendet, das als ATP-kompetitiver Inhibitor von Protein-Kinasen zu einem raschen Zelltod führt. ATP ist ein Nukleotid, das in den Zellen als universeller Energieträger genutzt wird und starke regulatorische Funktion in dem Zellstoffwechsel besitzt (Campbell et al. [17, S. 185 ff.]). STS bindet nun unspezifisch an die ATP-Bindungsstelle der Kinase und zeigt eine höhere Affinität als ATP selbst (Karaman et al. [53]), d.h. die Kinasen werden deaktiviert. Durch diesen Prozess wird die Aufnahme und der Umsatz des ATP vermindert bzw. verhindert, was schnell zu einem Zelltod führt. Das STS wurde ausgewählt, weil es vielfach für Apoptose-Experimente genutzt wurde (Chae et al. [18], Hasegawa et al. [46], Krohn et al. [61], Tamaoki and Nakano [114]).

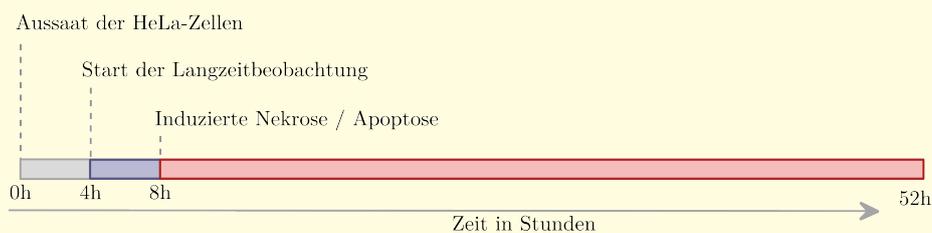
Weitere Versuche wurden mit Wasserstoffperoxid ( $H_2O_2$ ) durchgeführt, das bei hohen Konzentrationen die Zellmembran zerstört und einen Zelltod induziert, der als Nekrose bezeichnet wird. Der zeitliche Ablauf der Experimente ist in Abb. 2.12 dargestellt. In Kapitel 5.1 sind die folgenden Experimente ausgewertet worden (vgl. auch Abb. 5.2 auf Seite 94):

1. Für die Auswertung der Kontrollexperimente wurden die Bildserien vom 14. Dezember 2011, 23. und 25. Januar 2012 verwendet.
2. Die induzierte Apoptose wurde in den Bildserien der Experimente vom 21. Dezember 2011 und vom 3. und 19. Januar 2012 und 2012 beobachtet.
3. Die induzierte Nekrose wurde in den Bildserien der Experimente vom 30. Dezember 2012 und 9. und 27. Januar 2012 und analysiert.

**Steckbrief: Bildserien HeLa**

Mit Hilfe der HeLa-Zelllinie wurde der induzierte Zelltod in Langzeitaufnahmen charakterisiert. Hierzu wurde die Beobachtung nach vier Stunden unterbrochen, eine Apoptose oder Nekrose wurde induziert, und die Aufnahme fortgesetzt (vgl. Abb. 2.12).

Experiment	Kontrolle	Apoptose	Nekrose
Zeitraum	Dez. 2011 - Jan. 2012		Jan., Feb. 2012
Zellen	HeLa, P 55-64	HeLa, P 55-64	HeLa, P 55-64
Versuche	5	4	4
Mikroskop	Olympus ix8	Olympus ix81	Olympus ix81
Kontraste	PH1,PHL	PH1,PHL	PH1,PHL
Vergößerung	vierfach	vierfach	vierfach
Positionen	4	4	4
Zeitpunkte	288	288	288
Zeitintervall	10 Minuten	10 Minuten	10 Minuten
Aufnahmedauer	48 Stunden	48 Stunden	48 Stunden
Bilder gesamt	$4 \times 2 \times 288 = 2304$	$4 \times 2 \times 288 = 2304$	$4 \times 2 \times 288 = 2304$



**Abbildung 2.12.:** Die HeLa-Zellen wurden nach der Aussaat vier Stunden im Brutschrank inkubiert, bevor das Langzeitexperiment gestartet wurde. Nach weiteren vier Stunden wurde eine Nekrose (mit  $\text{H}_2\text{O}_2$ ) oder eine Apoptose (mit Staurosporin) induziert.

### 2.3. Bildserien von adulten Stammzellen und Zelllinien

Datum	Zellen	P.	Bilder	Dauer	Datum	Zellen	P.	Bilder	Dauer
18.01.2007	rPSC	1	209	52h	02.10.2009	rPSC	6	4797	66.5h
29.06.2010	rPSC	10	8284	71.3h	24.08.2010	CEder	4	2344	48.8h
27.08.2010	CEder	4	11288	235.2h	25.10.2010	CEder	4	10536	219.5h
16.12.2010	CEsd	4	12424	258.8h	16.12.2010	CEsd	4	12424	258.8h
27.08.2010	CEder	4	11288	235.2h	14.05.2010	rPSC	4	4800	100.0h
06.01.2011	CEder	4	7208	150.2h	13.01.2011	HeLa	4	11752	244.8h
01.02.2011	CEsd	4	6848	142.7h	07.02.2011	rPSC	4	4272	89.0h
01.03.2011	rPSC	5	8330	138.8h	18.11.2008	CEder	3	5518	75.2h
21.01.2011	CEder	4	4797	66.5h					
23.02.2011	HeLa	6	16476	228.8h	15.03.2011	rPSC	6	4968	69.0h
04.04.2011	rPSC	7	5348	63.7h	08.04.2011	rPSC	8	6176	64.3h
12.04.2011	rPSC	6	3276	45.5h	15.04.2011	CEsd	8	6448	67.2h
20.04.2011	CEsd	8	7120	74.2h	24.04.2011	CEsd	8	6720	70.0h
29.04.2011	CEsd	8	6336	66.0h	11.05.2011	CEsg	5	6660	111.0h
20.05.2011	CEsg	6	6528	90.7h	24.06.2011	CEsg	12	13968	97.0h
30.06.2011	CEsg	12	19824	137.7h	18.07.2011	CEsg	12	6744	46.8h
02.08.2011	CEsd	8	2256	23.5h	20.09.2011	rPSC	7	4494	73.0h
06.12.2011	HeLa	2	4892	10h	07.12.2011	HeLa	2	3376	7h
14.12.2011	HeLa	4	2304	48.0h	16.12.2011	HeLa	4	2304	48.0h
21.12.2011	HeLa	4	2304	48.0h	23.12.2011	HeLa	4	48	1.0h
03.01.2012	HeLa	4	2304	48.0h	09.01.2012	HeLa	4	2304	48.0h
19.01.2012	HeLa	4	2304	48.0h	23.01.2012	HeLa	4	2304	48.0h
25.01.2012	HeLa	4	2304	48.0h	27.01.2012	HeLa	4	2304	48.0h
30.01.2012	HeLa	4	2304	48.0h	03.02.2012	HeLa	4	2304	48.0h
01.08.2012	HeLa	3	1830	50.8h	07.08.2012	HeLa	3	1878	52.2h
14.08.2012	HeLa	3	1878	52.2h	14.08.2012	HeLa	1	624	52.0h
21.08.2012	HeLa	4	2504	52.2h	28.08.2012	HeLa	3	1878	52.2h
28.08.2012	HeLa	1	626	52.2h	04.09.2012	HeLa	4	2504	52.2h
12.09.2012	HeLa	4	2504	52.2h					

**Tabelle 2.1.:** Übersicht der Langzeitexperimente. Es wurden beobachtet: Die Krebszelllinie HeLa und verschiedene adulte Stammzellen, die aus Schweißdrüsen (CEsg), der Haut (CEsd,CEder) oder dem Rattenpankreas (rPSC) abgeleitet wurden. Der erste, kursiv dargestellte Datensatz vom 18.01.2007 wurde von Simone Schicktanz, M.Sc. aufgenommen. Alle weiteren, aufgelisteten Daten wurden von mir oder unter meiner Anleitung erstellt. Die Daten sind in insgesamt 30 Wochen Beobachtungszeit entstanden, in denen im Schnitt fünf Positionen gewählt wurden. Insgesamt ergibt dies einen Zeitraum von 2.7 Jahren.

## 2.4. Das Zelltracking-Framework

Um die aufgezeichneten Bildserien auszuwerten, wurde ein Zelltracking-Framework entwickelt, das es ermöglicht, die beschriebenen Experimente automatisiert zu analysieren und zu parametrisieren. In diesem Abschnitt werden die grundlegenden Konzepte beschrieben, die benötigt werden, um Zellen in den Bilddaten zu finden (Zelldetektion) und zu verfolgen (Zelltracking).

Das vorgestellte Framework wurde in Rapoport/ Becker et al. [8]) veröffentlicht und für die Mitose Detektion in Becker und Madany Mamlouk [2], Becker et al. [3, 4], für die Arbeiten zusammen mit Sandra Danner (Fraunhofer EMB, [6]) und Yousef Nazirizadeh (Uni Kiel, [7]) verwendet, die in dem späteren Kapitel 5 das breite Anwendungsspektrum der implementierten Methoden zeigen.

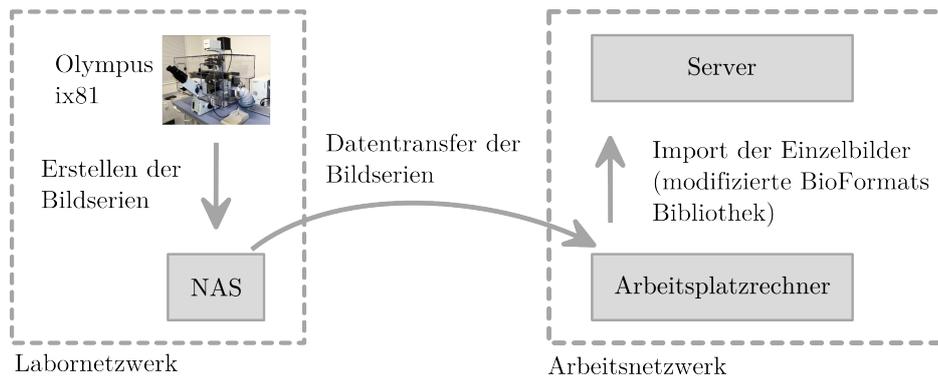
### 2.4.1. Automatisierung der Auswertung

Die Bilder der Zeitserien werden von der Olympus Software CELL<sup>M</sup> in einer Datenbank abgelegt. Für jede einzelne Position und Kontrasteinstellung werden die entsprechenden Bilder zusätzlich noch in einem Bildstapel zwischengespeichert. Um die Zeitserien möglichst effizient zu transferieren, werden die Bildstapel als Rohdaten kopiert und erst auf dem Arbeitsplatzrechner in Einzelbilder konvertiert (vgl. Abb. 2.13). Dies ist erforderlich, da der Export aus der Olympus Software CELL<sup>M</sup> heraus sehr lange dauert (>1Tag) und den gesamten Rechner bzw. das CELL<sup>M</sup> Programm blockiert. Um die Rohdaten einzulesen und in MATLAB verfügbar zu machen wird daher auf die Bio-Formats Java-Bibliothek zurückgegriffen (Linkert et al. [72]). Diese Bibliothek ermöglicht es, sowohl die verschiedensten Bildformate als auch die zugehörigen Metadaten auszulesen.

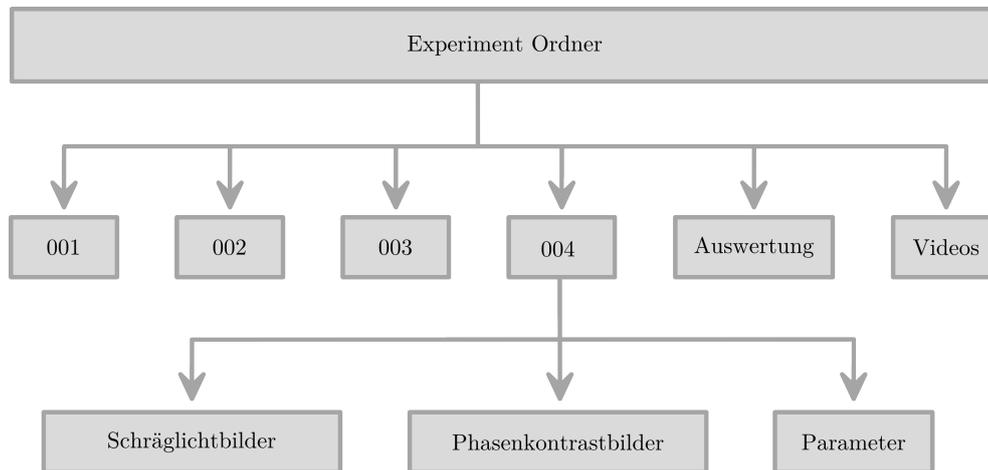
Da die ausgewerteten Dateien zu groß waren um direkt verarbeitet zu werden (> 10GB), wurden die MATLAB Routinen angepasst. Anstatt den gesamten Bildstapel auszulesen wird nur jeweils ein einzelnes Bild extrahiert, wodurch die geforderten Ressourcen an den Arbeitsspeicher erheblich gesenkt wurden.

Die Bilddaten wurden in der einheitlichen Ordnerstruktur abgelegt, die in Abb. 2.14 gezeigt ist. Hierzu werden die Bilder, die zu den einzelnen Positionen und Kontrasteinstellungen gehören, aus den Bildstapeln extrahiert und in 8-Bit Graustufenbilder umgewandelt. Zusätzlich werden Videosequenzen für alle Bildserien erstellt.

Bei dieser Verarbeitung der Rohdaten können für jede Position und für jeden Kontrast mehrere Bildstapel angegeben werden. Dies ist erforderlich, da die Langzeitexperimente zwischenzeitlich unterbrochen werden können, z.B. um Medium nachzufüllen. Ein Hilfsprogramm stellt eine Benutzeroberfläche zur Verfügung, mit der die Parameter für die Zelldetektion angepasst und gespeichert werden können. Für jede Aufnahme-Position kann ein einzelner Satz an Detektionsparametern gespeichert werden, was zum Teil erforderlich ist, da die unterschiedlichen Positionen starke Unterschiede in der Beleuchtung aufweisen können. Sind die Parameter abgespeichert, wird die Zelldetektion und das Zelltracking automatisiert ausgeführt, wobei verschiedene Ergebnisskurven und Auswertungen für jedes Experiment gespeichert werden. Eine detaillierte Auflistung dieser Beschreibung, die als parametrisierte Zellkultur bezeichnet wird, ist in Kap. 3.3 für einen der erstellten Referenzdatensätze gezeigt.



**Abbildung 2.13.:** Import der Bildserien. Die Bildserien werden auf einem Netzwerkspeicher abgelegt (NAS). Von dort werden die Bilddaten auf einen Arbeitsplatzrechner transferiert und in das Zelltracking-Framework importiert. Für den Import wird eine modifizierte Version der Bio-Formats-Bibliothek genutzt.



**Abbildung 2.14.:** Ordnerstruktur der Experimentdaten. Die Experimentdaten werden automatisiert in einer einheitlichen Daten- und Ordnerstruktur abgelegt. Jedes Experiment wird in einem separaten Ordner gespeichert, in dem für jede Aufnahmeposition ein einzelner Unterordner angelegt wird. Dieser enthält die Bilddaten und die zugehörigen Detektionsparameter. Exemplarisch ist ein (Standard-) Experiment gezeigt, in dem eine Zellpopulation an vier Positionen beobachtet wird. Für jede Position werden zwei Bilder aufgenommen: Ein Schräglichtbild und ein Phasenkontrastbild. In zwei weiteren Ordnern werden die Videodaten der Bildserien und die Auswertungen des Zelltracking gespeichert.

### 2.4.2. Zelldetektion

Durch die Optimierung der Mikroskopeinstellungen, die in Abschnitt 2.2.1 beschrieben sind, zeigen die verwendeten Schräglichtbilder der Zellen einen sehr guten Kontrast und werden daher für die Bildsegmentierung genutzt (vgl. Abb. 2.15(a)). Daher wird mit einer globalen Schwellwertsegmentierung ein sehr gutes Resultat erzielt (vgl. Abb. 3.9 auf Seite 53). Da der Fokus weiterhin auf dem Zelltracking und nicht auf der Entwicklung neuer Bildsegmentierungen liegt, wird die Zelldetektion im Folgenden mit der vorgestellten Methode durchgeführt. In dem modularen Framework selbst können leicht weitere Zelldetektionen genutzt werden.

#### Vorverarbeitung

Nachdem die Rohdaten in das Tracking-Framework importiert und in 8-Bit Graustufenbilder umgewandelt worden sind, werden die Bilder zunächst vorverarbeitet, um die inhomogene Beleuchtung der Mikroskopbilder zu korrigieren und den Kontrast zu erhöhen. Zuerst wird der Hintergrund der Bilder mit einem Gaußfilter geschätzt (Abb. 2.15(b)) und von den Mikroskopbildern subtrahiert (Abb. 2.15(c)). Anschließend werden die Bilder mit Hilfe einer Histogrammanpassung normalisiert (Abb. 2.15(d)).

#### Globale Schwellwertsegmentierung

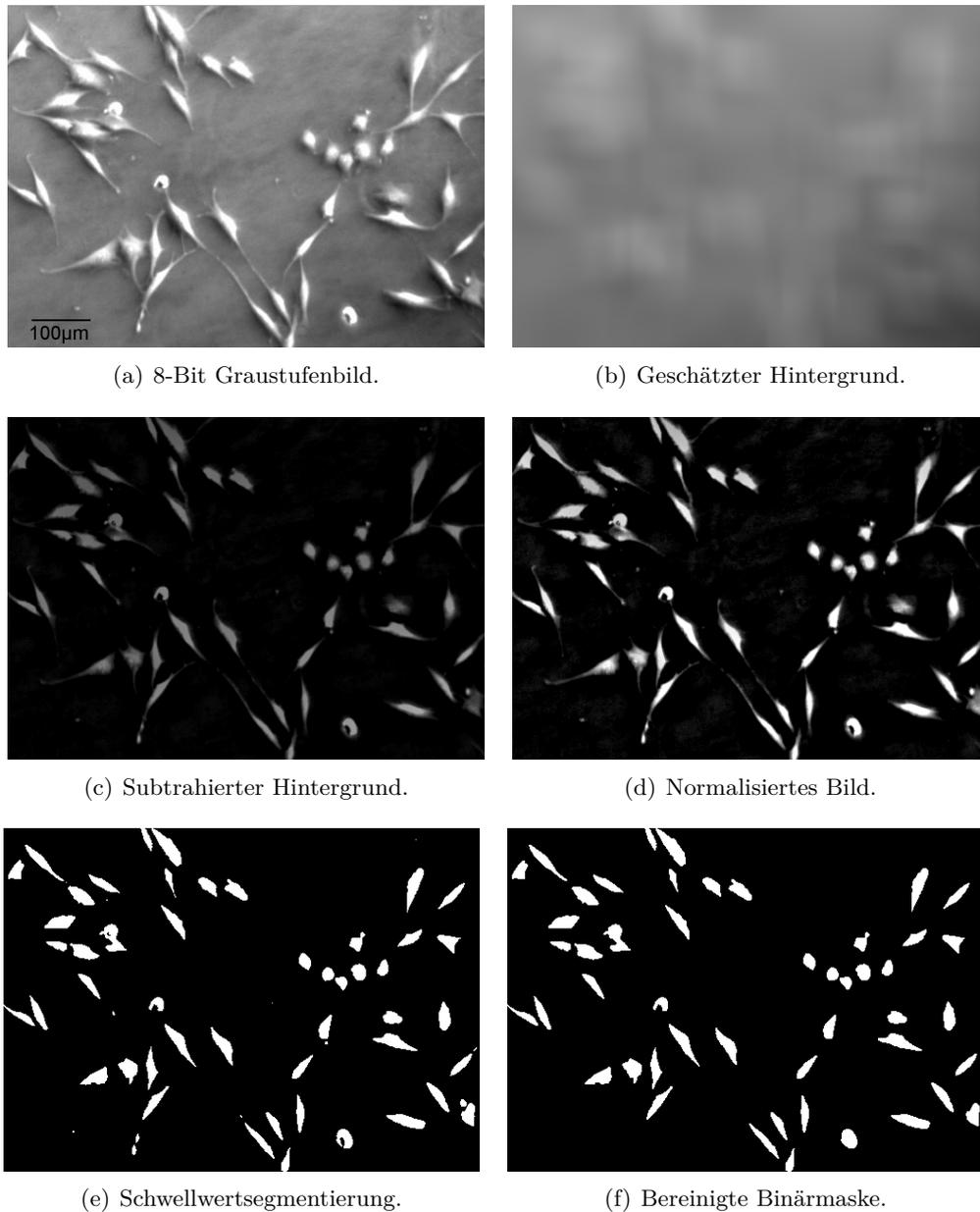
Um die Zellen in dem vorverarbeiteten Bild  $I$  zu detektieren, wird eine globale Schwellwertsegmentierung genutzt (Burger and Burge [15]). Mit Hilfe dieses Schwellwerts wird jeder Pixel  $x_i \in I$  mit

$$x_i \in \begin{cases} \text{Vordergrund,} & \text{falls } I(x_i) > \theta_{Otsu}, \\ \text{Hintergrund,} & \text{sonst,} \end{cases} \quad (2.1)$$

als Vorder- oder Hintergrund klassifiziert, wobei  $I(x_i)$  die Helligkeit (Intensität) von Pixel  $x_i$  angibt.  $\theta_{Otsu}$  ist der nach Otsu berechnete globale Schwellwert, eine genaue Erläuterung ist in der Infobox auf Seite 32 zu finden. Das Resultat dieser Operation ist ein Binärbild, siehe Abb. 2.15(e).

#### Korrektur der Binärmasken: Morphologische Operatoren

Anschließend werden die Binärdaten mit morphologischen Operatoren überarbeitet (Gonzalez and Woods [37]). Zuerst wird ein Erosionsoperator genutzt, um kleine Verbindungen zwischen den Zellkonturen zu entfernen und ein Dilationsoperator wird verwendet, um kleine Löcher und Lücken zu schließen. Als letztes werden alle Flächen gelöscht, die kleiner als eine gegebene Mindestgröße sind. So werden Staubpartikel und andere Artefakte entfernt (vgl. Abb. 2.15(f)). In dem so entstandenen Binärbild wird jede Zelle als ein Objekt von zusammenhängenden Pixeln (*connected regions*, Gonzalez and Woods [37]) dargestellt, die im Folgenden als Zellmaske bezeichnet wird. Alle Zellen in Bild  $t$  werden mit  $z_t^1, \dots, z_t^M$  bezeichnet, wobei  $M$  der Anzahl an Zellen in Bild  $I_t$  entspricht.

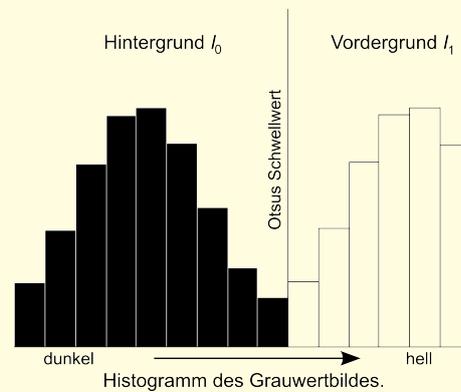


**Abbildung 2.15.:** Zelldetektion. Vor der Bildsegmentierung werden die Schräglichtbilder in einem Vorverarbeitungsschritt angeglichen. Die inhomogene Beleuchtung (a) wird korrigiert, indem der Hintergrund durch ein Gaussgefiltertes Bild geschätzt (b) und subtrahiert (c) wird. Anschließend wird das Bild mit einer Histogrammanpassung normalisiert (d). Die globale Schwellwertsegmentierung liefert ein Binärbild (e), das mit Hilfe von morphologischen Operatoren bereinigt wird. Zusätzlich werden alle Konturen, die kleiner als eine Mindestgröße sind, gelöscht (f).

### Globale Schwellwertberechnung nach dem Verfahren von Otsu

Das Verfahren von Otsu beschreibt eine robuste globale Schwellwertberechnung, die im Jahr 1979 von Nobuyuki Otsu vorgestellt wurde. Dieses kann genutzt werden, um Objekte in einem Bild automatisch zu segmentieren. Eine Voraussetzung ist, dass sich die Objekte, die detektiert werden sollen, in ihrer Helligkeit von dem Hintergrund abheben. Ist diese Bedingung erfüllt, kann ein Schwellwert berechnet werden, der beide Klassen optimal trennt.

Hierzu wird für das gegebene Graustufenbild  $I$  ein Schwellwert  $\theta_{otsu}$  bestimmt, der das Bild in zwei Klassen  $I_0$  und  $I_1$  aufteilt. Die Klasse  $I_0$  enthält alle Pixel mit einer Helligkeit kleiner  $\theta_{otsu}$ , Klasse  $I_1$  enthält die übrigen Pixel (vgl. Gl. (2.1)). Der Schwellwert wird so bestimmt, dass die Varianz innerhalb der Klassen minimiert ist; dieser Wert maximiert gleichzeitig die Varianz zwischen den Klassen. Um diesen optimalen Wert zu berechnen, wird die Varianz des Histogramms, d.h. die Verteilung der Grauwerte  $0, \dots, 255$  genutzt. Formal betrachtet, werden die Verteilungen  $p_i = \frac{n_i}{N}$ ,  $\sum_{i=1}^{255} p_i = 1$  der Grauwerte berechnet;  $N$  gibt hier die Gesamtzahl der Pixel an, 255 den höchsten Grauwert im Bild und  $n_i$  die Anzahl der Pixel mit dem Wert  $i$ , mit  $0 \leq i \leq 255$ . Diese Verteilung wird genutzt, um die Gewichte



**Abbildung 2.16.:** Anhand der Helligkeitsverteilung wird ein optimaler Schwellwert bestimmt.

$$\omega_0(\theta) = \sum_{i=0}^{\theta} p_i, \quad \text{und} \quad \omega_1(\theta) = \sum_{i=\theta+1}^{255} p_i$$

zu berechnen, wobei  $\omega_0 + \omega_1 = 1$  gilt. Anschließend kann die gewichtete Varianz zwischen den Klassen mit

$$\begin{aligned} \sigma_B^2(\theta) &= \omega_0(\theta)(\bar{I}_0 - \bar{I})^2 + \omega_1(\theta)(\bar{I}_1 - \bar{I})^2 \\ &= \omega_0(\theta)\omega_1(\theta)(\bar{I}_0 - \bar{I}_1)^2 \end{aligned}$$

bestimmt werden, wobei  $\bar{I}_0, \bar{I}_1$  und  $\bar{I}$  jeweils die Mittelwerte der Helligkeit des Bildes  $I$  und der Klassen  $I_0, I_1$  angeben. Der optimale Schwellwert kann als Maximum

$$\theta_{otsu} = \operatorname{argmax}_{\theta \in \{1, \dots, 255\}} \sigma_B^2(\theta)$$

angegeben werden. Eine ausführlichere Herleitung und ein entsprechender Beweis ist in [84] zu finden.

### Morphologische Charakterisierung der detektierten Zellen

Um die unterschiedlichen Formen und Eigenschaften einer Zelle bzw. ihrer Zellmaske zu beschreiben, wird der folgende Satz an Parametern analog zu Al-Kofahi et al. [9] berechnet. Zur Berechnung wird auf die Kontur der Zellmasken und die Zellbilder zurückgegriffen (vgl. Abb. 2.17):

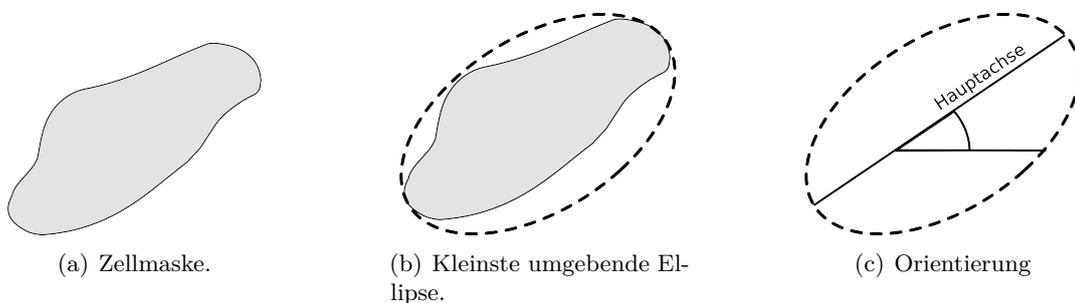
- Die Position einer Zelle wird als Schwerpunkt  $c = [f_x, f_y]$  der zugehörigen Zellmaske bestimmt.
- Die Größe  $f_A$  einer Zelle ist durch die Anzahl der Pixel der Zellmaske gegeben.
- Die Helligkeit wird als gemittelte Summe der Intensitäten der Pixel berechnet, die zu der Zellmaske gehören. Diese durchschnittliche Helligkeit wird als  $f_H$  gespeichert.
- Die Rundheit oder Kompaktheit  $f_K$  einer Zelle gibt die Ähnlichkeit der Zellkontur zu einem Kreis an. Sie wird als

$$f_K = \frac{U^2}{4\pi f_A}$$

berechnet, wobei  $f_A$  die Größe und  $U$  den Umfang der Zellkontur angibt. Hat eine Zelle eine Kompaktheit von 1, so hat sie eine kreisrunde Form.

- Um die Länge  $f_L$  einer Zelle zu bestimmen, wird die kleinste Ellipse berechnet, die die Zelle einschließt. Die längste Ausdehnung dieser Ellipse definiert anschließend die Länge  $f_L$  der Zelle (vgl. Abb. 2.17(b)).
- Um die Orientierung  $f_O$  zu berechnen, wird die kleinste Ellipse genutzt, die die Zelle einschließt. Anschließend wird  $f_O$  mit  $-180^\circ < f_O \leq 180^\circ$  als der Winkel angegeben, den die längste Gerade, die in die Ellipse passt, mit der x-Achse bildet (vgl. Abb. 2.17(c)).

Die genannten Eigenschaften werden für jede Zelle zum Zeitpunkt  $t$  berechnet und in dem Eigenschaftsvektor  $f^t = [f_x, f_y, f_A, f_H, f_K, f_L, f_O]$  gespeichert.



**Abbildung 2.17.:** Berechnung der Orientierung und der Länge einer Zellmaske.

### 2.4.3. Zelltracking

Die Aufgabe des Zelltracking besteht darin, den Weg einer Zelle über die Zeit zu rekonstruieren und die Verwandtschaft verschiedener Zellen zu bestimmen. Um das zu erreichen, werden die Konturen der Zellen in den Bildern mit Hilfe einer Bildsegmentierung extrahiert. Diese Konturen werden genutzt, um einen Zellpfad aufzubauen, der die zeitliche Bewegung einer Zelle über die Zeit beschreibt. Jede Zelle wird folglich von vielen Zellmasken repräsentiert, die als Zellpfad zusammengefasst werden.

#### (Naives) Überlaptracking

Mit den Zellmasken, d.h. den detektierten Zellen, besteht das Zelltracking-Problem darin, für eine gegebene Zelle zum Zeitpunkt  $t$  die Nachfolgerzelle in dem folgenden Bild zum Zeitpunkt  $t + 1$  zu bestimmen. Um diese zu identifizieren, wird der Schnitt oder der Überlapp der Zellkonturen der Bilder  $I_t$  und  $I_{t+1}$ , genutzt (vgl. Abb. 2.18). Das Verhältnis zwischen dem Schnitt zweier Zellmasken  $z_i^t, z_j^{t+1}$  und der Fläche  $A_i^t, A_j^{t+1}$  der Zellen wird als Vorwärts- und Rückwärtsüberlapp  $O^V(z_i^t, z_j^{t+1})$  und  $O^R(z_j^{t+1}, z_i^t)$  bezeichnet und mit

$$O^V(z_i^t, z_j^{t+1}) = \frac{A_i^t \cap A_j^{t+1}}{A_j^{t+1}} \quad \text{und} \quad O^R(z_j^{t+1}, z_i^t) = \frac{A_i^t \cap A_j^{t+1}}{A_i^t} \quad (2.2)$$

angegeben. Eine eindeutige Nachfolgerzelle wird identifiziert, indem das Zellpaar mit dem größten Überlapp der Zellkonturen bestimmt wird. Der Rückwärtsüberlapp und der Vorwärtsüberlapp wird dabei genutzt, um Fehldetektion wie z.B. die Verschmelzung von zwei Zellen zu detektieren (vgl. Abb. 2.18(b)). Im Falle einer Verschmelzung wäre z.B. der Vorwärtsüberlapp sehr groß ( $\approx 1$ ), der Rückwärtsüberlapp aber klein ( $\approx \frac{1}{2}$ ) (vgl. Abb. 2.19).

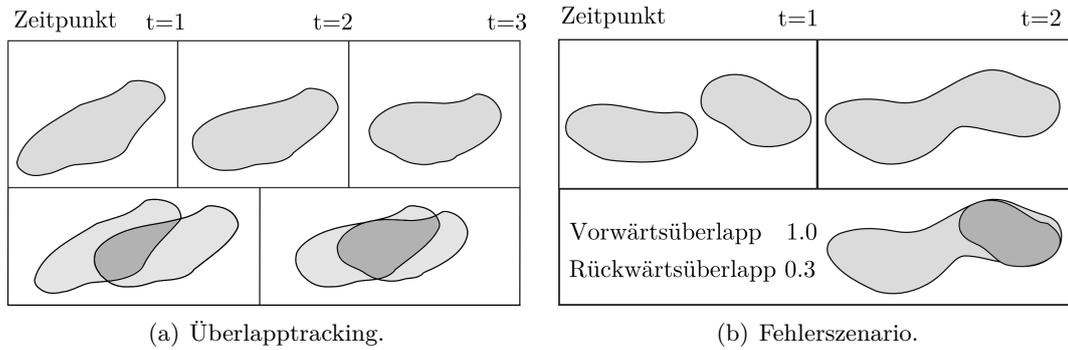
Mit diesem Überlaptracking wird ein naiver Tracking-Algorithmus implementiert. In diesem Ansatz wird ein Rückwärts-Tracking (*backward tracking*) genutzt, um die Wege der Zellen zu rekonstruieren. Für alle Zellen wird die Zellkontur mit dem größten Rückwärtsüberlapp als Vorgängerzelle gewählt und der Vorwärtsüberlapp wird nicht verwendet. Unter der naiven Annahme, dass alle Zellen korrekt erkannt werden und die zeitliche Auflösung hoch genug ist, liefert dieses Verfahren die gesuchten Zellpfade und somit die vollständige Genealogie. Daher der Name **naives Tracking**. Da die Annahme für echte Zellbilder nicht vorausgesetzt werden kann, wird das Überlaptracking genutzt, um sichere, kurze Pfadfragmente zu erstellen. Diese initialen Fragmente liefern die Grundlage für die Validierung, in der gezielt nach fehlerhaften Fragmenten oder Verknüpfungen von Fragmenten gesucht wird.

#### Initiale Pfadfragmente

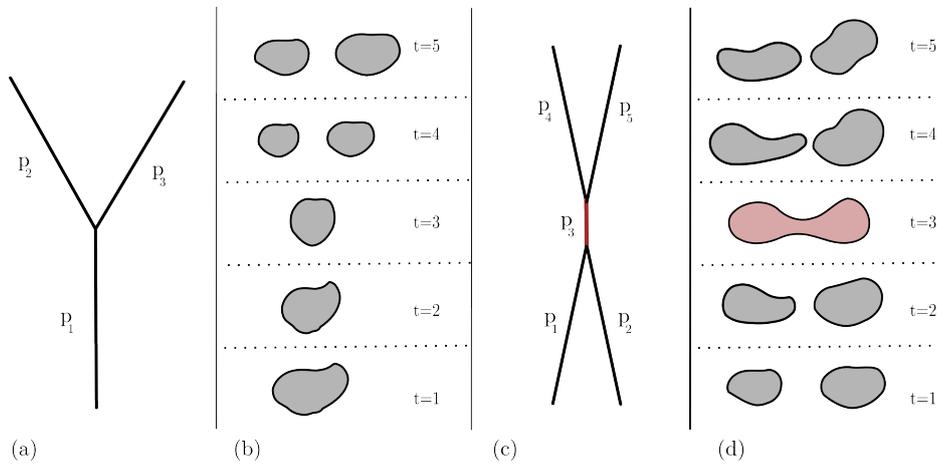
Um zuverlässige Pfadfragmente der Zellen zu extrahieren wird eine Mindestgröße für den Vorwärtsüberlapp  $O^V(z_i^t, z_j^{t+1}) > \theta^V$  und für den Rückwärtsüberlapp  $O^R(z_j^{t+1}, z_i^t) < \theta^R$  gefordert. Dieser Schwellwert hängt von einer Vielzahl von Faktoren ab wie z.B. dem Zelltyp, der Morphologie, der Geschwindigkeit und dem experimentellen Setup, wie z.B. die Zelldichte und die zeitliche Auflösung. In dem vorliegenden Framework wird

## 2.4. Das Zelltracking-Framework

ein Wert von  $\theta^V = \theta^R = 0.6$  gewählt. Bereits beim Erstellen der initialen Pfade wird verhindert, dass Detektionsfehler, wie in Abb. 2.18(b) dargestellt, zu falschen Tracking-Ergebnissen führen. Zusätzlich wird an jede Zelle eines sicheren Pfadfragmentes die Bedingung gestellt, dass sich eine zulässige Nachfolger- bzw. Vorgängerzelle nur mit exakt einer Zellkontur des vorherigen bzw. folgenden Bildes überschneiden darf.



**Abbildung 2.18.:** Überlappstracking. Für eine gegebene Zelle kann ein Nachfolger gefunden werden, indem der Schnitt (dunkelgrau) der Zellmasken (hellgrau) aus zwei folgenden Bildern berechnet wird (a). Das Verhältnis zwischen der Zellgröße und dem Schnitt zweier Zellen wird als Vorwärts- und Rückwärtsüberlapp definiert und liefert ein Fehlermaß für Fehldetektionen (b).



**Abbildung 2.19.:** Pfadfragmente des initialen Zelltracking. Solange eine Zelle einen eindeutigen und alleinigen Nachfolger hat, wird ihr Weg über die Zeit rekonstruiert und als Zellpfad gespeichert. Die Konstruktion dieses Zellpfades wird abgebrochen, falls es mehrere Kandidaten gibt (a-b), wobei die Konnektivität der Pfade in einer Adjazenzliste gespeichert wird. Anhand dieser Konnektivität es möglich, Fehlerquellen der Zelldetektion wie z.B. eine Verschmelzung zweier Zellkonturen automatisiert zu finden (b).

Nach der Erstellung der initialen Pfade wird jede Zellmaske  $z_i^t$  genau einem Pfadfragment  $z_i^t \in p_1, \dots, p_N$  zugeordnet, d.h. jedes Pfadfragment beschreibt den Weg einer Zelle über einen (kurzen) Zeitabschnitt. Üblicherweise wird die Konstruktion dieser Pfadfragmente dadurch beendet, dass mehrere konkurrierende Nachfolgerzellen, d.h. mehrerer überlappende Zellkonturen, vorhanden sind. Jede dieser Zellkonturen beschreibt eine Möglichkeit, wie der Zellpfad fortgeführt werden könnte. Diese Verknüpfungen werden in Form einer Adjazenzmatrix gespeichert. In der initialen Form speichert diese Matrix noch nicht die vollständigen Pfade und korrekten Verbindungen, sondern dient als Ausgangspunkt für die später folgende Validierung der Pfadfragmente.

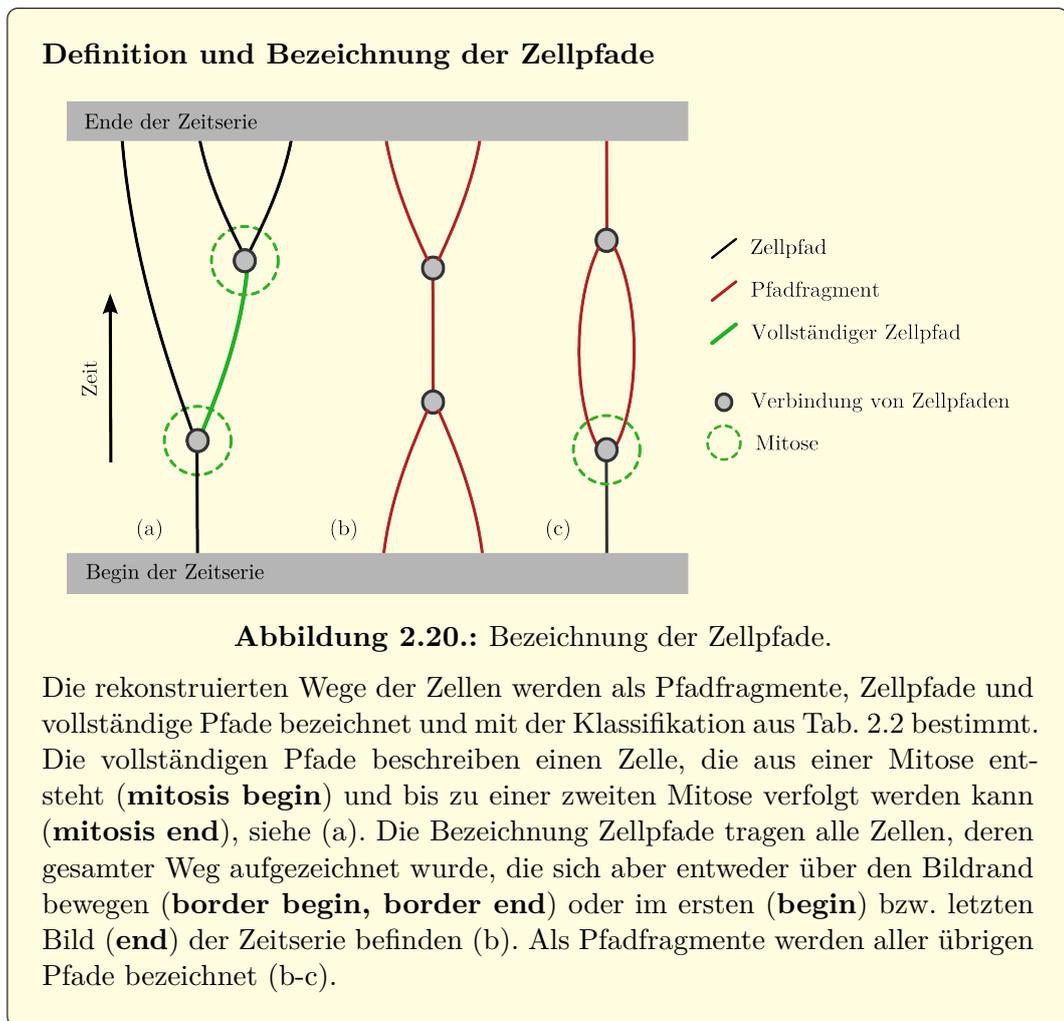
Klassifikation	Beschreibung
<b>begin</b>	Das Pfadfragment beginnt in dem ersten Zellbild $I_1$ .
<b>end</b>	Das Pfadfragment endet im letzten Zellbild $I_N$ .
<b>border begin</b>	Das Pfadfragment ohne Vorgänger beginnt in der Nähe des Bildrands.
<b>border end</b>	Das Pfadfragment ohne Nachfolger endet in der Nähe des Bildrands*.
<b>mitosis begin</b>	Das Pfadfragment entsteht aus einer Zellteilung.
<b>mitosis end</b>	Das Pfadfragment endet mit einer Zellteilung.
<b>cell death</b>	Das Pfadfragment endet mit einem Zelltod.
lost begin	Das Pfadfragment beginnt, aber weder <i>begin</i> noch <i>border begin</i> ist gesetzt.
lost end	Das Pfadfragment endet, aber weder <i>end</i> , <i>cell death</i> , <i>mitosis</i> oder <i>border end</i> ist gesetzt.
merged	Zwei oder mehr Zellen werden als eine Zellkontur detektiert.

**Tabelle 2.2.:** Die ersten sechs Klassifikationen (fett) bestimmen endgültige Start- oder Endpunkte der Pfadfragmente, während die folgenden beiden die Fragmente bezeichnen, an denen das Tracking noch nicht vollständig abgeschlossen ist. Die letzte Bezeichnung wird explizit für Fehldetektionen genutzt. Um eine einheitliche Bezeichnung zu nutzen, werden die englischen Begriffe aus Rapoport/ Becker et al. [8] verwendet. Quelle: doi: 10.1371/journal.pone.0027315.t001

\*: Die Nähe zu dem Bildrand ist als die minimale Entfernung des Mittelpunktes einer Zelle zu einem der vier Bildränder. Dieser Wert wird abhängig von den Zellen gewählt und auf 25 – 40 Pixel gesetzt, was ca.  $50\mu\text{m}$  entspricht.

### Klassifikation und Bezeichnung der Pfadfragmente

Im idealen Fall beschreibt ein Pfadfragment den zeitlich und räumlichen Weg einer Zelle von einer Mitose (der Geburt der Zelle) bis zu einer zweiten Zellteilung. Da die meisten Fragmente allerdings an weniger wünschenswerten Stellen beginnen, enden oder abreißen, werden die Pfadfragmente automatisch den Klassen in Tabelle 2.2 zugeordnet. Diese Klassifikation wird genutzt, um eine eindeutige Bezeichnung der verschiedenen rekonstruierten Wege der Zellen einzuführen. Die Pfade werden in **vollständige Pfade**, **Zellpfade** und **Pfadfragmente** unterteilt (vgl. Abb. 2.20). Ein vollständiger Pfad beschreibt eine Zelle, von der das gesamte Leben aufgezeichnet wurde. Hierfür ist es erforderlich, dass der gesamte Zellzyklus beobachtet wurde, d.h. von der Geburtsmitose bis zu der Teilung der Zelle selbst. Diese Pfade werden als **vollständige Pfade** bezeichnet und bilden die Bausteine, aus denen die Stammbäume zusammengesetzt werden. Eine valide Rekonstruktion der vollständigen Pfade ist daher notwendig für eine Extraktion der Genealogien.



**Anmerkung zu den vollständigen Pfaden:** Die Anzahl  $N_v$  der vollständigen Pfade kann mit der Anzahl der Zellen  $M_{\text{ende}}$  im letzten Bild abgeschätzt werden. Hierzu wird die Zahl aller Vorgängerzellen bestimmt. Die Zellen im letzten Bild haben  $\frac{M_{\text{ende}}}{2}$  Vorgänger, die wiederum  $d \frac{M_{\text{ende}}}{4}$  Vorgänger haben (die Großeltern). Werden nun alle Vorfahren aufsummiert, ergibt sich eine Zahl von

$$N_v < \frac{M_{\text{ende}}}{2} + \frac{M_{\text{ende}}}{4} + \dots + 1 \approx M_{\text{anfang}},$$

vollständigen Pfade, wobei davon ausgegangen wird, dass alle Zellen einen gemeinsamen Vorgänger haben. Geht man von  $M_{\text{anfang}}$  Zellen im ersten Bild aus, so erhält man  $N_v \approx M_{\text{ende}} - M_{\text{anfang}}$  vollständige Pfade. Diese Approximation liefert eine obere Schranke, da der Einfluss der Zellen, die über den Bildrand ein- oder ausgewandert sind, nicht gewertet wurde. Diese Approximation erlaubt es, die Genauigkeit der rekonstruierten Pfade und damit die Güte des Zelltracking abzuschätzen.

### Validierung der Zelltracking-Daten

Da es im Allgemeinen aber nicht möglich ist, eine fehlerfreie Bildsegmentierung und ein fehlerfreies Zelltracking durchzuführen, wird im folgenden Kapitel ein Ansatz vorgestellt, mit dem die Zelltracking-Daten validiert werden können. Mit der Validierung ist es möglich, die fehlerhaften und fehlerfreien Zellpfade zu identifizieren und zu extrahieren, d.h. aus den fehlerbehafteten Zelltracking-Daten werden verlässliche Daten generiert.

Um die komplexeste Aufgabe zu erfüllen, die in der Rekonstruktion von Verwandtschaften besteht, ist es weiterhin erforderlich, dass die Zellteilungen verlässlich detektiert werden, da hier neue Zellen in das Tracking-Szenario hinzugefügt werden. Die Detektion der Mitosen in *in vitro*-Zellkulturen ist insgesamt von großer Bedeutung [27, 42, 45, 48, 66, 69, 73, 76, 122]. In Kapitel 4 werden daher verschiedenen Detektionsansätze vorgestellt und ausführlich evaluiert.

# 3

## Validiertes Zelltracking

Das Zelltracking-Framework aus Abschnitt 2.4 liefert die grundlegenden Algorithmen und Methoden, um Zellen in einer gegebenen Bildserie zu erkennen und über die Zeit zu verfolgen. Nur solange die Bildsegmentierung (Zelldetektion) und die Zuweisung der Zellen zwischen den Bildern (Zelltracking) fehlerfrei funktionieren, kann eine vollständige Genealogie der beobachteten Zellen extrahiert werden.

In einer Auswertung von echten Bildserien ist es allerdings nicht möglich, eine fehlerfreie Bildsegmentierung zu garantieren, d.h. es muss davon ausgegangen werden, dass die detektierten Zellmasken und rekonstruierten Zellpfade falsche Konturen und Zuweisungen enthalten. Um die Auswirkung dieser Fehler abzuschätzen, wird eine systematische Analyse der verschiedenen Fehlerquellen durchgeführt. Anschließend wird die Fehlerfortpflanzung in der gegebenen Binärbaum-Struktur untersucht und es wird gezeigt, welche Auswirkung diese Fehler auf die Rekonstruktion der vollständigen Genealogien haben.

Um dennoch die Bildserien bzw. die Genealogien automatisiert auszuwerten, wird das Konzept des **validierten Zelltracking** vorgestellt, welches es erlaubt, eine fehlerfreie Extraktion von sicheren Ergebnissen durchzuführen. Hierzu wird ein **Genealogisches Modell** eingeführt, das die extrahierten Zellpfade anhand einer Integritätsprüfung validiert. Mit Hilfe dieser Prozesse ist es möglich, die extrahierten Zellpfade der Tracking-Algorithmen systematisch auf Fehler zu untersuchen und die fehlerhaften Pfade aus einer weiteren Auswertung auszuschließen.

Aufbauend auf diesem Modell sind zwei konsistente, validierte **Referenzdatensätze** entstanden, die sowohl korrigierte Zellmasken als auch korrigierte Genealogien enthalten, d.h. jede Zelle ist dem richtigen Stammbaum zugeordnet. Diese Datensätze nehmen in den folgenden zwei Kapiteln eine zentrale Rolle ein und werden verwendet, um die Zelltracking-Algorithmen aus Abschnitt 2.4 zu evaluieren und zu validieren.

Das vorgestellte validierte Zelltracking ist in dem Zelltracking-Framework aus Abschnitt 2.4 integriert und in der Arbeit von Rapoport/Becker et al. [8] veröffentlicht. Die vollständigen Referenzdaten wurden unter die **Creative Commons**-Lizenz gestellt, d.h. die Datensätze sind für jedermann frei zugänglich und können als Basis für weitere, vergleichende Auswertungen und Evaluationen genutzt werden.

### 3.1. Fehleranalyse und Validierung von Zelltracking-Daten

Um die Auswirkung einer fehlerhafter Zelldetektion abschätzen zu können, werden im Folgenden die Fehlerquellen und die Fehlerfortpflanzung ausführlich diskutiert. Es wird

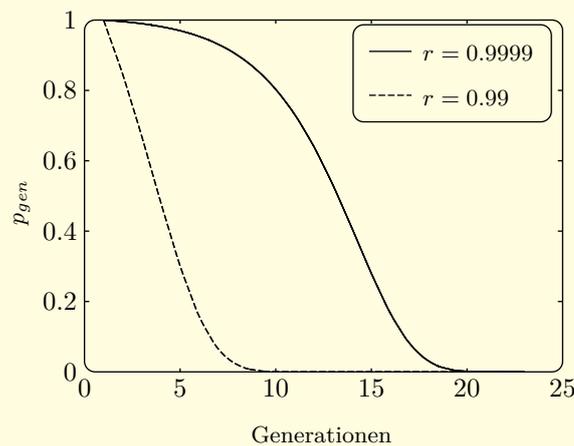
ein Genealogisches Modell vorgestellt, mit dem diese Fehler systematisch identifiziert werden können und das eine Validierung der rekonstruierten Stammbäume ermöglicht. Dieses Modell wird als Ausgangspunkt für einen validierten, konsistenten Referenzdatensatz genutzt.

### Fehlerfortpflanzung in zellulären Genealogien

Um die Auswirkung der Fehlerfortpflanzung zu veranschaulichen, wird zunächst die Fehleranfälligkeit in den gegebenen Stammbäumen betrachtet. Die natürliche Binärbaumstruktur bewirkt, dass bereits ein einzelner Detektionsfehler eine große Auswirkung auf die Rekonstruktion der zellulären Genealogien haben kann: Nur wenn von einer betrachteten Zelle alle Nachfolger in allen folgenden Bildern korrekt erkannt und zugeordnet werden, ist es möglich, alle Verwandten, d.h. die gesamte Zellfamilie, zu finden. Im Folgenden werden zwei Wahrscheinlichkeiten angegeben, die abhängig von der Genauigkeit der Zelldetektion  $r$  sind:

1. Die Wahrscheinlichkeit, mit welcher der Pfad einer Zelle über 40 Bilder rekonstruiert wird, ergibt sich zu  $p_{pfad}(p_i) = r^{40}$ .
2. Mit  $p_{pfad}$  kann nun die Wahrscheinlichkeit  $p_{gen}$  angegeben werden, mit der eine Genealogie einer Zellfamilie extrahiert wird. Ein Stammbaum mit  $n = 3$  Generationen enthält  $2^{n+1} - 1 = 15$  vollständige Pfade, woraus sich  $p_{gen} = (p_{pfad})^{15} = (r^{40})^{15}$  ergibt<sup>a</sup>.

In Abb. 3.1 ist die Wahrscheinlichkeit  $p_{gen}$  abhängig von der Generationstiefe angegeben. Für die Genauigkeit der Zelldetektion wurden zwei optimistische Werte  $r = 0.99$  und  $r = 0.9999$  angenommen. Die Resultate zeigen, welche Herausforderung an einen Zelltracking-Algorithmus gestellt werden.



**Abbildung 3.1.:** Rekonstruktionswahrscheinlichkeit  $p_{gen}$ , mit der ein vollständige Stammbäume abhängig von der Zelldetektionsrate  $r$  erstellt wird.

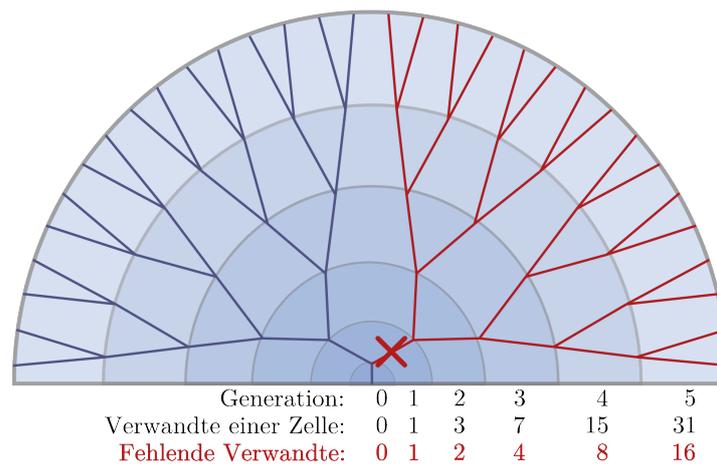
<sup>a</sup>Vorausgesetzt, jede Zelle ist in 40 Bildern sichtbar.

### 3.1.1. Fehler und Fehlerfortpflanzung in zellulären Stammbäumen

Die Wahrscheinlichkeit, mit der ein Zellpfad von einem naiven Zelltracking-Ansatz rekonstruiert wird, hängt direkt von der Genauigkeit der Zelldetektion ab. Um das zu veranschaulichen, wird eine hypothetische Zelle betrachtet, deren Weg über einen Zeitraum von zehn Stunden mit einem Aufnahmeintervall von 15 Minuten aufgezeichnet wurde. Um den zugehörigen Pfad zu rekonstruieren, muss die Zelle in 40 Bildern korrekt detektiert werden, was selbst bei einer angenommenen Detektionsrate einer einzelnen Zelle von  $r = 0.99$  eine geringe Rekonstruktionswahrscheinlichkeit  $p_{\text{pfad}} = 0.99^{40} = 0.669$  ergibt. Die resultierende Wahrscheinlichkeit  $p_{\text{gen}}$ , mit der ein vollständiger Stammbaum erstellt wird, ist in Abb. 3.1 für zwei optimistische gewählte Detektionsraten  $r = 0.99$  und  $r = 0.9999$  gezeigt und verdeutlichen, wie fehleranfällig bzw. schwer die Aufgabe des Zelltracking ist.

Während die bisher betrachteten Fehler die Genauigkeit der Rekonstruktionsraten von vollständigen Stammbäumen angeben, ist in Abb. 3.2 veranschaulicht, dass bereits ein einzelner Fehler einen großen Einfluss auf die extrahierte Genealogie haben kann. Gezeigt ist der Fall, dass eine Verbindung übersehen wird und die Genealogie als Folge zerbricht, d.h. Verbindungen und viele Verwandtschaften gehen verloren. Ein weitaus schwerwiegenderer Fehler ist es aber, wenn in zwei Zellfamilien aufgrund von einem Tracking-Fehler zwei Zellen bzw. die dazugehörigen Äste vertauscht werden. Dies kann dazu führen, dass die Interpretation der Bäume auf einer falschen Datengrundlage ausgeführt wird.

Insgesamt zeigen die Abbildungen 3.1 und 3.2, welche Anforderungen an eine automatisierte Rekonstruktion von zellulären Genealogien über viele Generationen hinweg gestellt wird.



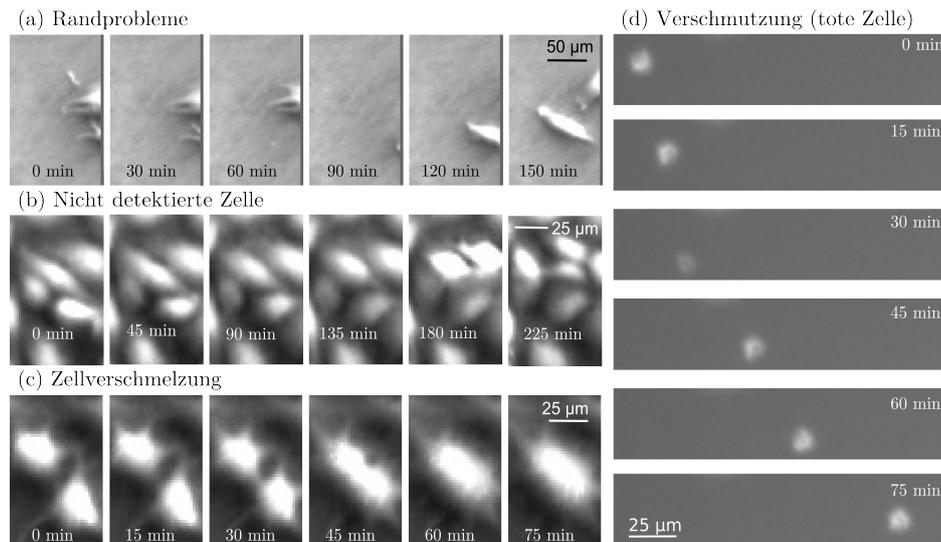
**Abbildung 3.2.:** Aufgrund der starken Fehlerfortpflanzung in den genealogischen Bäumen kann schon ein einzelner Fehler dazu führen, dass eine Genealogie zerbricht und die Hälfte aller Verwandtschaften nicht korrekt rekonstruiert werden kann.

## Fehlerquellen

Betrachten wir nun die Fehlerquellen, die in den Zelltracking-Szenarien auftreten können und in den ausgewerteten Datensätzen beobachtet wurden. Die folgende Liste bietet einen ausführlichen und vollständigen Überblick über die möglichen Fehlerarten und gibt an, wie diese mit Hilfe der Klassifikation der Pfadfragmente aus Tab. 2.2 des vorherigen Kapitels 2 identifiziert werden können.

1. **Randprobleme.** Da die Mikroskopbilder nur einen kleinen Ausschnitt der gesamten Zellpopulation zeigen, können an den Rändern des beobachteten Gebietes methodische Tracking-Fehler beobachtet werden: Es gibt Zellen, die aus dem Bild hinaus bzw. in das Bild hinein wandern, wie in Abb. 3.3(a) dargestellt, so dass neue Zellen in das Tracking-Szenario eingefügt bzw. daraus entfernt werden. Um diese Pfadfragmente zu identifizieren, werden sie als *border begin/end* klassifiziert.
2. **Nicht detektierte Zellen.** Zellen, die zu klein sind oder einen sehr geringen Kontrast aufweisen, werden von der Bildsegmentierung nicht erkannt (vgl. Abb. 3.3(b)). Diese Fehler führen zu Pfadfragmenten ohne gültigen Nachfolger bzw. ohne gültigen Vorgänger die als *lost end* bzw. *lost begin* markiert werden.
3. **Verschmolzene Zellen.** Zellen, die sehr dicht beieinander liegen, können von der Zelldetektion als eine einzelne Zelle erkannt werden (vgl. Abb. 3.3(c)). Diese Untersegmentierung ist eine der häufigsten Fehlerquellen in den betrachteten Bildserien und kann mit Hilfe der Pfadfragmente mit einem *lost end* Ende, der Adjazenzmatrix  $\mathbf{AM}$  und dem Überlapp gefunden werden (vgl. 2.4). Enden mehrere Pfadfragmente in einer großen Zellmaske, d.h. es wurde ein großer Vorwärtsüberlapp für alle verschmelzenden und ein kleiner Rückwärtsüberlapp für die verschmolzene Zellkontur beobachtet, so liegt eine Zellverschmelzung vor. Die entsprechenden Pfadfragmente werden mit dem *merged*-Flag versehen.
4. **Verschmutzung und Artefakte.** Neben den lebenden Zellen, die beobachtet werden, lässt es sich nicht verhindern, dass sich noch weitere Objekte in den Kulturschalen befinden. Es kann passieren, dass Teile toter Zellen, abgelöste Zellen oder Staubkörner von der Bildsegmentierung erkannt und fälschlicherweise als Zelle erkannt werden. Die meisten dieser Objekte zeigen ein typisches Muster: Sie bewegen sich im Vergleich zu adhärennten Zellen sehr schnell, da sie sich frei im Zellmedium bewegen und somit von der Fließgeschwindigkeit des Zellmediums direkt beeinflusst werden (vgl. Abb. 3.3). Diese Objekte werden in den Tracking-Daten als kurze Pfadfragmente ohne einen gültigen Anfang oder ein gültiges Ende (*lost begin* und *lost end*) erkannt.

Anhand der gegebenen Bildserien können diese Fehlerquellen leicht identifiziert und definiert werden. Um sie zu quantifizieren und zu analysieren werden aber vollständige Referenzdaten benötigt, in denen die korrekten bzw. korrigierten Zellkonturen aller Stammbäume enthalten sind. In dem nächsten Abschnitt wird ein Genealogisches Modell entworfen, welches es ermöglicht, die Fehlerfreiheit der Daten mit einer Konsistenzprüfung zu verifizieren. Weiterhin lässt die verwendete Datenstruktur eine Integritätsprüfung zu, die als Grundlage für die Erstellung von *validierten* Referenzdaten genutzt wird.



**Abbildung 3.3.:** Fehlerquellen in Zelltracking-Szenarien. **(a) Rand.** Zwei Zellen verlassen den beobachteten Bereich (der Bildrand ist auf der rechten Seite als grauer Balken dargestellt). Anschließend wandern zwei neue Zellen in die Aufnahmen, aber eine korrekte Zuweisung der Zellen ist nicht möglich. **(b) Nicht erkannte Zellen.** Einige Zellen werden sehr klein oder verlieren ihren Kontrast, wodurch sie nicht mehr von der Bildsegmentierung erkannt werden. **(c) Verschmolzene Zellkonturen.** Liegen einige Zellen sehr dicht beieinander, werden die Zellen als eine Zellkontur erkannt, so dass es nicht immer möglich ist, die Zellen eindeutig zu trennen. **(d) Verunreinigung und Artefakte.** Verschmutzungen wie Staubpartikel oder Zelldebris von toten Zellen können als Zellen erkannt werden und bewegen sich sehr schnell durch die Bilder. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g004.

#### Fehler vermeiden und erkennen: Evaluierung und Validierung

Die Analyse der Fehlerquellen zeigt, dass der Ursprung vieler Fehler in dem Verhalten der Zellen einerseits und den gegebenen Bilddaten bzw. dem eingeschränkten Beobachtungsgebiet andererseits liegt. Die Fehlerrate kann daher nur minimiert, es kann aber nicht ausgeschlossen werden, dass wenige Fehler auftreten. Bei einer Auswertung der Bilddaten und besonders bei der Extraktion von Genealogien muss daher davon ausgegangen werden, dass Teile der betrachteten Daten fehlerhaft sind. Das Ziel ist es nun, das vorhandene Zelltracking-Framework so zu erweitern, dass eine verlässliche Auswertung der Tracking-Daten auch bei den gegebenen Fehlern gewährleistet wird. Hierzu ist es erforderlich, die fehlerhaften Zellpfade automatisiert zu identifizieren und

- die Pfadfragmente und ihre Verknüpfungen zu korrigieren, **oder**
- die betroffenen Pfade aus der weiteren Auswertung zu entfernen.

Um das zu erreichen, wird eine Validierung vorgestellt, die das Ziel verfolgt, die Zelltracking- und Detektionsfehler zu identifizieren.

Die Grundidee dieser Validierung besteht darin, ein Modell für eine zelluläre Genealogie zu nutzen, die das Verhalten der Zellen beschreibt. Von diesem Modell wird

eine konsistente Datenstruktur abgeleitet, mit der jeder Fehler in den Genealogien als Inkonsistenz erkannt werden kann. Um diese zu finden, wird eine Integritätsprüfung der Datenstruktur und der Zellpfade durchgeführt, die auf dem Genealogischen Modell aufbaut. Diese Integritätsprüfung kann unabhängig von dem verwendeten Tracking-Algorithmus auf die erstellten Zellpfade und Stammbäume angewendet werden. Sie ist der Ausgangspunkt für einen konsistenten, manuell korrigierten Referenzdatensatz, der als Benchmark für weitere Evaluierungen verwendet wurde. Die vorgestellten Methoden und Prinzipien wurden genutzt, um das Zelltracking-Framework aus Abschnitt 2.4 zu erweitern.

### 3.1.2. Das Genealogische Modell

Das Genealogische Modell fasst die impliziten Annahmen zusammen, die bei jedem Zelltracking-Algorithmus vorausgesetzt werden: Es wird davon ausgegangen, dass die Objekte, die von der Bildsegmentierung erkannt werden, die biologischen Zellen in der Petrischale repräsentieren. Im Umkehrschluss lassen sich nun aus den Eigenschaften der biologischen Zellen Bedingungen an die rekonstruierten Zellpfade stellen, die als **Genealogisches Modell** definiert werden. Diese Zellen werden als zeitlich-räumliche Objekte in den Bildserien extrahiert: Ist eine biologische Zelle in  $n$  gegebenen Bildern sichtbar, so wird sie im Idealfall auch  $n$ -mal erkannt. Diese räumliche Repräsentation einer Zelle wird als Zellmaske bezeichnet und durch die Bildsegmentierung erstellt (vgl. Abschnitt 2.4.2). Die zeitliche Komponente liefert das Zelltracking, mit dem der Weg einer Zelle verfolgt wird. Die zeitliche Assoziation der einzelnen Zellmasken wird als Zellpfad zusammengefasst, wie in Abschnitt 2.4.3 beschrieben. Im Idealfall beschreibt ein Zellpfad eine gesamte Zelle von Mitose zu Mitose. Diese Pfade wurden in Abschnitt 2.4.3 als vollständige Pfade bezeichnet. Mit den Zellpfaden werden nun zwei grundlegende Eigenschaften der Zellen abgebildet. Erstens: Die Zellen sind persistent und können nicht einfach verschwinden oder entstehen. Zweitens: Sie bewegen sich und ändern sowohl ihre Form als auch ihren Ort. Die wichtigste Eigenschaft der Zellen ist aber die Fähigkeit, sich durch eine Zellteilung zu vermehren. Die so entstehenden Verwandtschaftsbeziehungen der Zellen werden als Verbindungen der Zellpfade extrahiert und gespeichert.

Die Konsistenzprüfung, die mit Hilfe des Genealogischen Modells durchgeführt wird, basiert auf der Konnektivität der Zellpfade. Hierbei wird vorausgesetzt, dass eine Mitose die einzige konsistente Verbindung zwischen zwei Zellpfaden ist und erklärt, wie neue Zellpfade in dem Tracking-Szenario entstehen. Da die aufgezeichnete Bildserie aber sowohl zeitlich durch die Beobachtungsdauer als auch räumlich durch die Bildgröße eingeschränkt wird, müssen mehrere Möglichkeiten zugelassen werden, die erklären, wie ein Zellpfad beginnt. Eine beobachtete Zelle kann

1. im ersten Bild bereits sichtbar sein,
2. durch eine Mitose entstanden sind, oder
3. vom Rand des Bildes in den Beobachtungsbereich einwandern.

Umgekehrt kann eine Zelle das Tracking-Szenario nur verlassen, indem sie

4. bis zum letzten Bild sichtbar ist,
5. sich durch eine Mitose in zwei Zellen aufteilt, oder

6. über den Bildrand wandert und den Beobachtungsbereich verlässt.

Es gibt keine andere Möglichkeiten für die Zellen, in das Tracking-Szenario hinzugefügt oder entfernt zu werden. Für das oben definierte Modell bedeutet das wiederum, dass jeder Fehler in der Zelldetektion oder dem Zelltracking zu einer Inkonsistenz der Datenstruktur führt. Die Integritätsprüfung basiert auf genau dieser Annahme und identifiziert automatisch alle Zellen, die keinen Ursprung oder kein Ende nach der gegebenen Definition haben und extrahiert diese als **inkonsistente** Zellpfade.

#### 3.1.3. Validiertes Zelltracking als Integritätsprüfung

Weiterhin wird eine Integritätsprüfung der Datenstrukturen genutzt, um Fehler in den Zelltracking-Daten zu identifizieren. Diese Prüfung basiert auf der Datenstruktur, die in Abschnitt 2.4.3 angegeben wurde und in der die Zellen  $z_i$ , die Zellpfade  $p_j$  und die Konnektivität (anhand der Adjazenzmatrix **AM**) gespeichert werden. Diese Definitionen werden genutzt, um die Konsistenz der Datenstruktur, d.h. die Detektionsergebnisse in den Bildern und die spatio-temporale Verknüpfung über die Zeit, anhand der folgenden Bedingungen zu überprüfen:

- Jeder Pixel  $x_i \in I$  gehört entweder zu dem Hintergrund oder zu genau einer Zellmaske  $z_i$ .
- Jede zusammenhängende Region wird als eine Zellmaske  $z_i$  zusammengefasst. Umgekehrt muss jede Zellmaske aus genau einer zusammenhängenden Region bestehen.
- Jede Zellmaske gehört zu genau einem Zellpfad und jeder Zellpfad verweist zu jedem Zeitpunkt, an dem er definiert ist, auf genau eine Zellmaske.

Jede Abweichung von diesen Bedingungen entspricht einem Fehler, wie exemplarisch in Abb. 3.4 dargestellt. Je nach Kontext werden diese Fehler unterschiedlich genutzt. Werden manuell erstellte Daten betrachtet, so repräsentiert jede Inkonsistenz eine falsch erstellte bzw. falsch korrigierte Zellmaske oder einen fehlerhaften Zellpfad. Diese Integritätsprüfung liefert die Grundlage für einen validierten und konsistenten Referenzdatensatz, der nur validierte und somit fehlerfreie Zellpfade und Stammbäume enthält. Eine weitere Anwendung besteht in der systematischen Evaluierung von Tracking-Algorithmen. Dieser Prozess nutzt die Konsistenzprüfungen um systematische Fehler in gegebenen Zelltracking-Algorithmen zu finden, d.h. es wird ein Werkzeug für die Analyse der Algorithmen selbst bereitgestellt.

Eine direkte Anwendung der Datenvalidierung ist das **validierte Zelltracking**, welches es erlaubt, fehlerhafte Zellpfade zu identifizieren und entweder zu korrigieren oder aus der Auswertung auszuschließen. In Abb. 3.5 ist das verwendete Schema dargestellt, das genutzt wird um fehlerfreie Zellpfade zu extrahieren. Die Validierung nutzt die Integritätsprüfungen aus 3.1.3 um die Zellpfade in konsistente, inkonsistente Pfade und Mitosekandidaten zu unterteilen, wie in Abb. 3.4 gezeigt. Der letzte und entscheidende Schritt, der für die Extraktion von Genealogien erforderlich ist, ist die Detektion der Zellteilungen. Da die Mitosedetektion einen komplexen und entscheidenden Schritt darstellt, wird sie ausführlich im folgenden Kapitel 4 thematisiert.

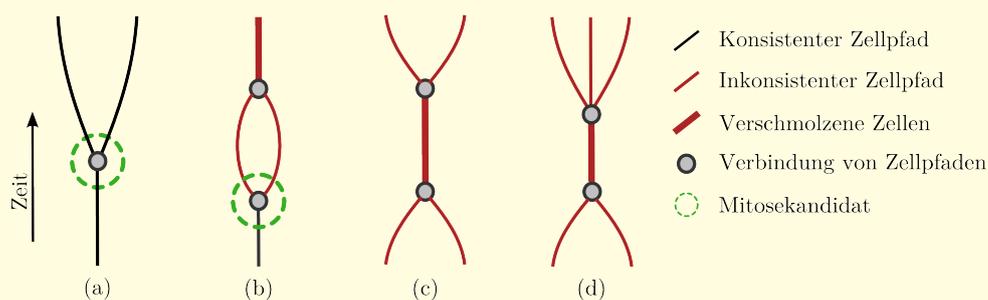
### Validierung der Zellfaden mit dem Genealogischen Modell

Das Ziel der Validierung besteht darin, alle fehlerhaften Zellpfade als Inkonsistenz in dem Genealogischen Modell zu bestimmen und alle konsistenten Zellpfade zu extrahieren. Mit Hilfe der Integritätsprüfung, die aus dem Genealogischen Modell abgeleitet wurde, werden alle Zellpfade mit einem gültigen Anfang bzw. mit einem gültigen Ende identifiziert. Anschließend werden die Zellpfade wie folgt anhand ihrer Konnektivitäten in konsistente Pfade, Mitosekandidaten oder kritische Pfade unterteilt:

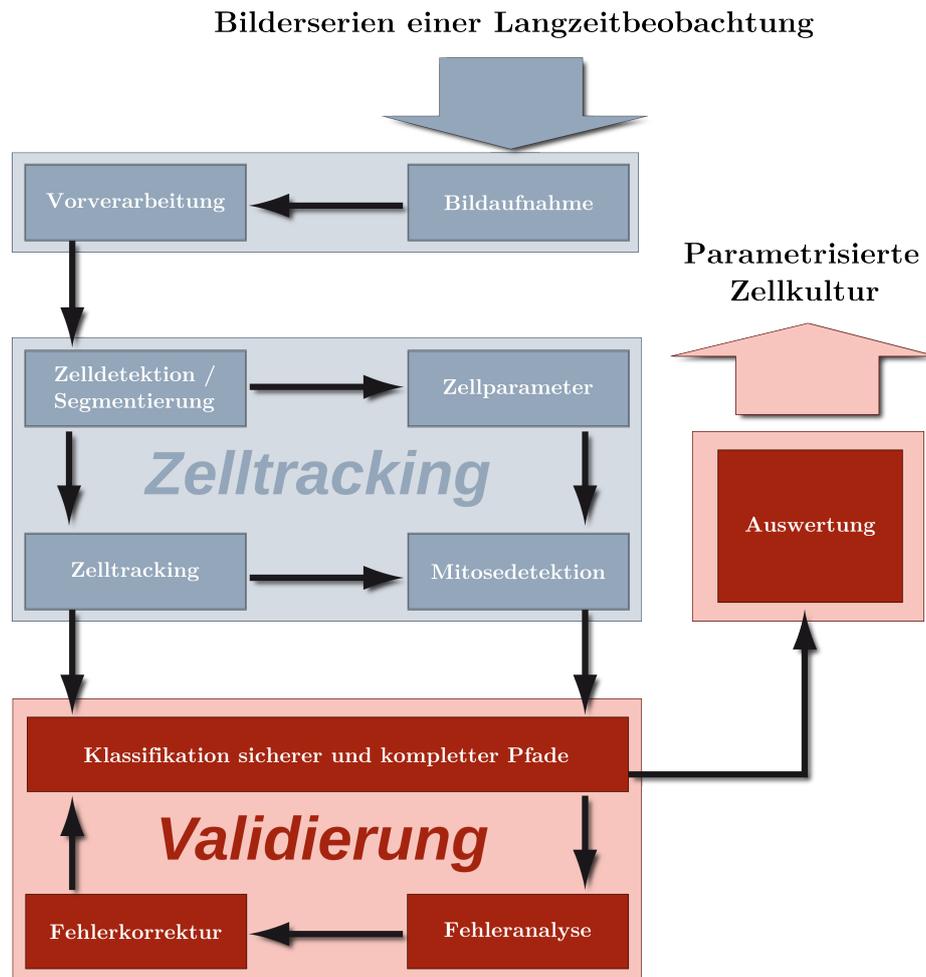
1. **Konsistente Pfade.** Diese Pfade haben einen gültigen Anfang und ein gültiges Ende.
2. **Mitosekandidaten.** Diese Zellpfade besitzen einen gültigen Ursprung **oder** ihre Vorgängerzelle ist selbst ein Mitosekandidat. Weiterhin müssen sie genau zwei nachfolgende Zellpfade aufweisen; diese beiden Zellpfade haben wiederum nur einen Pfad als Vorgänger, den Mitosekandidaten selbst.
3. **Inkonsistente Pfade.** Diese Pfade sind keine Mitosekandidaten und haben keinen gültigen Ursprung und / oder kein gültiges Ende.

Liefert ein idealer Tracking-Algorithmus korrekte Pfade, so würde die Integritätsprüfung nur konsistente Pfade und Mitosekandidaten finden, wobei jeder Mitosekandidat einer Zellteilung entsprechen würde. In der nicht-idealen Wirklichkeit werden aber viele kritische Pfade gefunden, siehe Abb. 3.4.

Die Klassifikation der Zellfaden aus Tab. 2.2 in Kap. 2.4 erleichtert es, die gültigen und ungültigen Zellpfade zu finden. Die Bezeichnung der Pfadklassen und die Einteilung der Integritätsprüfung sind zwar ähnlich, sie sind aber nur identisch und austauschbar, wenn eine vollständige und fehlerfreie Genealogie vorliegt. Aus diesem Grund wird auf eine strikte Unterscheidung der Mitosen (siehe Tab. 2.2) und Mitosekandidaten geachtet.



**Abbildung 3.4.:** Die Konnektivität der Zellfaden ist als Graph dargestellt. Anhand der Verbindungen wird in konsistente (a) und inkonsistenten Zellpfade (b-d) klassifiziert. Hierbei helfen die Pfadklassifikationen, die z.B. angeben, wenn eine oder mehrere Zellen verschmolzen sind (b-d).



**Abbildung 3.5.:** Schema des Zelltracking. Das validierte Zelltracking nutzt die Integritätsprüfungen der Zellpfade und Datenstruktur, um fehlerhafte Pfade zu identifizieren. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g001.

### 3.2. CeTReS: Ein Benchmark für Zelltracking-Algorithmen

In dem bisherigen Kapitel wurde die Fehleranfälligkeit des Zelltracking diskutiert und die möglichen Fehlerquellen wurden detailliert beschrieben. Mit dem Genealogischen Modell und der Validierung wurde ein Konzept vorgestellt, welches das Verhalten der Zellen nutzt, um Detektions- und Tracking-Fehler zu finden bzw. aus der Datenanalyse auszuschließen.

Der nächste logische Schritt besteht nun darin, das gegebene Zelltracking zu analysieren und die Fehler systematisch zu untersuchen bzw. die Auswirkung der Validierung zu beurteilen. Diese Evaluierung der Tracking-Algorithmen setzt voraus, dass es einen Referenzdatensatz gibt, d.h. eine Bildserie, in der bekannt ist, wie viele Zellen in den jeweiligen Bildern gezeigt sind, zu welcher Familie sie gehören und welche Kontur bzw. Form sie haben.

In dem folgenden Abschnitt werden zwei validierte Referenzdatensätze vorgestellt, die mit Hilfe der Genealogischen Modelle und der daraus abgeleiteten Validierung erstellt werden konnten. Weiterhin werden einheitliche Maße zur Beurteilung und Evaluation von Zelldetektions- und Tracking-Algorithmen vorgestellt. Diese Maße werden genutzt, um die Algorithmen mit Hilfe des Referenzdatensatzes CeTReS.A zu evaluieren und diesen Datensatz vorzustellen. Sowohl die Referenzdaten als auch die Fehlermaße wurden in der Arbeit von Rapoport/Becker et al. [8] veröffentlicht und sind unter der Creative Commons Lizenz frei verfügbar <sup>1</sup>.

### 3.2.1. Cell Tracking Reference Set (CeTReS)

Für die Evaluation der Zelldetektion und des Zelltracking wurden zwei Bildserien von adulten Stammzellen aus dem Rattenpankreas genutzt, um Referenzdatensätze zu erstellen. Die erste Bildserie enthält 209 Schräglichtbilder, die im Abstand von 15 Minuten aufgenommen wurden. Der zweite Datensatz enthält 399 Bilder, die im Abstand von 10 Minuten aufgenommen wurden und zeigt jedes Bild sowohl als Schräglicht- als auch als Phasenkontrastbild. Eine detaillierte Beschreibung der Experimente ist in Abschnitt 2.3.1 auf Seite 19 gegeben. Mit Hilfe dieser Langzeitaufnahmen wurden Referenzdaten erstellt, in denen

1. jede Zellmaske  $z_i$  kontrolliert und gegebenenfalls manuell korrigiert wurde,
2. jede Zellmaske  $z_i$  einem Zellpfad  $p_j$  zugeordnet ist, der eine Zelle repräsentiert,
3. alle Verwandtschaften der Zellen als Konnektivität der Zellpfade korrigiert

gespeichert sind. Um die Referenzdaten zu erstellen wurde zuerst die vorgestellte Zelldetektion und das vorgestellte Überlapptracking aus Abschnitt 2.4 genutzt, um die initialen Zellpfade zu erstellen. Nach diesem initialen Tracking wurde eine Integritätsprüfung genutzt um eine Liste mit allen fehlerhaften Pfadfragmenten zu erstellen, die manuell korrigiert wurden. Um die Korrektur weiter zu erleichtern, wurde ein graphisches Hilfsprogramm entwickelt, das den Anwender bei der Navigation durch die Bilder und Zellpfade bzw. Fehler unterstützt (vgl. Abb.3.7(a)). Nachdem alle Pfade manuell korrigiert wurden, ist die Integritätsprüfung der Datenstruktur aufgerufen worden, um mögliche Fehler in der manuellen Korrektur zu finden. Die Korrektur und die Integritätsprüfungen sind so oft wiederholt worden, bis (fast) keine Inkonsistenzen in den betrachteten Daten enthalten waren. Es hat sich gezeigt, dass einige wenige Verschmelzungen von Zellen in den Daten enthalten sind, die auch mit einer manuellen Korrektur nicht mit einer hundertprozentigen Sicherheit korrigiert werden konnten. Exemplarisch ist in Abb. 3.6 ein solcher Fehler gezeigt: Die beobachtete Zelle teilt sich, die beiden Tochterzellen verschmelzen aber kurz nach der Teilung wieder und sind als eine einzelne Zelle zu erkennen. In den beiden erstellten Referenzdatensätze CeTReS.A bzw. CeTReS.B konnten 21 bzw. 10 optische Verschmelzungen der Zellen nicht aufgelöst werden.

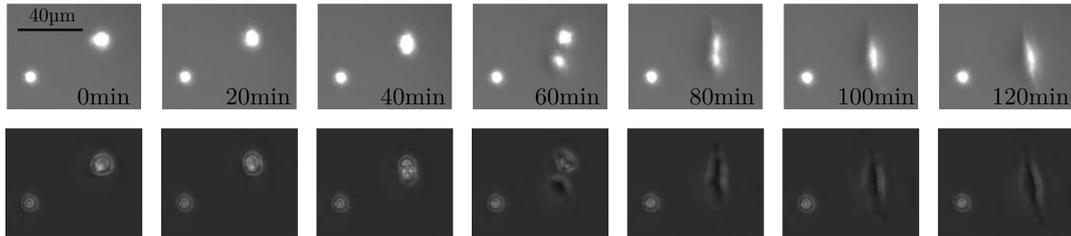
Bis auf diese Fehler ist ein vollständiger und konsistenter Datensatz entstanden, in dem jede Zelle bzw. jeder Zellpfad seinem Stammbaum zugeordnet ist. Insgesamt konnte der Arbeitsaufwand durch die Integritätsprüfungen erheblich reduziert werden, da

---

<sup>1</sup><ftp://cetres.emb.fraunhofer.de>

### 3.2. CeTReS: Ein Benchmark für Zelltracking-Algorithmen

die Anzahl der zu überprüfenden Positionen stark eingeschränkt wurde. Beispielsweise mussten für den ersten Datensatz CeTReS.A anstatt der 244.000 Zellmasken, die in der Bildserie enthalten sind, nur ca. 18.000 initiale Pfadfragmente überprüft werden. Der Zeitaufwand, der benötigt wurde, um die letzten 6000 Pfade zu korrigieren, ist in Abb.3.7(b) dargestellt. Eine Zusammenfassung der Referenzdaten ist in Tab. 3.1 gegeben.



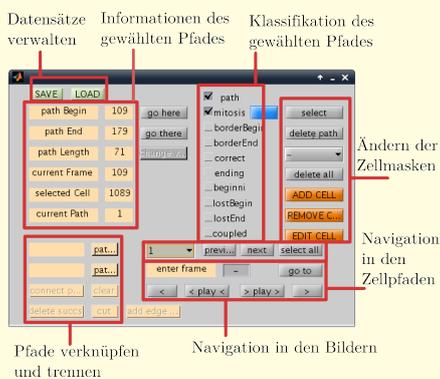
**Abbildung 3.6.:** Zellverschmelzung. Einige Inkonsistenzen in den Referenzdaten konnten auch mit einer manuellen Korrektur nicht aufgelöst werden. Die gezeigten Zellen scheinen nach der Zellteilung wieder zu fusionieren und eine (eindeutige) Aufteilung ist nicht möglich.

	CeTReS.A		CeTReS.B	
Bilder	1 × 209		2 × 399	
Aufnahmeintervall	15 Minuten		10 Minuten	
Zellen	rPSC		rPSC	
Kontraste	Schräglich		Schräglich +Phasenkontrast	
	Rohdaten	Referenz	Rohdaten	Referenz
Summe Zellkonturen	234974	244580	75333	80491
Pfade / Pfadfragmente	17952	4689	4134	1431
Zellen im ersten Bild	273	282	47	49
Zellen im letzten Bild	2134	2343	618	673

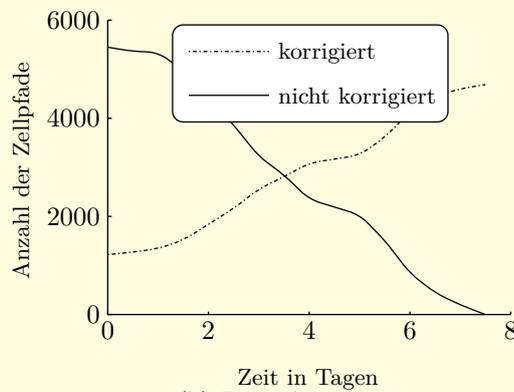
**Tabelle 3.1.:** Übersicht der Referenzdatensätze CeTReS.A und CeTReS.B. Es wurden zwei Bildserien verwendet um die Referenzdaten zu erstellen, die adulte Stammzellen zeigen, die aus dem Rattenpankreas abgeleitet wurden (rPSC). Die Anzahl der Zellkonturen und Zellpfade ist für beide Bildserien vor der Korrektur (Rohdaten) und für die korrigierten Referenzdaten (Referenz) gezeigt.

### Cell Tracking Reference Set (CeTReS)

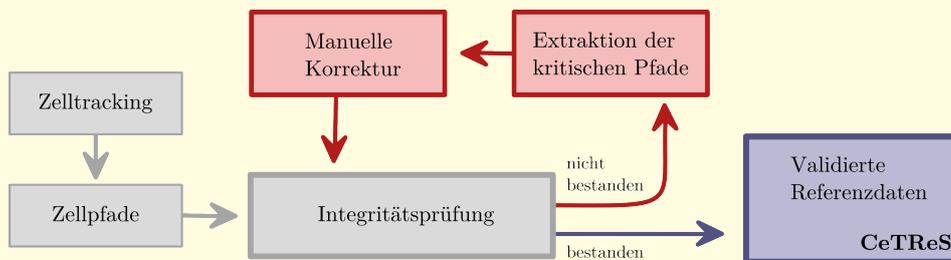
Die erstellten Referenzdatensätze CeTReS.A und CeTReS.B zeigen eine Population von rPSC in Passage 48 und 39. Die bearbeiteten Bildserien enthalten 209 und 399 Bilder, die zusammen  $\approx 320.000$  einzelne Zellen zeigen. Bei einer Fehldetektion von 5% ergaben sich geschätzte 16.000 Fehldetektionen, die gefunden und korrigiert werden mussten. Da eine manuelle Korrektur der einzelnen Masken in jedem einzelnen Bild sehr langwierig und fehleranfällig war, wurde die vorgestellte Datenvalidierung genutzt, um die Referenzdaten zu erstellen. Mit Hilfe der Integritätsprüfung der Pfad- und Datenstrukturen wurden alle Fehler als Inkonsistenzen extrahiert, so dass alle fehlerhaften Pfade bzw. Zellmasken direkt korrigiert werden konnten. Als Folge verkleinerte sich der Arbeitsaufwand erheblich. So war es möglich, die großen Datenmengen schnell zu verarbeiten und gleichzeitig einen validierten Referenzdatensatz zu erstellen.



(a) Grafisches Interface.

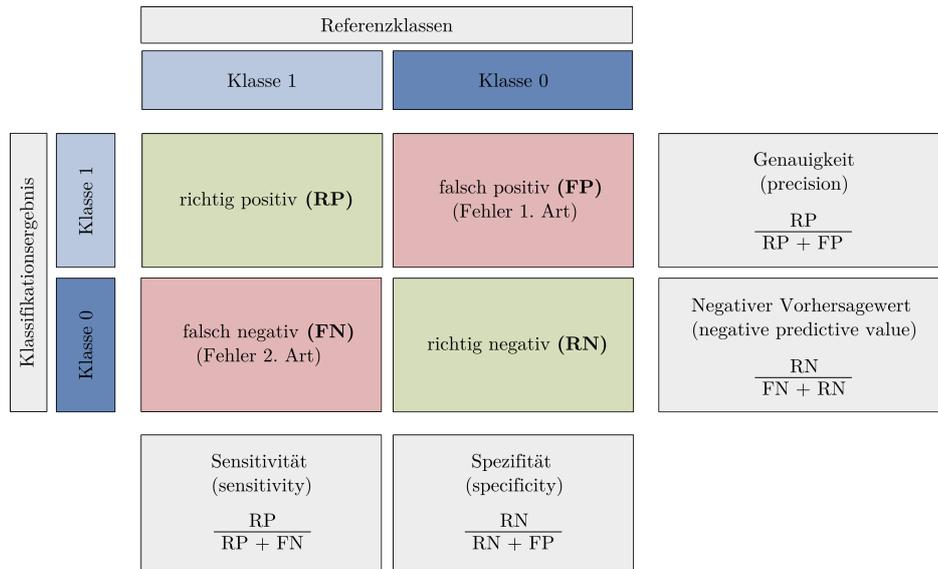


(b) Zeitaufwand.



(c) Schematische Darstellung der Korrekturschritte.

**Abbildung 3.7.:** Die Referenzdaten wurden mit Hilfe einer automatischen Fehlerdetektion erstellt, die auf der Validierung und den Integritätsprüfungen aufbaut (a). Um dem Anwender die Navigation durch die Daten und die Korrektur der Zellen zu erleichtern wurde ein grafisches Interface implementiert (b). So konnte der Arbeitsaufwand erheblich reduziert werden (c). Insgesamt wurden zwei validierte Referenzdatensätze (CeTReS.A und CeTReS.B) erstellt.



**Abbildung 3.8.:** Fehlermaße. Die Detektions- und Trackingfehler werden als Spezifität, Sensitivität und Genauigkeit angegeben. Die geläufigsten englischen Bezeichnungen sind angegeben; eine Übersicht alternativer Namen und Definitionen ist in Tab. 3.2 gegeben.

#### 3.2.2. Einheitliche Fehlermaße für Zelltracking-Algorithmen

Die Fehler der Zelldetektion und des Zelltracking werden mit Hilfe der Fehlermaße angegeben, die in Abb. 3.8 dargestellt sind. Als Grundlage für die verwendeten Fehler dienen die möglichen vier Ausgänge einer Klassifikation. Um diese zu veranschaulichen wird ein hypothetischer Detektor betrachtet, der für ein gegebenes Objekt entscheidet, ob es eine Zelle zeigt oder nicht:

1. Wenn eine Zelle gegeben war und erkannt wurde, hat der Detektor richtig entschieden (richtig positiv, RP).
2. Falls das Objekt jedoch keine Zelle war, handelt es sich um einen Fehler 1. Art (falsch positiv, FP).
3. Wird eine Zelle nicht erkannt, so handelt es sich um einen Fehler 2. Art (falsch negativ, FN).
4. Wird ein Objekt, welches keine Zelle ist, korrekt als Nicht-Zelle erkannt, so hat der Detektor wieder richtig entschieden (richtig negativ, RN).

Aufbauend auf diesen Ereignissen werden die verschiedenen Fehlermaße Sensitivität, Spezifität und Genauigkeit definiert und für die Bewertung der Detektions- und Tracking-Algorithmen genutzt.

Die Genauigkeiten der unterschiedlichen Detektionen und Klassifikatoren werden mit Hilfe der folgenden beiden Paare von Fehlermaßen angegeben:

1. **Detektion und Tracking.** Die Detektionsgenauigkeit der Zellen und die Rekonstruktionsrate der vollständigen Pfade wird mit Hilfe der **Sensitivität** und der

**Genauigkeit** angegeben, wobei in der Literatur oftmals die Bezeichnung *recall* und *precision* verwendet wird. Die Sensitivität gibt den Anteil der richtig detektierten Objekte (RP) in Bezug zu allen Objekten an, die erkannt werden könnten (RP + FN). Die Genauigkeit dagegen gibt den Anteil der richtig erkannten Objekten (RP) von allen erkannten Objekten (RP + FP) an. Diese beiden Maße sind unabhängig von der negativen Klasse (RN) definiert und eignen sich z.B. für den Vergleich der rekonstruierten vollständigen Pfade, die abhängig von den verwendeten Tracking-Algorithmen auf unterschiedlich großen Negativ-Klassen (unvollständige Pfade) berechnet werden.

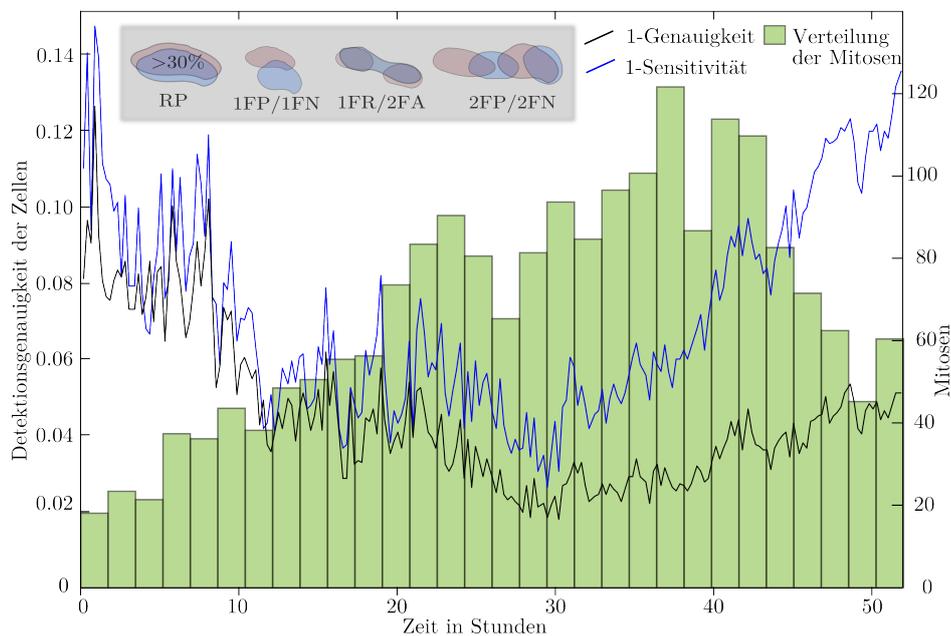
2. **Klassifikation.** Die Genauigkeit von Klassifikationsalgorithmen wird mit Hilfe der **Sensitivität** und der **Spezifität** angegeben; eine geläufige Bezeichnung dieser beiden Fehlermaße ist *sensitivity* und *specificity* oder *true positive rate* und *true negative rate*, siehe Tab. 3.2. Die Spezifität liefert den Anteil der richtig erkannten Negativklasse (RN) von der gesamten Negativklasse (RN+FP). Mit Hilfe dieser Fehlermaße ist es daher möglich, unterschiedliche Klassifikatoren und Detektoren zu vergleichen, wobei eine Grundvoraussetzung darin besteht, dass dieselben Referenz- oder Benchmarkdaten verwendet werden. Diese Daten müssen sowohl eine positive Klasse als auch eine negative Klasse enthalten. Diese Fehlermaße dienen als Grundlage für die Analyse der Mitosedetektionen, die in Kap. 4 vorgestellt werden.

Name	Definition	Alternative Bezeichnungen
<b>Sensitivität</b>	$\frac{RP}{RP+FN}$	Empfindlichkeit, Richtig-Positiv-Rate (RPR), sensitivity, recall, true positive rate (TPR)
<b>Spezifität</b>	$\frac{RN}{RN+FP}$	Richtig-Negativ-Rate (RNR) specificity, true negative rate (TNR)
<b>Genauigkeit</b>	$\frac{RP}{RP+FP}$	positiver Vorhersagewert, precision, positive predictive value (PPV)
<b>Negativer Vorhersagewert</b>	$\frac{RN}{RN+FN}$	negative predictive value (NPV)

**Tabelle 3.2.:** Alternative Bezeichnungen für die verwendeten Fehlermaße.

### Zelldetektion

Die Genauigkeit der Bildsegmentierung wird in jedem Bild als Fehler mit 1–Sensitivität und 1–Genauigkeit angegeben. Hierbei wird die Zellmaske einer Zelle aus dem Testbild  $I^T$  als korrekt erkannt (richtig positiv, RP), wenn sie im Referenzbild  $I^R$  mit nur einer einzigen Zellmaske überlappt. Die Referenzmaske darf sich wiederum nur mit der Zellmaske der Testzelle überschneiden. Zusätzlich wird der Überlapp der Zellkonturen als  $\frac{f_A^T \cap f_A^R}{f_A^T}$  und  $\frac{f_A^T \cap f_A^R}{f_A^R}$  definiert, wobei  $f_A^T$  und  $f_A^R$  die Fläche der Zellmasken aus den Test- und Referenzdaten angeben (vgl. Gl. (2.2)). Der Wert beider Überlappwerte muss größer als 0.3 sein, damit eine Zellmaske als korrekt erkannt wird. Abbildung 3.9 zeigt diese Fehlerraten für jedes Bild der Zeitserien.

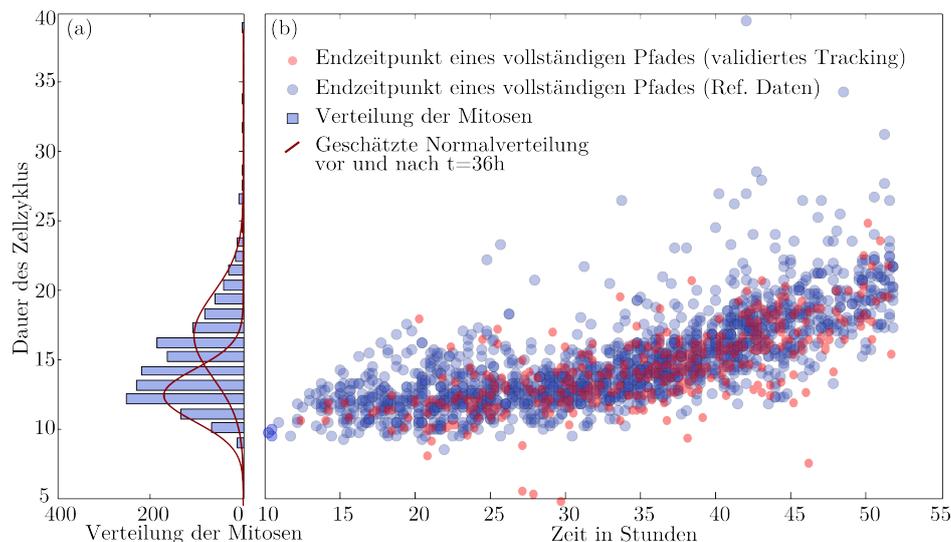


**Abbildung 3.9.:** Genauigkeit der Zelldetektion. Die gemittelte Fehlerrate der Zelldetektion liegt bei einem Fehler (1-Genauigkeit)  $< 5\%$  und (1-Sensitivität)  $< 8\%$ . Es ist aber eine zeitliche Änderung der Genauigkeit zu erkennen: Da die Zellen in den ersten Bildern noch sehr mobil sind, werden sie im ersten Drittel mit einem erhöhten Fehler gefunden. In der Phase des exponentiellen Wachstums wird die Zelldetektion mit einer sehr hohen Genauigkeit ausgeführt. An der Verteilung der Mitosen ist zu erkennen, dass sich in dieser Zeit viele Zellen teilen, d.h. dieser Zeitabschnitt ist bedeutend für die Rekonstruktion der zellulären Genealogien. Im letzten Teil der Zeitserie verschmelzen viele Zellkonturen durch die hohe Zelldichte, was zu einem erhöhten Fehleraufkommen führt. Daher ist der Anstieg der Fehler in dem letzten Drittel auf die hohe Zelldichte zurückzuführen. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g007.

## Zelltracking

Die vorgestellten Zelltracking-Algorithmen werden anhand der vollständigen Zellpfade evaluiert, die korrekt rekonstruiert werden können. Diese Pfade folgen einer Zelle von Mitose bis zu Mitose und sind in dem vorherigen Kapitel 2 in Abb. 2.20 veranschaulicht. Die Rekonstruktionsrate wird mit der Genauigkeit und der Sensitivität gemessen.

Um zu entscheiden, ob ein Zellpfad in den Testdaten einem Zellpfad in den Referenzdaten entspricht, wird ein Maß benötigt, das angibt, ob zwei Zellpfade identisch sind. Ein Referenz- und ein Testpfad werden dabei als identisch erkannt, wenn alle Zellmasken, die zu den Zellpfaden gehörten, richtig erkannt wurden. Hierbei werden dieselben Methoden genutzt, die zuvor für die Evaluation der Zelldetektion verwendet wurden. Die kritischen Punkte dieser Definition sind das Ende und der Anfang der Zellpfade, da die räumliche Trennung der Tochterzellen, die nach einer Mitose geboren werden, in den Mikroskopbildern oftmals nicht eindeutig bestimmt werden können. Ein Testpfad und ein Referenzpfad werden daher als identisch angesehen, wenn die Zellen der kleinsten gemeinsamen räumlichen Schnittmenge identisch sind. Eine weitere Möglichkeit besteht darin, die vollständigen Pfade über korrekt erkannte Mitosen zu definieren (vgl. Kap. 4). In Tab. 3.3 und Abb. 3.10 sind die Ergebnisse für das naive Tracking, die der initialen Pfadfragmente und des validierte Tracking für den Datensatz CeTRE.S.A zusammengefasst.



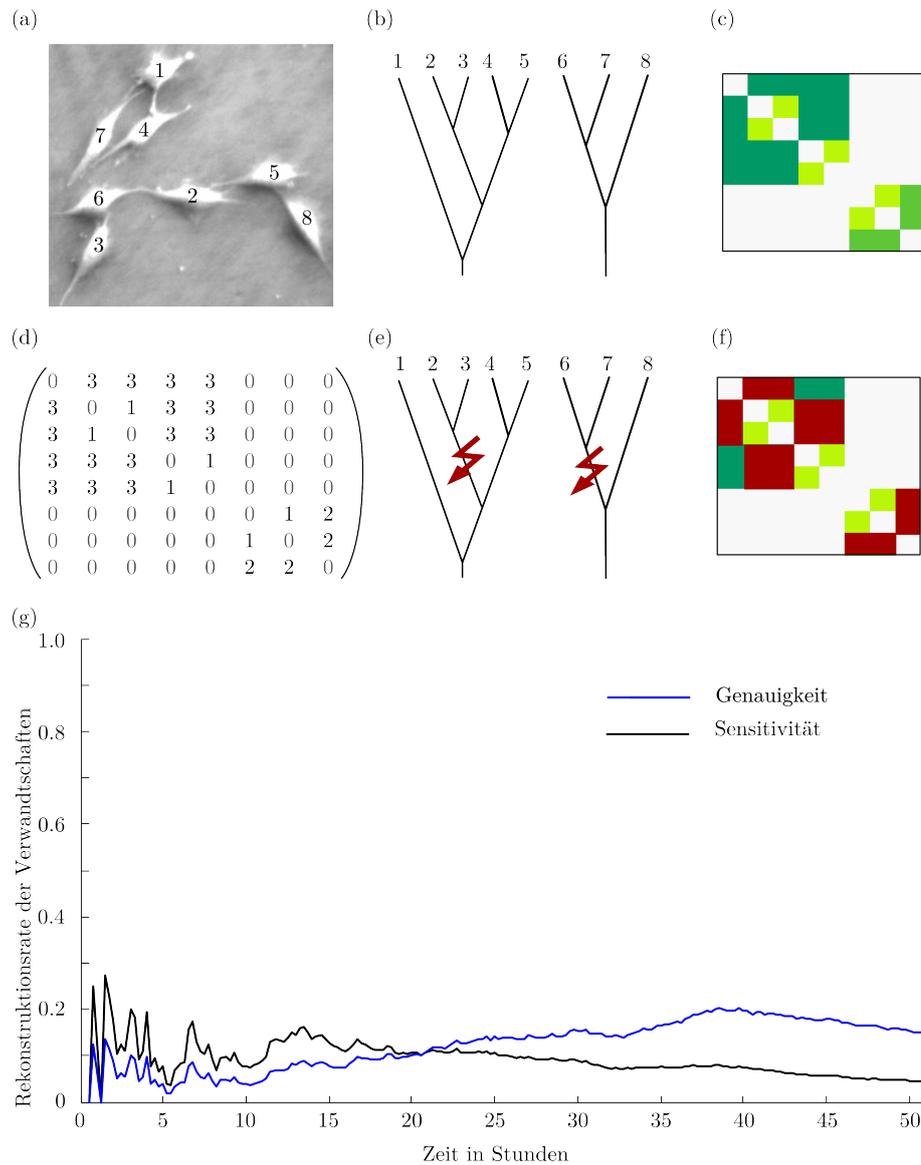
**Abbildung 3.10.:** Evaluation von Zelltracking-Ergebnissen. Ein gutes Maß für die Evaluation von Zelltracking-Algorithmen liefert die Genauigkeit, mit der vollständige Pfade rekonstruiert werden. Diese Pfade folgen einer Zelle über ihre gesamte Lebenszeit, d.h. von ihrer Geburtsmitose bis zu ihrer eigenen Zellteilung. Die vollständigen Pfade sind für die Referenzdaten und für die Ergebnisse des validierten Tracking angegeben. Hierbei gibt jeder Punkt den genauen Zeitpunkt und die Lebenszeit bzw. Zellzykluszeit einer Zelle an. Quelle (b): Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g011.

	Referenz	Naives Tracking	Validiertes Tracking
Zellpfade	4684	8704	11455
Vollständige Pfade	1635	1400	377
Korrekte vollständige Pfade	1635	789	360
Genauigkeit	1.0	0.61	0.95
Sensitivität	1.0	0.69	0.17

**Tabelle 3.3.:** Die erste Spalte zeigt die Referenzdaten CeTReS.A. In der zweiten sind die Ergebnisse des naiven Tracking aus Abschnitt 2.4.3 gezeigt und die dritte Spalte gibt die Resultate an, die mit dem validierten Tracking erzielt wurden. Es ist deutlich zu erkennen, dass die Validierung die Pfade mit einer wesentlich höheren Genauigkeit rekonstruiert, allerdings auch einige Zellpfade übersieht. **Anmerkung:** Die Ergebnisse des validierten Tracking wurden mit Hilfe der morphologischen Mitosedetektion aus Abschnitt 4.2 des folgenden Kapitels erstellt. Die Auswertung wurde vorgezogen, um die einheitlichen Fehlermaße zu veranschaulichen. Eine ausführlichere Analyse der Mitosedetektion folgt im nächsten Kapitel, in dem u.a. verschiedene Tracking-Ansätze und Mitosedetektionen mit dem zweiten Datensatz CeTReS.B verglichen werden (vgl. Abb. 4.16).

### Zelluläre Genealogie

Die rekonstruierten Genealogien der Zellen werden mit Hilfe der korrekten Verwandtschaftsgrade analysiert, die die Anzahl der Mitosen angeben, die zwischen zwei Zellen liegen. Es wird die Eigenschaft genutzt, dass zwei verwandte Zellen zu einem vorherigen Zeitpunkt einen gemeinsamen Vorfahren haben. Aus Sicht der Graphentheorie kann der Verwandtschaftsgrad für zwei Zellen in einem gegebenen Stammbaum über die Anzahl der Mitosen bestimmt werden, die auf dem kürzesten Pfad zwischen den beiden Zellen liegen. Nicht verwandte Zellen haben dabei einen Verwandtschaftsgrad von 0. In Abb. 3.11 ist dieses Maß veranschaulicht und es werden verschiedene Darstellungsmöglichkeiten gegeben. Dieses Maß verdeutlicht die starke Fehlerfortpflanzung in den Stammbäumen, da bereits zwei Fehler dazu führen, dass 61% der Verwandtschaftsgrade nicht korrekt angegeben werden können. In Abb. 3.11(g) sind die Fehler als Sensitivität  $\frac{RP}{RP+FN}$  und Genauigkeit  $\frac{RP}{RP+FP}$  für Datensatz CeTReS.A angegeben.



**Abbildung 3.11.:** Fehlermaß für zelluläre Genealogien. Der Verwandtschaftsgrad zweier Zellen gibt die Anzahl der Mitosen an, die beide Zellen verbindet. Die acht Zellen aus (a) bilden zwei Zellfamilien deren Verwandtschaft in (b) gezeigt ist. Die daraus resultierenden Verwandtschaftsgrade sind in (c) und (d) farbkodiert und als Matrix visualisiert. Die Auswirkung von einzelnen Fehlern ist in (e-f) dargestellt und verdeutlicht die starke Fehlerfortpflanzung. Die gesamte Rekonstruktionsrate des Zelltracking ist als Sensitivität  $\frac{RP}{RP+FN}$  und Genauigkeit  $\frac{RP}{RP+FP}$  in (g) für CeTReS.A gezeigt.

## 3.3. Der Referenzdatensatz CeTReS.A

Die automatisierte Auswertung von Zeitserien ermöglicht es, Zellkulturen mit biologischen Parametern zu charakterisieren. Im folgenden Abschnitt werden verschiedene biologische Eigenschaften angegeben, definiert und exemplarisch für den Referenzdatensatz CeTReS.A dargestellt. Die Parametrisierung umfasst dabei konventionelle Eigenschaften wie die Proliferation oder die Konfluenz, die eine valide Bildsegmentierung voraussetzen, aber auch komplexere Maße wie die Migration der Zellen, die anhand der Zellbewegung geschätzt wird. Ein wichtiger Parameter, der nur mit Hilfe von Zelltracking-Algorithmen extrahiert werden kann, ist die exakte Verteilung der Lebenszeiten bzw. der Zellzykluszeit. Diese können in dem Zelltracking-Framework als vollständige Pfade extrahiert werden (vgl. Abschnitt 2.4.3).

### 3.3.1. Proliferation und Morphologie

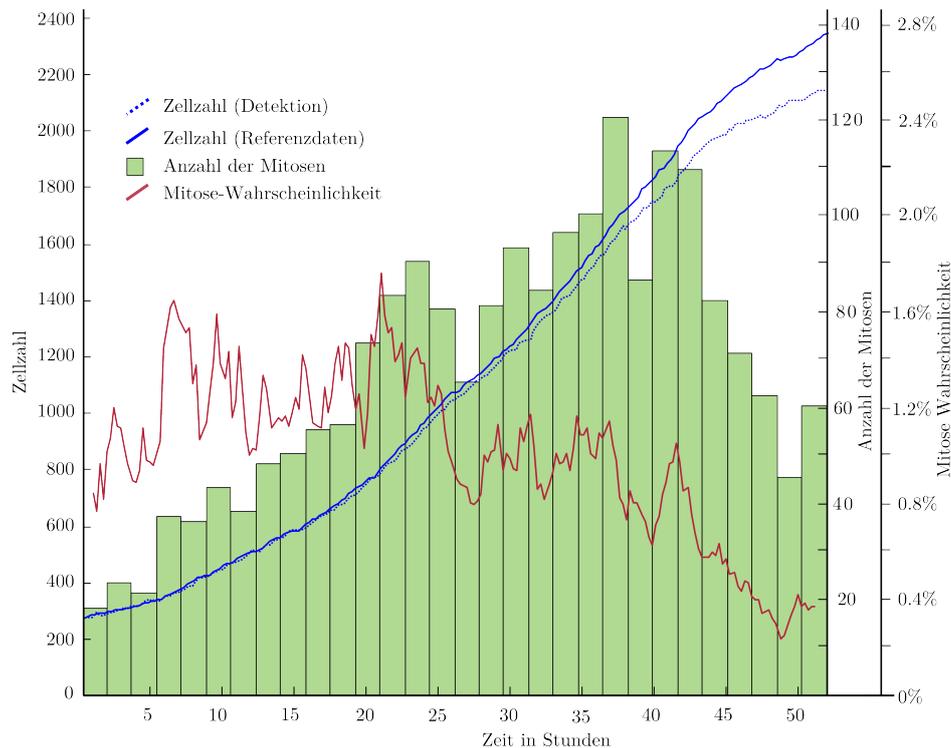
Ein wichtiger Parameter für *in vitro*-kultivierte Zellen ist die Wachstums- bzw. Proliferationskurve, die die Zunahme der Zellzahl über die Zeit beschreibt. Diese Daten können in einer gegebenen Zeitserie mit Hilfe der Zelldetektion erstellt werden. Hierzu wird die Zahl der detektierten Zellen in jedem Bild bestimmt. Es wird vorausgesetzt, dass der beobachtete Ausschnitt repräsentativ für die gesamte Zellkulturschale ist. Aus diesem Grund werden bei einem Langzeitexperiment vier oder mehr unterschiedliche Positionen beobachtet (vgl. Abschnitt 2.2.2). In Abb. 3.12 sind die Wachstumskurven für die automatische Auswertung und für die manuell korrigierten Referenzdaten gezeigt. Die automatische Zelldetektion zeigt dabei einen Fehler kleiner als 5%.

Die Genauigkeit der Zellzahlbestimmung ist sehr hoch und wird nur von invasiven Methoden zur Bestimmung der Zellzahl erreicht. Bei dieser Zählung werden die Zellkerne mit dem DAPI Farbstoff angefärbt, wodurch die Zellen getötet und nicht weiter verwendet werden können (vgl. Abschnitt 2.1.2). So kann die Zellzahl zwar sehr genau bestimmt werden, es ist aber nicht möglich, die Wachstumskurve einer einzelnen Zellpopulation zu bestimmen.

Als weitere nicht-invasive Methoden stehen impedanzbasierte Messungen zur Verfügung, mit denen die Änderung der Impedanz einer proliferierenden Zellpopulation angegeben werden (Atienza et al. [11], Ke et al. [54], Solly et al. [109]). Diese Methoden sind nicht in der Lage, die Änderung der Zellzahl von der Größenänderung der Zellen zu unterscheiden, was zu einer systematischen Ungenauigkeit führt. Dieses Problem wird im nächsten Abschnitt ersichtlich, wenn gezeigt wird, dass die Größenänderung der Zellen mit der Zelldichte korreliert ist.

#### Vollständige Wachstumskurve

Die Proliferation der Zellen in Abb. 3.12 zeigt die typische sigmoidale Form einer Wachstumskurve, in der alle Phasen des Zellwachstums aufgezeichnet wurden: Die initiale Lag-Phase, in der die Zellen kaum eine Proliferation zeigen (0-10h), gefolgt von der Phase des exponentiellen Wachstums (10-40h). Erreicht die Anzahl der Zellen eine kritische Dichte, so verlangsamt sich die Proliferation durch eine Kontaktinhibition der Zellen (Schmitz [101]). Diese letzte Phase wird als Stationäre- oder Plateauphase be-

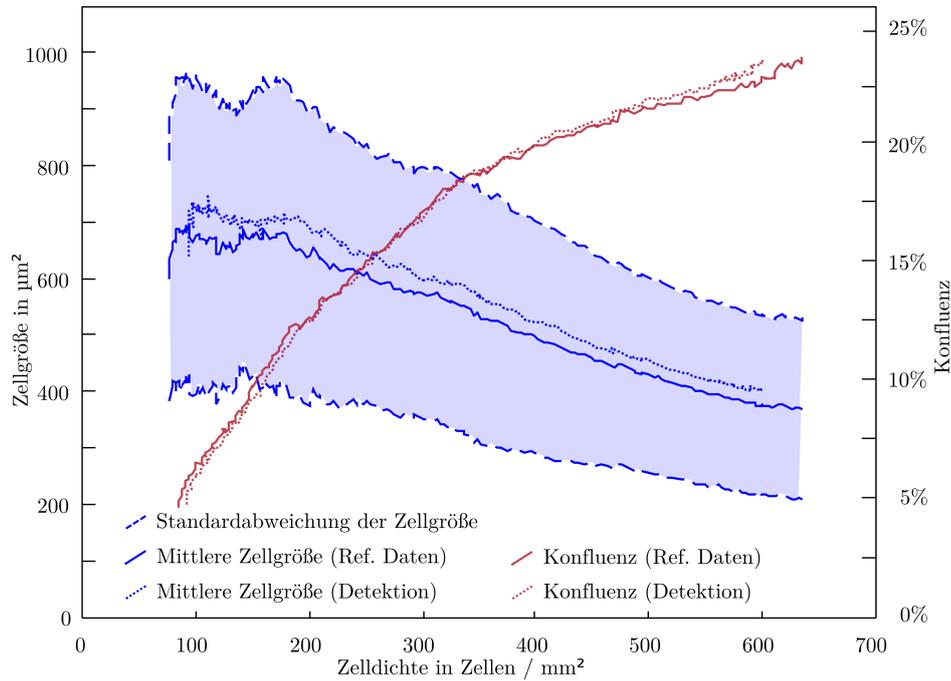


**Abbildung 3.12.:** Wachstumskurve. Die Zellzahl der aufgenommenen Zellpopulation verzehnfacht sich in dem Beobachtungszeitraum. Die Referenzdaten (durchgezogene Linie) und die automatisierte Auswertung (gestrichelte Linie) zeigen nur eine geringe Abweichung  $< 5\%$ . Die Verteilung der Mitosen (Histogramm) bzw. die Mitosehäufigkeit (rote Linie) korreliert mit dem Zellwachstum und verdeutlicht die Phasen des Wachstums. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g009.

zeichnet (40-52h). Die Verteilung der Mitosen in den Referenzdaten ist in Abb. 3.12 als Histogramm gezeigt. Mit Hilfe dieser Verteilung kann in jedem Bild die Mitosehäufigkeit  $p_M = \frac{N_{Mitose}}{N_{Zellen}}$  bestimmt werden. Hierbei gibt  $N_{Mitose}$  die Anzahl der Mitosen in einen Zeitraum von zwei Stunden an. Wie erwartet, steigt diese Wahrscheinlichkeit zuerst langsam bis zu einem Wert von 1.6% an und nimmt zum Ende der Phase des exponentiellen Wachstums ab und sinkt bis auf 0.4%. Diese starke Abnahme der Mitosehäufigkeit in den letzten zehn Stunden ist eine direkte Auswirkung der Kontaktinhibition der Zellen.

### Zellgröße und Konfluenz

Ein weiterer Standardparameter in der Zellkultur ist die Zelldichte oder Konfluenz, die den Bedeckungsgrad der Kulturschale mit Zellen angibt. In Abbildung 3.13 ist die Konfluenz als Funktion der Zelldichte gezeigt; durch die Kontaktinhibition sinkt die Zellgröße mit steigender Zellzahl. Dieser Effekt bzw. der Verlauf der Größenänderung kann nur aufgezeichnet werden, wenn mit Hilfe einer zuverlässigen Zelldetektion die Größe der Zellen bestimmt wird.



**Abbildung 3.13.:** Zelldichte. Die Wachstumskurven (Proliferationskurven) und die Größenänderung der Zellen sind für die Referenzdaten (durchgezogene Linie) und für eine automatisierte Detektion (gestichelte Linie) gezeigt. Die Konfluenz ist als Bedeckungsgrad der Petrischale in Prozent angegeben. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g010.

**Anmerkung zu der Berechnung der Zelldichte.** In der Zellkultur bzw. im Labor wird die Konfluenz einer Petrischale in der Regel subjektiv bestimmt, indem die Zellen unter dem Mikroskop betrachtet werden, d.h. es fehlt ein objektives Maß für den Bedeckungsgrad. Die Konfluenz liefert z.B. ein Maß dafür, ob eine Zellpopulation passagiert werden kann (vgl. Abschnitt 2.1). Hier wird die Konfluenz der Petrischale oftmals manuell am Mikroskop geschätzt und eine Zellzahlbestimmung erfolgt erst, nachdem die Zellen trypsiniert, d.h. geerntet wurden.

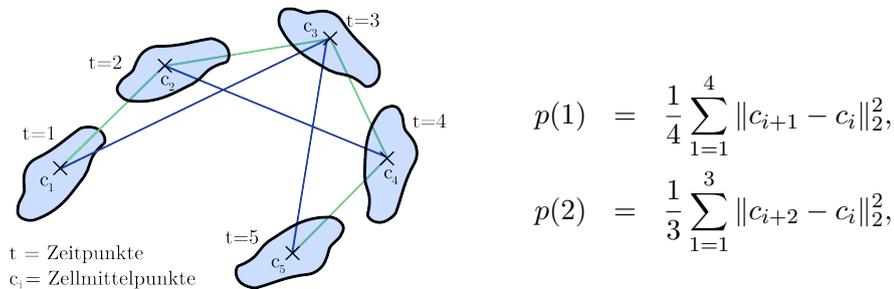
Mit der Zelldetektion ist es dagegen möglich, mit dem Bedeckungsgrad ein objektives Maß für die Zelldichte anzugeben. Der Bedeckungsgrad ist in Abb. 3.13 angegeben und wird als  $\frac{\sum_{i=1}^N f_A^i}{S}$  definiert, wobei  $N$  die Zahl der Zellen und  $f_A^i$  die Größe der Zellen und  $S$  die Fläche des Zellbildes angibt. Dieser objektive Wert liegt immer unter der manuellen geschätzten Konfluenz: Abb. 3.13 zeigt eine Zellpopulation, die vollgewachsen ist und in der das Zellwachstum durch die Kontaktinhibierung gebremst bzw. gestoppt wird. Nach manueller Einschätzung liegt die Konfluenz also bei 100%, das definierte objektive Maß liefert allerdings einen maximale Wert von  $\approx 25\%$ .

### 3.3.2. Migration

Mit Hilfe der Zellpfade können der räumliche Weg der Zellen rekonstruiert und die Bewegung der Zellen geschätzt werden. Eine entscheidende Frage bei der Definition der

Bewegung ist die Bestimmung der exakten Position einer Zelle, da es zwei wesentliche Faktoren gibt, die eine Änderung und Verschiebung der Zellen bewirken. Zum einen kann sich eine Zelle mehr oder weniger gerichtet bewegen (Lo et al. [74]). Diese Bewegung wird als Migration bezeichnet und führt zu einer linearen Verschiebung der Mittelpunkte über die Zeit (Stephens [110]). Der zweite Faktor, der die Position verändert, ist die Verformung der Zellkontur. Zeigt eine Zelle keine gerichtete Bewegung, so würde die Änderung der Zellform zu einer zufälligen Verschiebung des Zellmittelpunktes führen. Diese zufälligen Bewegungen können mit dem mathematischen Modell des Randomwalks beschrieben und analysiert werden (Berg [13]).

In den betrachteten Bildserien ist ein Zeitabstand zwischen zwei Bildern von 10 bis 15 Minuten gewählt worden. Diese Zeitspanne ist sehr klein im Vergleich zu der Strecke, die eine Zelle in dieser Zeitspanne zurücklegen kann. Das führt dazu, dass die Bewegung und die Verformung der Zellen in den gegebenen Bildserien in derselben Größenordnung von  $1 - 2\mu\text{m}$  liegen.



**Abbildung 3.14.:** Bewegung und Migration der Zellen. Die Zellmigration wird als mittlere quadratische Verschiebung der Zellschwerpunkte berechnet (*mean square displacement*, MSD). Dieser Parameter erlaubt es, eine gerichtete Bewegung von einer ungerichteten Bewegung zu unterscheiden. In (a) sind die Strecken über einen (grün) bzw. zwei Zeitpunkte (blau) dargestellt, die für die Berechnung der MSD  $p(1)$  (grün) bzw.  $p(2)$  (blau) genutzt werden, wie in der Formel auf der rechten Seite angegeben wird.

### Mittlere quadratische Verschiebung (mean square displacement (MSD))

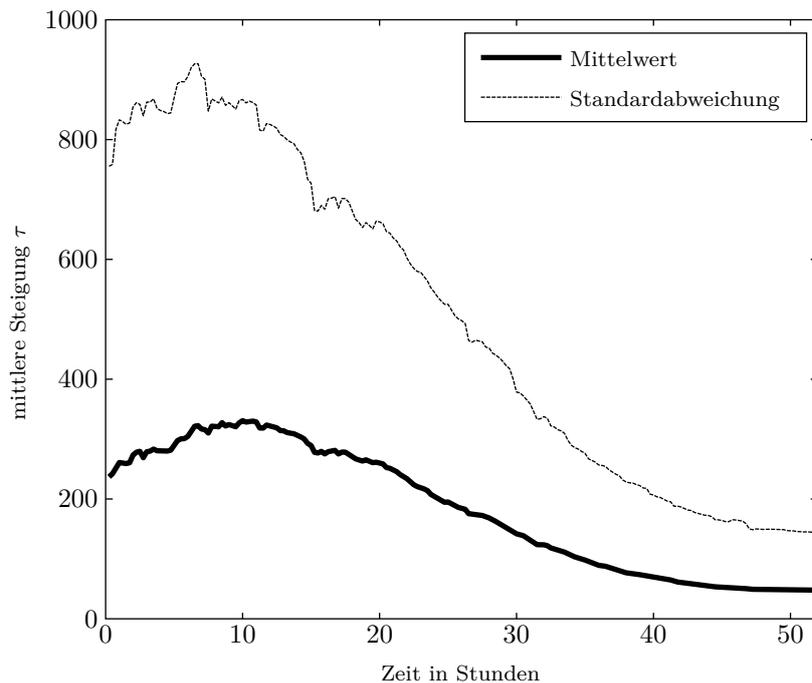
Um den Unterschied einer gerichteten oder einer ungerichteten Zellbewegung zu untersuchen, wird die mittlere quadratische Verschiebung genutzt (*mean square displacement* (MSD)). Die MSD ist ein Maß für die Diffusion einer Bewegung und gibt die mittlere quadrierte Strecke an, die eine Zelle über einen bestimmten Zeitraum zurücklegt (Qian et al. [90]). Für einen gegebenen Zellpfad mit  $N$  Zellen und zugehörigen Mittelpunkten  $c_1, \dots, c_N$  ist die MSD als

$$p(n) = \frac{1}{N-n} \sum_{i=1}^{N-n} \|c_{i+n} - c_i\|_2^2, \quad n = 1, \dots, N-1 \quad (3.1)$$

definiert. Zwei Eigenschaften der MSD-Kurven beschreiben die Beweglichkeit der Zellen. Die Form der Kurve kann genutzt werden, um Rückschlüsse auf das Verhalten zu ziehen:

Während eine gerichtete Zellbewegung zu einer linear ansteigenden MSD führt, zeigt eine zufällige Zellbewegung eine gekrümmte Form. Bei einer gerichteten Bewegung kann der Diffusionskoeffizient  $\tau$  als Steigung der MSD angegeben werden, der eine Aussage über die Migrationsgeschwindigkeit zulässt (Qian et al. [90], Takagi et al. [112]).

In Abb. 3.15 sind die Bewegungsprofile aller Zellpfade als gemittelte Diffusionskoeffizienten  $\tau$  angegeben. Diese Abbildung zeigt, dass die Beweglichkeit der Zellen mit wachsender Zellzahl abnimmt. Die Analyse der Migration und Zellbewegung ermöglicht weitere Ansätze, auf die in dieser Arbeit nicht weiter eingegangen wird (Mehrayin et al. [77], Mitchison and Cramer [80], Takagi et al. [112, 112]).



**Abbildung 3.15.:** Zellbewegung. Die Bewegung der Zellen ist als mittlerer Diffusionskoeffizient  $\tau$  angegeben. Während die zurückgelegte Strecke der Zellen in den ersten zehn Stunden noch zunimmt, führt die wachsende Zellzahl ab Stunde zehn zu einer eingeschränkten Beweglichkeit. Zum Ende der Beobachtung zeigen die Zellen die geringste Migration. Die Standardabweichung zeigt die große Inhomogenität der Zellbewegung.

### Zellzykluszeiten

Eine Beschreibung der Zellpopulation die nur mit Zelltracking-Algorithmen extrahiert werden kann und von Rapoport/Becker et al. [8] eingeführt wurde, ist die Verteilung der Zellzykluszeit. In dem vorgestellten Framework werden die Lebenszeiten der Zellen als vollständige Pfade beschrieben und extrahiert (vgl. 2.4.3). Um die Bedeutung dieses Parameters zu veranschaulichen, wird der hypothetische Fall betrachtet, dass eine Zellpopulation mit identischen Zellen beobachtet wird. Das angenommene symmetrische Teilungsverhalten führt dazu, dass die Zellzykluszeit aller Zellen identisch ist.

Die Heterogenität einer Zellpopulation kann daher direkt mit der Verteilung und der Varianz der Lebenszeit der Zellen angegeben werden. Dieser Parameter ist charakteristisch für (adulte) Stammzellen, da die Differenzierung und das Teilungsverhalten direkt verknüpft sind: Es wird angenommen, dass sich eine Stammzelle langsam teilt und in der Lage ist, Vorgänger- oder Progenitorzellen zu bilden, die eine stärkere Proliferation zeigen. Nach einigen Generationen werden ausdifferenzierte Zellen gebildet, die keine proliferierenden Eigenschaften mehr besitzen (Dingli et al. [31], Watt and Jensen [118]).

Die mit Hilfe des validierten Zelltracking extrahierten Zellzykluszeiten sind in der Abb. 3.16 dargestellt. Die Auswertungen zeigen drei verschiedene Experimente: Es wurden die Referenzdatensätze CeTReS.A, CeTReS.B und ein weiteres Langzeitexperiment ausgewertet, in dem eine Zellpopulation an zehn unterschiedlichen Positionen beobachtet wurde. Alle drei Experimente wurden unter den gleichen Kulturbedingungen ausgeführt, aber sowohl die Anfangskonfluenz als auch Aufnahmedauer unterschieden sich. Eine ausführliche Beschreibung der Experimente bzw. resultierenden Datensätze ist im vorherigen Kapitel 2 in Abschnitt 2.3.1 gegeben und die jeweils ersten Bilder der Zeitserien sind in Abb. 2.6 gezeigt.

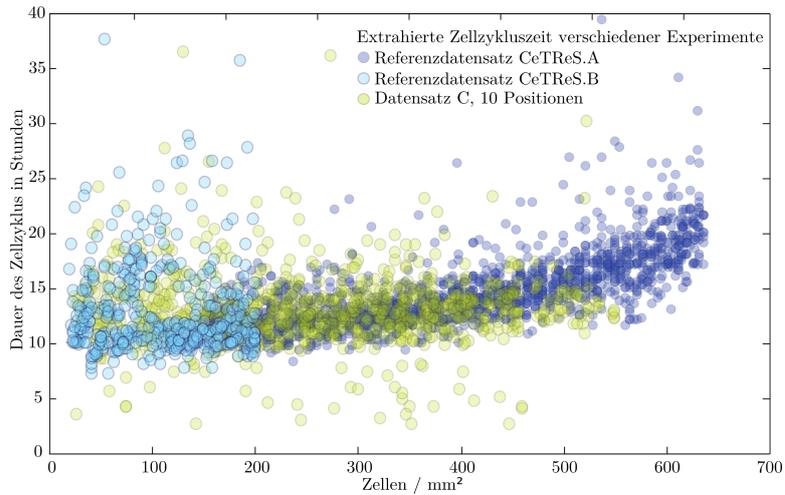
Die Abb. 3.16 zeigt, dass die Lebenszeiten der Zellen in allen drei Experimenten stark mit der vorhandenen Konfluenz korrelierten, d.h. dass die Zellen bei dergleichen Zelldichte identische Lebenszeit besitzen. Diese Eigenschaft ist durch die Kontaktinhibition der RPC zu erklären, die dazu führt, dass die Zellen bei einer hohen Zelldichte keine Proliferation mehr zeigen. Der typische Charakter einer Stammzelle, der eine asymmetrische Teilung beschreibt, ist in diesen Zellen nicht zu finden.

### 3.3.3. Genealogien

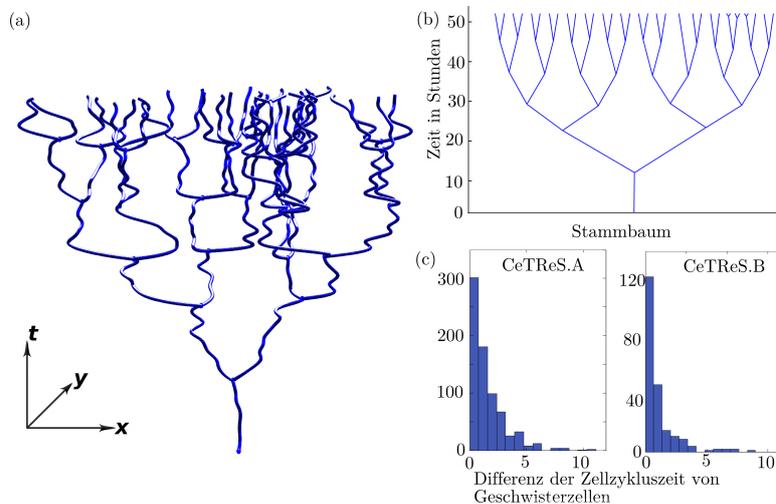
Die schwierigste Aufgabe des Zelltracking ist die Rekonstruktion vollständiger Genealogien, die nur möglich ist, wenn alle Zellen bzw. Zellpfade der Stammbäume korrekt erkannt werden. Um diese Bäume analysieren, unterscheiden und nach spezifischen Mustern, wie z.B. einer Zelldifferenzierung, suchen zu können, wurden verschiedene topologische Maße beschrieben (Al-Kofahi et al. [9], Braun et al. [14], Glauche et al. [36], Roeder et al. [97]). Der Nachteil dieser Parameter ist, dass sie nur für große Bäume angegeben werden können, die eine Zellfamilie über mindestens sieben oder acht Generationen beschreiben. Zu diesem Zeitpunkt hat eine einzelne Zelle bereits hunderte von Nachkommen erzeugt, was es erschwert, einen vollständigen Stammbaum in der beobachteten Zellpopulation aufzuzeichnen. Zum einen ist das Beobachtungsgebiet durch die Bildgröße und der Auflösung der verwendeten Kamera eingeschränkt, zum anderen ist die Dauer eines Experimentes begrenzt (vgl. Abschnitt 2.2.2).

In den ausgewerteten Aufnahmen werden kleinere Stammbäume beobachtet und eine Zelle wird nur über wenige Generationen verfolgt. Typischerweise wird eine konfluente Zellpopulation im Verhältnis 1:2 bis 1:10 aufgeteilt, was zwei bis vier Generationen entspricht (Schmitz [101]). Daher können die oben erwähnten Maße nicht auf die extrahierten Bäume angewendet werden. Stattdessen wird ein einfaches Maß genutzt, um die Stammbäume zu beschreiben: Die Teilungssymmetrie.

### 3.3. Der Referenzdatensatz CeTReS.A



**Abbildung 3.16.:** Vergleich der Zellzykluszeiten verschiedener Experimente. Die Verteilung der Zellzykluszeit ist als Länge der vollständigen Pfade für drei verschiedene Langzeitbeobachtungen gezeigt. In den Experimenten wurden Zellpopulationen mit unterschiedlichen Zelldichten im Startbild beobachtet und mit dem validierten Zelltracking ausgewertet. Das Bild zeigt die Verteilung der Referenzdaten CeTReS.A, CeTReS.B und die Ergebnisse des validierten, automatisierten Zelltracking von einem weiteren Datensatz. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g013.



**Abbildung 3.17.:** Genealogien. Die Referenzdaten CeTReS.A und CeTReS.B enthalten 386 bzw. 85 Stammbäume mit einer durchschnittlichen Tiefe von 3.5 bzw. 2.8 Generationen. In (a) ist der größte Stammbaum in  $xyt$ -Koordinaten angegeben. Dieselbe Genealogie ist in (b) als Binärbaum gezeigt. Die hohe Symmetrie der Teilungen aus (a) und (b) kann mit Hilfe der Teilungssymmetrie (c) statistisch untersucht werden. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.g012.

Diese gibt die Differenz der Lebenszeit von zwei Zellen an, die aus einer gemeinsamen Mitose entstanden sind. Der Wert kann für alle Schwesterzellen einer Zellpopulation angegeben und analysiert werden. Je niedriger die Differenzen sind, desto symmetrischer teilen sich die Zellen.

In Abb. 3.17 sind verschiedene Darstellungen der genealogischen Informationen der Referenzdaten CeTReS.A gezeigt. Das Zelltracking ermöglicht es, den exakten spatio-temporalen Weg der Zellen zu rekonstruieren (vgl. Abb. 3.17(a)). Diese Daten erlauben eine Analyse der räumlichen Verteilung, der Umgebung und der Zell-Zellkontakte. Die Symmetrie und die Topologie der rekonstruierten Pfade kann sowohl als Binärbaum wie in Abb. 3.17(b) oder als Teilungssymmetrie 3.17(c) dargestellt werden. In der letztgenannten Abbildung ist die Verteilung dieser Symmetrien für die Datensätze CeTReS.A und CeTReS.B gezeigt.

### 3.3.4. Zusammenfassung Zelltracking

Die vorgestellten Ergebnisse sind in der gemeinsamen Arbeit von Rapoport/Becker et al. [8] publiziert worden. Die veröffentlichten Daten sind in diesem dritten Kapitel um die explizite Darstellung des Genealogischen Modells in Abschnitt 3.1.2 und eine ausführlichere Beschreibung der Zellmigration aus Abschnitt 3.3.2 erweitert worden. Die wichtigsten Aussagen sind in der folgenden Liste nochmals zusammengefasst:

1. Die Validierung, die mit Hilfe des Genealogischen Modells durchgeführt wird, liefert die entscheidende Grundlage für eine automatisierte Auswertung von Zelltracking-Ergebnissen.
2. Mit den CeTReS-Daten ist der erste, öffentliche Referenzdatensatz für Zelltracking-Algorithmen veröffentlicht worden.
3. Mit dem Fehlermaß, das einen Zelltracking-Algorithmus anhand der extrahierten vollständigen Pfade evaluiert, ist eine einheitliche Beurteilung von verschiedenen Algorithmen und Datensätzen möglich.

In der bisherigen Arbeit wurde die Datenvalidierung ausführlich beschrieben und es wurden einheitliche Fehlermaße für Zelltracking-Algorithmen eingeführt. Da die genaue Auswirkung der Datenvalidierung direkt von der Mitosedetektion abhängt, werden in dem nächsten Kapitel verschiedene Methoden vorgestellt, mit denen die gesuchten Zellteilungen zuverlässig identifiziert werden können. Mit Hilfe dieser Mitosedetektion wird die genaue Auswirkung der Validierung und der Mitosedetektion am Ende des folgenden Kapitels in Abschnitt 4.4 anhand der rekonstruierten vollständigen Pfade analysiert.

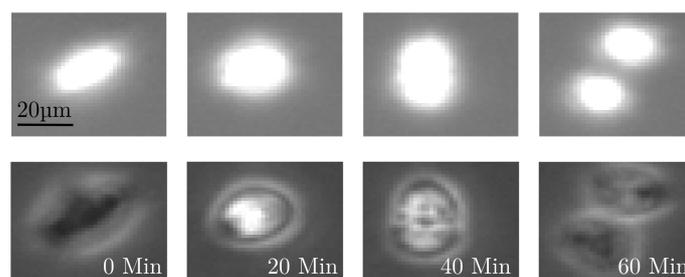
# 4

## Mitosedetektion

Das wichtigste Ereignis in den beobachteten, proliferierenden *in vitro*-Zellpopulationen ist die Mitose, die den biologischen Prozess der Fortpflanzung bzw. der Zellteilung beschreibt. Da sich eine Zelle in zwei eigenständige Nachfolger teilt, wird der Pfad der Mutterzelle durch diese Mitose beendet und die beiden Pfade der Tochterzellen werden begonnen. Somit definieren die Mitosen die Start- und Endpunkte der vollständigen Pfade. Diese wiederum sind die Bausteine, aus denen die Stammbäume und Genealogien zusammengesetzt werden.

Um ein verlässliches Zelltracking durchzuführen, ist es daher erforderlich, die Mitosen in einer Bildserie mit möglichst hoher Genauigkeit zu identifizieren ([27, 42, 45, 48, 66, 69, 73, 76, 122]).

Ziel dieses Kapitels ist es, das Zelltracking aus Kapitel 2 und die Validierung aus Kapitel 3 um eine Mitosedetektion zu erweitern. Hierzu werden zwei Ansätze vorgestellt, die entweder die Mitosen in den Schräglicht- oder Phasenkontrastbildern finden. Die Detektion in den Schräglichtbildern nutzt die vorher erstellten Zellpfade, um eine Mitose an der Änderung der Zellmorphologie zu erkennen. In dem zweiten Ansatz wird nach dem charakteristischen Muster der Mitosen in den Phasenkontrastbildern gesucht (vgl. Abb. 4.1). Die vorgestellten Mitosedetektionen wurden in den Arbeiten von Becker and Madany Mamlouk [2], Becker et al. [3, 4] und Rapoport/Becker et al. [8] veröffentlicht.



**Abbildung 4.1.:** Eine Zellteilung ist in den Schräglichtbildern (obere Reihe) durch eine morphologische Änderung charakterisiert: Die Zelle verlangsamt ihre Bewegung und wird kleiner und runder. Dieselbe Zelle im Phasenkontrast (untere Reihe) zeigt eine deutliche Änderung der Helligkeit, die durch die physikalischen Eigenschaften bzw. die Dicke des Zellkörpers verursacht wird. Für eine bessere Darstellung wurde die Helligkeit der Bilder in der unteren Reihe erhöht. Quelle: Becker et al. [3, Fig.1].

## 4.1. Analyse von Detektions- und Klassifikationsproblemen

Bevor die morphologische Mitosedetektion in den Schräglichtbildern in Abschnitt 4.2 und in den Phasenkontrastbildern in Abschnitt 4.3 vorgestellt wird, wird zunächst eine allgemeine Formulierung der Mitosedetektion als Klassifikationsproblem betrachtet. Ausgangspunkt ist ein möglicher Mitosekandidat  $x$ , der in den Schräglichtbildern durch einen Zellpfad oder in den Phasenkontrastbildern durch einen Bildausschnitt gegeben ist. Für diesen Kandidaten  $x$  wird die Mitosedetektion als Klassifikationsproblem

$$K(x) = \begin{cases} \text{Mitose} & \text{falls } p(x) > \theta, \\ \text{Nicht - Mitose} & \text{sonst.} \end{cases} \quad (4.1)$$

formuliert, bei dem der Klassifikator  $K$  für jeden Mitosekandidaten  $x$  entscheidet, ob er eine Zellteilung zeigt oder nicht. Hier wird abhängig von einem Schwellwert  $\theta$  entschieden, welcher Klasse ein gegebener Kandidat angehört. Aufbauend auf dieser Formulierung werden die verwendeten Fehlermaße definiert und veranschaulicht, mit der die Klassifikations- bzw. Detektionsgenauigkeit analysiert wird. Anschließend folgt eine kurze Übersicht der verwendeten Methoden des Maschinellen Lernens, die zum Teil genutzt werden, um die Klassifikatoren zu trainieren.

### 4.1.1. Fehlermaße zur Beurteilung von Klassifikatoren

Die Genauigkeit der Mitosedetektion wird mit den Fehlermaßen analysiert, die in Abschnitt 3.2.2 bzw. in Abb. 3.8 für die Evaluation der Zelldetektion eingeführt wurden. Diese werden noch einmal zusammengefasst und an die vorgestellten Klassifikationsprobleme angepasst.

Um die Klassifikatoren zu evaluieren, wird die Sensitivität  $= \frac{RP}{RP+FN}$ <sup>1</sup> und das Fehlermaß  $1 - \text{Spezifität} = 1 - \frac{RN}{RN+FP} = \frac{FP}{RN+FP}$  betrachtet (vgl. Tab. 3.2). Hierbei gibt  $1 - \text{Spezifität}$  das Fehlermaß an, wie viele der erkannten Mitosen falsch als Mitose erkannt wurden. Die Sensitivität dagegen gibt den Anteil von allen Mitosen an, die korrekt gefunden wurden. Um nun das Verhalten von einem Klassifikator  $K(x)$  zu untersuchen, werden diese Fehlerraten abhängig von einem Schwellwert  $\theta$  betrachtet. Bei  $\theta = 0$  werden alle Kandidaten  $x$  als Mitose erkannt ( $FN = 0$ ), aber es werden alle Nicht-Mitosen falsch klassifiziert ( $RN = 0$ ) und es folgt  $1 - \text{Spezifität} = 1$ , Sensitivität  $= 1$ . Bei einem Schwellwert  $\theta = \infty$  wird keine Mitose korrekt erkannt ( $RP = 0$ ), aber alle Nicht-Mitosen werden korrekt klassifiziert ( $FP = 0$ ) und es gilt  $1 - \text{Spezifität} = 0$ , Sensitivität  $= 0$ . Wird der Schwellwert  $\theta$  von 0 bis auf  $\infty$  angehoben und die Werte für  $1 - \text{Spezifität}$  und der Sensitivität gegeneinander aufgetragen, so ergibt sich für jeden Klassifikator eine charakteristische Kurve, die als Receiver Operating Characteristic(ROC)-Kurve bezeichnet wird (vgl. Abb. 4.2).

Der charakteristische Parameter dieser Kurve ist die **Area Under the Curve** (AUC), der das Integral bzw. die Fläche unter der Kurve angibt. Bei einem optimalen Klassifikator beträgt der Wert der AUC 1; der schlechteste Klassifikator, der mit einer Wahrscheinlichkeit von 50% richtig rät, erreicht eine AUC von  $\frac{1}{2}$ . Anhand des

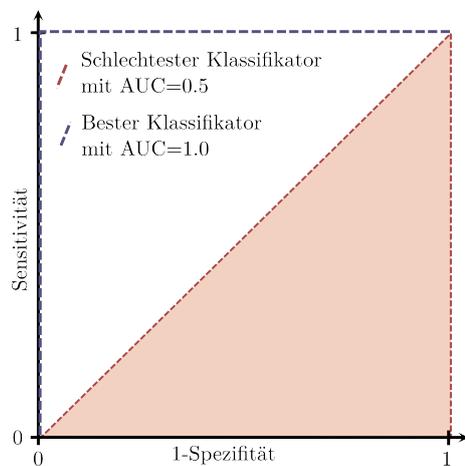
<sup>1</sup>RP = richtig positiv, FP = falsch positiv, FN = falsch negativ, RN = richtig negativ

AUC-Wertes und der ROC-Kurven werden die Mitosedetektionen verglichen. Die angegebenen Werte wurden mit der MATLAB-Funktion `perfcurve.m` erstellt.

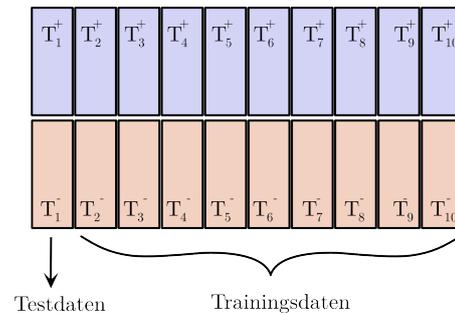
### Maschinelles Lernen von Klassifikatoren

Um die gesuchten Klassifikatoren zu erzeugen, wird auf verschiedene Methoden des maschinellen Lernens zurückgegriffen, mit denen einem Computerprogramm beigebracht werden kann, was eine Mitose von einer Nicht-Mitose unterscheidet. Ausgangspunkt dieser Methoden ist ein repräsentativer Trainingsdatensatz, der Beispiele der Mitosen und Nicht-Mitosen enthält. Nun wird ein Computerprogramm genutzt, um die Unterschiede der beiden gegebenen Klassen zu untersuchen und die Merkmale zu extrahieren, in denen sich die Klassen am leichtesten unterscheiden lassen. Diese Merkmale werden gespeichert und können fortan genutzt werden, um neue Daten zu klassifizieren, d.h. für ein gegebenes Bild wird entschieden, ob es eine Mitose oder eine Nicht-Mitose zeigt. Dieses Verfahren wird als überwachtes Lernverfahren bezeichnet, mit dem ein Klassifikator  $K$  erstellt werden kann.

Die Genauigkeit von einem trainierten Klassifikator  $K$  wird mit Hilfe einer zehnfach-Kreuzvalidierung untersucht. Hierbei werden die gegebenen Trainingsdaten  $T$  in zehn gleichgroße, disjunkte Mengen  $T_1, \dots, T_{10}$  unterteilt (vgl. Abb. 4.2(b)). In zehn Durchläufen wird der Klassifikator  $K$  mit den Daten  $T \setminus T_i, i = 1, \dots, 10$  trainiert, d.h. es wird jeweils eine Gruppe aus den Trainingsdaten entfernt und mit den übrigen neun Gruppen wird gelernt. Die ausgelassenen Daten  $T_i$  werden als Testdaten genutzt, um die Detektionsgenauigkeit zu bestimmen.

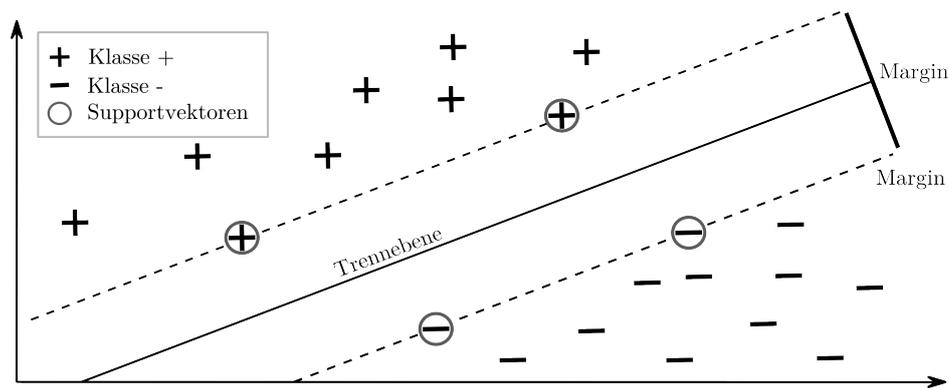


(a) ROC-Kurve und AUC.



(b) Schema der Kreuzvalidierung.

**Abbildung 4.2.:** Die Charakteristik eines Klassifikators wird als ROC-Kurve dargestellt, in der die Spezifität gegen die Sensitivität aufgetragen wird (a). Die Genauigkeit der Klassifikation wird als AUC angegeben, die mit einer zehnfach-Kreuzvalidierung bestimmt wird (b). In diesem Beispiel bezeichnen  $T^+, T^-$  die verschiedenen Klassen.



**Abbildung 4.3.:** Für ein gegebenes Klassifikationsproblem mit zwei Klassen  $T^+$ ,  $T^-$  bzw. +, -, berechnet eine Support Vector Machine (SVM) eine optimale Trennebene, die den minimalen Margin (Abstand der Datenpunkte zu der Trennebene) maximiert.

#### 4.1.2. Klassifikation mit einer Support Vector Machine (SVM)

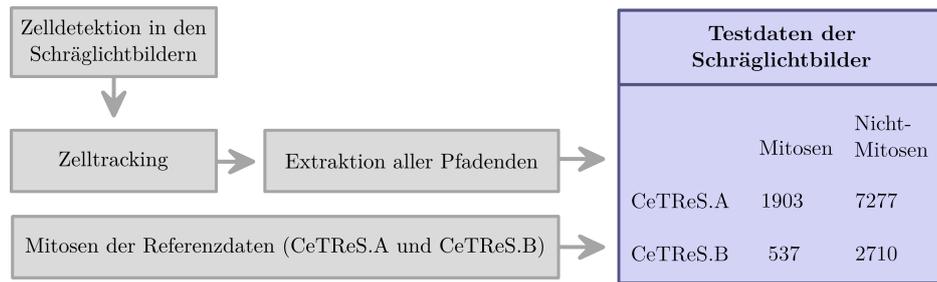
Eine Support Vector Machine (SVM) ist ein überwachtes Lernverfahren des Maschinellen Lernens, mit dem ein Zwei-Klassen-Problem gelöst wird (Cortes and Vapnik [23]). Für die Datenpunkte der beiden Klassen  $T^+$ ,  $T^-$  wird eine optimale (Hyper-) Ebene bestimmt, die beide Klassen voneinander trennt und den minimalen Margin maximiert, d.h. dass die kürzeste Distanz der Datenpunkte zu der Trennebene maximiert wird (vgl. Abb. 4.3). Eine SVM kann um eine Kernelfunktion erweitert werden, mit der die Daten in einen höher dimensional Raum projiziert werden (Schölkopf and Smola [103]). Diese Erweiterung ermöglicht es, Klassifikationsprobleme für nicht linear separierbare Daten zu lösen. Im Folgenden werden drei verschiedene Kernel genutzt: ein linearer Kernel (L-SVM), ein quadratischer Kernel (Q-SVM) und ein Gauss-Kernel (G-SVM).

## 4.2. Mitosedetektion in den Schräglichtbildern

Die charakteristischen Merkmale einer Zellteilung sind in Abb. 4.1 gezeigt: Vor einer Mitose verlangsamt sich die Bewegung einer Zelle und sie wird kleiner und runder, d.h. sie ändert ihre Morphologie. Die Ursache dieser Veränderung liegt in dem biologischen Prozess, der während einer Zellteilung abläuft. Eine intrazelluläre Struktur, die als Spindelapparat bezeichnet wird, zieht die Zelle zusammen, um die replizierte DNA der beiden Tochterzellen zu trennen (Campbell et al. [17, Kap. 12]). Aufbauend auf diesem Muster einer Mitose werden verschiedene Methoden vorgestellt, um die Zellteilungen zu detektieren (Abb. 4.1). Hierzu wird ein Klassifikator implementiert, der die Änderung der Zellmorphologie direkt aus den Zellpfaden extrahiert, d.h. dass die Mitosedetektion direkt von den Ergebnissen des Zelltracking abhängig ist. Die Aufgabe besteht darin, für jeden gegebenen Pfad zu entscheiden, ob er in einer Mitose endet oder nicht.

Mit dem eingeführten Klassifikator werden verschiedene morphologische Parameter genutzt, um eine Mitose zu detektieren. Weiterhin wird analysiert, welche Auswirkung die Änderung der Zelldichte in den verschiedenen Wachstumsphasen auf die Mitosedetektion hat (vgl. Zusammenfassung auf Seite 70). Zunächst werden jedoch die Referenz-

## 4.2. Mitosedetektion in den Schräglichtbildern



**Abbildung 4.4.:** Die Mitosedetektion in den Schräglichtbildern basiert auf den Änderungen der Zellmorphologie, die aus den Zellpfaden extrahiert werden. Als Testdaten werden die Zellpfade bzw. Pfadenden genutzt, die mit dem Zelltracking aus Abschnitt 2.4 erstellt wurden und mit Hilfe der Referenzdaten CeTReS.A und CeTReS.B als Mitose oder Nicht-Mitose eingeteilt wurden.

daten CeTReS.A und CeTReS.B genutzt, um einen Satz von Referenzmitosen für die Evaluation der morphologischen Mitosedetektion zu erstellen.

### 4.2.1. Testdaten der Schräglichtbilder

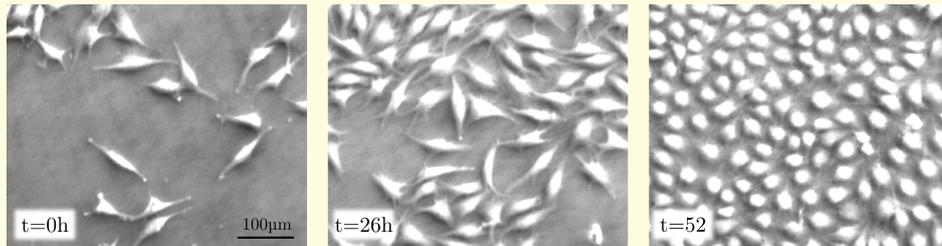
Die Mitosedetektion wurde mit den Zellteilungen aus den Referenzdaten CeTReS evaluiert (vgl. Abschnitt 3.2.1 und Rapoport/Becker et al. [8]). Diese stellen zwei gelabelte Bildserien zur Verfügung, die im Schräglichtkontrast aufgenommen wurden (CeTReS.A + CeTReS.B). Die Bildserie CeTReS.B enthält zusätzlich Phasenkontrastaufnahmen, da hier eine doppelte Kontrasteinstellung verwendet wurde (vgl. Abschnitt 2.2.2).

#### Schräglichtbilder (CeTReS.A + CeTReS.B)

Um die morphologischen Änderungen einer Zelle zu extrahieren, werden die Konturen von aufeinanderfolgenden Zellen genutzt, die in den Zellpfaden gespeichert sind. Somit ist diese Mitosedetektion direkt von dem zugrundeliegenden Zelltracking abhängig. Die erstellten Zellpfade können daher als Ausgangspunkt für die verwendeten Testdaten genutzt werden: Für jeden Pfad wird entschieden, ob er in einer Zellteilung endet oder nicht. Um Referenzdaten für die morphologische Mitosedetektion zu generieren, wurde der Zelltracking-Algorithmus aus Abschnitt 2.4 genutzt, um in den Bildserien von CeTReS.A und CeTReS.B initiale Zellpfade zu erstellen. Mit Hilfe der gelabelten Mitosen wurden diese Zellpfade anschließend als Mitose oder Nicht-Mitose klassifiziert (vgl. Abb. 4.4). Zu den 2019 Mitosen, die in CeTReS.A enthalten sind, wurden 1903 korrespondierende Zellpfade gefunden. Die fehlenden 116 Zellteilungen sind nicht in den Testdaten enthalten, weil die zugehörigen mitotischen Zellen nicht korrekt erkannt werden konnten, was detailliert in der Zusammenfassung auf Seite 77 beschrieben wird. Die Anzahl der nicht-mitotischen Pfade liegt bei 7277. Für Referenzdatensatz CeTReS.B ergaben sich analog 537 mitotische und 2710 nicht-mitotische Zellpfade. In diesen Daten wurden 36 Mitosen nicht erkannt, was einem Fehler von 6.3% entspricht.

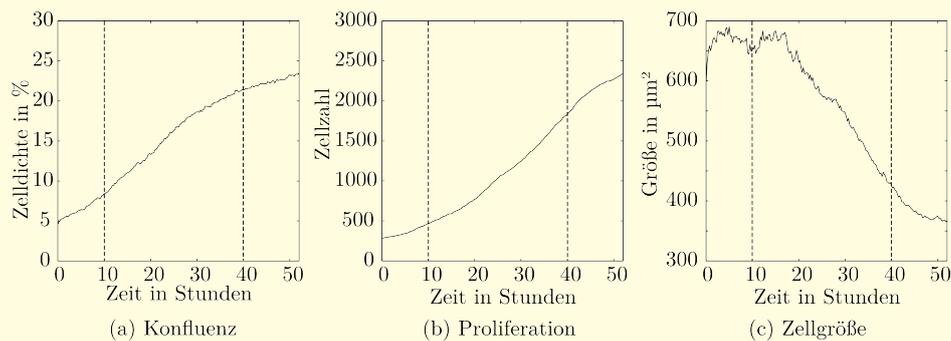
### Zeitliche Änderung der Zellmorphologie

Eine charakteristische Eigenschaft der beobachteten Zellpopulationen ist die Änderung ihrer Zellmorphologie über die Zeit: Die Größe bzw. Form der Zellen korreliert stark mit der Zelldichte und ändert sich im Laufe eines Langzeitexperimentes. Diese Änderung beeinflusst das typische morphologische Muster einer Zellteilung, da sich z.B. die länglichen Zellen zu Beginn einer Zeitserie stärker zusammenziehen als die Zellen zum Ende der Zeitserie (vgl. Abb. 4.5).



**Abbildung 4.5.:** Änderung der Zellmorphologie über die Zeit. Die Bilder zeigen einen Ausschnitt der Bildserie aus dem Datensatz CeTReS.A, Quelle: Becker and Madany Mamlouk [2, Fig. 1].

Der Referenzdatensatz CeTReS.A liefert einen guten Ausgangspunkt, um diese Änderungen zu untersuchen, da die beobachtete Zellpopulation alle Stadien einer Wachstumskurve durchläuft. So ist es möglich, das Verhalten derselben Zellpopulation bei unterschiedlichen Zelldichten zu analysieren. Die Zelldichte, die Anzahl der Zellen und die Fläche der Zellen ist in Abb. 4.6 gegeben. Im Folgenden wird die Zeitserie in drei Wachstumsphasen 0 – 10h, 10 – 40h und 40 – 52h eingeteilt, um den Einfluss der morphologischen Änderung auf die Mitosedetektion zu untersuchen.



**Abbildung 4.6.:** Änderung der Zellparameter von Datensatz CeTReS.A über die Zeit, Quelle: Becker and Madany Mamlouk [2, Fig. 2]. Die zeitlichen Grenzen der Wachstumsphasen sind als gestrichelte Linien eingezeichnet.

### 4.2.2. Morphologische Mitosedetektion

In dem folgenden Abschnitt werden verschiedene Klassifikatoren vorgestellt, die eine Zellteilung anhand der morphologischen Änderung der Zellkonturen finden können. Die Verformung der Zellkontur wird mit Hilfe der morphologischen Eigenschaften einer Zelle berechnet, die in Abschnitt 2.4 definiert wurden. In dem Vektor  $f^t = [f_x, f_y, f_A, f_H, f_K, f_L, f_O]$  werden Position, Fläche, Helligkeit, Kompaktheit, Länge und Orientierung einer Zelle zum Zeitpunkt  $t$  gespeichert. Mit Hilfe dieses Vektors wird die morphologische Charakterisierung eines gesamten Zellpfades  $\mathbf{p}_i$  der Länge  $N$  als  $\mathbf{f}_i = \{f^{t_1}, \dots, f^{t_N}\}$  gespeichert, wobei  $f^{t_1}, \dots, f^{t_N}$  die Eigenschaften der Zellen aus Pfad  $\mathbf{p}_i$  angeben. Da die Position und die Orientierung einer Zelle für die Mitosedetektion unerheblich sind, wird der reduzierte Vektor  $\tilde{f}^t = [f_A, f_H, f_K, f_L]$  genutzt, um den Differenzenvektor

$$\mathbf{d}_i = \{\Delta \tilde{f}^{t_1}, \dots, \Delta \tilde{f}^{t_N}\} \text{ mit } \Delta \tilde{f}^{t_j} = \tilde{f}^{t_{j+1}} - \tilde{f}^{t_j} = [\Delta f_A, \Delta f_H, \Delta f_K, \Delta f_L], j = 1, \dots, N - 1$$

zu berechnen. Hier gibt  $\Delta \tilde{f}^{t_j}$  die Änderung in einem Zeitschritt an und  $N$  die Länge des Zellpfades. Für jeden gegebenen Zellpfad  $\mathbf{p}_i$  wird ein Differenzenvektor  $\mathbf{d}_i$  berechnet.

Die Grundlage der nun folgenden Mitosedetektion ist die Änderung der Merkmale in  $\tilde{f}^t$ . Um diese messbar zu machen, wird die Verteilung der Änderungen der Größe  $\Delta f_A$ , der Intensität  $\Delta f_H$ , der Kompaktheit  $\Delta f_K$  und der Länge  $\Delta f_L$  als mehrdimensionale Normalverteilung mit der Kovarianz  $\Sigma \in \mathbb{R}^{4 \times 4}$  und dem Mittelwert  $\mu \in \mathbb{R}^4$  geschätzt. Diese Verteilung wird anhand aller fehlerfreien Pfade ermittelt, die automatisiert aus den Zelltracking-Ergebnissen extrahiert werden. Somit ist kein separater Trainingsschritt notwendig. Mit diesen Parametern wird die Wahrscheinlichkeit

$$p_m(\mathbf{d}_i) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} e^{\left(\frac{-(\mathbf{d}_i - \mu)\Sigma^{-1}(\mathbf{d}_i - \mu)}{2}\right)} \quad (4.2)$$

angegeben, die als Grundlage für die folgenden Klassifikatoren dient. Die folgende Liste liefert einen Überblick über die implementierten, morphologischen Mitosedetektionen.

- **Naive Detektion**  $K_Y$ . Der naive Klassifikator  $K_Y$  detektiert jeden Zellpfad  $\mathbf{p}_i$  als Mitose, der zwei Nachfolger hat. Bei einer fehlerlosen Zelldetektion und einem fehlerlosen Zelltracking ist dieser Klassifikator in der Lage, alle Mitosen zu extrahieren, d.h. wenn die Annahmen des naiven Tracking erfüllt wären, würde die naive Mitosedetektion eine vollständige Genealogie extrahieren.
- **Single Feature Maximum Likelihood** ( $K_{sfml}$ ). Für jeden gegebenen Pfad  $\mathbf{p}_i$  kann mit Hilfe der Differenzenvektoren  $\mathbf{d}_i$  ein Klassifikator  $K_{sfml}$  angegeben werden, der mit

$$K_{sfml}(\mathbf{p}_i) := \mathbf{p}_i \in \begin{cases} \text{Mitose} & \text{falls } \min_t p_m(\mathbf{d}_t) < \theta, \\ & \text{für } t = N - 5, \dots, N - 1 \\ \text{Nicht - Mitose} & \text{sonst,} \end{cases} \quad (4.3)$$

entscheidet, ob der Pfad  $\mathbf{p}_i$  in einer Zellteilung endet, d.h. es wird eine einzelne, morphologische Eigenschaft genutzt, um die Wahrscheinlichkeit für eine Mitose in jedem der letzten fünf Zeitschritte zu berechnen. Diese fünf Zeitschritte sind notwendig, da das mitotische Muster sichtbar wird, bevor sich eine Zelle in zwei separate Tochterzellen teilt und der Zellpfad endet.

- **(3-Phase) Multi Feature Maximum Likelihood ( $K_{mfml}$  und  $K_{3p-mfml}$ ).** Der Klassifikator  $K_{mfml}(\mathbf{p}_i, \omega)$  nutzt dieselbe Klassifikation aus Gl. (4.3), die mit dem Klassifikator  $K_{sfml}$  eingeführt wurde. Der Unterschied zwischen den beiden Versionen ist die Anzahl der verwendeten Eigenschaften: Während der  $K_{sfml}$  nur ein einzelnes Feature nutzt, kann der  $K_{mfml}$  mit  $\omega \in [0, 1]^4$  eine von insgesamt 15 statischen Kombinationen auswählen, die in Tab. 4.1 aufgelistet sind. In einer zweiten Version wird eine adaptive Version verwendet, die in jeder der drei Wachstumsphasen eine andere Kombination auswählt. Dieser adaptive Klassifikator wird als 3-Phasen adaptierter  $K_{mfml}$  bzw. als  $K_{3p-mfml}$  bezeichnet.
- **Linear- / Quadratic- / Gaussian Support Vector Machine ( $K_{morph-svm}$ ).** In einem weiteren Ansatz wurden die gelabelten Zellpfade genutzt, um mit einer SVM das Muster einer Mitose bzw. den Unterschied zwischen einer Mitose und einer Nicht-Mitose zu lernen. Hierzu wurde von allen 1903 Testmitosen die Differenz  $\Delta \tilde{f}^{tj} \in \mathbb{R}^4$  berechnet. Somit enthält die Mitose-Klasse  $T^{mitose}$  1903 Vektoren. Die Klasse der Nicht-Mitosen wird als  $T^{zelle}$  bezeichnet und wurde mit den 7277 Zellpfaden erstellt, die nicht mit einer Mitose enden. Für jeden dieser Pfade wurden die letzten fünf Differenzen  $\Delta \tilde{f}^{zelle} \in \mathbb{R}^4$  berechnet. Anschließend wurde für jeden Pfad die Differenz mit der kleinsten Wahrscheinlichkeit ausgewählt, d.h. es wurden die Differenzen gewählt, die die größte Ähnlichkeit zu einer Mitose zeigen. Die beiden Klassen  $T^{zelle}, T^{mitose}$  wurden genutzt, um eine SVM mit einem linearen ( $K_{morph-lsvm}$ ), einem quadratischen ( $K_{morph-qsvm}$ ) und einem Gauss-Kernel ( $K_{morph-gsvm}$ ) zu trainieren. Anschließend wird mit  $K_{morph-svm}(\tilde{f}_{kand}) > 0$  entschieden, ob  $\tilde{f}_{kand}$  eine Mitose zeigt oder nicht. Es wurde für jede Wachstumsphase ( $0h - 10h, 10h - 40h, 40h - Ende$ ) ein separater Klassifikator berechnet. In dem Lernschritt wurde ein Soft-Margin genutzt, d.h. es wurden minimale Fehlklassifikationen in den Trainingsdaten zugelassen. Vor dem Trainingsschritt wurden die Daten normalisiert, indem sie vom Mittelwert befreit und auf Einheitsvarianz skaliert wurden.

### 4.2.3. Evaluation der morphologischen Mitosedetektionen

Für die Auswertung und Analyse der morphologischen Mitosedetektion wurde der Referenzdatensatz CeTRE.S.A genutzt. Die vollständig aufgezeichnete Wachstumskurve

ermöglicht es, den Einfluss der verschiedenen Parameter auf die Mitosedetektion detailliert zu untersuchen (vgl. Abb. 4.5 und Abb. 4.6). Für die Evaluation wurden die erstellten Referenzmitosen verwendet, die aus 1903 Mitosen und 7277 Nicht-Mitosen bestehen (vgl. Abb. 4.4).

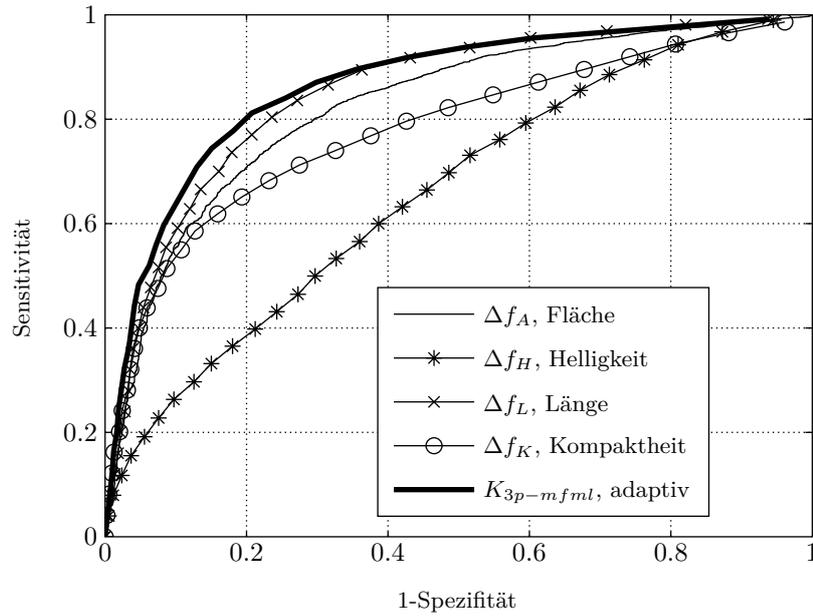
### Auswertung für den $K_{sfml}$ und $K_{mfml}$ Klassifikator

Die Detektionsgenauigkeit des  $K_{sfml}$  Klassifikators ist für die verschiedenen Parameter in Abb. 4.7 als ROC-Kurve gezeigt. Hier ist deutlich zu erkennen, dass die besten Ergebnisse mit der Änderung der Zelllänge erzielt werden. Das schlechteste Detektionsergebnis wurde mit der Änderung der Helligkeit der Zellen erreicht. Mit den Parametern Länge, Größe und Kompaktheit werden die Mitosen mit unterschiedlichen Genauigkeiten gefunden, obwohl diese Parameter sehr ähnliche Eigenschaften der Zellkontur beschreiben: Die Kompaktheit wird z.B. aus der Fläche berechnet. Eine Übersicht der Detektionsgenauigkeit aller Parameterkombinationen der  $K_{sfml}$  und  $K_{mfml}$  Klassifikatoren ist in Tab. 4.1 aufgelistet. Insgesamt zeigt sich, dass die Länge der verlässlichste Parameter ist, um eine Mitose zu detektieren.

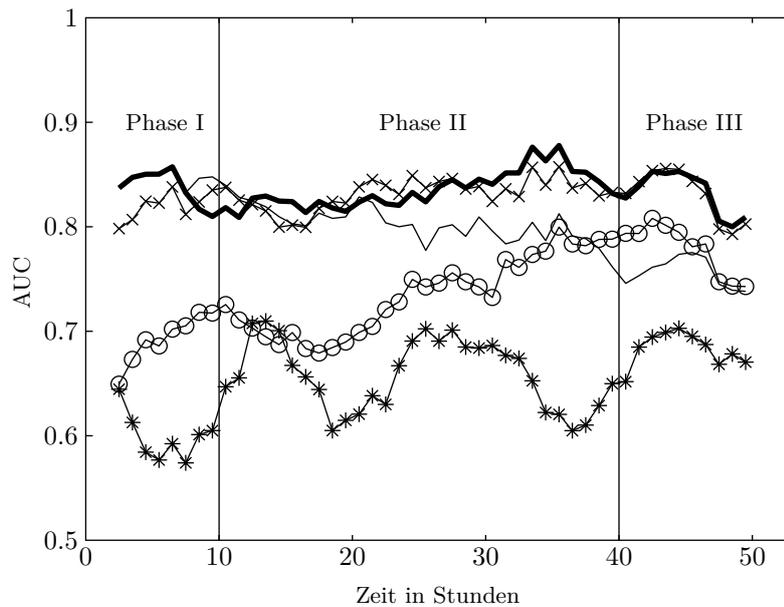
In Abb. 4.7 wird die Detektionsgenauigkeit der einzelnen Parameter des  $K_{sfml}$  Klassifikators über die Zeit verglichen. Hierzu wurden die AUC-Werte für alle Zeitfenster von 5 Stunden berechnet. Wieder ist zu erkennen, dass der  $K_{sfml}$  Klassifikator anhand der Länge die besten Ergebnisse liefert. Während die Parameter Länge und Helligkeit durchgehend am besten bzw. schlechtesten geeignet sind, um Zellteilungen zu finden, ändert sich die Detektionsgenauigkeit für die Parameter Kompaktheit und Fläche: Während die Fläche einer Zellkontur in der ersten Hälfte sehr gute Resultate liefert ( $AUC > 0.8$ ) und am Ende abfällt ( $AUC < 0.75$ ), steigt die Detektionsrate des  $K_{sfml}$ , der die Kompaktheit nutzt von einer  $AUC \approx 0.65$  zu  $AUC > 0.8$ . Diese Änderung ist mit der Zellmorphologie zu erklären, die sich abhängig von der Zelldichte stark ändert (vgl. Abb. 4.5 und Abb. 4.6).

$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_K$	$AUC$	$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_K$	$AUC$
X	-	-	-	0.825	-	X	-	-	0.661
-	-	X	-	0.853	-	-	-	X	0.781
X	X	-	-	0.818	X	-	X	-	0.841
X	-	-	X	0.804	-	X	X	-	0.852
-	X	-	X	0.784	-	-	X	X	0.853
X	X	X	-	0.835	X	X	-	X	0.807
X	-	X	X	0.831	-	X	X	X	0.854
X	X	X	X	0.828					

**Tabelle 4.1.:** Vergleich aller Kombinationen von statischen, morphologischen Eigenschaften als AUC-Wert. Die Einteilung zeigt die Resultate des  $K_{sfml}$  und des  $K_{mfml}$  Klassifikators.



(a) ROC-Kurve.



(b) Zeitabhängige AUC-Werte.

**Abbildung 4.7.:** Die Genauigkeit der einzelnen, morphologischen Parameter  $\Delta f_A$ ,  $\Delta f_H$ ,  $\Delta f_L$ ,  $\Delta f_K$  des  $K_{sfml}$  und des adaptierten  $K_{3p-mfml}$  Klassifikators sind als ROC-Kurve (a) und zeitabhängige AUC-Werte angegeben, die in einem Zeitfenster von fünf Stunden berechnet wurden (b). Die Zellmorphologie ändert sich mit zunehmender Konfluenz und bewirkt eine abnehmende (Fläche  $\Delta f_A$ ) oder zunehmende Genauigkeit (Kompaktheit  $\Delta f_K$ ) der Mitosedetektion.

**Adaption an die Wachstumsphasen ( $K_{3p-mfml}$ )**

Die Auswertung in Abb. 4.7 zeigt, dass die Änderung der Zelldichte und der damit korrelierenden Änderung der Morphologie zu einer schwankenden Genauigkeit der Mitosedetektionen führt. Um die Detektionsperformance über die Zeit auszugleichen, können die verwendeten Parameter an die jeweilige Wachstumsphase angepasst werden. Diese Adaption wurde in dem  $K_{3p-mfml}$  Klassifizierer realisiert, der zeitabhängig die beste Parameterkombination der jeweiligen Wachstumsphase auswählt. Es ist leicht möglich, die zeitliche Adaption zu erweitern und die Parameterwahl an die Konfluenz, Zelldichte oder Zellgröße zu koppeln.

Die besten Parameterkombinationen der jeweiligen Wachstumsphase wurde mit Hilfe der Referenzdaten CeTReS.A ermittelt. Hierfür wurden die Referenzmitosen in drei Gruppen eingeteilt ( $0h - 10h$ ,  $10h - 40h$ ,  $40h - Ende$ ), die den Wachstumsphasen entsprechen. Anschließend wurden alle fünfzehn Parameterkombinationen in jeder der drei Phasen getestet und die Kombination mit den höchsten AUC-Werten bestimmt. Die gewählten Parameter, die für die Adaption des  $K_{3p-mfml}$  verwendet wurden, sind in der Abb. 4.8 angegeben.

Wie erwartet lieferte der adaptierte Klassifikator  $K_{3p-mfml}$  bessere Ergebnisse als der  $K_{sfml}$  oder der  $K_{mfml}$  (vgl. Abb. 4.7). Eine vollständige Liste aller Parameterkombinationen in jeder der drei Phasen ist im Anhang in Tab. A.1 gegeben.

Wachstumsphase	Zeitintervall	Parameter
I. (Lag Phase)	$0h - 10h$	$\Delta f_A$
II. (Exp. Wachstums)	$10h - 40h$	$\Delta f_L, \Delta f_K, \Delta f_H$
III. (Stationäre Phase)	$40h - Ende$	$\Delta f_K, \Delta f_L$

**Abbildung 4.8.:** Parameterkombination für den Klassifikator  $K_{3p-mfml}$ .

**4.2.4. Grenzen der morphologischen Klassifikatoren**

Die untersuchten Klassifikatoren  $K_{sfml}$  und  $K_{mfml}$  zeigen, dass eine Mitose über die Änderung der Differenzen  $\Delta f_A, \Delta f_H, \Delta f_K, \Delta f_L$  gefunden werden kann. Allerdings ist das Muster einer Mitose nicht statisch, sondern hängt von der Zelldichte ab. Hier kann eine Verbesserung der Genauigkeit erzielt werden, indem die Parameter abhängig von der Wachstumskurve und die damit verbundenen Änderungen der Zellmorphologie adaptiert werden, was mit dem  $K_{3p-mfml}$  umgesetzt wurde. Allerdings ist auch die adaptierte Mitosedetektion nicht in der Lage, alle Mitosen zu erkennen, was die Detektionsgenauigkeit des  $K_{3p-mfml}$  mit einer  $AUC \approx 0.85$  zeigt.

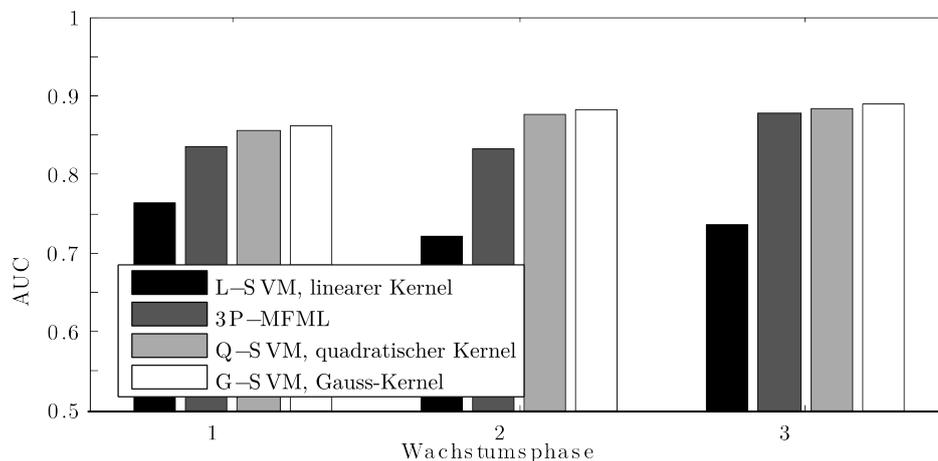
Um die Frage zu beantworten, mit welcher maximalen Genauigkeit eine morphologische Mitosedetektion durchgeführt werden kann, wurden mit dem Klassifikator  $K_{morph-svm}$  verschiedene SVMs mit einem linearen, quadratischen und Gauss-Kernel genutzt, um die bestmöglichen Klassifikatoren zu trainieren.

Die Detektionsergebnisse in Abb. 4.9 zeigen die Resultate des  $K_{morph-sum}$  Klassifikators mit den unterschiedlichen Kernelfunktionen. Bei der Auswertung wurde eine zehnfach-Kreuzvalidierung genutzt, um die mitotischen Pfade in den verschiedenen Wachstumsphasen zu evaluieren. Der gezeigte Fehler kann so interpretiert werden, dass er die bestmögliche Klassifikation beschreibt. Dies bedeutet, dass es mitotische und nicht-mitotische Pfade gibt, die sich sehr ähnlich sind und nicht anhand des morphologischen Musters der richtigen Klasse zugeordnet werden können.

### Systematisch übersehene Mitosen

Eine weitere Schwäche der morphologischen Detektion liegt in der impliziten Voraussetzung, dass die mitotischen Zellen korrekt detektiert werden müssen: Falls der zugehörige Zellpfad nicht richtig rekonstruiert wurde, kann das gesuchte Mitose-Muster nicht extrahiert werden. Umgekehrt bedeutet dies, dass bei einer Detektionsrate von 95% ungefähr 5% aller Mitosen nicht gefunden werden können, wie in der Zusammenfassung auf Seite 70 beschrieben. Insgesamt zeigt sich, dass die Genauigkeit der Mitosedetektion von der Genauigkeit der Zelldetektion beschränkt wird. Da diese wiederum direkt von der gegebenen Bildqualität abhängt, wurde mit dem verwendeten Langzeitmikroskop nach einer besseren Darstellung der Mitose gesucht und mit den verwendeten Phasenkontrastbildern gefunden (vgl. Abb. 4.1). Diese gute Repräsentation einer Mitose liefert den Grund, warum in dem verwendeten Setup der Langzeitexperimente zwei Kontrasteinstellungen genutzt werden, um die Zellen abzufotografieren (vgl. Abschnitt 2.2.1).

Der nächste Abschnitt beschreibt, wie die Phasenkontrastbilder genutzt werden können, um die Mitosen unabhängig von den extrahierten Zellpfaden mit einer sehr hohen Genauigkeit zu detektieren.

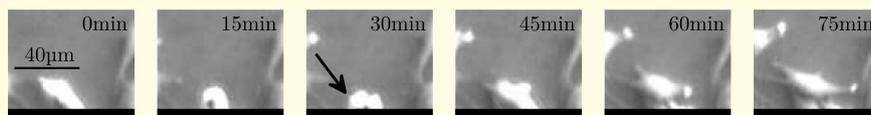


**Abbildung 4.9.:** Die Detektionsgenauigkeit ist für die unterschiedlichen Klassifikatoren als AUC für die jeweiligen Wachstumsphasen angegeben. Die Werte der SVM-basierten Klassifikation wurden mit einer zehnfach-Kreuzvalidierung ermittelt.

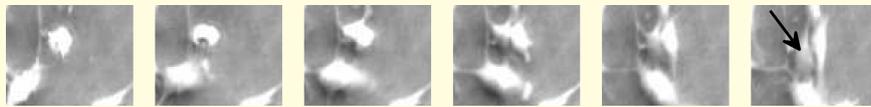
### Fehlerhafte Zellpfade

Die morphologische Mitosedetektion nutzt die Zellpfade, um anhand der Änderung der Zellkontur zu entscheiden, ob ein Zellpfad in einer Mitose endet oder nicht. Somit hängt die Genauigkeit, mit der eine Zellteilung gefunden wird, direkt von der Güte der Zelldetektion ab: Sobald eine mitotische Zelle in den letzten Aufnahmen vor ihrer Teilung nicht korrekt erkannt wurde, kann diese Zellteilung nicht mehr gefunden werden.

In Abb. 4.10 sind einige Detektionsfehler gezeigt, die aus den Bildserien vom Referenzdatensatz CeTReS.A extrahiert wurden. Dieser Datensatz enthält 2019 Mitosen, von denen 1903 richtig erkannt wurden. Von den fehlenden 116 Zellen liegt in 59 Fällen ein Detektionsfehler vor der eigentlichen Mitose vor, so dass der mitotische Pfad nur aus einem Zeitschritt bestand. Hier war es nicht möglich, den Differenzvektor zu berechnen. Die übrigen 57 Fehler setzen sich aus 39 Zellverschmelzungen während der Mitose, aus 13 Zellen, die zu dicht am Rand lagen, drei Trackingfehlern und zwei nicht erkannten Zellen zusammen.



(a) Rand des Beobachtungsgebietes.



(b) Nicht erkannte Tochterzelle.



(c) Zelle, die nicht verfolgt werden konnte.



(d) Verschmelzung der Zellkonturen.

**Abbildung 4.10.:** Beispiele für fehlerhafte Zellpfade. Ein Fehler kann durch den Rand (a), eine nicht detektierte Zelle (b), einen Trackingfehler (c) oder eine Falschdetektion (c) verursacht werden und verhindert, dass eine Mitose gefunden wird. Die Fehlerquellen sind mit Pfeilen markiert.

### 4.3. Mitosedetektion in Phasenkontrastbildern

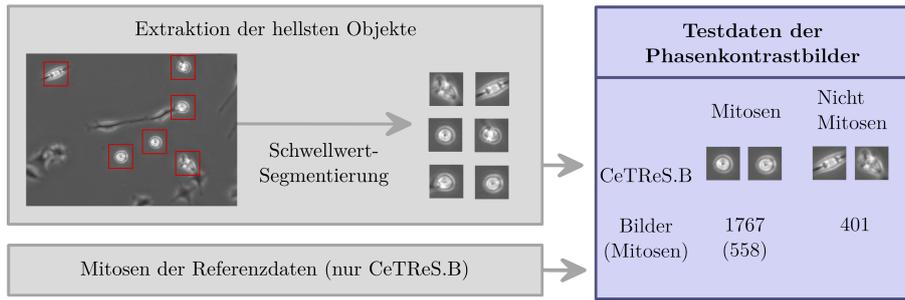
In diesem Abschnitt werden verschiedene Methoden vorgestellt, mit denen das eindeutige Muster einer Mitose im Phasenkontrast genutzt wird, um eine Zellteilung zu finden. Hierbei wird die typische Charakteristik genutzt, die in Abb. 4.1 gezeigt ist: Während einer Mitose erhöht sich die Helligkeit einer Zelle und sie nimmt eine fast kreisförmige, runde Kontur an, wodurch sie sich deutlich von den Nicht-Mitosen abhebt (vgl. Abb. 4.1). Der große Vorteil gegenüber der morphologischen Mitosedetektion besteht darin, dass dieses Muster genutzt werden kann, um die Zellteilungen unabhängig von den Zellpfaden zu finden, d.h. es wird keine Detektion der Zellen im Schräglichtkontrast vorausgesetzt.

Um das Mitose-Muster zu finden, wird in den folgenden Ansätzen ein zweistufiges Verfahren genutzt. In dem ersten Schritt werden die hellsten Zellen bzw. Objekte aus den Bildern als mögliche Mitosekandidaten extrahiert (vgl. Abb. 4.11). Da sowohl die nicht-mitotischen Zellen als auch der Bildhintergrund sehr dunkel sind, wird eine schwelwertbasierte Segmentierung genutzt, um alle Mitosen zu finden. Hierbei werden neben den Mitosen noch weitere, helle Objekte gefunden, die z.B. durch Verschmutzung, tote oder sich überlagernde Zellen entstehen. Von allen Kandidaten wird ein Bild der Größe  $41 \times 41px$  ausgeschnitten, wobei alle Objekte verworfen werden, die zu nah am Rand liegen. Diese Bilder liefern den Ausgangspunkt für die zweite Stufe, in der verschiedene Charakteristika der Mitosen genutzt werden, um die Zellteilungen zu identifizieren. Im folgenden Abschnitt werden zwei Klassifikatoren vorgestellt, die anhand der Rundheit der Zellkontur oder der Helligkeit der Zellen entscheiden, ob eine mitotische Zelle gefunden wurde. In einem dritten Ansatz werden die erstellten Testdaten genutzt, um ein repräsentatives Muster einer Mitose mit Hilfe einer Support Vector Machine (SVM) zu lernen und die Zellteilungen zu finden.

#### 4.3.1. Referenzmitosen der Phasenkontrastbilder (CeTReS.B)

Um die Mitosekandidaten bzw. einen Satz von Referenzbildern für die Mitosen im Phasenkontrast zu erstellen, wurde die Bildserie von CeTReS.B genutzt. Dem zweistufigen Verfahren entsprechend wurde in den Phasenkontrastbildern eine schwelwertbasierte Bildsegmentierung genutzt, um alle hellen Objekte zu extrahieren. Jedes Objekt wurde als ein Bild mit der Größe von  $41 \times 41px \approx 66 \times 66\mu m^2$  ausgeschnitten und abgespeichert. Anschließend wurden diese Bilder mit Hilfe der CeTReS.B-Daten in Mitosen und Nicht-Mitosen eingeteilt (vgl. Abb. 4.11). Die so extrahierten Daten enthalten 558 der 573 Mitosen der Referenzdaten CeTReS.B und werden als  $I_i^m, i = 1, \dots, 558$  bezeichnet. Die fehlenden fünfzehn Mitosen lagen zu dicht am Bildrand und konnten nicht vollständig ausgeschnitten werden. Folglich wurden sie aus der Auswertung ausgeschlossen.

Von den 558 enthaltenen Mitosen sind insgesamt 1767 verschiedene Bilder extrahiert worden. Diese große Anzahl ist bedingt durch die Länge einer typischen Mitose: Sie dauert drei- bis viermal so lange wie das gewählte Aufnahmeintervall von 10 Minuten. Daher ist jede Mitose durchschnittlich auf 3 – 4 Bildern festgehalten. Von jeder Mitose ist manuell das beste Bild ausgewählt worden, so dass die Klasse der Mitosen insgesamt 558 Bilder enthält. Die Klasse der Nicht-Mitosen enthält 401 Bilder, die als  $I_i^z, i = 1, \dots, 401$  bezeichnet werden. Zusätzlich wurden noch 594 Bilder  $I_i^d, i = 1, \dots, 594$  von



**Abbildung 4.11.:** Die mitotischen Zellen werden in den Phasenkontrastbildern mit Hilfe eines zweistufigen Verfahrens gesucht. Zuerst werden mögliche Mitosekandidaten über die Helligkeit der Zellen extrahiert, die anschließend als Mitose bzw. Nicht-Mitose klassifiziert werden. Dieses Schema wurde genutzt, um die Testdaten der Phasenkontrastbilder zu erstellen. Die so erstellten Referenzdaten der Mitosen enthalten 558 Mitosen und 401 Nicht-Mitosen, die als Bilder der Größe  $41 \times 41px \approx 66 \times 66\mu m^2$  abgespeichert wurden.

toten Zellen erstellt.

### Detektion über die Rundheit und die Helligkeit

Der erste Klassifikator nutzt die Helligkeit  $f_H$  und die Rundheit bzw. Kompaktheit  $f_K$  eines Mitosekandidaten um zu entscheiden, ob eine Zellteilung gefunden wurde. Um diese Parameter zu berechnen, wird eine globale Schwellwertsegmentierung genutzt, um die Zellkonturen  $z_i$  in den gegebenen Bildern zu finden. Der Schwellwert wird mit der Methode von Otsu bestimmt (Otsu [84]). Anschließend wird mit den Klassifikatoren

$$K_H(z_i) := z_i \in \begin{cases} \text{Mitose} & \text{falls } f_H(z_i) > \theta_I, \\ \text{Nicht - Mitose} & \text{sonst,} \end{cases}$$

und

$$K_C(z_i) := z_i \in \begin{cases} \text{Mitose} & \text{falls } f_K(z_i) > \theta_C, \\ \text{Nicht - Mitose} & \text{sonst,} \end{cases}$$

entschieden, ob eine Zellteilung beobachtet wurde oder nicht;  $f_H(z_i)$  und  $f_K(z_i)$  liefern die Helligkeit bzw. die Kompaktheit der Zellkontur  $z_i$  in den Phasenkontrastbildern. Die schlechtesten Mitosen und besten Nicht-Mitosen der Testdaten sind in Abb. 4.13 abgebildet, d.h. die gezeigten Mitosen haben nach den genutzten Kriterien die niedrigste Helligkeit bzw. die größte Kompaktheit; entsprechendes gilt für die gezeigten Nicht-Mitosen.

### SVM basierte Klassifikation

In einem weiteren Ansatz wurden die erstellten Referenzbilder der Mitosen und Nicht-Mitosen genutzt, um ein repräsentatives Muster einer Mitose zu lernen. Hierbei werden

die Referenzbilder der Mitosen  $I^m$  und Nicht-Mitosen  $I^z$  genutzt, um mit Hilfe einer linearen SVM den Klassifikator  $K_{svm}$  zu trainieren. Für die Berechnung und Klassifikation wurde auf die MATLAB Funktionen `svmtrain.m` und `svmclassify.m` zurückgegriffen. Mit Hilfe des Klassifikators  $K_{svm}$  ist es nun möglich, für jeden gegebenen Mitosekandidaten zu entscheiden, ob eine mitotische Zelle gezeigt ist oder nicht.

### 4.3.2. Evaluation der Mitosedetektion auf den Phasenkontrastbildern

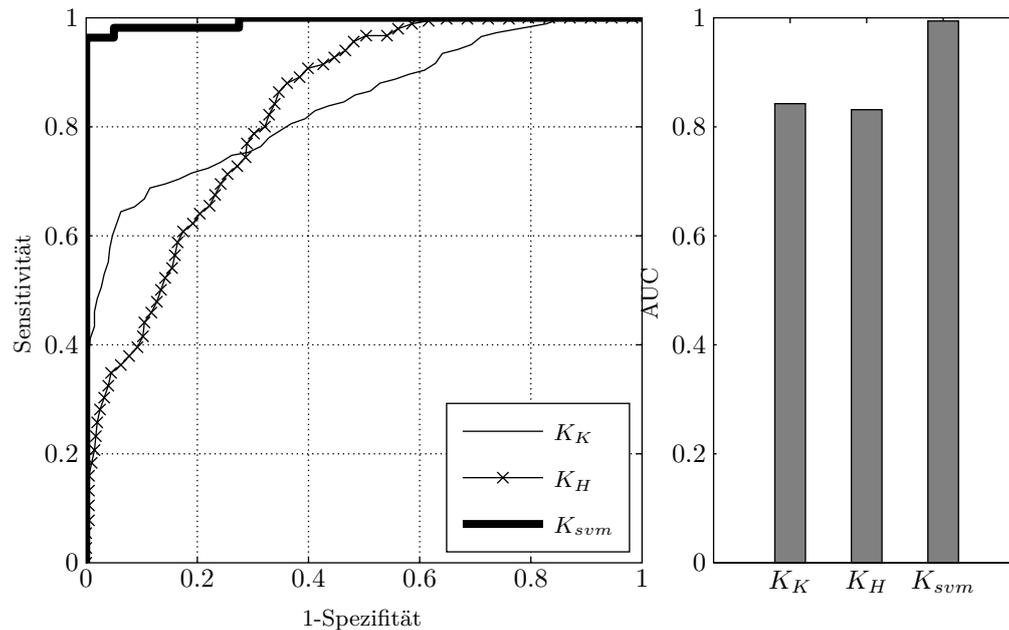
In Abb. 4.12 ist ein Vergleich der Klassifikatoren  $K_K$ ,  $K_H$  und  $K_{svm}$  gezeigt, die auf den gegebenen Testbildern  $I_i^z, i = 1 \dots, 401$  und  $I_i^m, i = 1, \dots, 558$  evaluiert wurden. Die ROC-Kurven zeigen, dass die Klassifikatoren  $K_K$  und  $K_H$ , die auf der Kompaktheit und der Helligkeit basieren, dem SVM-basierten Klassifikator  $K_{svm}$  unterlegen sind. Während  $K_K$  und  $K_H$  eine AUC von 0.843 bzw. 0.831 erreichen, klassifiziert  $K_{svm}$  die Daten fast fehlerfrei (die AUC ist der gemittelte Wert einer zehnfach-Kreuzvalidierung). Im Folgenden wird daher nur noch der Klassifikator  $K_{svm}$  betrachtet.

#### Detektion im Phasenkontrast mit dem $K_{svm}$ Klassifikator

Der direkte Vergleich der verschiedenen Mitosedetektionen zeigt, dass die Repräsentation einer Mitose im Phasenkontrast deutlich besser geeignet ist, um Zellteilungen zu finden als die morphologische Mitosedetektion in den Schräglichtbildern. Anhand der vorgestellten Detektion  $K_{svm}$  konnte gezeigt werden, dass eine kleine Trainingsmenge von ca. 50 Mitosen ausreichend ist, um einen guten Klassifikator mit einer  $AUC > 0.99$  zu trainieren. Dieses sehr gute Resultat zeigt, dass eine Mitose in den gewählten Bild-daten ein eindeutiges, charakteristisches Muster besitzt, welches gut geeignet ist, um eine Zellteilung zu finden.

Um das charakteristische Muster der Mitosen, Nicht-Mitosen und toten Zellen zu visualisieren, wurde eine Hauptkomponentenanalyse (*principal component analysis*, PCA) genutzt, um die stärksten Merkmale der gegebenen Bilder zu extrahieren. Die PCA zerlegt die gegebenen Bilder in sogenannte Hauptkomponenten, die die Merkmale mit der größten Varianz zeigen, d.h. es werden die charakteristischen Muster extrahiert, aus denen sich die Mitosebilder zusammensetzen. Die PCA wurde 1901 von Pearson [87] beschrieben und liefert den modernen Wissenschaften ein robustes Werkzeug für die Multivariate Datenanalyse (Kessler [57]). Die Bilder der Hauptkomponenten werden in Anlehnung an die *Eigenfaces* (EM), die in der Gesichtserkennung genutzt werden (Turk and Pentland [116]), als Eigenmitosen bezeichnet.

Die Eigenmitosen und Hauptkomponenten der 558 Mitosen, der 594 toten Zellen und 401 Nicht-Mitosen zu den 10 größten Eigenwerten sind in Abb. 4.14 gezeigt, d.h. es sind die zehn Muster gegeben, die die Daten am besten repräsentieren. Die Bilder der Eigenmitosen zeigen dominante, runde Strukturen, die unterschiedliche Helligkeitsverläufe aufweisen. In den Hintergründen der Hauptkomponenten sind die unterschiedlichen Kontrastverläufe erkennbar, aus denen die Bilder zusammengesetzt werden: Es ist z.B. ein Helligkeitsverlauf von innen nach außen (EM 1), von links nach rechts (EM 2), von unten nach oben (EM 3) und über die Diagonale (EM 5) zu erkennen. Diese Bilder verdeutlichen, wie einheitlich das Muster einer mitotischen Zelle ist.



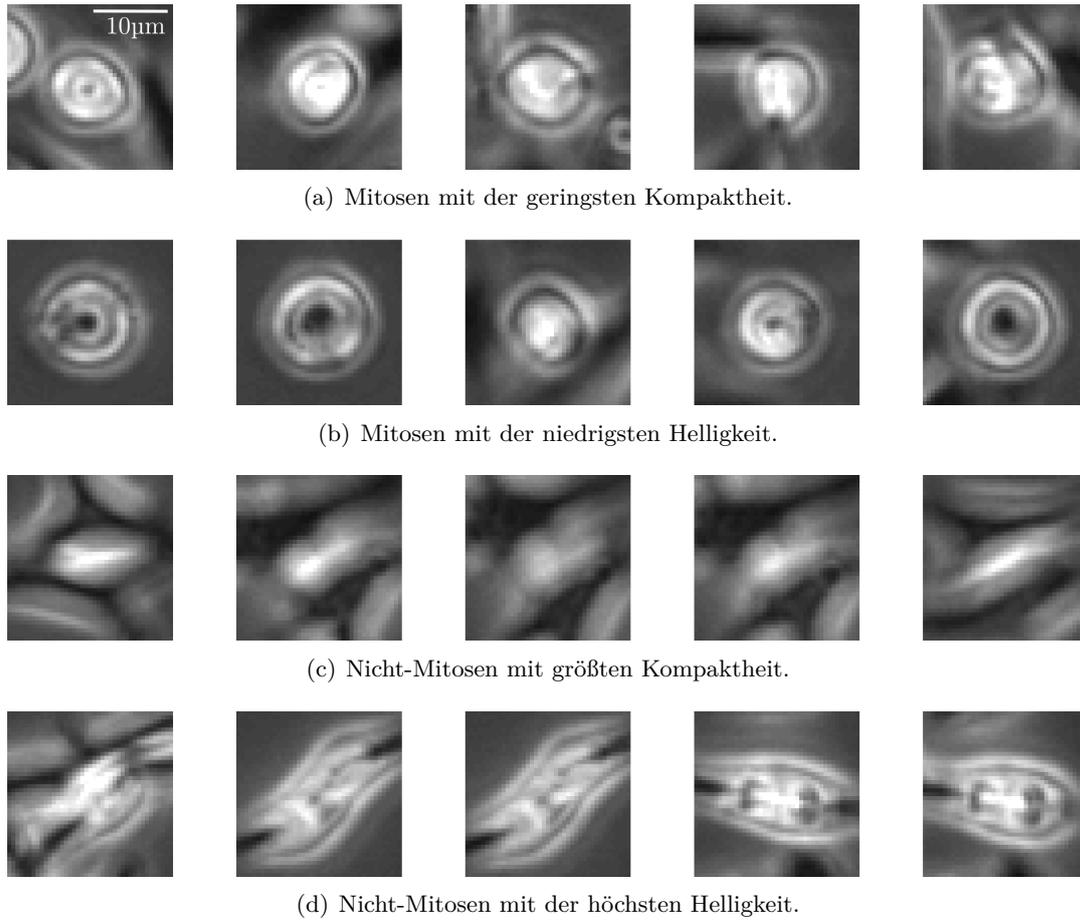
**Abbildung 4.12.:** Die ROC-Kurven (links) und die AUC-Werte (rechts) sind für die vorgestellten Mitosedetektionen in den Phasenkontrastbildern gezeigt. Die Klassifikatoren  $K_K$  und  $K_H$  nutzen die Kompaktheit und die Helligkeit der Zellkonturen, während  $K_{svm}$  ein Mitose-Muster aus den Bildern gelernt hat. Für die Berechnung der AUC-Werte von  $K_{svm}$  ist eine zehnfach-Kreuzvalidierung genutzt worden.

### Schräglichtbilder vs. Phasenkontrastbilder

Ein Vergleich der verschiedenen Mitosedetektionen in den Schräglicht- und in den Phasenkontrastbildern ist in Tab. 4.2 gegeben. Hier zeigt sich, dass die Repräsentation der Mitosen im Phasenkontrast eine sehr gute Detektionsgenauigkeit von  $AUC = 0.993$  erzielt. Der Vergleich zu der morphologischen Mitosedetektion mit einer  $AUC = 0.85$  illustriert, dass die Phasenkontrastbilder besser geeignet sind, um eine Mitose zu finden.

Bei der Interpretation sollte berücksichtigt werden, dass die verwendeten Testdaten nicht dieselben Informationen nutzen: Die Testdaten der Schräglichtbilder enthalten alle Pfadenden (2710 nicht-mitotische Pfade), während die Daten der Phasenkontrastbilder alle hellen Objekte zeigen (995 Bilder von hellen Artefakten und toten Zellen).

Trotz der kleineren Testdaten können die Zellteilungen in den Phasenkontrastbildern mit einer wesentlich höheren Genauigkeit extrahiert werden, als es in den Schräglichtbildern mit der morphologischen Detektion möglich ist.



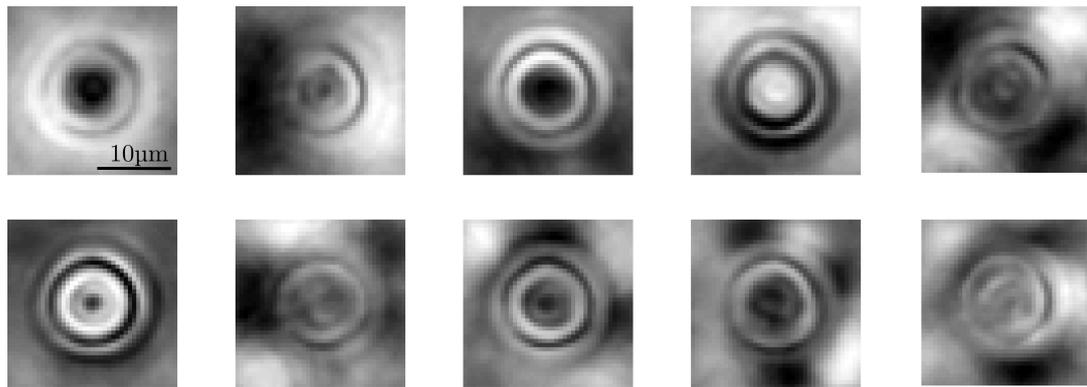
**Abbildung 4.13.:** Die vorgestellten Klassifikatoren  $K_K$  und  $K_H$  nutzen die Kompaktheit der Zellkonturen und die Helligkeit der Zellen, um eine Mitose zu detektieren. Anhand dieser Maße sind die schlechtesten Mitosen (a+b) und die besten Nicht-Mitosen (c+d) gezeigt.

$K_K$	$K_H$	$K_{svm}$	$K_{sfml}$				$K_{mfml}$
			$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_C$	$(\Delta f_A, \Delta, f_H, \Delta f_L)$
0.843	0.831	0.993	0.820	0.768	0.828	0.763	0.843

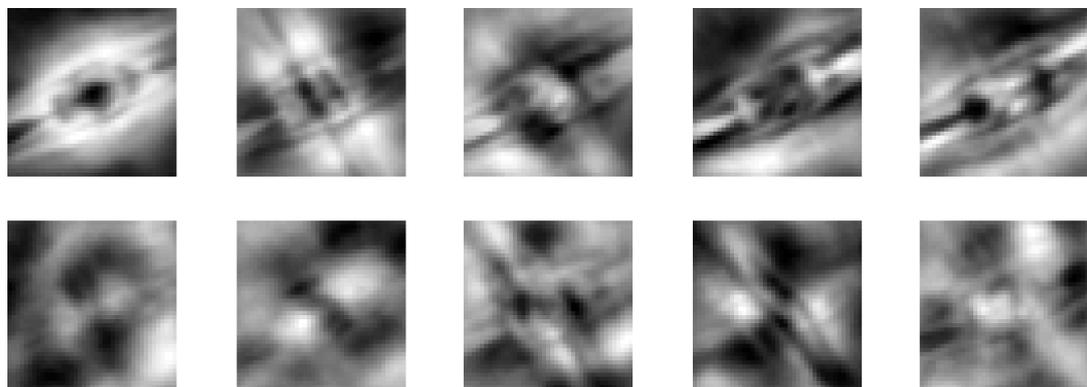
**Tabelle 4.2.:** Die erzielten Klassifikationsgenauigkeiten zeigen, dass die Mitosedetektion in den Phasenkontrastbildern mit  $K_K$ ,  $K_H$  und  $K_{svm}$  mit einer höheren Genauigkeit ausgeführt werden kann als in den Schräglichtbildern. Für den Klassifikator  $K_{mfml}$  ist die beste Parameterkombination ausgewählt worden. Eine vollständige Übersicht der Resultate aller Kombinationen ist in Tab. A.2 im Anhang gegeben.

### 4.3. Mitosedetektion in Phasenkontrastbildern

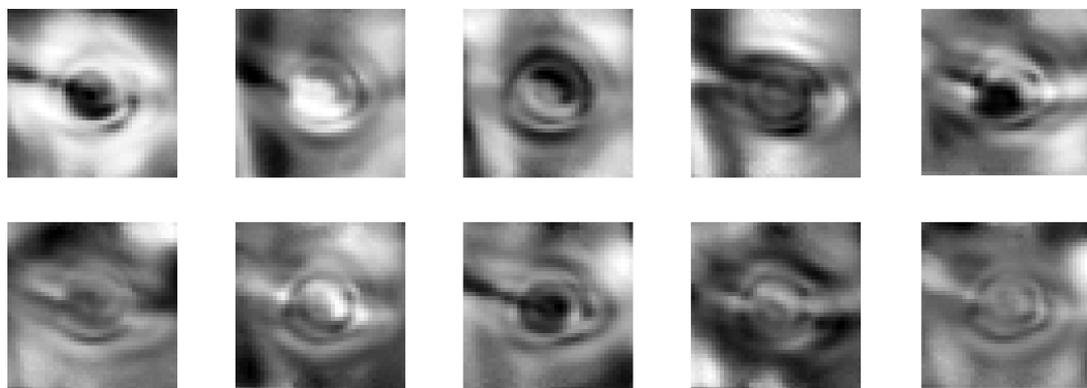
---



(a) Hauptkomponenten der Mitosen (Eigenmitosen).



(b) Hauptkomponenten der Nicht-Mitosen.



(c) Hauptkomponenten der toten Zellen.

**Abbildung 4.14.:** Mit Hilfe einer Hauptkomponentenanalyse (*principal component analysis*, PCA) sind die Hauptkomponenten der Mitosen (a), Nicht-Mitosen (b) und toten Zellen (c) berechnet worden. Es sind jeweils die Hauptkomponenten zu den 10 größten Eigenwerten abgebildet, d.h. es sind die zehn Merkmale gegeben, welche die Daten am besten repräsentieren. Die Eigenmitosen in (a) zeigen das typische runde Muster einer mitotischen Zelle und verdeutlichen das einheitliche Muster einer Zellteilung.

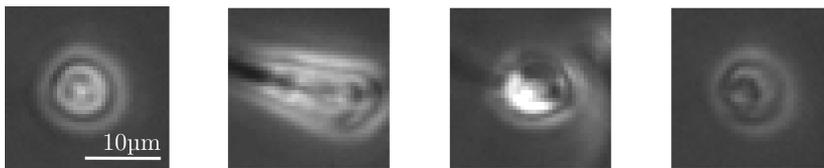
### Detektion abgestorbener Zellen

In Abb. 4.15 sind vier Beispielbilder der toten Zellen aus der Bildserie vom Datensatz CeTReS.B gezeigt, die ein Problem der Mitosedetektion im Phasenkontrast verdeutlichen: Wenn eine Zelle stirbt, zieht sie sich zusammen, wird kleiner und runder, wodurch sie ein ähnliches Muster wie eine mitotische Zelle zeigt.

Um zu verhindern, dass nun eine tote Zelle als Zellteilung erkannt wird, kann das Muster einer toten Zelle mit Hilfe von repräsentativen Bildern gelernt werden, d.h. die Idee besteht darin, erneut eine SVM als Klassifikator zu nutzen, um die toten Zellen auszusortieren. Hierzu wurde mit der Bildserie von Referenzdatensatz CeTReS.B ein Trainingsdatensatz  $I_i^d, i = 1, \dots, 594$  erstellt, der die Bilder der toten Zellen zeigt. Anschließend wurde analysiert, wie gut sich Mitosen von abgestorbenen Zellen unterscheiden lassen, indem eine SVM mit den mitotischen Bildern  $I^m$  und den toten Zellen in  $I^d$  trainiert wurde. Dieser Klassifikator  $K_d$  wurde anschließend mit einer zehnfach-Kreuzvalidierung evaluiert und war in der Lage, die beiden gegebenen Bilddaten mit einer hohen Genauigkeit von  $AUC = 0.996$  zu trennen.

Es sollte aber beachtet werden, dass sämtliche Testbilder der toten Zellen von insgesamt nur zwei toten Zellen stammen. Bei dem Training der SVM könnte es daher passiert sein, dass sich die SVM das typische Muster der beiden Zellen selber gemerkt hat und diese leicht wiederfindet. Um diesen Trainingsfehler auszuschließen, wurde in einem weiteren Test der Klassifikator  $K_{svm}$  genutzt um herauszufinden, wie gut ein Klassifikator eine tote von einer mitotischen Zelle unterscheiden kann, wenn er nur das Muster der Mitosen und normalen Zellen kennt. Der Klassifikator  $K_{svm}$  wurde erneut mit den Bildern der Mitosen  $I^m$  und Nicht-Mitosen  $I^z$  trainiert, aber mit den Bildern der toten Zellen  $I^d$  evaluiert. Auch hier ergab sich eine hohe Klassifikationsgenauigkeit von  $AUC = 0.992$  was zeigt, dass die Muster der Mitosen und toten Zellen gut unterschieden werden können.

Die Ergebnisse zeigen, dass umgekehrt die SVM sehr gut auf das typische Mitosemuster adaptiert ist und selbst andere, helle Objekte, wie z.B. tote Zellen, zuverlässig von dem Klassifikator  $K_{svm}$  ausgeschlossen werden.



**Abbildung 4.15.:** Beispielbilder für die toten Zellen, die in der Bildserie zu CeTReS.B enthalten sind.

## 4.4. Ergebnisse der Mitosedetektion und des Zelltracking

Während in den vorherigen zwei Abschnitten die Eigenschaften der einzelnen Mitosedetektionen mit den erstellten Referenzdaten analysiert worden sind, werden die Algorithmen nun anhand der gesamten Referenzdaten CeTReS.B verglichen. Hierzu wird das Fehlermaß genutzt, das in Abschnitt 3.2 definiert wurde und die Genauigkeit eines Zelltracking-Algorithmus mit Hilfe der vollständigen Pfade bewertet, die erfolgreich rekonstruiert werden konnten. Dieses Maß hängt sowohl von dem verwendeten

- **Zelltracking** als auch von
- der **Mitosedetektion**

ab und liefert ein sehr gutes Kriterium für die Extraktion von zellulären Genealogien und Verwandtschaften. In Abb. 4.16 sind die Sensitivitäten und die Genauigkeiten der verschiedenen Ansätze angegeben. Diese Werte geben an, wie viele der vollständigen Pfade extrahiert wurden (Sensitivität) und wie viele der Pfade, die als vollständig erkannt wurden, auch korrekt klassifiziert wurden (Genauigkeit). Die folgenden Methoden bzw. Kombinationen wurden analysiert und verglichen:

- **Naive Auswertung.** In einem ersten Ansatz wurde das naive Zelltracking genutzt, das in Abschnitt 2.4 definiert wurde. Um die Pfade zu verknüpfen wurde die naive Mitosedetektion  $K_Y$  verwendet.
- **Validiertes Tracking.** Als zweite Methode wurde das validierte Zelltracking genutzt, um gültige Zellpfade zu finden und alle fehlerhaften Pfade zu entfernen, wie in Abschnitt 3.1 beschrieben. In den so generierten Daten wurde zuerst die naive Mitosedetektion  $K_Y$  genutzt, die anhand der Konnektivität der Pfade die Verwandtschaften extrahiert. Anschließend wurde eine weitere Auswertung mit der morphologischen Mitosedetektion durchgeführt; hier ist das Ergebnis des  $K_{m, fml}$  Klassifikators gezeigt, der die Parameterkombination mit der besten Detektionsrate verwendet.
- **Kombination Schräglicht und Phasenkontrast I.** Die Mitosedetektion  $K_{svm}$  in den Phasenkontrastbildern wurde mit den Zellpfaden des validierten Zelltracking kombiniert. Hierzu wurde der Klassifikator  $K_{svm}$  mit einem ähnlichen Schema trainiert und evaluiert, wie es bei der zehnfach-Kreuzvalidierung genutzt wurde. Aber anstatt mit 9 von 10 Gruppen zu trainieren, wurde diesmal nur mit jeweils einer Gruppe gelernt, also einem Zehntel der Trainingsdaten. Anschließend wurden alle hellen Objekte (einschließlich aller toten Zellen) als Mitosekandidaten extrahiert und von  $K_{svm}$  klassifiziert. Zu jeder Mitose wurde der korrespondierende Zellpfad herausgesucht und als mitotisch markiert. Die Auswertung zeigt den Mittelwert der zehn durchgeführten Auswertungen, die Varianz der Genauigkeiten und Sensitivität lag bei  $\approx 2.0 \times 10^{-5}$ .
- **Fluides Tracking.** Mit dem *fluiden Tracking* aus (Scherf et al. [99, 100]) wurde ein weiterer Zelltracking-Ansatz genutzt, wobei die erstellten Zellpfade mit Hilfe der naiven Mitosedetektion verknüpft wurden. Eine Beschreibung des fluiden

Tracking und der Evaluierung und Validierung ist im Anhang auf Seite 176 zu finden. Alle benötigten Schnittstellen um die Ergebnisse des fluiden Tracking zu importieren, zu validieren und zu evaluieren sind in dem Zelltracking-Framework implementiert.

- **Fluides Tracking (evaluiert und validiert).** In dieser Version des fluiden Tracking wurden die Referenzdaten CeTReS.A genutzt, um systematische Fehler in den Tracking-Daten zu identifizieren und die Algorithmen bzw. Parameter anzupassen. Die Ergebnisse des Zelltracking wurden mit Hilfe der naiven Mitosedetektion  $K_Y$  kombiniert, um die vollständigen Pfade zu extrahieren.
- **Kombination Schräglicht und Phasenkontrast II.** Die letzte Auswertung kombiniert die Mitosedetektion  $K_{svm}$  im Phasenkontrast mit den Ergebnissen des fluiden Trackings. Die Ergebnisse wurden wieder mit einer zehnfach-Kreuzvalidierung evaluiert, wobei die Varianz der Sensitivität bzw. der Genauigkeit bei  $2.8 \times 10^{-5}$  bzw.  $2.0 \times 10^{-5}$  lag.

Die Ergebnisse der Evaluation sind in Abb. 4.16 und in Tab. 4.3 gezeigt. Wie erwartet liefert das naive Zelltracking die schlechtesten Ergebnisse, da hier nur ca. 50% aller Pfade gefunden werden. Es zeigt sich auch, dass von allen gefundenen Zellpfaden nur 34% korrekt extrahiert wurden. Dagegen zeigt sich eine deutliche Verbesserung der Genauigkeit, wenn das validierte Überlaptracking genutzt wird.

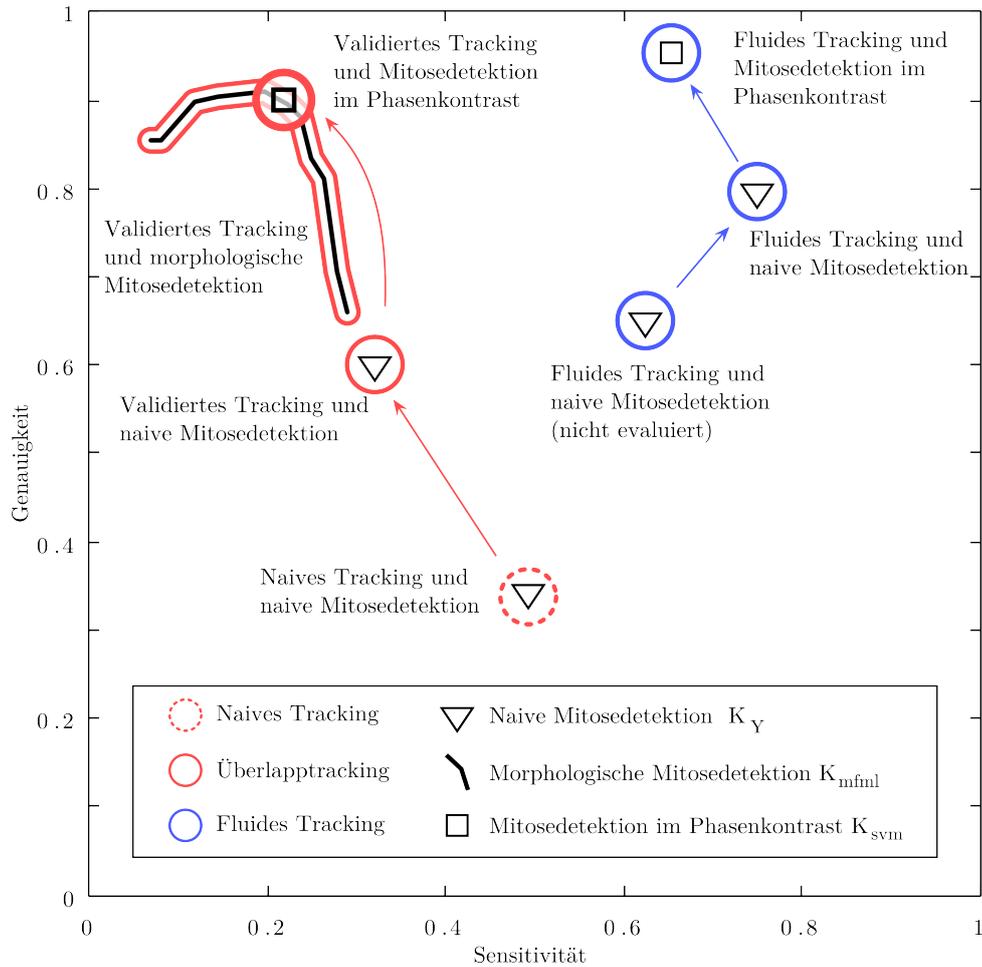
Wird die Auswirkung der Mitosedetektion betrachtet, so liefert die Kombination des Zelltracking in den Schräglichtbildern mit der Mitosedetektion in den Phasenkontrastbildern die besten Ergebnisse, sowohl bei dem getesteten Überlaptracking als auch bei dem fluiden Tracking. Auffällig ist hier, dass die Kombination des validierten Überlaptracking und der Mitosedetektion in den Schräglicht- und Phasenkontrastbildern zu sehr ähnlichen Ergebnissen mit einer nahezu identischen Genauigkeit führt. Hier zeigt sich, dass das Tracking zu fehlerhaft ist, wodurch die erreichbare Genauigkeit limitiert wird.

Die besten Ergebnisse werden mit der Kombination aus dem fluiden Tracking in den Schräglichtbildern und der Mitosedetektion in den Phasenkontrastbildern erzielt. Insgesamt zeigt sich, dass die Validierung der Daten zu einer höheren Rekonstruktionsgenauigkeit der vollständigen Pfade führt. Dieses Ergebnis zeigt sich sowohl für das Überlaptracking als auch für das fluide Tracking.

Tracking	Naiv	Überlaptracking			Fluides Tracking		
		$K_Y$	$K_Y$	$K_{mfml}$	$K_{svm}$	nicht evaluiert	validiert + eval.
Mitosedetektion	$K_Y$	$K_Y$	$K_{mfml}$	$K_{svm}$	$K_Y$	$K_Y$	$K_{svm}$
Sensitivität	0.49	0.32	0.19	0.22	0.62	0.75	0.66
Genauigkeit	0.34	0.60	0.91	0.91	0.65	0.80	0.96

**Tabelle 4.3.:** Genauigkeiten der verschiedenen Detektions- und Tracking-Algorithmen für die Auswertungen aus Abb. 4.16.

#### 4.4. Ergebnisse der Mitosedetektion und des Zelltracking



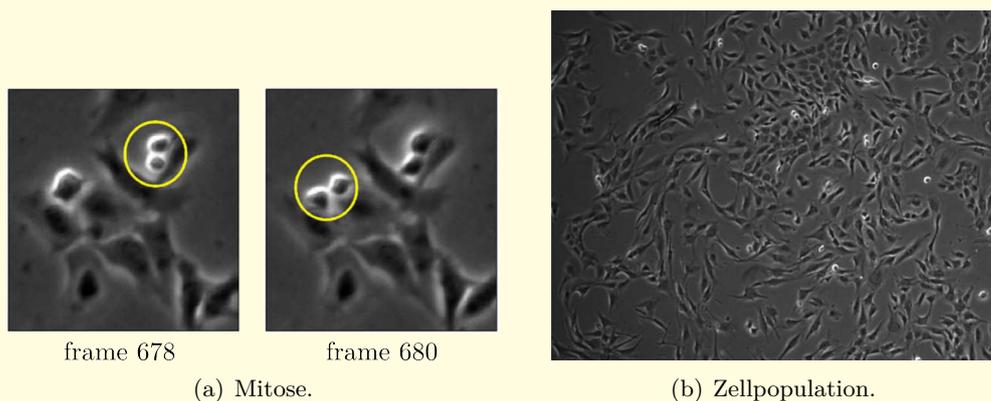
**Abbildung 4.16.:** Die vorgestellten Zelltracking-Algorithmen und Mitosedetektionen sind anhand der Genauigkeit der rekonstruierten, vollständigen Zellpfade evaluiert worden. Das Überlapptracking (blau) und das fluide Tracking (rot) wurden jeweils mit mehreren Mitosedetektionen ausgewertet. Hierbei wurden zwei Versionen des fluiden Tracking verwendet: Eine nicht evaluierte Version, die nicht modifiziert wurde, und eine evaluierte. Bei der durchgeführten Evaluation wurden die Referenzdaten CeTReS.A genutzt, um den Algorithmus systematisch auf Fehler zu untersuchen. Bei der verwendeten Detektion  $K_{svm}$  wurde jeweils das Zelltracking in den Schräglichtbildern ausgeführt und mit der Mitosedetektion  $K_{svm}$  in den Phasenkontrastbildern kombiniert. Die SVM des Klassifikators  $K_{svm}$  wurde in 10 Durchläufen mit  $\frac{1}{10}$  der Mitosen trainiert (jeweils 54 Mitosen). Anschließend wurden die Mitosen in allen 2762 Mitosekandidaten gesucht, d.h. es wurde in den Bildern der toten Zellen, der Mitosen und der Nicht-Mitosen gesucht. Die Werte für die morphologische Detektion  $K_{mfml}$  sind für die beste Parameterkombination  $[\Delta f_A \Delta f_H \Delta f_L]$  angegeben und für einen variierenden Schwellwert  $\theta = 10^0, \dots, 10^{-30}$  gezeigt (vgl. Formel (4.3)). Die exakten Werte sind in Tab. 4.3 zu finden.

### Vergleich mit anderen Methoden

Mit Hilfe der erstellten Referenzdatensätze CeTReS.A und CeTReS.B war es möglich, eine ausführliche Evaluation des Zelltracking und der Mitosedetektion durchzuführen. Mit der Auswertung des fluiden Tracking konnte gezeigt werden, dass sich die Datenvalidierung und die Mitosedetektion auch auf andere Methoden übertragen lässt. Es wäre aber weiterhin wünschenswert, die erzielten Ergebnisse in einem größeren Kontext mit weiteren Arbeiten zu vergleichen. Das wird aber durch die folgenden Punkte erschwert:

1. Die Rohdaten und Bildserien hängen sehr stark von den verwendeten Zellen und von dem verwendeten Equipment ab. Auch wenn in vielen experimentellen Aufbauten Phasenkontrastmikroskope genutzt werden, so ist die Qualität der Bilder jedoch stark unterschiedlich. Beispielhaft ist in Abb. 4.17 das Bildmaterial aus der Arbeit von Huh et al. [48] gezeigt. Während die Bilder eine sehr große Ähnlichkeit zu den verwendeten Phasenkontrastbildern aus CeTReS.B aufweisen, so weicht das Muster der Mitosen doch deutlich voneinander ab (vgl. Abb. 4.1).
2. Die verwendeten Zellen und Zelllinien haben unterschiedliche Eigenschaften, so dass viele Algorithmen direkt auf die gegebene Problemstellung angepasst wurden. Da weiterhin viele Bildserien nicht publiziert werden, ist es nicht immer möglich, die unterschiedlichen Ergebnisse zu vergleichen.
3. Sehr wenige der publizierten Zelltracking-Algorithmen und Zelltracking-Frameworks sind frei zugänglich und in vielen Evaluationen werden unterschiedliche Fehlermaße genutzt, was einen Vergleich weiter erschwert.

Die Referenzdaten CeTReS und die vorgestellten Fehlermaße aus Abschnitt 3.1 liefern einen ersten Schritt zu einer einheitlichen und vergleichbaren Evaluation von Zelltracking Algorithmen.



**Abbildung 4.17.:** Phasenkontrastbilder aus Huh et al. [48, Abb. 8 und 9].

##### 4.4.1. Zusammenfassung der Mitosedetektion

Die Voraussetzung für die vorgestellten Mitosedetektionen und die ausführliche Evaluation waren die Referenzaufnahmen aus Abschnitt 3.2.1: Mit den validierten Datensätzen CeTReS.A und CeTReS.B standen zwei Bildserien zur Verfügung, in denen alle Zellpfade korrigiert waren und alle Mitosen bzw. mitotischen Zellpfade direkt extrahiert werden konnten. Die folgenden zwei Punkte fassen die wichtigsten Ergebnisse zusammen, die mit Hilfe der gelabelten Bildserien erzielt werden konnten:

- Die Auswirkung der Zelldichte auf das morphologische Muster einer Mitose konnte mit dem Klassifikator  $K_{(3p-)mfml}$  detailliert untersucht werden. Mit der Bildserie zu dem Datensatz CeTReS.A konnte gezeigt werden, dass sich dieses Muster abhängig von den Wachstumsphasen der Zellpopulation ändert.
- Die Bildserien zu dem Datensatz CeTReS.B zeigen die beobachtete Zellpopulation als Schräglicht- und Phasenkontrastaufnahme. Daher war es mit diesem Datensatz erstmals möglich, eine morphologische Mitosedetektion mit einer Detektion des Mitosemusters  $K_{svm}$  in Phasenkontrastbildern zu vergleichen.

Da sowohl die verwendeten Bilddaten als auch die erstellten Referenzdaten unter der Creative Commons Lizenz veröffentlicht wurden und frei zugänglich sind, können diese leicht für eine vergleichbare Evaluation von weiteren Algorithmen genutzt werden.

##### Die morphologische Mitosedetektion in den Schräglichtbildern

Mit den vorgestellten Mitosedetektionen  $K_{sfml}$  und  $K_{(3p-)mfml}$  wurde untersucht, wie verlässlich die Änderung der Zellmorphologie genutzt werden kann, um eine Zellteilung zu finden. Hierzu wurden mit den Klassifikatoren  $K_{sfml}$  und  $K_{mfml}$  die Rundheit, Länge, Größe und Helligkeit einer Zelle genutzt, um eine Mitose zu detektieren (Abb.4.7 und Tab. 4.1). Mit Hilfe der Bildserie CeTReS.A und dem Klassifikator  $K_{3p-mfml}$  wurde gezeigt, dass das Muster einer Mitose abhängig von der Zelldichte der beobachteten Zellpopulation ist (Abb. 4.7 und Abb. 4.8). Diese Ergebnisse zeigen, dass mit einer Adaption der verwendeten Parameter an die Wachstumsphasen die Detektionsrate verbessert werden kann.

##### Detektion des Mitosemusters in Phasenkontrastbildern

Die Doppelkontrastaufnahmen aus dem erstellten Benchmark CeTReS.B wurde genutzt, um die morphologische Mitosedetektion mit der Detektion des typischen Mitosemusters im Phasenkontrast zu vergleichen, wie auf Seite 81 zusammengefasst. Hier zeigt sich, dass es ein eindeutiges Muster einer Mitose gibt, dass sehr gut mit einer SVM gelernt werden kann: Selbst mit einer kleinen Trainingsmenge von  $\approx 50$  Mitosen war es möglich, die vorhandenen Mitosen mit einer sehr hohen Genauigkeit von  $AUC > 0.99$  zu finden (vgl.4.12).

Um dieses Muster zu visualisieren, wurden mit einer Hauptkomponentenanalyse die stärksten Merkmale der Mitosen extrahiert, die sogenannten Eigenmitosen (vgl. Abb. 4.14). Diese Zerlegung liefert zusätzlich einen weiteren, allgemeinen Ansatz für

eine Mitosedetektion: Die Eigenmitosen können genutzt werden, um mit Hilfe einer musterbasierten Suche (*pattern matching*) Zellteilungen in Bildserien zu finden. Aufbauend auf den Eigenmitosen kann entweder eine eigene Mitosedetektion implementiert werden oder diese Muster werden genutzt, um mögliche Mitosekandidaten aus einer unbekannt Bildserie zu extrahieren.

### Optimierung von Zelltracking-Algorithmen

Wie zu Beginn des Kapitels beschrieben, lieferte die Mitosedetektion den letzten, fehlenden Baustein der vorgestellten Zelltracking-Algorithmen: Nur mit einer zuverlässigen Detektion der Zellteilungen ist es möglich, die vollständigen Pfade mit einer hohen Genauigkeit zu extrahieren. Die Auswirkungen, die sowohl die Mitosedetektion als auch die vorgestellte Validierung aus Kap. 3 auf die Genauigkeit der rekonstruierten Zellpfade haben, sind in Abb. 4.16 zusammengefasst.

In dieser Auswertung ist auch die Evaluation des fluiden Trackings zusammengefasst, die zwei wesentliche Ergebnisse enthält: Zum einen wurde gezeigt, dass sich die vorgeschlagene Datenvalidierung leicht auf andere Tracking-Algorithmen übertragen lässt. Die zweite Erkenntnis ist, dass auch für das fluide Tracking die Mitosedetektion in den Phasenkontrastbildern entscheidend für die Rekonstruktion der Pfade bzw. Stammbäume ist. Dies wird besonders deutlich, wenn der Fehler betrachtet wird, mit dem die Zellpfade gefunden werden: Wird das fluide Tracking mit der naiven Mitosedetektion verwendet, werden die vollständigen Pfade mit einer Genauigkeit von 80% gefunden, d.h. mit einem Fehler von 20%. In Kombination mit der Mitosedetektion im Phasenkontrast sinkt dieser Fehler auf 4%, was einem Fünftel des vorherigen Wertes entspricht.

Insgesamt lassen sich die folgenden drei Punkte zusammenfassen, die für beide analysierten Tracking-Algorithmen gezeigt wurden:

1. Mit den vorgestellten Referenzdaten kann eine Evaluation der eingesetzten Zelltracking-Algorithmen durchgeführt werden, wodurch es leicht möglich ist, systematische Fehler zu identifizieren, zu beheben und die benötigten Parameter anzupassen.
2. Mit Hilfe der Validierung ist es leicht möglich, die fehlerhaften Zellpfade zu identifizieren und aus der Auswertung zu entfernen.
3. Die Kombination von einer Zelldetektion bzw. einem Zelltracking in den Schräglichtbildern mit einer Mitosedetektion in den Phasenkontrastbildern verbessert die erzielte Genauigkeit, mit der die Zellpfade rekonstruiert werden.

Mit dem Zelltracking, der Validierung, den Referenzdaten und der Mitosedetektion ist insgesamt ein leistungsfähiges Framework vorgestellt worden, das direkt für eine sehr robuste Auswertung von Langzeitaufnahmen eingesetzt werden kann.

# 5

## Anwendungen

Als Schwerpunkt dieser Arbeit wurde eine bildbasierte Rekonstruktion von zellulären Genealogien erstellt, die auf einer Zelldetektion in einer Serie von Mikroskopbildern aufbaut. In diesem Kapitel werden verschiedene wissenschaftliche Anwendungsszenarien gezeigt, in denen diese Methode eingesetzt wurde.

Zuerst werden zwei Langzeitexperimente vorgestellt, die mit dem automatisierten Zelltracking aus Abschnitt 3.1.3 ausgewertet wurden. Die letzten beiden Anwendungen zeigen dagegen die Vielseitigkeit der Zelldetektion aus Abschnitt 2.4.

Die erste Anwendung beschäftigt sich mit der morphologischen Charakterisierung von apoptotischen bzw. nekrotischen Zellen. Hierzu wurde die HeLa-Zelllinie aus Abschnitt 2.3.5 genutzt, um die zeitliche Veränderung der Zellmorphologie mit einer fluoreszenzbasierten Apoptosedetektion zu vergleichen (vgl. Abschnitt 2.3.5). Die Resultate zeigen, dass die Zelldetektion in den Langzeitexperimenten optimal für toxikologische Untersuchungen und als Vitalitätstest eingesetzt werden kann.

In dem darauf folgenden Abschnitt wird die Zellmigration von humanen, adulten Stammzellen und von Stammzellen aus der Ratte untersucht. Der Fokus der Auswertung liegt auf der Bewegungsrichtung der Zellen vor und nach einer Teilung. Die Idee besteht darin, mit diesem Muster eine Vorhersage der Bewegung der Tochterzellen zu erstellen. Diese Information kann wiederum genutzt werden, um die richtigen Zellen vor und nach einer Mitose zu verknüpfen, d.h. die richtigen Verwandtschaften zuzuordnen.

Im dritten Abschnitt wird die Hautregeneration in einem Wundmodell mit einer automatisierten Auswertung von Mehrkanal-Fluoreszenzbildern quantifiziert. Die ausgewerteten Bilder sind in einem Tierversuch an der Fraunhofer EMB entstanden, in dem erstmalig humane *in vitro*-kultivierte, Schweißdrüsen-abgeleitete adulte Stammzellen zur Unterstützung der Wundregeneration einer Maus eingesetzt wurden. Die Ergebnisse dieser Studie sind in der gemeinsamen Arbeit mit Danner et al. [6] zu finden.

Abschließend wird ein Kooperationsprojekt mit der Arbeitsgruppe für Integrierte Systeme und Photonik der Christian-Albrechts-Universität zu Kiel (CAU) vorgestellt. Prof. M. Gerken und Dr. Y. Nazirizadeh haben hier eine neuartige Mikroskopietechnik zur Kontrastverbesserung entwickelt, die als Oberflächenkontrastmikroskopie bezeichnet wird (Nazirizadeh et al. [83]). Mit Hilfe dieser Technik sind Aufnahmen von adhären wachsenden Stammzellen erstellt und mit den implementierten Algorithmen zur Zelldetektion aus Abschnitt 2.4 systematisch ausgewertet und evaluiert worden. Die Ergebnisse sind in der gemeinsamen Arbeit von Nazirizadeh et al. [7] veröffentlicht.

## 5.1. Zelltracking zur Vitalitätsanalyse von Zellpopulationen

Eine Einsatzmöglichkeit der Langzeitexperimente liegt in der Echtzeitüberwachung von proliferierenden Stammzellpopulationen. In diesem Anwendungsszenario wird eine fortlaufende Bildauswertung genutzt, um verschiedene Zellparameter wie z.B. Proliferation, Zellgröße, Migration oder das Teilungsverhalten der Zellen zu extrahieren. Diese Analyse erlaubt es, das Zelltracking als ein Online-Monitoring-System zu benutzen, das immer den aktuellen Zustand der Zellen beschreibt und mit dem z.B. die Konfluenz einer Zellpopulation durchgehend überwacht werden kann (Ker et al. [55]). Eine weitere Anwendung liegt in toxikologischen Untersuchungen, in denen die Auswirkung von Substanzen und Substraten auf *in vitro*-kultivierte Zellen untersucht wird (Eisenbrand et al. [34]). Mit Hilfe von Langzeitbeobachtungen ist es möglich, die ausgelesenen Parameter um die Migration, Proliferation oder Morphologie der Zellen zu erweitern (Lin et al. [71]), oder gezielt nach toten, apoptotischen Zellen zu suchen (Huh et al. [49]). Der folgende Abschnitt beschäftigt sich mit der Charakterisierung bzw. der Analyse eines induzierten Zelltodes auf das Verhalten von Zellpopulationen und untersucht die Auswirkung einer Apoptose bzw. Nekrose auf die HeLa-Krebszelllinie.

### 5.1.1. Induktion und Nachweis des Zelltodes (Apoptose und Nekrose)

Der Tod einer Zelle unter *in vivo*-Bedingungen, d.h. in einem lebenden Organismus, ist ein essentieller Bestandteil der Entwicklung (Ontogenese), der Aufrechterhaltung und dem Funktionieren des Organismus selbst (Homöostase) (Fadeel and Orrenius [35]). Mit dem programmierten Zelltod, der **Apoptose**, wird ein Mechanismus bereitgestellt, der es erlaubt, krankhafte, veränderte oder nicht mehr benötigte Zellen zu entfernen (Kerr et al. [56]). Die Apoptose wird durch die Zelle selbst als Teil des Stoffwechsels durchgeführt und kann durch externe Signale (z.B. das Andocken einer Immunzelle) oder durch interne Prozesse (z.B. wenn eine starke Schädigung der DNA festgestellt wird) ausgelöst werden. Dagegen ist der Tod einer Zelle in einer *in vitro*-Kultur von proliferierenden, adulten Stammzellen oftmals ein Anzeichen ungeeigneter Kulturbedingungen oder einer krankhaften Veränderung der Zellen, die z.B. während einer toxikologischen Untersuchung induziert wird. Diese unregulierte Form des Zelltodes wird als **Nekrose** bezeichnet.

Im Folgenden wird das Zelltracking als eine nicht-invasive Methode zur Bestimmung der Zellvitalität eingesetzt. Um herauszufinden, wie sensitiv die Auswertung der Bildserien ist und in welcher zeitlichen Auflösung eine toxikologische Reaktion erkannt wird, wurde eine induzierte Apoptose und eine induzierte Nekrose beobachtet. Mit der Auswertung dieser Langzeitbeobachtungen sollte herausgefunden werden, wie schnell eine Apoptose nachgewiesen werden kann bzw. ob der Nachweis schneller gelingt, als es mit einer Standardmethode möglich ist. Hierzu wurden die Ergebnisse mit einer Annexin V-FITC Färbung verglichen, die sich als Standardnachweis einer Apoptose etabliert hat. Ein Vergleich der beiden Methoden ist schwierig, da eine Annexin V-FITC Färbung nur zu konkreten Zeitpunkten durchgeführt werden kann, während das Zelltracking eine fortlaufende Überwachung ermöglicht. In dem verwendeten Aufbau wurden daher zeitgleich mit den Langzeitexperimenten weitere Zellpopulationen kultiviert, in denen ebenfalls eine Apoptose bzw. Nekrose induziert wurde. Diese Populationen wurden für

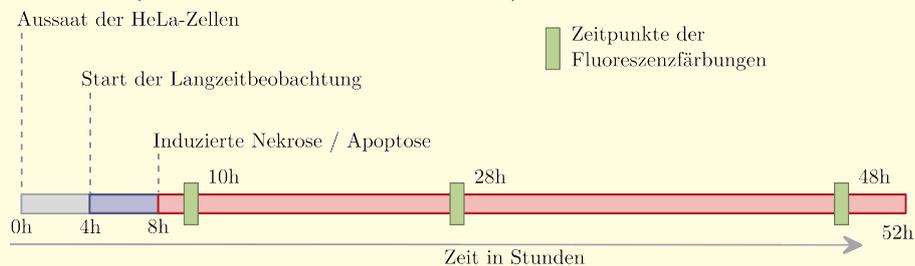
die Annexin V-FITC Färbungen verwendet.

### Induktion der Apoptose und Nekrose

Die Untersuchungen zu dem induzierten Zelltod wurden mit Hilfe der HeLa-Zelllinie durchgeführt, die als Grundlage für viele Zellexperimente genutzt wurde (Chen et al. [20], Hadjidemetriou et al. [42], Zhou et al. [125]). Die durchgeführten Langzeitbeobachtungen sind in Abb. 2.12 bzw. auf Seite 26 in Kap. 2 beschrieben, werden hier nochmals zusammengefasst und um eine Fluoreszenzanalyse erweitert.

In Abb 5.1 ist der schematische Ablauf der Experimente dargestellt. Die Zellen wurden zum Zeitpunkt  $t = 0h$  ausgesät und in einem Brutschrank vier Stunden lang inkubiert, bevor das Langzeitexperiment gestartet wurde. Zum Zeitpunkt  $t = 8h$  wurde die Aufnahme gestoppt, um eine Apoptose durch Zugabe von Staurosporin (STS) bzw. eine Nekrose mit  $H_2O_2$  zu induzieren. Anschließend wurde die Langzeitbeobachtung weitere  $40h$  fortgeführt. Insgesamt stand ein Zeitraum von 48 Stunden zur Verfügung, in denen die Änderung der Morphologie, Migration und Proliferation untersucht werden konnte.

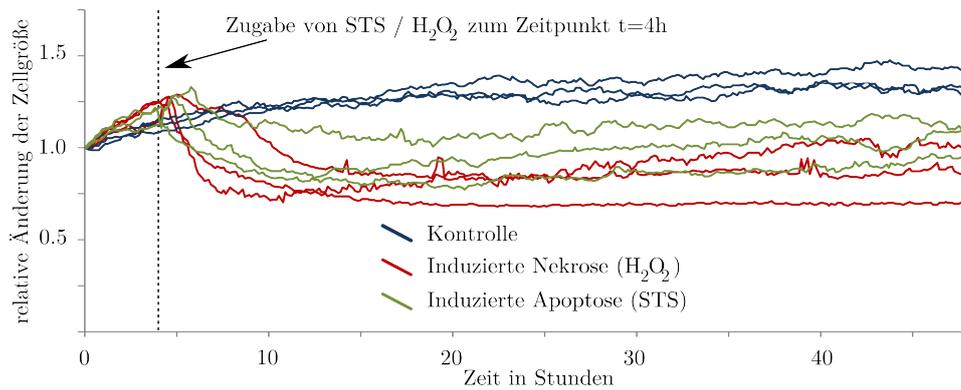
Das Ziel der Experimente war es herauszufinden, ob eine Apoptose in diesen Langzeitbeobachtungen schneller nachgewiesen werden kann als es mit etablierten Standardmethoden möglich ist. Hierzu wurde eine invasive Annexin V-FICT-Färbung genutzt, um die Zahl der toten Zellen zu unterschiedlichen Zeitpunkten ( $t = 10h$ ,  $t = 28h$  und  $t = 48h$ ) zu bestimmen.



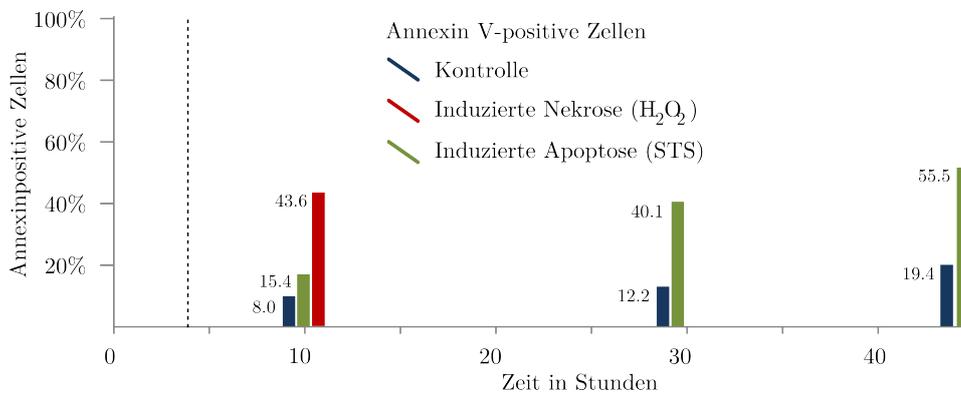
**Abbildung 5.1.:** Zeitlicher Ablauf der Experimente. Jede Langzeitbeobachtung ist für die induzierte Apoptose, Nekrose und das Kontrollexperiment dreimal wiederholt worden. Es wurden jeweils vier Aufnahmepositionen gewählt, so dass insgesamt 36 Bildserien ausgewertet wurden.

### Fluoreszenzfärbung apoptotischer Zellen

Um eine Apoptose oder Nekrose bei Zellen nachzuweisen, hat sich die Fluoreszenzfärbung mit Annexin V als Standardmethode etabliert. (Koopman et al. [59], Vermes et al. [117]). Dieses Protein bindet in Anwesenheit von Calcium mit einer großen Spezifität reversibel an Phosphatidylserine (PS), das sich bei vitalen Zellen auf der Innenseite der



(a) Auswertung der Langzeitbeobachtung (Morphologieänderung).



(b) Fluoreszenzanalyse (Annexin V-FITC Färbung).

**Abbildung 5.2.:** Die Zelldetektion bzw. das Zelltracking ermöglicht es, die beobachteten Zellen als parametrisierte Zellkultur zu beschreiben. In (a) ist die (Änderung der) Zellgröße dargestellt, die direkt nach der Zugabe von STS oder H<sub>2</sub>O<sub>2</sub> einen markanten Abfall zeigt. Die Auswertung der Annexin V-FITC-Färbung ist in (b) zu drei unterschiedlichen Zeitpunkten angegeben (10h, 28h, 48h nach Aussaat der Zellen).

Zellmembran befindet und im Frühstadium der Apoptose auf die Außenseite transloziert wird (Haanen and Vermees [41]). Durch eine Kopplung von Annexin V mit dem Fluoreszenzfarbstoff Fluoresceinisothiocyanat (FITC) werden alle apoptotischen Zellen mit einer Annexin V-FITC Färbung markiert. Die Auswertung der gefärbten Zellen kann mit einer Fluoreszenzanalyse (Fluoreszenzmikroskopie oder quantitative Auswertung mittels Durchflusszytometrie, FACS) durchgeführt werden (vgl. Abschnitt 2.1.2). In Abb. 5.2(b) sind die Ergebnisse einer Fluoreszenzanalyse angegeben, die zu unterschiedlichen Zeitpunkten durchgeführt worden sind (10h, 28h, 48h nach Aussaat der Zellen). Die Fluoreszenzanalyse der nekrotischen Zellen wurde nur zum Zeitpunkt  $t = 10h$  durchgeführt. Zu den späteren Messpunkten hatten sich bereits alle Zellen vom Boden des Kulturgefäßes gelöst, d.h. die Zellmembran war zerstört, die Zellen abgestorben und eine weitere Kultivierung war somit nicht mehr möglich.

### 5.1.2. Vergleich Fluoreszenzfärbung vs. Morphologieänderung

Um die aufgenommenen Experimente auszuwerten und als parametrisierte Zellkultur zu visualisieren, ist aufbauend auf dem validierten Zelltracking aus Abschnitt 3.1.3 eine grafische Benutzeroberfläche erstellt worden, die es erlaubt, alle benötigten Berechnungen durchzuführen und die gewünschten Parameter zu visualisieren (vgl. Abb A.3 im Anhang). Dieses Programm wurde genutzt, um die aufgenommenen Experimente auszuwerten. Die Experimente (Kontrolle, induzierte Apoptose und induzierte Nekrose) wurden dreimal wiederholt und es wurden jeweils vier Aufnahmepositionen gewählt, so dass insgesamt 36 Bildserien ausgewertet wurden. In Abb. 5.2(a) ist die relative Zellgröße abhängig von der Zeit für die verschiedenen Experimente gezeigt, d.h. dass die Zellgrößen zum Startzeitpunkt der Langzeitexperimente auf 1 normiert wurde. Ein Vergleich mit der Fluoreszenzanalyse aus Abb. 5.2(b) zeigt einen deutlichen Unterschied in der Reaktionszeit, bis die Auswirkung der induzierten Apoptose bzw. Nekrose gemessen werden konnte. Die Zugabe von STS oder  $H_2O_2$  hat direkt Einfluss auf die Zellmorphologie: Unmittelbar nach der Zugabe fällt die mittlere Zellgröße um 30% ab. Dagegen zeigt die Fluoreszenzanalyse bei der induzierten Apoptose einen langsameren Anstieg der Annexin V-positiven Zellen als bei der induzierten Nekrose. Dieser unterschiedlich starke Anstieg ist mit der  $H_2O_2$  induzierten Nekrose zu erklären, da das  $H_2O_2$  die Zellmembranen zerstört hat. Somit wurde das Protein PS an der Innenseite der Zellmembran zugänglich, d.h. die Annexin V-Färbung hat jede nekrotische Zelle nachgewiesen.

### Zelltracking zur Vitalitätsmessung

Mit den präsentierten Analysen wurde gezeigt, dass sich das Zelltracking-Framework als eine nicht-invasive Methode zur Messung der Zellvitalität einsetzen lässt. Es ist leicht möglich, die Auswertung der Langzeitexperimente um weitere Parameter wie z.B. die Zellmigration zu erweitern. Doch auch schon diese vorläufigen Ergebnisse zeigen, dass die automatisierte Auswertung der Bildserien als sehr schneller Indikator genutzt und somit als Basistechnologie für einen nicht-invasiven Vitalitätstest oder für toxikologische Untersuchungen eingesetzt werden kann.

## 5.2. Migrationsstudie mitotischer Zellen

Die zweite Studie beschäftigt sich mit dem wichtigsten Ereignis einer *in vitro*-kultivierten und proliferierenden Zellpopulation: Der Teilung von Zellen. Mit jeder Zellteilung wächst die Population und die Komplexität des betrachteten Szenarios vergrößert sich, indem ein neuer Familienzweig im Stammbaum begonnen wird. Die zuverlässige Detektion einer Zellteilung ist daher ein essentieller Bestandteil in dem vorgestellten Zelltracking-Framework und wurde ausführlich in Kapitel 4 behandelt. Mit der gegebenen Detektion der Mitosen in den Phasenkontrastbildern aus Abschnitt 4.3 ist es möglich, die Zellteilungen mit einer sehr hohen Genauigkeit und Zuverlässigkeit zu identifizieren. Neben der Detektion der Zellen liegt aber eine weitere, fehleranfällige Aufgabe in der korrekten Zuweisung der beiden Tochterzellen. Da sich eine Zelle während der Teilung zusammenzieht und zum Teil von der Oberfläche der Kulturschale ablöst, zeigen viele

Zellen schnelle, ruckartige Bewegungen. Diese hohe Motilität führt dazu, dass es gerade vor und während einer Zellteilung zu vielen Fehlsegmentierungen kommt. Besonders in Zellpopulationen mit einer hohen Konfluenz, wenn die Zellen sehr dicht beieinander liegen, ist es leicht möglich, dass mitotische Zellen mit Nachbarzellen optisch *verschmelzen* und nicht mehr als disjunkte Objekte erkannt werden können. Das folgende Kapitel geht der Frage nach, welchen Einfluss die Bewegungsrichtung einer Zelle auf das Migrationsverhalten der Tochterzellen hat. Oder anders formuliert: Kann die Bewegungsrichtung der Tochterzellen aus der Bewegung der Mutterzellen vorhergesagt werden? Um das zu beantworten wird die Bewegung einer Zelle vor und nach einer Teilung systematisch untersucht. Die bisher unveröffentlichten Ergebnisse zeigen, dass es eindeutige Muster der Zellen gibt: So teilt sich z.B. eine Zelle entlang der Bewegungsrichtung und die beiden Tochterzellen bewegen sich oftmals in die entgegengesetzte Richtung der Mutterzelle. In der folgenden Untersuchung wurden zwei Bildserien ausgewertet, die eine humane (HDF) und eine Zellpopulation der Ratte (rPSC) zeigen.

### 5.2.1. Verwendete Datensätze

Für die Analyse der Migrationsrichtungen wurde eine Bildserie von humanen Fibroblasten (*human dermal fibroblast*, HDF, vgl. Abschnitt 2.3.4) verwendet, die bereits in der Arbeit von Rapoport/Becker et al. [8, Suppl. Fig. 2] <sup>1</sup> ausgewertet wurde. Die Zellpopulation wurde über einen langen Zeitraum von 10 Tagen beobachtet und es wurden insgesamt 1411 Bilder erstellt; eine genaue Beschreibung des durchgeführten Langzeitexperimentes ist in 2.3.4 zu finden. Die Aufnahmen wurden manuell korrigiert und es wurde ein validierter Zelltracking-Datensatz erstellt, analog zu den Referenzdaten CeTReS.A und CeTReS.B, vgl. Abschnitt 3.2.1. Die beobachteten HDF-Zellen werden durch eine langsamere Proliferation und einen größeren Zellkörper als die bisher genutzten Stammzellen (rPSC) charakterisiert. Daher war es möglich, die ursprüngliche Aufnahmefrequenz von zehn Minuten auf 30 Minuten auszudehnen. Da die ursprüngliche Bildserie mit zehn Minuten Abstand aufgenommen wurde, musste nur  $\frac{1}{3}$  der Bilder korrigiert werden. Somit wurde die Anzahl der zu überprüfenden Bilder von 1411 auf 471 reduziert und der Arbeitsaufwand entsprechend verringert. Der korrigierte Datensatz enthielt insgesamt 1137 Zellen und 274 Zellteilungen, die als Grundlage für die Analyse der Migrationsrichtungen genutzt wurden.

Als weiterer Datensatz wurde der Referenzdatensatz CeTReS.A genutzt, der adulte Stammzellen aus dem Rattenpankreas zeigt, vgl. Abschnitt 2.3.1. Diese Bildserien wurden gewählt, da hier auf eine große Anzahl von Mitosen zurückgegriffen werden konnte (2019 Mitosen). In Tab. 5.1 ist eine Übersicht der beiden ausgewerteten Bildserien gegeben.

Bei der Interpretation der Daten sollte beachtet werden, dass die Bildserien zwei Zellpopulationen mit stark unterschiedlichen Zelldichten zeigen. So erreichten die Zellen der CeTReS.A-Bildserie eine fast vollständige Konfluenz, in der eine gerichtete Migration nicht mehr möglich war. Die Population der HDF-Zellen erreichte aufgrund der langsameren Proliferation selbst nach zehn Tagen Beobachtungszeit keine Konfluenz, so dass sich diese Zellen ungehinderter bewegen konnten.

---

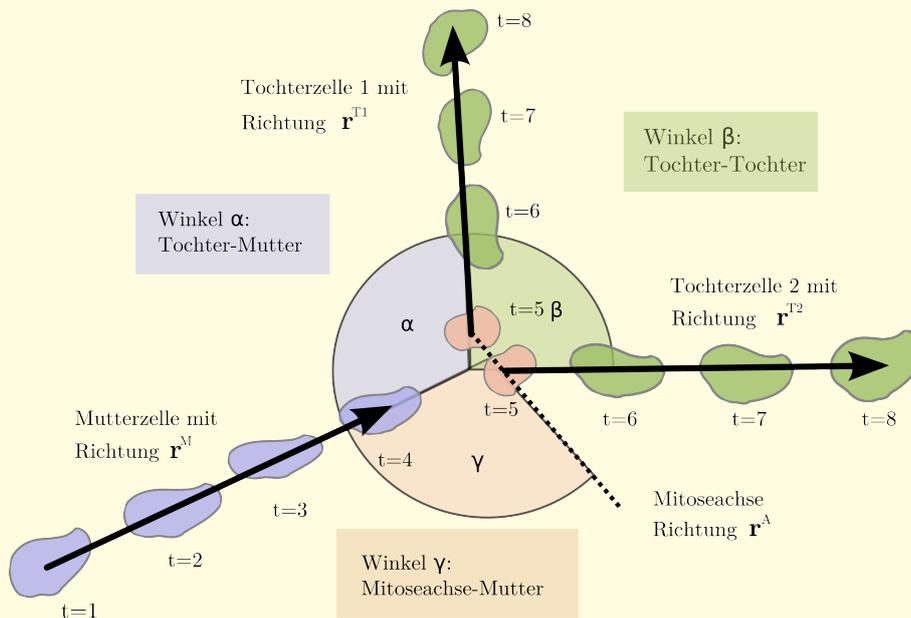
<sup>1</sup>doi: 10.1371/journal.pone.0027315.s002

### Bestimmung der Migrationsrichtung einer mitotischen Zelle

Die rekonstruierten Zellpfade werden genutzt, um die Bewegungsrichtung der Zellen vor und nach einer Mitose zu ermitteln. Hierbei wurde die Migration als Richtung der Strecke berechnet, die eine Zelle in  $n$  Zeitschritten vor und nach einer Mitose zurücklegt, wie in Abb. 5.3 dargestellt. Die Richtungen für die Mutter- bzw. Tochterzellen werden als  $\mathbf{r}^M$ ,  $\mathbf{r}^{T_1}$  und  $\mathbf{r}^{T_2}$  bezeichnet. Zusätzlich wird die Mitoseachse  $\mathbf{r}^A$  als die Gerade definiert, die durch die Mittelpunkte der Tochterzellen verläuft, wenn diese erstmalig als einzelne Zellen detektiert werden, siehe Abb. 5.3. Basierend auf den Bewegungsrichtungen  $\mathbf{r}^M$ ,  $\mathbf{r}^{T_1}$ ,  $\mathbf{r}^{T_2}$  und  $\mathbf{r}^A$  wird eine Mitose über die Winkel  $\alpha, \beta, \gamma \in [0^\circ, 180^\circ]$  charakterisiert, die folgenden Winkel angeben:

- $\alpha$  gibt den Winkel zwischen einer Tochterzelle und der Mutterzelle an,
- $\beta$  gibt den Winkel zwischen den Tochterzellen an,
- $\gamma$  bezeichnet den Winkel zwischen der Mitoseachse und der Mutterzelle.

Aufbauend auf dieser Migrationsanalyse wurde das Verhalten der mitotischen Zellen von zwei korrigierten Datensätzen analysiert. Hierzu wurde eine humane Zellpopulation und der Datensatz CeTRE.S.A aus Abschnitt 3.2.1 genutzt.



**Abbildung 5.3.:** Definition der Migrationsrichtungen, der Mitoseachse und der daraus resultierenden charakteristischen Winkel.

### 5.2.2. Gerichtete Migration mitotischer Zellen

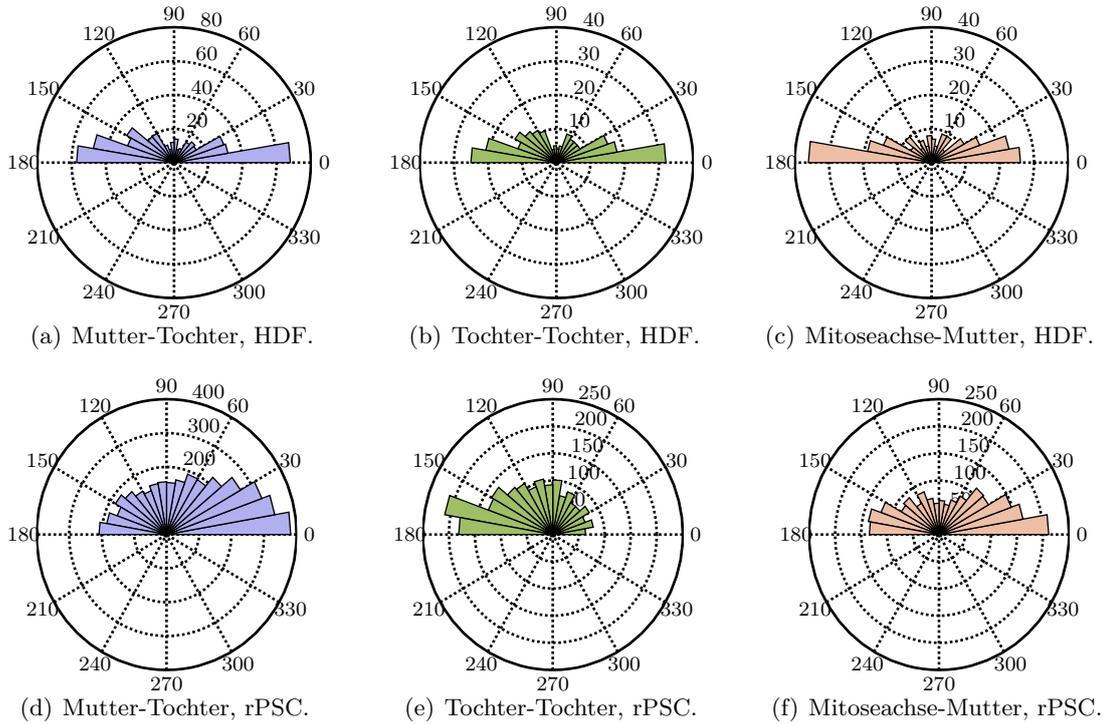
Um die Migration von mitotischen Zellen zu untersuchen, wurden die rekonstruierten Stammbäume und Zellpfade der korrigierten HDF-Bildserie und der Referenzdaten CeTReS.A genutzt. Für alle Zellen, die vor und unmittelbar nach der Mitose über einen Zeitraum von fünf Stunden / zehn Bilder (HDF-Daten) bzw. zwei Stunden / acht Bilder (CeTReS.A-Daten) verfolgt werden konnten, wurden die charakteristischen Winkel  $\alpha$ ,  $\beta$  und  $\gamma$  berechnet, wie in Abb. 5.3 angegeben. Insgesamt wurden so 222 Mitosen der Bildserie HDF und 1856 Mitosen der Bildserie CeTReS.A analysiert. Im Folgenden eine Auflistung der Resultate:

1. Der Winkel  $\alpha$  zwischen der Mutter- und den Tochterzellen zeigt in den HDF-Daten zwei eindeutige Häufungspunkte der Winkelverteilungen bei  $0^\circ$  und  $180^\circ$ , vgl. Abb. 5.4(a). Diese Verteilung sagt aus, dass die Tochterzellen mit großer Wahrscheinlichkeit die Bewegung der Mutter fortsetzen, oder in genau die entgegengesetzte Richtung wandert. Die Verteilung von Winkel  $\alpha$  ist für die rPSC-Zellen der CeTReS.A-Daten leicht unterschiedlich: Hier wandern die Zellen meist in die Richtung, aus der die Mutterzelle kam (vgl. Abb. 5.4(d)). Ein Grund hierfür könnte die hohe Zelldichte sein, die eine (gerichtete) Bewegung erschwert. Bei einer vorausgesetzten hohen Konfluenz wäre es sogar eine logische Schlussfolgerung, dass sich die Zelle in die Richtung bewegt, aus der die Mutterzelle kam: Hier hat die Mutter schon etwas Platz geschaffen bzw. hinterlassen.
2. Die Verteilung der Werte von Winkel  $\beta$  der HDF-Zellen ist in Abb. 5.4(b) gezeigt. Hier wird deutlich, dass sich die beiden Tochterzellen mit gleicher Wahrscheinlichkeit in dieselbe Richtung bewegen, d.h. es gilt  $\beta = 0^\circ$ , oder sie bewegen sich mit einem Winkel von  $\beta = 180^\circ$  in genau die entgegengesetzte Richtung. Diese Winkelverteilung ist bei den rPSC-Zellen anders, siehe Abb. 5.4(e). Hier liegt der Häufungspunkt bei  $180^\circ$ , d.h. die Zellen bewegen sich mit Vorliebe in entgegengesetzte Richtungen.
3. Der Winkel  $\gamma$  gibt die Teilungsachse der Mitose an, wie in Abb. 5.4(c+f) anschaulich dargestellt. Hier zeigen die beiden analysierten Bildserien ein sehr ähnliches Verhalten: Die Teilungsachse verläuft wie erwartet genau entlang der Bewegungsrichtung der Mutterzelle.

#### Zusammenfassung der Migrationsanalyse

Die analysierten Winkelverteilungen in Abb. 5.4 zeigen, dass die Bewegungsrichtung einer Zelle vor und nach einer Mitose korreliert. Die Bewegungsprofile der HDF-Population zeigen eine eindeutige Winkelverteilung der drei Winkel  $\alpha$ ,  $\beta$  und  $\gamma$  mit jeweils zwei Häufungspunkten bei  $0^\circ$  und  $180^\circ$ . Das bedeutet, dass eine der Tochterzellen mit großer Wahrscheinlichkeit die Bewegung der Mutterzelle fortsetzt, während die andere in die entgegengesetzte Richtung läuft.

## 5.2. Migrationsstudie mitotischer Zellen



**Abbildung 5.4.:** In (a-c) sind die Verteilungen der verschiedenen Winkel einer Mitose für eine Bildserie mit humanen, dermalen Fibroblasten (HDF, obere Reihe) und in (d-f) für eine Bildserie mit pankreatischen Stammzellen der Ratte (rPSC, Datensatz CeTReS.A, untere Reihe) angegeben. Die HDF-Bildserie enthält 274 Mitosen und zeigt die Zellen bei einer geringen Konfluenz, während der rPSC-Datensatz 2019 Mitosen enthält, die zum Ende der Aufnahme einer sehr hohen Konfluenz aufweisen. Die Auswertungen der HDF-Zellen (obere Reihe) waren in ihrer Bewegung folglich weniger eingeschränkt als die rPSC (untere Reihe). Die Migration wurde für alle Mitosen ausgewertet, die über einen Zeitraum von fünf Stunden (HDF) bzw. zwei Stunden (rPSC) verfolgt werden konnten. Hierdurch ergab sich eine Zahl von 222 (HDF) und 1856 (rPSC) analysierten Mitosen.

Zellen	Bilder (korr. Daten)	Zeitintervall (korr. Daten)	Mitosen	mittlere Geschw.	mittlere Zellgröße
HDF	1411 (471)	10 (30) Min.	274	$16.8\mu\text{m}/\text{h}$	$1040\mu\text{m}^2$
rPSC (CeTReS.A)	209 (209)	15 (15) Min.	2019	$9.3\mu\text{m}/\text{h}$	$484\mu\text{m}^2$

**Tabelle 5.1.:** Ausgewertete Datensätze der Migrationsstudie.

Dagegen ist das Verhalten der rPSC-Zellen schwieriger zu interpretieren: In Abb. 5.4(d) ist gezeigt, dass sich die Tochterzellen hauptsächlich in die Richtung bewegen, aus der die Mutterzelle kam. Da es aber kaum Zellen gibt, die die entgegengesetzte Richtung  $\alpha = 180^\circ$  einschlagen, hat es den Anschein, als ob sich die Tochterzellen in dieselbe Richtung bewegen. Diese Annahme widerspricht wiederum den Daten aus Abb. 5.4(d), da hier gezeigt ist, dass sich die Tochterzellen in einem Winkel von  $180^\circ$  trennen und sich in genau die entgegengesetzte Richtung bewegen.

Wie oben erwähnt, ist die Interpretation der Daten sehr stark von der Zelldichte beeinflusst. Bei der hohen Zellzahl, die die rPSC-Population erreichte, war eine gerichtete Bewegung kaum noch möglich (vgl. Abb. 3.15 auf Seite 61). In diesem Szenario wäre es denkbar, dass die beobachtete Winkelverteilung ein Artefakt der Zellen selber darstellt: Nach einer Zellteilung wachsen die beiden Tochterzellen und vergrößern ihr Volumen, wodurch sich die Schwerpunkte der Zellen verschieben. Da beide Tochterzellen zwangsläufig sehr dicht nebeneinander liegen, führt die Vergrößerung der Zellen konsequenterweise dazu, dass sich die Schwerpunkte der Zellen entfernen. Das wiederum kann den beobachteten Häufungspunkt des Winkel  $\beta$  bei  $180^\circ$  erklären. Aus diesem Grund wurde der weitere Winkel  $\gamma$  betrachtet, der den Winkel zwischen der Mutterzellen und der Mitoseachse beschreibt. Dieser Winkel aus Abb. 5.4(f) zeigt eine recht ähnliche Anordnung der Mitoseachse zu der Bewegungsrichtung der Mutterzellen und bestätigt die getroffene Annahme, dass die Beobachtung in Abb. 5.4(e) eine Auswirkung der hohen Zelldichte war.

Insgesamt zeigt sich, dass neben der Änderung der Zellmorphologie (Abschnitt 4.2) und dem eindeutigen Mitosemuster im Phasenkontrast (Abschnitt 4.3) noch ein Bewegungsprofil einer Zellteilung extrahiert werden konnte. Ob dieses Muster auch in anderen Zellpopulationen zuverlässig gefunden werden kann und wie genau der Zusammenhang zu der Zelldichte der beobachteten Population ist, sollte in weiterführenden Experimenten untersucht werden.

### 5.3. Oberflächenkontrastmikroskopie zur Zelldetektion

Der folgende Abschnitt stellt eine Studie zur Detektion von Zellen mit einer neuartigen Mikroskopietechnik vor, die in Kooperation mit Dr. Y. Nazirizadeh durchgeführt wurde (Nazirizadeh et al. [7]).

Das Ziel dieser Studie war, kontrastreiche Mikroskopbilder zu erzeugen, die als Alternative zu der genutzten Schräglichtmikroskopie aus Abschnitt 2.2.1 eingesetzt werden können. Der Vorteil der verwendeten Oberflächenkontrastmikroskopie besteht darin, dass der Bildkontrast direkt auf der verwendeten Oberfläche erzeugt wird, was zu einem sehr homogenen Bild mit einem gleichmäßigen Kontrast führt.

Vor der Einführung der neuen Mikroskopietechnik wird zunächst das Grundproblem der mikroskopischen Darstellung von Zellen betrachtet. Da die Zellen einen sehr geringen Kontrast aufweisen und unter einem normalen Lichtmikroskop fast durchsichtig sind, müssen verschiedene kontrasterzeugende oder kontrastverstärkende Techniken verwendet werden, um sie zu visualisieren. Hierbei haben sich für die nicht-invasive Beobachtung von biologischen Proben der Phasenkontrast (Zernike [123]) und der Differentialinterferenzkontrast (*differential interference contrast*, DIC)(Allen et al. [10]) als Standardmethode etabliert. Beide Methoden ermöglichen es, die beobachteten Zellen anhand der Unterschiede der Brechungsindizes des umgebenden Mediums und der Zellkörper selbst sichtbar zu machen. Beide Methoden können folglich transparente Phasenobjekte mit einem Hell-Dunkel-Kontrast darstellen. Der Phasenkontrast nutzt hierbei den Phasenunterschied einer Lichtwelle, die ein Phasenobjekt durchläuft, während der DIC die Unterschiede der optischen Weglängen zweier Lichtstrahlen nutzt. Der Nachteil beider Methoden liegt jedoch zum einen in der Qualität der Zellbilder, die durch eine inhomogene Ausleuchtung und einen geringen Kontrast charakterisiert sind. Sie können verschiedene Artefakte beinhalten, die u.a. durch die verwendete Kulturschale oder die Bewegung des Zellmediums entstehen. Ein weiterer Nachteil liegt in dem benötigten Aufbau der Mikroskope, die komplexe und kostspielige Optiken erfordern.

Mit der Oberflächenkontrastmikroskopie (*surface contrast microscopy*, SCM) steht eine weitere Methode zur Verfügung, die es erlaubt, Phasenobjekte wie z.B. Zellen mit einem hohen Kontrast zu visualisieren. Dieses Verfahren wurde von Y. Nazirizadeh und M. Gerken an der Christian-Albrechts-Universität zu Kiel entwickelt und nutzt eine nano-optische Modifizierung der Kulturschalenoberfläche, um den Kontrast der beobachteten Zellen zu erzeugen (Nazirizadeh et al. [83]). Während die Phasenkontrastmikroskopie bzw. der Differentialinterferenzkontrast mit Hilfe der Optiken erzielt wird, liegt der kontrasterzeugende Mechanismus des SCM in der genutzten Oberfläche, wie in Abb. 5.5 dargestellt. Im Gegensatz zu den bisherigen Standardmethoden (Phasenkontrast- bzw. DIC-Mikroskopie, s.o.) besitzt diese Technik somit zwei wichtige Vorteile:

1. Der Kontrast entsteht in bzw. auf der verwendeten Oberfläche und kann mit einem herkömmlichen Lichtmikroskop kombiniert werden. Daher ist der Einsatz von komplexen (und teuren) Optiken nicht erforderlich.
2. Die SCM-Bilder sind durch einen sehr hohen Kontrast und einer guten Homogenität charakterisiert, die als Grundlage für eine zuverlässige Zelldetektion verwendet werden könnten.

Die folgenden Abschnitte beschreiben die Funktionsweise der SCM und stellen die Analyse der Zellbilder vor, die mit dieser Technik aufgenommen wurden. Die verwendeten Bilddaten sind an der CAU von Dr. Y. Nazirizadeh aufgenommen und mit den Algorithmen zur Zelldetektion aus Abschnitt 2.4 ausgewertet worden. Aufgrund des verwendeten experimentellen Aufbaus wurden die Mikroskopaufnahmen mit einer 40-fachen Vergrößerung als SCM- und Phasenkontrastbilder erstellt (ein Schräglichtkontrast wie in Abschnitt 2.2.1 beschrieben stand nicht zur Verfügung).

### 5.3.1. Zelldetektion in SCM Bildern

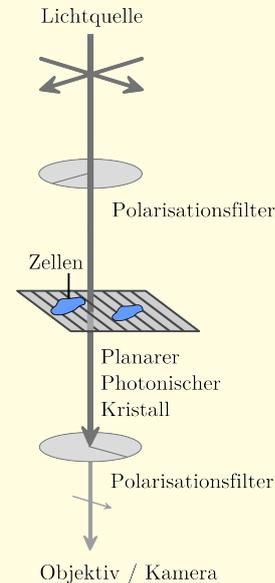
Um die Zelldetektion in SCM-Bildern zu untersuchen wurden Aufnahmen von embryonalen Fibroblasten der Ratte erstellt und mit Phasenkontrastbildern verglichen. Hierzu wurden die Fibroblasten auf einer PCS-Oberfläche (*photonic crystal slab*, PCS) ausgesät (vgl. Zusammenfassung auf Seite 103). Bevor die Aufnahmen mit der SCM erstellt werden konnten, mussten die Zellen fixiert werden, d.h. es war nicht möglich, lebende Zellen zu verwenden. Anschließend wurden 94 Positionen bzw. Ausschnitte auf der PCS-Oberfläche gewählt, an denen die fixierten Zellen abfotografiert wurden. Für die Aufnahmen wurde eine 20-fache Vergrößerung genutzt und von jeder Position wurden zwei Bilder mit einer Größe von  $1600 \times 1200$ px erstellt: Ein Phasenkontrast- und ein SCM-Bild. Da der verwendete Aufbau des Experiments eine höhere Vergrößerung nutzt als die Vergrößerung in Langzeitexperimenten ( $20\times$  in den SCM-Bildern und  $4\times$  in den Langzeitexperimenten, vgl. Abschnitt 2.2.1), wurde die Auflösung der Aufnahmen mit Hilfe der MATLAB Funktion `impyramid.m` skaliert, in der eine Gaußsche-Pyramiden Reduktion verwendet wird (Burt et al. [16]). In den so erstellten Bildern wurden die folgenden Segmentierungen genutzt, um die Zellen zu detektieren:

1. Da sowohl die SCM- als auch die Phasenkontrastbilder einen guten Kontrast zeigten, wurde die **Globale Schwellwertsegmentierung** aus Abschnitt 2.4.2 genutzt, um die Zellen zu detektieren (der Schwellwert wurde nach der Methode von Otsu berechnet (Otsu [84])).
2. Eine zweite Detektion wurde mit dem Chan-Vese-Algorithmus durchgeführt, der eine **aktive Kontur** zur Segmentierung verwendet (Chan and Vese [19]). Bei diesem Algorithmus wird die Kontur von einem Objekt als parametrisierte Kurve beschrieben und das Mumford-Shah-Funktional wird genutzt, um eine optimale Segmentierung bzw. Trennung des Vorder- und Hintergrunds zu erreichen (Mumford and Shah [82]). Dieser Algorithmus wurde als weitere Segmentierung ausgewählt, da aktive Konturen vielfach für die Detektion von Zellen in Mikroskopbildern eingesetzt wurden (Mehrayin et al. [77], Pecreaux et al. [88], Ray and Acton [93], Shen et al. [105]).

Die ausgewerteten Zellbilder enthalten durchschnittlich sieben Zellen, die manuell ausgezählt wurden um die Referenzzellzahl zu bestimmen. Aufgrund von Fehlern in der nanostrukturierten Oberfläche war es in 12 der insgesamt 94 Bildern nicht möglich, die Zellzahl eindeutig zu bestimmen, so dass diese Bilder aus der Evaluation entfernt wurden. Die Detektionsgenauigkeit der Algorithmen wurde mit dem Fehlermaß  $\left| \frac{N_d - N_r}{N_r} \right|$

### Oberflächenkontrastmikroskopie

Im Gegensatz zu den gebräuchlichen Standardmethoden wie Phasenkontrast oder Differentialinterferenzkontrast wird der optische Kontrast der Oberflächenkontrastmikroskopie (*Surface Contrast Microscopy*, SCM) nicht durch die verwendete Optik, sondern direkt auf bzw. von der Oberfläche erzeugt. Hierzu verwendet die SCM eine Oberfläche mit einer periodischen Nanostruktur auf einer Dünnschicht, die als **Planarer Photonischer Kristall** (*Photonic Crystal Slab*, PCS) bezeichnet wird (Johnson et al. [52]). Abhängig von der verwendeten Nanostruktur sind diese PCS nur für bestimmte Frequenzbänder des sichtbaren Lichts durchlässig, die als Moden bezeichnet werden. Die Frequenz dieser Moden ist wiederum eine Funktion des Brechungsindex, d.h. das jede Brechungsindexänderung an der Oberfläche zu einer Frequenzverschiebung führt. Werden gekreuzte Polarisationsfilter vor und nach dem PCS eingesetzt, ist in der Transmission nur die Resonanz dieser Mode sichtbar und die Frequenzverschiebung wird als eine Farb- und Intensitätsveränderung wahrgenommen. D.h., es werden die (Zell-) Körper sichtbar, die einen anderen Brechungsindex zeigen. Der gesamte Aufbau wird als Oberflächenkontrastmikroskopie oder *Surface Contrast Microscopy* bezeichnet und ist in der Arbeit von Nazirizadeh et al. [83] beschrieben.



**Abbildung 5.5.:** Oberflächenkontrastmikroskopie (SCM).

evaluiert, wobei  $N_d$  die Anzahl der detektierten Zellen in jedem segmentierten Bild angibt und  $N_r$  die Zahl der Zellen in den Referenzbildern. In Abb. 5.6(b) ist dieser Fehler für beide Zelldetektionen in den Phasenkontrast- und SCM-Bildern für verschiedene Skalierungen angegeben (Binningfaktor 1, 2, 4, 8, ..., 128).

### 5.3.2. Auswertung und Vergleich der Zelldetektionen

Ein Vergleich der verwendeten Zelldetektionen hat gezeigt, dass die schwellwertbasierte Segmentierung auf den gegebenen Bilddaten robustere Ergebnisse erzielt. Während die Zellen in allen Bildern mit der Schwellwertsegmentierung erkannt werden konnten, schaffte es die aktive Kontur-basierte Segmentierung nicht, in allen Bildern zu konvergieren: In zehn SCM-Bildern und 22 Phasenkontrastbildern wurden keine gültigen Zellkonturen gefunden. Die Ursache dieser hohen Fehlerrate liegt in der verwendeten initialen Maske, die als Ausgangspunkt für die Detektion genutzt wurde und darauf

optimiert war, gleichverteilte Zellen in den Bildern zu finden. Das führte dazu, dass Bilder mit nur wenigen Zellen ( $<3$ ) oder Randzellen sehr fehlerhaft waren, während die Bilder mit mehreren Zellen eine gute Genauigkeit erzielten. Der direkte Vergleich der Fehlerraten aus Abb. 5.6(b) zeigt, dass die Detektion der Zellen in den SCM-Bildern mit beiden Algorithmen eine höhere Genauigkeit erzielte: Die Schwellwertsegmentierung hatte in den SCM-Bildern einen um den Faktor  $2.5 - 3.5\times$  und die aktive Kontur Detektion einen um den Faktor  $1.2 - 2\times$  niedrigere Fehlerrate.

### Segmentierung in den skalierten Bilddaten

Die Detektion in den skalierten Bilddaten zeigt, dass die Zellen bis zu einer Verkleinerung um den Faktor 32 mit einer konstanten Fehlerrate gefunden werden. Die Genauigkeit der Segmentierung in den Bildern mit einem Binningfaktor von 64 und 128 zeigt eine sehr hohe Ungenauigkeit, was auf die Größe der Bilder zurückzuführen ist: Während die Originalbilder eine Auflösung von  $1600 \times 1200\text{px}$  haben, hat ein skaliertes Bild mit einem Binningfaktor von 32 eine Auflösung von nur noch  $50 \times 38\text{px}$ . In Tab. 5.2 sind die vollständigen Angaben zu der Zelldetektion in den skalierten Bildern zu finden. Insgesamt zeigt sich, dass die SCM-Bilder sehr gut für die Zelldetektion geeignet sind. Die Resultate der Segmentierung in den skalierten Bildern verdeutlichen, dass eine Zelldetektion auch in Bildern mit einer niedrigeren Auflösung gute Ergebnisse liefert. Die erzielte Fehlerrate in der Zelldetektion impliziert, dass eine Objekterkennung in SCM-Mikroskopbildern den bisherigen Standardmethoden wie Phasenkontrast oder Differentialinterferenzkontrast hoch überlegen ist.

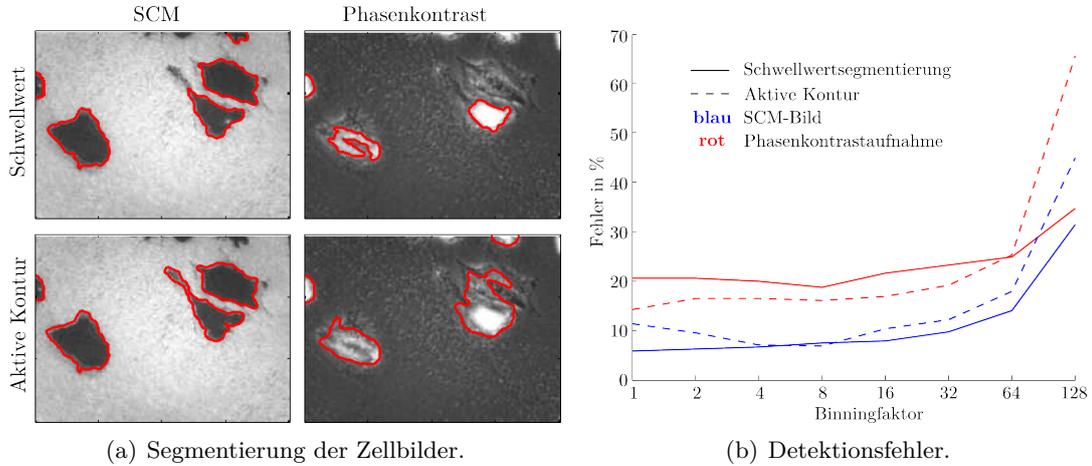
Mit der Bildsegmentierung aus Abschnitt 2.4 konnten sehr gute Resultate erzielt werden. Da diese Algorithmen bereits für die Zelldetektion in den Schräglichtbildern verwendet wurden, könnten die SCM-Bilder direkt von dem Zelltracking-Framework ausgewertet werden. Weiterhin illustriert die hohe Detektionsgenauigkeit der skalierten Bilder aus Tab. 5.2, dass mit der neuen Kontrasttechnik eine niedrigere Vergrößerung gewählt werden kann, ohne einen Qualitätsverlust der Bilder hinnehmen zu müssen. Als direkte Konsequenz könnten daher größere Bereiche und somit mehr Zellen pro Bild aufgezeichnet werden.

### Zusammenfassung der Zelldetektion in SCM-Bildern

Mit der Oberflächenkontrastmikroskopie (*surface contrast microscopy*, SCM) wurde eine markerfreie und nicht-invasive Methode zur Kontrastverbesserung von biologischem Material vorgestellt (Nazirizadeh et al. [7]). Diese neu entwickelte Technik wurde genutzt, um Aufnahmen von fixierten Stammzellen zu erstellen, die systematisch mit den Algorithmen zur Zelldetektion ausgewertet wurden. Die Resultate zeigen, dass die SCM-Bilder durch einen sehr guten Kontrast charakterisiert sind und bessere Detektionsergebnisse liefern als vergleichbare Phasenkontrastbilder. Ein Nachteil der verwendeten Technik ist die limitierte Einsatzmöglichkeit, da bisher nur der Kontrast von fixierten Zellen dargestellt werden kann. Um die Technik so zu erweitern, dass auch lebende Zellen beobachtet werden können, müssen weitere Optimierungen der nanostrukturierten Oberfläche durchgeführt werden. Insgesamt illustrieren die vielversprechenden Resultate, dass diese Technik ein großes Potential besitzt um sich als wichtige Methode für die

### 5.3. Oberflächenkontrastmikroskopie zur Zelldetektion

Analyse von adhären wachsenden Zellen zu etablieren.



**Abbildung 5.6.:** Genauigkeit der Zelldetektion in SCM- und Phasenkontrastbildern. Eine schwellwertbasierte Segmentierung und eine Detektion, die eine aktive Kontur verwendet, wurden genutzt, um die Zellen in beiden Bildtypen zu segmentieren (a). Die Zelldetektion in Bildern mit unterschiedlicher Skalierung zeigt, dass eine konstante Fehlerrate bis zu einem Binning von 32 erzielt wird, was einer effektiven Bildgröße von  $50 \times 38$ px entspricht (b), Quelle: Nazirizadeh et al. [7, Fig. 5].

Binning Faktor	Auflösung	Fehler (SCM)		Fehler (Phasenkontrast)		Pixel pro Zelle
		Schwellw.	akt. Kont.	Schwellw.	akt. Kont.	
1	$1600 \times 1200$	0.0585	0.1133	0.2057	0.1427	21613
2	$800 \times 600$	0.0618	0.0944	0.2057	0.1658	5402
4	$400 \times 300$	0.0666	0.0707	0.1992	0.1645	1342
8*	$200 \times 150$	0.0751	0.0678	0.1871	0.1599	339
16	$100 \times 75$	0.0781	0.1034	0.2151	0.1693	88
32	$50 \times 38$	0.0964	0.1223	0.2315	0.1908	24
64	$25 \times 19$	0.1397	0.1786	0.2480	0.2518	8
128	$12 \times 9$	0.3138	0.4496	0.3476	0.6563	3

**Tabelle 5.2.:** Zelldetektion in skalierten Bildern. Der Fehler wird mit  $\left| \frac{N_d - N_r}{N_r} \right|$  angegeben, wobei  $N_d$  die Anzahl der detektierten Zellen und  $N_r$  die Referenzzahl liefert.

\* Diese Auflösung entspricht der Vergrößerung der CeTReS-Daten.

## 5.4. Fluoreszenzbilder zur Quantifizierung von Wundregeneration

Der folgende Abschnitt stellt eine weitere Kooperation vor, in der die vorgestellte Zelldetektion für eine Quantifizierung der Wundregeneration in einem Tiermodell genutzt wurde.

An der Fraunhofer EMB wird die direkte Anwendung von adulten Stammzellen erforscht, die aus verschiedenen exokrinen Geweben wie Schweißdrüsen oder Speicheldrüsen gewonnen werden (Gorjup et al. [38], Petschnik et al. [89]). Hier wurde nachgewiesen, dass die Zellen ein großes Potential im Bereich der regenerativen Medizin bzw. der Hautregeneration besitzen (Egaña et al. [32]). Es wurde ein Tiermodell etabliert, um die regenerativen Fähigkeiten von humanen *in vitro*-kultivierten, adulten Stammzellen in einem *in vivo*-Modell für eine Matrix-gestützte, dermale Regeneration zu untersuchen.

Für die Analyse der untersuchten Gewebe wurde die Zelldetektion aus 2.4 erweitert und für eine quantifizierte Auswertung der Wundregeneration angepasst. Im Folgenden wird die automatisierte Analyse von Mehrkanal-Fluoreszenzbildern vorgestellt, mit der die Bildauswertung durchgeführt wurde. Zunächst wird ein Überblick über die verwendeten Methoden und ausgewerteten Experimente gegeben; eine weiterführende Beschreibung ist in der gemeinsamen Arbeit mit Danner et al. [6] zu finden.

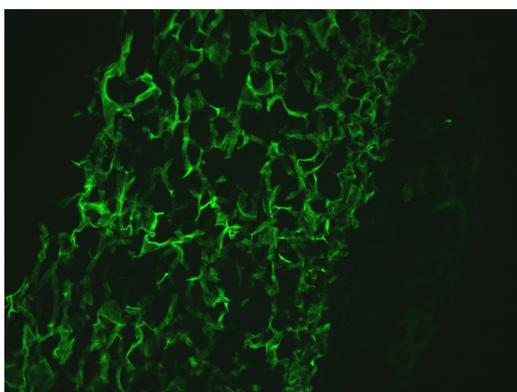
### 5.4.1. Quantitative Auswertung

Die Grundlage für die quantitative Auswertung der Hautregeneration liefern histologische Kryoschnitte, die einen Querschnitt des Wundareals zeigen. Insgesamt sind 95 Mehrkanal-Fluoreszenzbilder dieser Schnitte erstellt worden, die den Regenerationsgrad der Wunden nach Abschluss des Tierversuchs dokumentieren, wie in der Beschreibung auf Seite 108 erläutert. Das Ziel der quantifizierten Auswertung war es, die Regeneration einer Wunde objektiv zu beurteilen. Hierzu wurde die Besiedlungsdichte der Zellen in der Wunde bzw. auf der Matrix bestimmt (vgl. Beschreibung des Tierversuchs auf Seite 108). Anhand der ermittelten Zellzahlen konnte die Regeneration abgeschätzt werden: Je stärker die Matrix zellularisiert war, desto weiter war die Regeneration fortgeschritten. Für diese Quantifizierung waren folgende Schritte notwendig:

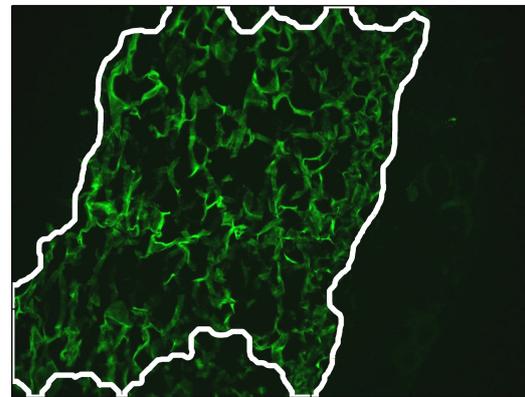
1. Der von der Matrix abgedeckte Wundbereich musste vermessen werden. Hierzu wurde die Eigenfluoreszenz der Matrix genutzt, siehe Abb. 5.7(a-b).
2. Die Anzahl der Zellen wurde bestimmt, die sich in der Wunde bzw. der Matrix befanden. Hierzu wurde eine DAPI-Zellkernfärbung verwendet, so dass die Zellkerne im blauen Fluoreszenzkanal gut erkannt werden konnten.

In den 95 aufgenommen Bildern wurde zuerst der Wundbereich ermittelt, indem mit einer Bildsegmentierung im grünen Kanal die Matrix extrahiert wurde. Hierzu wurden die Aufnahmen mit einem Gaußfilter geglättet, damit die Eigenfluoreszenz eine möglichst homogene Färbung zeigt. Anschließend wurde der abgedeckte Wundbereich mit der schwellwertbasierten Segmentierung aus Abschnitt 2.4 detektiert, wobei ein globaler Schwellwert mit der Methode von Otsu [84] berechnet wurde. Ein Resultat dieser

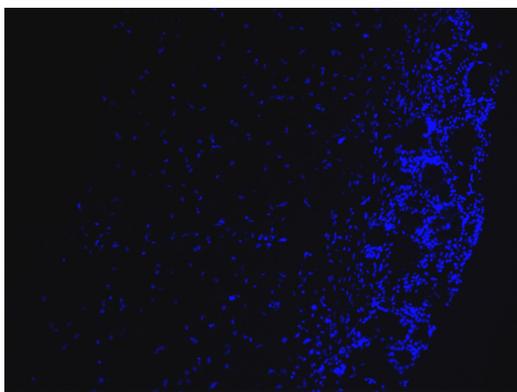
Detektion ist in Abb. 5.7(c) dargestellt. Diese Segmentierung wurde als Maske verwendet, um den relevanten Bereich der DAPI-Bilder auszuschneiden. Ein entsprechender Ausschnitt ist in Abb. 5.7(d) gegeben. Die Detektion der Zellkerne erfolgte mit einer weiteren globalen, schwellwertbasierten Segmentierung (Schwellwert nach der Methode von Otsu), mit der die genaue Anzahl der Zellen auf der Wunde bestimmt wurde. Um die Besiedlungsdichte auszuwerten, wurde die errechnete Zellzahl mit der Größe der Maske normalisiert, so dass die Zellzahl als Zellen pro  $mm^2$  angegeben wurde. Mit dieser automatisierten Prozedur sind 95 aufgenommene Bilder ausgewertet worden, wovon 53 Bilder zu der Testgruppe und 42 Bilder zu der Kontrollgruppe gehörten. Die Auswertung ergab eine durchschnittlich höhere Zellzahl auf den Bildern der Testgruppe, die statistisch signifikant war (doppelter T-Test,  $P < 0.05$ ) (Danner et al. [6]).



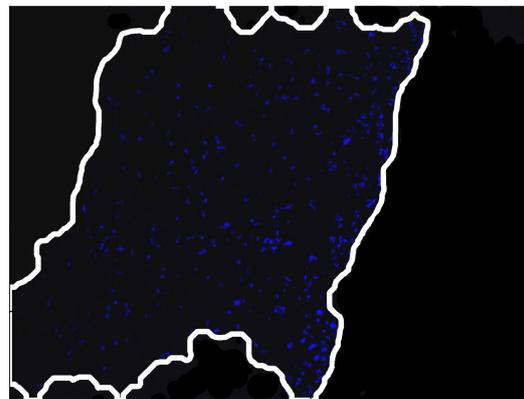
(a) Eigenfluoreszenz der Matrix (grüner Kanal).



(b) Bildsegmentierung.



(c) DAPI gefärbte Zellkerne (blauer Kanal).



(d) Zellzahlbestimmung auf der Matrix.

**Abbildung 5.7.:** Die mit Stammzellen besiedelte Kollagenmatrix ist in den Mikroskopbildern an ihrer Eigenfluoreszenz im grünen Kanal zu erkennen (a). Eine schwellwertbasierte Segmentierung wurde genutzt, um die Region zu detektieren, in der die Matrix auf der Wunde auflag (b). Mit einer DAPI-Zellkernfärbung wurden die Zellen gefärbt (c) und mit der segmentierten Maske aus (b) konnte die exakte Zellzahl auf der Wunde bestimmt werden (d).

Das angestrebte Ziel war es, eine objektive Quantifizierung anhand der gegebenen Bilddaten der Wundregeneration durchzuführen. Die entsprechenden Auswertungen konnten mit der Bildsegmentierung aus Abschnitt 2.4 leicht umgesetzt werden und haben den Nutzen der verwendeten Zellen im Tiermodell gezeigt.

### **Ein Tierversuch zur Hautregeneration mit adulten Stammzellen**

Die vollständige Regeneration der Haut ist ein komplexer Prozess, der bisher nur zum Teil verstanden ist, der aber im Bereich der regenerativen Medizin und für die *in vitro*-Gewebezucht von großer Bedeutung ist (Jackson et al. [50], Lewis [65]). Verschiedene Studien haben gezeigt, dass (adulte) Stammzellen ein großes Potential besitzen, um diesen Heilungsprozess zu unterstützen und zu beschleunigen (Egaña et al. [32]). In dem hier analysierten Tierversuch wurde der Einfluss von adulten Schweißdrüsen-abgeleiteten Stammzellen auf die Hautregeneration untersucht. In dem genutzten Wundmodell wurden acht (athymische) immundefiziente Nacktmäuse verwendet. Diese besitzen aufgrund einer Mutation keine Körperbehaarung und sind ein viel genutzter Modellorganismus in der Hautforschung (Lee et al. [64], Michael et al. [79], Skardal et al. [108]).

Die Mäuse wurden in zwei Gruppen zu je vier Tieren eingeteilt und jeweils mit zwei Vollhautdefekten von 10mm Durchmesser versehen. Um die *in vitro*-kultivierten Stammzellen auf die Wunde zu bringen, wurde eine Auflagematrix auf Basis von Rinderkollagen verwendet. Diese Matrices wurden mit *in vitro*-kultivierten, adulten Stammzellen aus der Schweißdrüse besiedelt und zu Beginn des Experimentes auf die Wunde aufgebracht (in der Kontrollgruppe wurde eine nicht-besiedelte Matrix aufgebracht).

Nach zwei Wochen Regenerationszeit wurden die Mäuse getötet und die Matrix wurde entfernt. Anschließend wurden Gewebeschnitte des Wundareals erstellt und histologisch untersucht, d.h. es wurden fluoreszenzmikroskopische Aufnahmen von dem Gewebe erstellt. Die Ergebnisse der folgenden Untersuchung verdeutlichte, dass die aufgebrachten Stammzellen eine positive Auswirkung auf die Hautregeneration haben, eine stärkere Vaskularisierung der Haut und beschleunigte Wundregeneration bewirken (Danner et al. [6]).

# 6

## Kombination von Genealogien und Fluoreszenzanalysen

In der bisherigen Arbeit wurde mit dem validierten Zelltracking eine nicht-invasive Methode entwickelt, die es ermöglicht, die Stammbäume einzelner Zellen bzw. gesamter Zellpopulation aus Langzeitbeobachtungen zu extrahieren. In diesem Kapitel werden die dynamischen Informationen des Zelltracking um eine statische Analyse der Proteinexpression einzelner Zellen erweitert. Die grundlegende Idee ist, dass mit den Tracking-Algorithmen aus Abschnitt 3.1.3 für jede Zelle angegeben werden kann, mit welchen anderen Zellen sie verwandt ist, woher sie kam und welche Zellen durch Zell-Zell-Kontakte Informationen ausgetauscht haben könnten. Mit diesem Wissen kann allerdings keine Aussage über die Zelldifferenzierung oder die Funktionalität einer Zelle ausgesagt werden. Somit kann die Frage beantwortet werden, welche Zellen verwandt sind, nicht aber die Frage, was aus den einzelnen Zellen geworden ist.

Mit der Immunfluoreszenzanalyse steht wiederum eine sehr genaue und etablierte Methode zur Verfügung, mit der die Proteinexpression und die Differenzierungsmuster der Zellen extrahiert werden können. Diese Methode liefert jedoch nur eine Momentaufnahme der Zellpopulation, da die Zellen vor der Färbung fixiert, d.h. getötet werden.

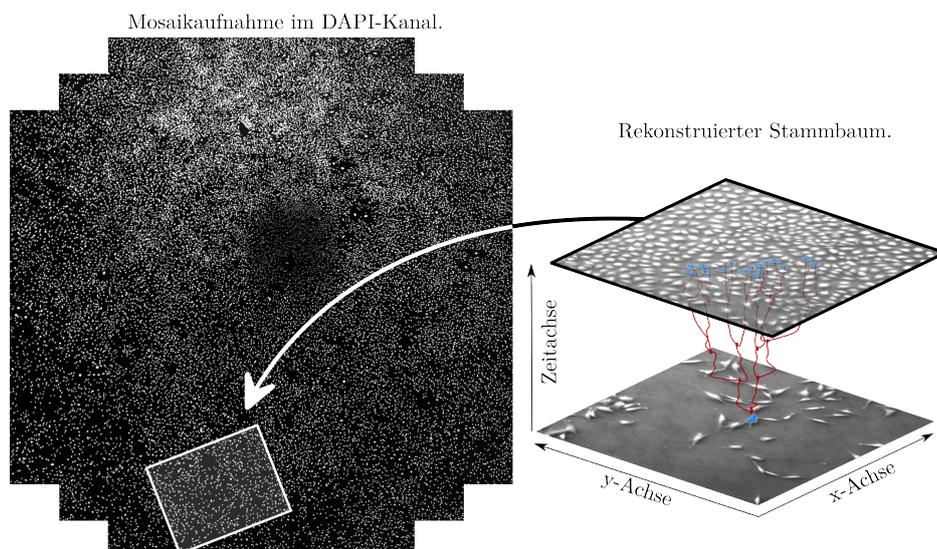


Abbildung 6.1.: Kombination von Genealogien und Immunfluoreszenzanalysen.

Mit der Fluoreszenzanalyse ist es nicht möglich, die Zellen in einem dynamischen Kontext zu analysieren, wie er von dem validierten Zelltracking bereitgestellt wird. Das heißt, es kann genau bestimmt werden, welche Zellen z.B. zu einer Nervenzelle wurden, nicht aber, ob sie verwandt sind oder ob eine Differenzierung durch einen Zellkontakt induziert wurde. Um die Lücke zwischen der genealogischen Auswertung und der Fluoreszenzanalyse zu schließen, wird in dem folgenden Kapitel eine Methode vorgestellt, die in einem neuen experimentellen Ansatz beide Techniken kombiniert. Mit dieser **funktionalen Fluoreszenzanalyse** (*functional fluorescence analysis*, FFA) kann der Einfluss der Zellverwandtschaften und der Zell-Zell-Kontakte auf die Differenzierung der einzelnen Zellen bestimmt werden.

Die grundlegende Idee der Kombination ist in Abb. 6.1 illustriert. Zuerst wird eine Zellpopulation in einem Langzeitexperiment beobachtet und anschließend wird die extrahierte Genealogie mit einer Fluoreszenzfärbung kombiniert. Da es technisch nicht möglich ist, die Langzeitbeobachtung und die Färbung in einem Mikroskop durchzuführen, wird ein großes Mosaikbild der gefärbten Zellen aufgenommen, das als **Referenzbild** bezeichnet wird. Dieses Referenzbild wird als Mehrkanal-Fluoreszenzaufnahme erstellt und enthält u.a. eine Aufnahme, in der die DAPI-gefärbten Zellkerne sichtbar sind. Die letzte Langzeitaufnahme wird dagegen als Testbild bezeichnet und zeigt einen kleinen Ausschnitt des Referenzbildes im Phasen- oder Schräglichtkontrast. Die Aufgabe besteht nun darin, die exakte Position des Testbildes im Referenzbild zu bestimmen, wie in Abb. 6.1 gezeigt. Eine Möglichkeit besteht darin, in beiden Bildern nach markanten, identischen Bildpunkten zu suchen wie sie z.B. eine skaleninvariante Merkmalstransformation (*scale-invariant feature transform*, SIFT) liefert (Lowe [75]). Da die beiden Bilder jedoch mit unterschiedlichen Kontrastmechanismen erstellt wurden und entweder die Zellen oder die Zellkerne zeigen, existieren keine ähnlichen oder identischen Bereiche in den Bildern (vgl. Abb. 6.3).

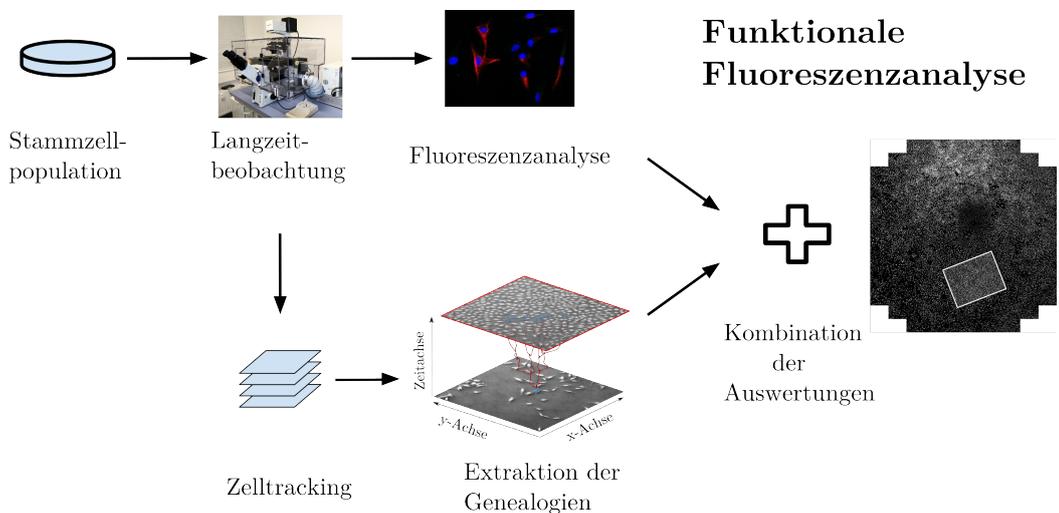
Um dieses Problem zu lösen, wird ein Algorithmus implementiert, der die eindeutigen Anordnungen von drei, vier oder fünf Zellen nutzt, um die exakte Bildposition wiederzufinden. Hierzu werden verschiedene geometrische Hashfunktionen definiert, die einen eindeutigen Wert für jede dieser Zellkonstellationen angeben, die invariant gegenüber linearen Transformationen sind. Diese Eigenschaft besagt, dass eine Konstellation von Zellen immer mit einem identischen Hashwert beschrieben wird, unabhängig von der Rotation (Ausrichtung der Petrischale), der Skalierung (Vergrößerung der verwendeten Aufnahme) oder der Translation (Position der Zellen in einem Bildausschnitt).

Der Fokus der folgenden Abschnitte liegt zunächst auf der Entwicklung der mathematischen Formalismen und Algorithmen die benötigt werden, um die Ergebnisse der verschiedenen Auswertungen zu kombinieren. Anschließend werden die Eigenschaften der Hashfunktionen und Modelle genau analysiert und es wird gezeigt, wie für ein gegebenes Zellbild robuste und adaptive Parameter bestimmt werden können.

Die Methoden wurden anhand von echten Bilddaten und mit Hilfe von künstlich generierten Testmodellen evaluiert. Die Ergebnisse zeigen, dass mit den verwendeten Methoden eine robuste Kombination von Phasenkontrast und Fluoreszenzbildern möglich ist. Dies ist selbst dann möglich, wenn die benötigte Zelldetektion mit einer geringen Genauigkeit ( $\approx 50\%$ ) durchgeführt wurde. Die vorgestellten Algorithmen, die eine Suche basierend auf geometrischen Hashfunktionen für Dreiecke nutzt, sind auf der IEEE Konferenz ISBI im Jahr 2013 vorgestellt worden (Becker et al. [5]).

## Funktionale Fluoreszenzanalyse

Bevor die mathematischen Grundlagen und Methoden eingeführt werden, wird zunächst der experimentelle Ablauf beschrieben, mit dem eine FFA durchgeführt werden kann, der in Abb. 6.2 gezeigt ist. Zuerst wird die Zellpopulation in einem Langzeitexperiment beobachtet, wie in Abschnitt 2.2.2 beschrieben. Unmittelbar nach der letzten Aufnahme werden die Zellen fixiert, um eine weitere Migration der Zellen zu stoppen. Danach wird eine Fluoreszenzfärbung durchgeführt, mit der eine spezifische Proteinexpression der Zellen untersucht wird. Anschließend wird eine große Mehrkanal-Fluoreszenzaufnahme der Zellen erstellt, die aus mehreren einzelnen Bildern besteht, die zu einem großen Mosaikbild zusammengesetzt werden. Ein wesentlicher Schritt für die vorgestellten Algorithmen ist die DAPI-Zellkernfärbung, die bei jeder Fluoreszenzanalyse durchgeführt wird und es erlaubt, die Zellen in einer Fluoreszenzaufnahme zu finden. Diese Kernfärbung erlaubt eine sehr robuste Detektion (Genauigkeit und Sensitivität  $> 0.998$ , vgl. Tab. 6.1) und wird verwendet, um die Zellen in der Langzeitaufnahmen in den Mosaikbildern zu finden. Nun ist die herausfordernde Aufgabe, die verwendeten Mosaikbilder einen großen Ausschnitt der Petrischale zeigen und aus einigen hundert Einzelaufnahmen zusammengesetzt wurden. So setzte sich das Bild, das in der ersten FFA ausgewertet wurde, aus 360 einzelnen Aufnahmen zusammen. Das resultierende Mosaikbild hatte eine Auflösung von 320MP ( $320 \times 10^6$  px), in dem mehr als  $10^4$  Zellkerne gefunden wurden. Das Problem liegt nun darin, in diesen großen Bildern jene Ausschnitte zu finden, die in den Langzeitexperimenten beobachtet wurden.



**Abbildung 6.2.:** Die funktionale Fluoreszenzanalyse kombiniert die Auswertung des Zelltracking mit einer (Immun-) Fluoreszenzfärbung der Zellen. Hierzu wird die Zellpopulation unmittelbar nach der letzten Aufnahme der Bildserie fixiert und gefärbt. Anschließend wird ein Übersichtsbild (Mosaikbild) im Fluoreszenzkanal aufgenommen.

### Geometrische Hashverfahren

Das geometrische Hashing beschreibt eine modellbasierte Methode zur Identifizierung von Objekten, die durch mehrere Punktkoordinaten definiert werden (Schwartz and Sharir [102], Wolfson and Rigoutsos [120]). In dem betrachteten Fall sind die Punkte durch die Schwerpunkte der Zellen und Zellkerne gegeben und die Modelle entsprechen den Konstellationen, die diese bilden. Um dieselben Zellen in zwei gegebenen Bildern zu identifizieren, wird nach denselben Zellkonstellationen gesucht. Diese Konstellationen werden als geometrische Figuren interpretiert und mit einem Hashwert beschrieben, der

1. in der Lage ist, die Ähnlichkeiten der Konstellationen zu erhalten, d.h. dass ähnliche Konstellationen auch ähnliche Hashwerte erhalten. So ist es möglich, verzerrte bzw. verrauschte Konstellationen zu finden.
2. unabhängig von der Rotation, Skalierung und Translation der Objekte ist, d.h. invariant gegenüber einer linearen Transformation. Das ermöglicht es, die gesuchten Objekte unabhängig von der Aufnahmeposition und der verwendeten Vergrößerung zu erkennen.

Ein Verfahren, das diese Informationen nutzt, um Objekte zu identifizieren, wird als **geometrisches Hashverfahren** bezeichnet (Schwartz and Sharir [102]).

## 6.1. Geometrisches Hashing

In dem ersten Abschnitt 6.1 wird die Motivation für die Wahl eines geometrischen Hashverfahrens gegeben und es werden die benötigten mathematischen Modelle hergeleitet und methodische Grundlagen geliefert. Die gesamte Suche wurde mit dem geometrischen Hashverfahren in einer MATLAB-Implementierung umgesetzt.

### 6.1.1. Geometrisches Hashing zur Positionssuche von Zellbildern

In Abb. 6.3 ist das gegebene Problem dargestellt, das für einen menschlichen Betrachter nur schwer zu lösen ist: Aufgabe ist, die Position des gegebenen Testbildes  $I^T$ , welches in der unteren Reihe dargestellt ist, in dem Referenzbild  $I^R$  wiederzufinden. Nur anhand der gegebenen Konstellationen ist es sehr schwer, den richtigen Ausschnitt zu finden. Werden allerdings Dreiecke als Konstellationen von drei Zellen eingezeichnet, so vereinfacht sich die Suche und man findet die Bildposition „auf einen Blick“. Nach diesem Prinzip arbeitet auch ein geometrisches Hashverfahren, nur dass hier die Ähnlichkeit der Dreiecke durch einen eindeutigen Hashwert ersetzt wird. Dieser Hashwert muss die Bedingungen erfüllen, die in der obigen Zusammenfassung gegeben sind, d.h. er muss invariant gegenüber linearen Transformationen und Ähnlichkeitserhaltend sein.

Die Konstellationen, die von den Zellen und Zellkernen gebildet werden, sind nicht identisch (vgl. Abb. 6.4 auf Seite 115). Das hat zur Folge, dass es für einen gesuchten

## 6.1. Geometrisches Hashing

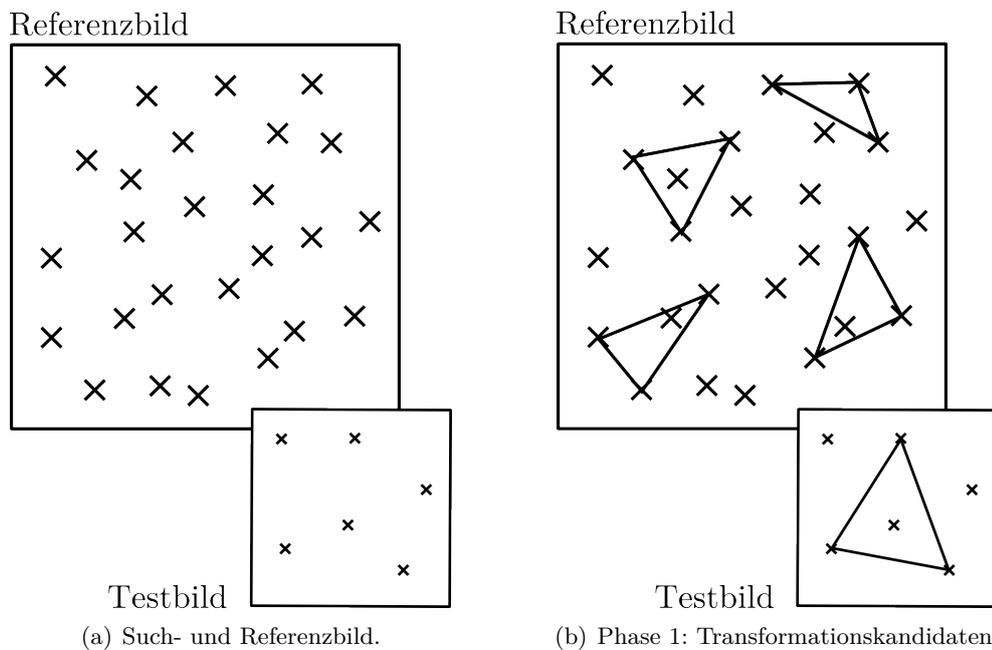
---

Hashwert viele ähnliche Konstellationen gibt, die alle eine Positionszuordnung, d.h. Transformation, beschreiben.

Um zu unterscheiden, welche der gefundenen Transformationen korrekt sind, wird das geometrische Hashverfahren in zwei Phasen unterteilt:

1. In der ersten Phase werden die Hashwerte der Bilder  $I^R, I^T$  genutzt, um mögliche Transformationen zu erstellen.
2. In der zweiten Phase werden die erstellten Transformationen validiert und ein Bayes-Test wird genutzt, um zu entscheiden, welche richtig sind.

In dem folgenden Abschnitt werden die Grundlagen für die Berechnung der Transformationen der ersten Phase gelegt. Hierzu werden die verwendeten Hashfunktionen beschrieben und die Invarianz der Hashwerte gezeigt. Aufbauend auf diesen Hashwerten werden Hashtabellen für das Test- und Referenzbild eingeführt, mit deren Hilfe die Transformationen aufgestellt werden.



**Abbildung 6.3.:** Positionsbestimmung. Die Suche der richtigen Bildposition ist als Beispiel für wenige Punkte illustriert. Die Aufgabe besteht darin, die Position des kleinen Testbildes in dem großen Referenzbild wiederzufinden (a). Das geometrische Hashing sucht nach identischen Konstellationen von mehreren Zellen. Um das zu veranschaulichen, sind in (b) mehrere Konstellationen aus drei Zellen eingezeichnet. Die Suche nach diesen Dreiecken kann leichter gelöst werden als die Suche nach den einzelnen Punkten.

### Von der Astronomie zur Petrischale

Die Idee, ein geometrisches Hashverfahren für die Suche der Zellbilder zu verwenden, hat ihren Ursprung in der Ähnlichkeit einer mikroskopischen Fluoreszenzaufnahme mit einem Sternenhimmel. Als Beispiel kann die Mosaikaufnahme aus Abb. 6.1 betrachtet werden, in der DAPI-gefärbte Zellkerne gezeigt sind. Neben der großen Ähnlichkeit der Bilddaten werden in der Astronomie auch sehr ähnliche Fragestellungen bearbeitet. So ist es z.B. nicht trivial, die Sterne in einer astronomischen Aufnahme zu identifizieren: Der aufgenommene Ausschnitt des Himmels hängt von der verwendeten Vergrößerung, der Ausrichtung des Fernglases und der Position des Beobachters ab. Um dieses Problem zu lösen, ist vielfach eine Suche verwendet worden, die auf ein geometrisches Hashverfahren zurückgreift. Diese Verfahren wurden z.B. genutzt, um die Sterne in gegebenen Aufnahmen zu identifizieren (Groth [40]) oder um die exakte Position und Ausrichtung der Kamera zu bestimmen (Liebe [70]).

#### **astrometry.net**

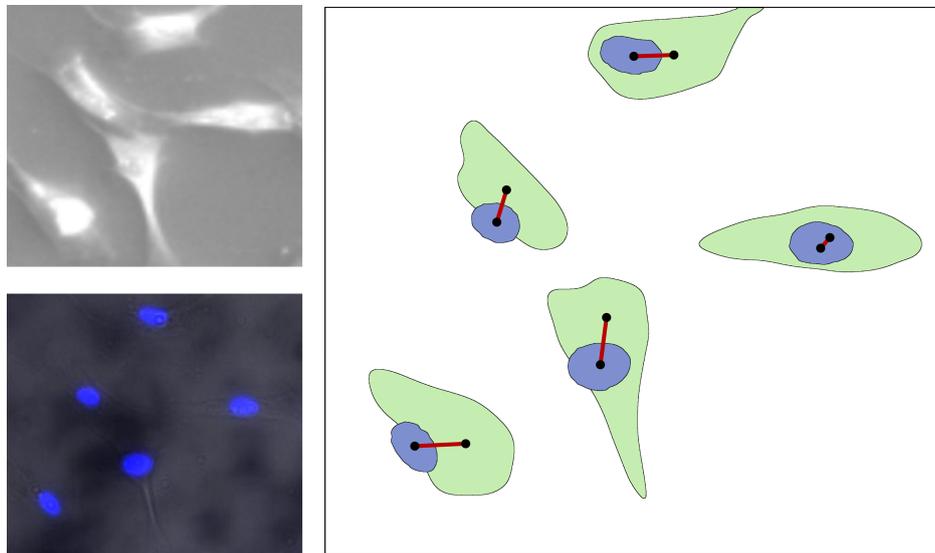
Im Jahr 2010 wurde von Lang et al. das Projekt *astrometry.net*<sup>a</sup> vorgestellt, das in einer gegebenen astronomischen Aufnahme alle Sterne identifiziert und den exakten Ausschnitt des Sternenhimmels bestimmt, der auf dem Bild gezeigt ist (Lang et al. [63]). Die Grundidee dieser Methode besteht darin, mit einem (fast) kompletten Katalog von astronomischen Objekten jede Konstellation von vier Sternen mit einem eindeutigen, invarianten geometrischen Hashwert zu beschreiben und diesen in einer Referenz-Hashtabelle zu speichern. Aufbauend auf diesen Hashtabellen wurde eine Suche implementiert, die in zwei Phasen abläuft.

Um die Sternkonstellationen zu bestimmen, die auf einem gegebenen Testbild zu sehen sind, wird in der ersten Phase nach ähnlichen Hashwerten in der Referenztable gesucht. Das größte Problem der astronomischen Daten ist die riesige Anzahl an Sternen, die eine noch größere Anzahl an Vierecken liefert. Das wiederum hat zur Folge, dass es zu jeder gesuchten Konstellation sehr viele ähnliche Vierecke gibt, d.h. für jedes gegebene Bild gibt es sehr viele mögliche Positionen. Um zu entscheiden, ob nun eine Konstellation die richtige Position beschreibt, wird in der zweiten Phase ein Bayes-Test genutzt (vgl. Rigoutsos und Hummel [95]). Dieser Bayes Test nutzt die Transformation, die durch die Stern- bzw. Zellkoordinaten beschrieben wird, um das Testbild in den Sternenhimmel zu projizieren. Anschließend wird die Wahrscheinlichkeit berechnet, mit welcher der korrekte Bildausschnitt gefunden wurde. Die Modelle erlauben es, die gesuchten Sternbilder mit einer sehr hohen Genauigkeit zu finden ([63]).

---

<sup>a</sup><http://astrometry.net>

Aufgrund der großen Anzahl an sichtbaren Sternen war eine Bearbeitung oder Suche der Bilder jedoch nur eingeschränkt möglich. So musste z.B. die ungefähre Position der Aufnahme bekannt sein (Pal und Bakos [86]) oder es wurde mit einer festen Brennweite bzw. Vergrößerung gearbeitet (Liebe [70]). Um alle bekannten Sterne wiederfinden zu können war es notwendig, Hashwerte für alle sichtbaren Sternkonstellationen zu berechnen, was nur mit dem Einsatz von großen Ressourcen umgesetzt werden konnte. Darauf aufbauend wurde ein Suchalgorithmus vorgestellt, der alle Sterne in einem gegebenen astronomischen Foto identifizieren kann (Lang et al. [63]). Die Suche und die verwendeten Hashwerte, die auf einem fast vollständigen Katalog von bekannten Sternen erstellt wurden, sind in dem Projekt *astrometry.net* bereitgestellt. Die letztgenannte Publikation lieferte den gedanklichen Anstoß, ein geometrisches Hashverfahren für die Zellbilder zu implementieren. Eine Zusammenfassung der Arbeit von Lang et al. [63] ist auf Seite 114 gegeben.



**Abbildung 6.4.:** Versatz der Zellkerne. Oben links: Die Mikroskopaufnahme zeigt die Zellen vor der Fixierung im Schräglichtkontrast. Unten links: Mit der anschließenden DAPI-Färbung werden die Zellkerne sichtbar. Rechte Seite: Die detektierten Zellmasken zeigen den Versatz, d.h. die Differenz, der Schwerpunkte der Zellen (vor der Färbung) und Zellkerne (nach der Färbung, rot markiert). Die Verteilung dieser Differenzen wird als Rauschmodell beschrieben.

### Unterschiede zwischen der Petrischale und der Astronomie

Wird eine astronomische Aufnahme, d.h. werden Sterne betrachtet, kann davon ausgegangen werden, dass sich die Positionen bzw. die Konstellationen der gesuchten Objekte nur minimal ändern.

In den Zellbildern werden dagegen unterschiedliche Objekte gesucht. Während in den Mosaikaufnahmen die DAPI-gefärbten Zellkerne sichtbar sind, zeigen die Aufnahmen der Langzeitbeobachtung die Zellen als Phasenkontrast- oder Schräglichtbild, wie in

Abb. 6.4 gezeigt. Das bedeutet, dass die Konstellationen in den Schräglichtbildern durch die Schwerpunkte der Zellen und in den Mosaikbildern durch die Schwerpunkte der Zellkerne definiert werden. Es ist daher davon auszugehen, dass sich die Positionen der Zellen und Zellkerne unterscheiden, wobei zwei Faktoren einen wesentlich Einfluss auf diese Verschiebung haben:

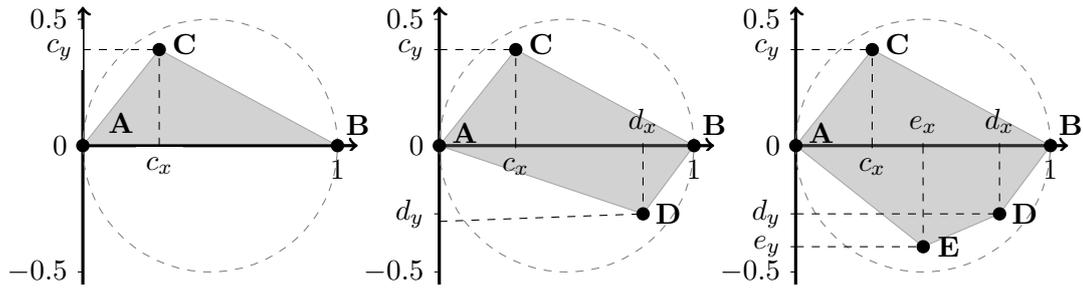
- Die Bewegung der Zellen. Die Langzeitaufnahmen werden von lebenden Zellen aufgenommen, die in der Lage sind, ihre Position zu verändern.
- Verschiebung der Zellkerne. Die Position der Zellen und Zellkerne wird über ihre Mittelpunkte bestimmt. Hierbei ist nur sichergestellt, dass sich der Zellkern in der Zelle befindet, nicht aber, dass er in dem Zentrum der Zelle liegt.

Der resultierende Versatz der Zellen und Zellkerne wird als Rauschen interpretiert, mit dem die gesuchten Muster verzerrt werden (siehe Abb. 6.4). Mit dem Muster werden auch die eindeutigen Beschreibungen, die geometrischen Hashwerte, verrauscht (Wolfson und Rigoutsos[120]). Das hat zur Folge, dass nach unähnlicheren Konstellationen als in astronomischen Ansätzen gesucht werden muss und sich der Suchraum entsprechend vergrößert. Es gehen auch einige Hashwerte verloren bzw. werden ungültig. Viel folgenreicher aber ist, dass mit den Hashwerten auch die Transformationen verrauscht werden, die als Projektion von einer Konstellation aus dem Testbild auf eine Projektion in das Referenzbild erstellt werden. Das bedeutet, dass sich die Verschiebung der Zellkerne in dem projizierten Bild vergrößert, da die zugrundeliegende Transformation einen systematischen Fehler enthält. Um eindeutig zwischen dem Rauschen der Zellkerne in den Referenzaufnahmen und dem Rauschen der projizierten Daten zu unterscheiden, wird das Referenzrauschen über die Varianz  $\sigma^2$  der zugrundeliegenden Normalverteilung angegeben. Die Varianz der Verteilung der projizierten Daten wird dagegen als  $\sigma_F^2$  bezeichnet. Die letztgenannte Varianz liefert den entscheidenden Parameter für das spätere Vordergrundmodell  $F$ , mit dem die Verteilung der Zellkerne modelliert wird. Hierbei wird die Abschätzung  $\sigma_F \approx 2\sigma$  genutzt. Eine genauere Erklärung ist in Gl. (6.13) auf Seite 131 gegeben. Zusammengefasst verschlechtert sich das gestellte Problem im Vergleich zu den astronomischen Ansätzen.

Vorteilhaft in dem vorgestellten Szenario ist, dass die Anzahl der Zellen in einer Petrischale im Gegensatz zu den Sternen sehr begrenzt ist: Im sichtbaren Universum befinden sich  $\approx 6 \times 10^{22}$  Sterne (Gott et al. [39]), während in einer Petrischale maximal  $10^6$  Zellen beobachtet werden. Das wiederum erlaubt es, die benötigten Berechnungen auf einem Standard-PC auszuführen und es wird kein Rechencluster benötigt, wie er im Projekt *astrometry.net* verwendet wird (Lang et al. [63]).

### 6.1.2. Hashfunktionen für Drei-, Vier-, und Fünfecke

Zuerst werden die verwendeten Hashfunktionen vorgestellt, die für eine gegebene Konstellation aus drei, vier oder fünf Zellen einen eindeutigen Hashwert berechnen. Diese Hashwerte sind invariant gegenüber Rotation, Skalierung und Translation, d.h. dieselben drei, vier oder fünf Zellen werden mit demselben Hashwert beschrieben, unabhängig von der verwendeten Vergrößerung des Mikroskops oder der Position und Ausrichtung der Petrischale. Die Berechnung der Hashwerte ist in Abb. 6.5 dargestellt und wird nach dem folgenden Schema durchgeführt:



**Abbildung 6.5.:** Die geometrischen Hashfunktionen für Drei-, Vier-, und Fünfecke sind durch das lokale Koordinatensystem definiert, welches die Zellen  $A, B$  mit dem größten Abstand durch  $A = (0, 0)$  und  $B = (1, 0)$  aufspannt. In diesem Koordinatensystem können die Hashwerte als Koordinaten der Zellen  $C, D$  und  $E$  als  $h = [c_x, c_y, d_x, d_y, e_x, e_y]$  abgelesen werden. Für drei, vier und fünf Zellen ergeben sich entsprechend zwei-, vier- oder sechs-dimensionale Hashwerte  $h$ .

- Fünfecke.** Zuerst wird für fünf gegebene Zellen das Paar mit dem größten Abstand gesucht. Diese Zellen werden mit  $A$  und  $B$  bezeichnet und definieren ein lokales Koordinatensystem mit  $A = (0, 0)$  und  $B = (1, 0)$ . Die übrigen Zellen müssen sich in einem Kreis mit Durchmesser  $\|B - A\|_2$  befinden, der auch die Zellen  $A$  und  $B$  selbst enthält, siehe Abb. 6.5. Diese Zellen werden als  $C = (c_x, c_y), D = (d_x, d_y), E = (e_x, e_y)$  mit  $c_x < e_x < d_x$  bezeichnet. Mit dieser Anordnung ist noch keine eindeutige Beschreibung gegeben, da die Zellen  $A$  und  $B$  getauscht werden könnten und somit zwei gültige Hashwerte generiert werden würden. Um einen eindeutigen Hashwert zu erhalten, muss die Bedingung  $|c_y| > |d_y|$  erfüllt sein, d.h. die Zelle  $D$  muss sich näher an der  $x$ -Achse befinden als Zelle  $C$ . Ist das nicht der Fall, werden die Zellen  $A$  und  $B$  getauscht (und mit ihnen  $C$  und  $D$ ), um eine gültige Konstellation zu erhalten. In dem so generierten, lokalen Koordinatensystem wird der 6-dimensionale Hashwert durch die Zellkoordinaten  $h = [c_x, c_y, d_x, d_y, e_x, e_y]$  definiert.
- Vierecke.** Die Berechnung des Hashwertes für vier Zellen erfolgt analog. Es wird dieselbe Bezeichnung gewählt und es werden dieselben Bedingungen an die Zellen  $A, B, C$  und  $D$  gestellt wie zuvor. Da diesmal die zwei Koordinaten der fünften Zelle  $E$  fehlen, wird ein Viereck mit einem vierdimensionalen Hashwert  $h = [c_x, c_y, d_x, d_y]$  beschrieben.
- Dreiecke.** Für drei gegebene Zellen wird dieselbe Definition der Zellen  $A, B$  und  $C$  wie zuvor gewählt. Einen eindeutigen Hashwert  $h = [c_x, c_y]$  erhält man mit der Bedingung  $c_x < \frac{1}{2}$ .

Diese Hashfunktionen werden genutzt, um vollständige Hashtabellen für die gegebenen Zellbilder zu berechnen, in denen die Hashwerte aller Konstellationen mit drei, vier oder fünf Zellen enthalten sind. Eine genaue Beschreibung dieser Hashtabellen ist in der Definition auf Seite 119 zu finden.

Die Hashfunktionen, die in den Arbeiten von Lang et al. [63] und Wolfson and Rigoutsos [120] vorgestellt wurden, greifen auf ähnliche Hashfunktionen zurück, wie die hier verwendete Methode. Wolfson gibt eine generelle Berechnung für  $N$ -Ecke an, in der ein lokales Koordinatensystem durch zwei Punkte in  $(-\frac{1}{2}, 0)$  und  $(\frac{1}{2}, 0)$  aufgespannt wird. In der Methode von Lang et al. [63] wird eine explizite Berechnung für eine Hashfunktion angegeben, in der vier Sterne genutzt werden. Für die Berechnung der entsprechenden Hashwerte wird ein lokales Koordinatensystem durch zwei Sterne in  $(0, 0)$  und  $(1, 1)$  bestimmt. Die hier verwendete Methode wurde gewählt, weil sie eine schnelle und einfache Berechnung der Hashwerte ermöglicht.

### Invarianz der Hashwerte

Eine wichtige Voraussetzung der verwendeten Hashfunktionen ist die Invarianz gegenüber linearen Transformationen, d.h. dass die Hashwerte unabhängig von der Skalierung, Rotation und Translation der Zellbilder angegeben werden. Die geforderte Invarianz ist für die betrachteten geometrischen Figuren durch das lokale Koordinatensystem gegeben, welches in Abb. 6.5 dargestellt ist. Die Zellen  $A, B$  mit dem größten Abstand definieren mit  $A = (0, 0), B = (1, 0)$  das lokale Koordinatensystem, womit eine eindeutige Ausrichtung (Rotationsinvarianz) und einheitliche Normierung (Skalierungs- und Translationsinvarianz) angegeben wird. Um die Invarianz nachzuweisen, wird eine explizite Rechenvorschrift für die Hashwerte betrachtet. Hierzu werden drei Punkte  $A, B$  und  $C$  mit der obigen Bezeichnung gewählt. Mit den Vektoren  $\overline{AB} = \frac{B - A}{\|B - A\|_2}$  und  $\overline{AC} = \frac{C - A}{\|B - A\|_2}$  können die Hashwerte  $h = [c_x, c_y]$  über die Skalarprodukte

$$\begin{aligned} c_x &= \langle \overline{AB}, \overline{AC} \rangle & \text{und} \\ c_y &= \langle \overline{AB}, \overline{AC}^\perp \rangle \end{aligned}$$

berechnet werden, wobei  $\overline{AC}^\perp$  der zu  $\overline{AC}$  orthogonale Vektor gleicher Länge ist, der um  $90^\circ$  im Uhrzeigersinn gedreht ist. Die Berechnung der Hashwerte für Vier- bzw. Fünfecke erfolgt analog. Aus den Eigenschaften des Skalarproduktes folgt direkt die Invarianz gegenüber linearen Transformationen.

### Die Hashtabellen

Mit den gegebenen Hashfunktionen kann ein vollständiger Satz von Hashwerten für ein gegebenes Zellbild berechnet werden, der alle Konstellationen von drei, vier oder fünf Zellen enthält. Die Hashwerte für das Referenzbild  $I^R$  und das Testbild  $I^T$  werden als **Hashtabellen**

$$\begin{aligned} H^T & \text{ mit } h_l^T \in H^T, l = 1, \dots, N^T & \text{ und} \\ H^R & \text{ mit } h_k^R \in H^R, k = 1, \dots, N^R \end{aligned}$$

bezeichnet, wobei  $N^R$  und  $N^T$  die Größe der Hashtabellen bzw. die Anzahl der berechneten Hashwerte angeben. Diese Größe ist entscheidend für die Laufzeit der Suche (nach ähnlichen Konstellationen) und hängt direkt von der Zahl  $M$  der Zellen im Bild ab:  $N^R$  bzw.  $N^T$  wachsen mit  $\frac{M!}{M^k!(M-M^k)!} = \binom{M}{M^k}$ , wobei  $M^k \in 3, 4, 5$  die Zahl der verwendeten Zellen bzw. die Konstellationsgröße (Drei-, Vier-, und Fünfecke) angibt.

In dem vorgestellten Ansatz ist es nicht erforderlich, eine vollständige Hashtabelle zu berechnen: Solange sichergestellt wird, dass die Konstellationen  $h_l^T$  aus dem Testbild auch in der Referenztabelle  $H^R$  enthalten sind, kann die Größe dieser Tabellen eingeschränkt werden. In dem vorgestellten Ansatz wird die Größe begrenzt, indem mit

$$d_{min} < \|B - A\|_2 < d_{max} \quad (6.1)$$

eine minimale und maximale Länge der Konstellationen vorgegeben wird, wobei  $A$  und  $B$  die Position der Zellen im Bildraum angeben. Hierbei ist zu beachten, dass die Wahl von  $d_{min}, d_{max}$  nur vorgenommen werden kann, wenn die Skalierung der Bilddaten bekannt ist. Da diese Bedingung für Mikroskopbilder vorausgesetzt werden kann, stellt die Wahl von  $d_{min}$  und  $d_{max}$  keine Einschränkung dar, sondern reduziert den benötigten Rechenaufwand enorm.

### Die Transformationen

Die Suche nach ähnlichen Konstellationen kann mit Hilfe der gegebenen Hashtabellen  $H^R, H^T$  aus der obigen Zusammenfassung als eine Suche nach ähnlichen Hashwerten ausgeführt werden. Für eine gegebene Konstellation aus dem Testbild  $I^T$  mit dem Hashwert  $h_l^T \in H^T$  sind ähnliche Konstellationen des Referenzbildes über die Hashwerte

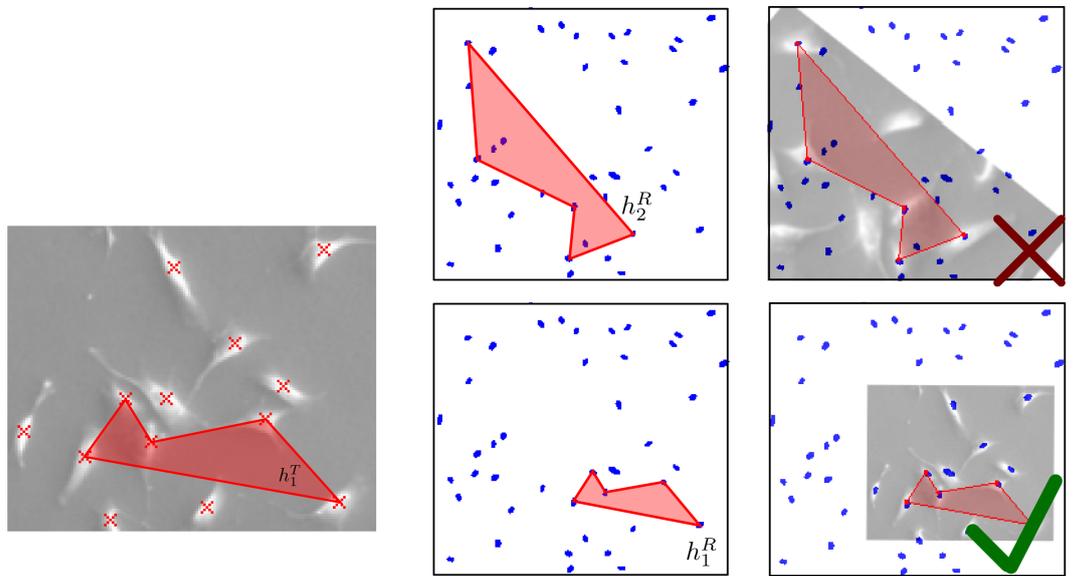
$$\mathcal{H}_l := \left\{ h_k^R \mid \|h_l^T - h_k^R\|_2 < \epsilon, h_k^R \in H^R \right\}, \quad (6.2)$$

gegeben. Hierbei beschreibt  $\mathcal{H}_l$  die Menge aller Hashwerte der Referenztabelle  $H^R$ , die eine ähnliche Konstellation wie  $h_l^T$  beschreiben. Jedes Tupel von Hashwerten  $(h_k^R, h_l^T)$ ,  $h_k^R \in \mathcal{H}_l$  definiert eine mögliche Transformation  $T_{k,l} : I^T \rightarrow I^R$  vom Testbild  $I^T$  in das

Referenzbild  $I^R$ . Diese Transformation  $T_{k,l}$  wird aus den Koordinaten der Zellen bzw. Zellmittelpunkte der Konstellationen  $h_k^R, h_l^T$  definiert. Mit  $T_{k,l}$  ist es möglich, alle Zellen des Testbildes  $c_m^Z, m = 1, \dots, M^T$  mit

$$T_{k,l}(c_m^Z) = c_m^T \quad (6.3)$$

in das Koordinatensystem des Referenzbildes  $I^R$  zu transferieren. Für eine Konstellation  $h_l^T \in H^T$  existieren daher viele mögliche Transformationen  $T_{k,l}$ , von denen maximal eine die richtige Bildposition beschreibt. Um zu entscheiden, welches  $k$  die richtige Transformationen angibt, wird im Folgenden ein Bayes-Test genutzt, um eine **globale Prüfung** mit allen projizierten Zellkernen  $c_m^T, 1, \dots, M^T$  durchzuführen.



(a) Referenzbild  $I^T$ .

(b) Ähnliche Konstellationen.

(c) Mögliche Transformationen.

**Abbildung 6.6.:** Mit Hilfe der Schwerpunkte der Zellen und Zellkerne werden Zellkonstellationen mit einem eindeutigen und invarianten Hashwert beschrieben. Für die Konstellation mit dem Hashwert  $h_1^T$  des Testbildes  $I^T$  in (a) gibt es zwei ähnliche Konstellationen  $\mathcal{H}_1 = \{h_1^R, h_2^R\}$  in dem Referenzbild  $I^R$  (b). Jedes Tupel  $(h_1^T, h_1^R)$  und  $(h_1^T, h_2^R)$  definiert eine Transformation von Bild  $I^T$  in den Raum des Referenzbildes  $I^R$  (c).

### 6.1.3. Validieren der Transformationen

Für die zweite Phase wird eine Validierung in Form eines Bayes-Tests eingeführt, der entscheidet, ob eine gegebene Transformation die richtige Bildposition beschreibt. Es wird dem Ansatz von Rigoutsos et al. [95, 96] gefolgt, in dem eine Bayessche Formulierung für das Klassifikationsproblem von geometrischen Hashverfahren gegeben wurde.

Im Folgenden werden die benötigten Modelle eingeführt und aufbauend auf dem Satz von Bayes wird das Entscheidungsproblem auf ein Schwellwertproblem zurückgeführt und eine robuste Validierung der Transformation erlaubt.

### Bayessche-Modellierung

In der Bayesschen-Modellierung wird mit einem Vordergrundmodell  $F$  die Wahrscheinlichkeit berechnet, mit der eine mögliche Transformation  $T_{k,l}$  die richtige Bildposition angibt. Das Hintergrundmodell  $B$  beschreibt dagegen die Wahrscheinlichkeit dafür, dass  $T_{k,l}$  keine richtige Position zeigt. Die entsprechenden Wahrscheinlichkeiten werden mit dem Satz von Bayes als

$$\begin{aligned} p(F|T_{k,l}) &= \frac{p(T_{k,l}|F)P(F)}{p(T_{k,l})} && \text{für das Vordergrundmodell und} \\ p(B|T_{k,l}) &= \frac{p(T_{k,l}|B)P(B)}{p(T_{k,l})} && \text{für das Hintergrundmodell} \end{aligned}$$

angegeben. Hierbei beschreiben  $p(F|T_{k,l})$  und  $p(B|T_{k,l})$  die bedingten oder *a-posteriori* Wahrscheinlichkeiten, wenn die Testzellen mit der Transformation  $T_{k,l}$  projiziert wurden.  $p(F)$ ,  $p(B)$  und  $p(T_{k,l})$  bezeichnen die (unabhängigen) *a-priori* Wahrscheinlichkeiten (Barber [12]).

Um die Modelle  $F$  und  $B$  zu vergleichen und das wahrscheinlichere Modell auszuwählen, werden mit Hilfe der *a-posteriori* Wahrscheinlichkeiten  $p(F|T_{k,l})$  und  $p(B|T_{k,l})$  die datenabhängigen, bedingten Wahrscheinlichkeiten

$$\underbrace{\frac{p(F|T_{k,l})}{p(B|T_{k,l})}}_{a\text{-posteriori}} = \underbrace{\frac{p(T_{k,l}|F)}{p(T_{k,l}|B)}}_{\text{Bayes-Faktor}} \underbrace{\frac{p(F)}{p(B)}}_{a\text{-priori}}$$

angegeben. Die *a-priori* Wahrscheinlichkeiten  $p(F)$  und  $p(B)$  geben an, wie wahrscheinlich es ist, zufällig eine gültige bzw. ungültige Transformationen zu finden. Da es für jede Konstellation nur eine richtige, aber sehr viele falsche  $T_{k,l}$  gibt, gilt  $p(F) \ll p(B)$ . Die konkurrierenden Modelle  $F$  und  $B$  können anhand des **Bayes-Faktors**  $K = \frac{p(T_{k,l}|F)}{p(T_{k,l}|B)}$  verglichen werden, der von Jeffreys [51] eingeführt und bezeichnet wurde. Somit wird das Entscheidungsproblem, ob eine Transformation  $T_{k,l}$  richtig ist oder nicht, auf das Schwellwertproblem

$$K > \theta_K \tag{6.4}$$

zurückgeführt. Wenn der Bayes-Faktor  $K$  über einem vordefinierten Schwellwert  $\theta_K$  liegt, wird die durch  $T_{k,l}$  beschriebene Position als korrekt akzeptiert.

## Modellierung

Die verwendeten Grundlagen der Modellierung sind in der Beschreibung der Bayesschen Formulierung auf Seite 121 zusammengefasst. Hier wird die entscheidende Frage, ob eine Transformation  $T_{k,l}$  die korrekte Position zeigt, mit Hilfe der bedingten Wahrscheinlichkeiten  $p(F|T_{k,l})$  und  $p(B|T_{k,l})$  modelliert. Hierbei bezeichnet  $F$  das Vordergrundmodell (richtige Transformation) und  $B$  das Hintergrundmodell (falsche Transformation). Anschließend wird der Satz von Bayes genutzt, um diese Frage auf das Schwellwertproblem aus Gl. (6.4) zu reduzieren. In der gegebenen Formulierung hängt die Entscheidung, ob die Transformation  $T_{k,l}$  richtig ist oder nicht, von dem Bayes-Faktor  $K$  ab. Dieser wiederum wird von den Modellen  $F$  und  $B$  bestimmt, die die Wahrscheinlichkeiten für eine korrekte Position oder eine ungültige Transformation beschreiben. Um diese entscheidenden Modelle explizit anzugeben, wird der Ansatz von Lang et al. [63] verfolgt.

Neben dem Bayes-Faktor  $K$ , ist der Schwellwert  $\theta_K$  der zweite entscheidende Parameter. Eine erste, sehr grobe Abschätzung für  $\theta_K$  liefert das Suchszenario selbst: Es ist bekannt, dass für jede Konstellation  $h_l^T$  sehr viele ähnliche Hashwerte  $h_k^R$  existieren, von denen höchstens eine richtig sein kann. Das bedeutet wiederum, dass die *a-priori* Wahrscheinlichkeit  $p(F)$  für eine richtige Transformation sehr viel kleiner ist als die für eine falsche, d.h. es gilt  $p(F) \ll p(B)$ . Als Konsequenz sollte der Schwellwert  $\theta_K$  entsprechend groß gewählt werden. Wie eine gute Wahl für  $\theta_K$  getroffen werden kann, wird in dem späteren Abschnitt 6.2.3 gezeigt.

## Vordergrundmodell $F$

Das Vordergrundmodell  $F$  gibt die Wahrscheinlichkeit an, mit der eine gegebene Transformation die richtige Bildposition beschreibt. Das Modell zu  $F$  setzt sich aus zwei Termen zusammen. In dem ersten Term werden alle Zellkerne zusammengefasst, die in beiden Bildern gefunden wurden. Hierzu wird für jede transformierte Zelle  $c_m^T$  die Wahrscheinlichkeit berechnet, mit der sie durch einen Referenzkern  $c_n^R, n = 1, \dots, M^R$  generiert wird. Es wird angenommen, dass die Verschiebung der Schwerpunkte der Zellen und Zellkerne einer Normalverteilung mit Varianz  $\sigma_F^2$  folgt.

Im dem zweiten Term werden alle Zellen zusammengefasst, die keinem Zellkern aus dem Referenzbild zugeordnet werden konnten. Hierzu wird die Wahrscheinlichkeit aller Zellen berechnet, die nicht durch die Referenzkerne erklärt werden und die als verlorene Zellen  $C^V$  bezeichnet werden. Eine verlorene Zelle kann durch verschiedene Ereignisse erklärt werden. Selbst wenn eine richtige Transformation betrachtet wird, gibt es zwei Fehlerquellen, die eine verlorene Zelle generieren können. Zum einen ist es möglich, dass durch eine Fehldetektion ein Fremdpartikel (Zellreste, Staubpartikel etc.) im Testbild  $I^T$  als Zelle erkannt wurde, zu der es keinen Referenzzellkern gibt. Weiterhin ist es möglich, dass ein DAPI-gefärbter Zellkern nicht detektiert wurde oder dass sich eine richtig erkannte Zelle während der Färbung abgelöst hat und in dem Referenzbild  $I^R$  nicht zu sehen ist. Diese Fehldetektionen im Test- oder Referenzbild werden in dem Term der verlorenen Zellen zusammengefasst. Eine Zelle  $c_m^T$  wird von keinem Referenzzellkern  $c_n^R$  generiert, wenn die Entfernung zwischen  $c_m^T$  und dem nächstgelegenen Zellkern  $c_n^R$  größer als die dreifache Standardabweichung  $\sigma_F$  der Zellkernverschiebung ist,

$$c_m^T \in C^V \Leftrightarrow \min_{n=1, \dots, M^R} \|c_m^T - c_n^R\| > 3\sigma_F. \quad (6.5)$$

Der Anteil dieser verlorenen Zellen  $c_n^V \in C^V, n = 1, \dots, M^V \leq M^T$  wird als  $\nu = \frac{M^V}{M^T}$  bezeichnet. Die gesamte Wahrscheinlichkeit für das Vordergrundmodell wird mit

$$p(c_m^T|F) = \frac{\nu}{S} + \frac{1-\nu}{M^R} \sum_{n=1}^{M^R} \mathcal{N}(c_m^T|c_n^R, \sigma_F^2), \quad (6.6)$$

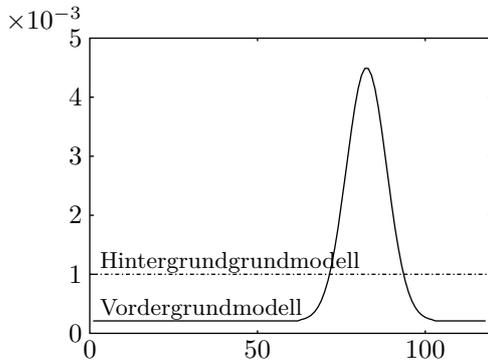
angegeben, wobei  $\mathcal{N}(c_m^T|c_n^R, \sigma_F^2)$  die Wahrscheinlichkeit angibt, mit der die Referenzzelle  $c_n^R$  den Zellkern  $c_m^T$  erklärt, wenn eine Normalverteilung mit Varianz  $\sigma_F^2$  angenommen wird;  $S$  bezeichnet die Fläche von Testbild  $I^T$ .

### Hintergrundmodell $B$

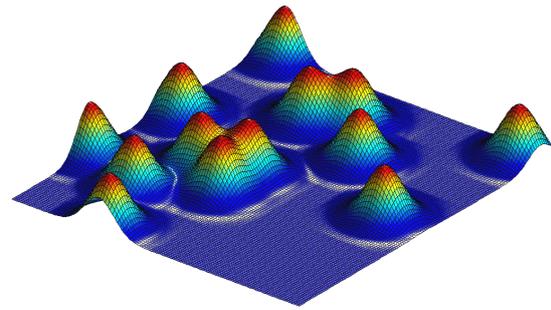
Das Hintergrundmodell  $B$  liefert die Wahrscheinlichkeit für den Fall, dass die durchgeführte Transformation  $T_{k,l}$  nicht die richtige Bildposition zeigt. Hierzu wird in dem Hintergrundmodell  $B$  angenommen, dass die Testzellen  $c_m^T$  nicht von den Referenzkernen  $c_n^R$  generiert wurden, sondern im gesamten Testbild  $I^T$  mit derselben Wahrscheinlichkeit

$$p(c_m^T|B) = \frac{1}{S} \quad (6.7)$$

verteilt sind, wobei  $S$  die Fläche des Testbildes  $I^T$  angibt, d.h., dass die Wahrscheinlichkeit für jede Zelle  $c_m^T$  als Gleichverteilung modelliert wird.



(a) Dichtefunktionen in 2D-Darstellung.



(b) Dichtefunktion von  $F$  in 3D-Darstellung.

**Abbildung 6.7.:** Dichtefunktionen der Wahrscheinlichkeiten der Modelle  $F$  und  $B$ . Das Hintergrundmodell  $B$  nimmt eine Gleichverteilung der Zellpositionen im Testbild an, während das Vordergrundmodell  $F$  jede Zelle im Testbild als verschobene Referenzzelle interpretiert, wobei die Verschiebung als Gaußverteilung modelliert wird. In (b) ist die Dichtefunktion des Vordergrundmodells für die Zellen aus Abb. 6.6 gezeigt.

## Übersicht der Parameter des Bayes-Tests

Mit den vorgestellten geometrischen Hashfunktionen aus Abschnitt 6.1.1 und dem Bayes-Test aus Gl. (6.12) ist es nun möglich, die richtige Position eines Bildes anhand der Zellkonstellationen zu bestimmen. Die Berechnung der Hashtabellen, das Vordergrundmodell und der Bayes-Tests hängen dabei von verschiedenen Parametern ab, die von den beobachteten Zellen bzw. deren Eigenschaften beeinflusst werden. In Gl. (6.10) und Gl. (6.12) sind alle Werte und Parameter angegeben, die Einfluss auf den Bayes-Test haben: Die Werte  $M^R, M^T$  geben die Zellzahl in dem Referenzbild  $I^R$  bzw. dem Testbild  $I^T$  an,  $S$  gibt die Fläche des Testbildes an,  $\nu$  ist ein Maß dafür, wie viele Zellen wiedergefunden werden konnten und die Varianz  $\sigma_F^2$  beschreibt den Versatz der Zellkerne. Die vier Werte  $M^R, M^T, S, \nu$ , die für die Berechnung der Modelle  $F, B$  verwendet werden, sind direkt von den gegebenen Bilddaten abhängig. Nur die beiden folgenden Parameter können oder müssen frei gewählt werden:

- Der Versatz der Zellkerne wird als Normalverteilung mit Varianz  $\sigma_F^2$  modelliert. Dieses Rauschmodell wird genutzt, um die Wahrscheinlichkeitsdichte des Vordergrundmodells in Gl. (6.6) aufzustellen.
- Als zweiter Parameter muss der Schwellwert  $\theta_K$  des Bayes-Tests aus Gl. (6.12) angegeben werden, der bestimmt, ab wann eine Transformation korrekt ist.

## Abschätzung der Genauigkeit der Zelldetektion

Der Parameter  $\nu$  wird für die Berechnung des Vordergrundmodells in Gl. (6.6) verwendet und gibt an, wie viele Zellen des Testbildes im Referenzbild gefunden wurden. Wird eine falsche Transformation  $T_{k,l}$  analysiert, so ist dieser Wert sehr klein. Wird  $\nu$  dagegen für eine gültige Transformation berechnet, so liefert dieser Wert eine gute Abschätzung für die Genauigkeit der Zelldetektion. Voraussetzung hierfür ist, dass die genaue Anzahl der Zellen bekannt ist. In der folgenden Tab. 6.1 auf Seite 130 wird gezeigt, dass die Detektion der DAPI-gefärbten Zellkerne mit einer sehr hohen Genauigkeit ausgeführt wird (Genauigkeit und Sensitivität  $> 0.998$ ), d.h. es kann davon ausgegangen werden, dass in dem Referenzbild alle Zellen enthalten sind. Ist also eine richtige Transformation gefunden, kann leicht die richtige Zahl der Zellen  $M^Z$  in dem Testbild ermittelt werden. Anschließend kann mit dem Faktor  $\nu$  und der Zellzahl  $M^T$  im Testbild die Genauigkeit und Sensitivität der Zelldetektion über

$$\text{Sensitivität} = \frac{RP}{RP + FN} = \nu \frac{M^T}{M^Z} \quad \text{und} \quad (6.8)$$

$$\text{Genauigkeit} = \frac{RP}{RP + FP} = \nu \frac{M^T}{M^T} = \nu \quad (6.9)$$

berechnet werden (vgl. Abschnitt 3.2.2). Wird also die richtige Position gefunden, so kann gleichzeitig abgeschätzt werden, mit welcher Genauigkeit die Bildsegmentierung ausgeführt wurde.

**Bayes-Faktor**

Mit Hilfe der vorgestellten Modelle  $F$  und  $B$  kann für jede Transformation  $T_{k,l}$  eine Wahrscheinlichkeit angegeben werden, mit der bestimmt wird, ob die richtige Bildposition gefunden wurde. Hierzu werden die Zellen des Testbildes  $I^T$  in die Koordinaten des Referenzbildes projiziert. Anschließend wird für jede transformierte Zelle  $c_m^T$  überprüft, ob sie einer Zelle im Referenzbild zugeordnet werden kann. Hierzu wird mit dem Vordergrundmodell  $F$  die Wahrscheinlichkeit berechnet, mit der die Zelle  $c_m^T$  von einem der Referenzkerne  $c_n^R, n = 1, \dots, M^R$  generiert wird (vgl. Gl. (6.6)). Das Hintergrundmodell  $B$  gibt wiederum die Wahrscheinlichkeit an, mit der  $T_{k,l}$  keine gültige Transformation beschreibt (vgl. Gl. (6.7)). Mit Hilfe dieser Wahrscheinlichkeiten kann der Bayes-Faktor  $K$  angegeben werden, der mit dem Satz von Bayes aus Gl. (6.4) und der Unabhängigkeit der Wahrscheinlichkeiten  $p(c_m^T|F)$  und  $p(c_m^T|B)$  als Produkt

$$K = \frac{p(T_{k,l}|F)}{p(T_{k,l}|B)} = \prod_{m=1}^{M^T} \frac{p(c_m^T|F)}{p(c_m^T|B)} \quad (6.10)$$

$$= \prod_{m=1}^{M^T} \left( \frac{\frac{v}{S} + \frac{1-v}{M^R} \sum_{n=1}^{M^R} \mathcal{N}(c_m^T|c_n^R, \sigma_F^2)}{\frac{1}{S}} \right) \quad (6.11)$$

zusammengefasst werden. Somit wird die Entscheidung, ob eine korrekte Transformation gefunden wurde, als

$$T_{k,l} \text{ ist gültig, falls } K > \theta_K, \quad (6.12)$$

bestimmt. Die Parameter dieser Gleichung sind das Rauschen  $\sigma_F^2$  und der Schwellwert  $\theta_K$ , die weiteren Werte werden durch die Zellbilder selbst bestimmt.

#### 6.1.4. Implementierung und Suchstrategie

Die vorgestellten Algorithmen sind für Konstellationen mit drei, vier und fünf Zellen als MATLAB-Implementierung (Version R2011a) umgesetzt worden. Das verwendete Schema ist in Abb. 6.8 dargestellt. Im Folgenden wird die verwendete Diskretisierung der Hashtabellen beschrieben und die Positionssuche erläutert.

##### Diskretisierung der Hashtabelle

Ein kritischer Punkt für die Laufzeit der Extraktion der möglichen Transformationen ist die Suche nach ähnlichen Hashwerten  $h_k^R \in H^R$ . Um diese Konstellationen zu finden, werden für einen gegebenen Hashwert  $h_l^T \in H^T$  aus dem Testbild alle Referenz-Hashwerte  $h_k^R \in H^R$  mit

$$\|h_l^T - h_k^R\|_2 < \epsilon$$

gesucht, vgl. Gl. (6.2). Um die Laufzeit zu optimieren, wurde in der vorliegenden Implementierung eine Diskretisierung der Hashtabelle  $H^R$  verwendet. Diese teilt den  $n$ -dimensionalen Zahlenraum  $[0, 1]^n$  in insgesamt  $M$  gleichgroße Teilmengen  $H^R = H_1 \cup \dots \cup H_M$  ein, in denen jeweils ähnliche Hashwerte liegen. Für jeden Hashwert  $h_l^T$  bzw.  $h_k^R$  wird mit einer Hashfunktion  $\text{hash} : H^T \rightarrow \{1, \dots, M\}$  ein eindeutiger Index  $m$  berechnet. Mit Hilfe dieser Hashfunktion werden die Indexmengen  $H_i^R, i = 1, \dots, N$  für die Referenzhashtabellen vorberechnet. Anschließend können die möglichen Transformationen für einen gegebenen Hashwert  $h_l^T$  über

$$\mathcal{H}_l = H_m^R, \quad \text{mit} \quad m = \text{hash}(h_l^T)$$

angegeben werden. Die Diskretisierung ist in Gl. (A.5) im Anhang explizit angegeben.

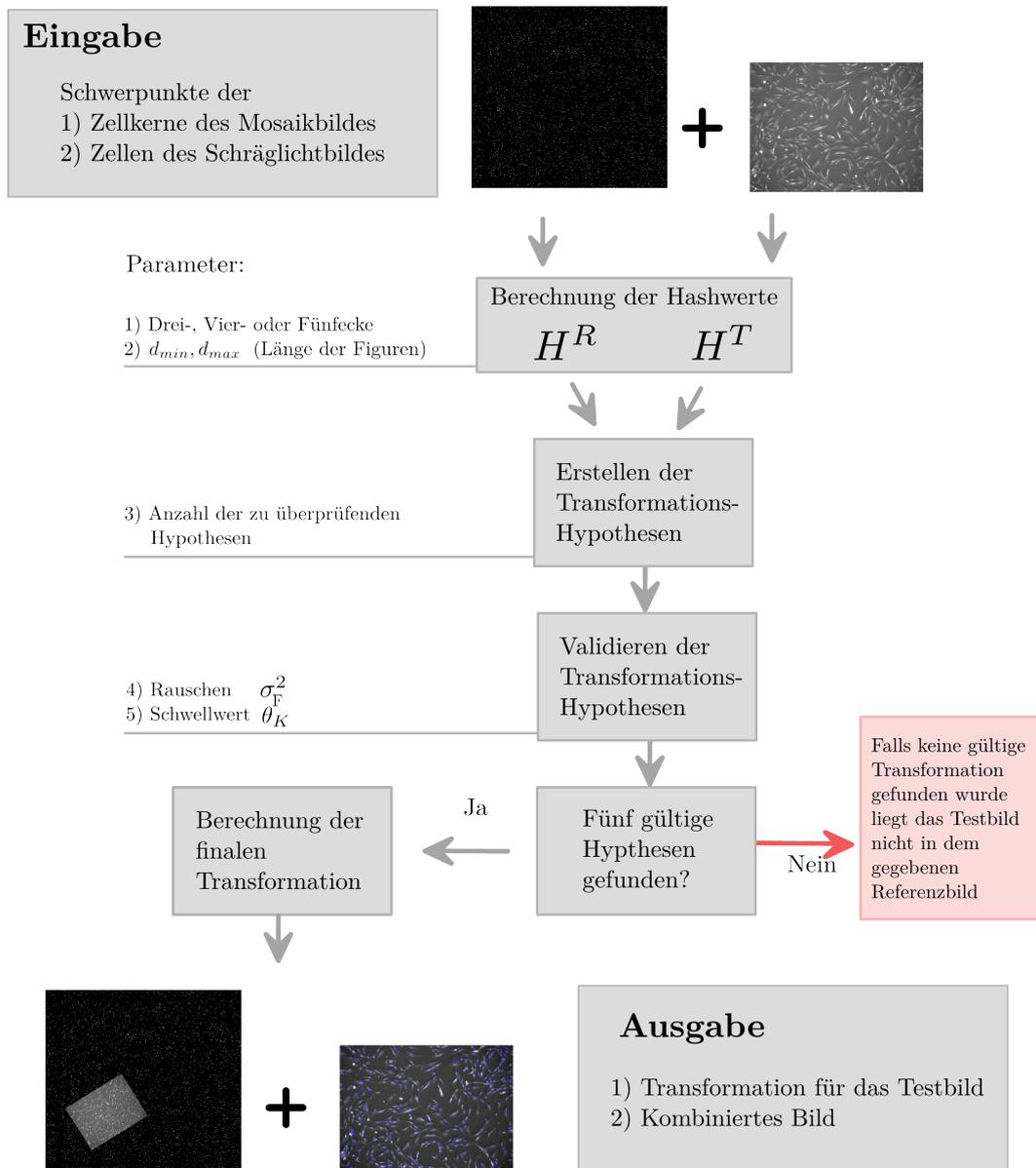
##### Parallelisierung

In MATLAB wurde eine parallele Implementierung umgesetzt, in der die Umkreissuche und die kombinatorische Suche nach gültigen Konstellationen zeitgleich ausgeführt werden kann. Auf einem Computer mit einem Intel® Core™ i7 konnten so acht Kerne genutzt werden und die Berechnung beschleunigte sich abhängig von der Größe der Hashtabellen um einen Faktor zwischen 3 und 6 gegenüber einer Implementierung auf einem Kern.

##### Die Positionssuche und die finale Transformation

In dem implementierten Suchschema aus Abb. 6.8 wird zuerst eine beliebige Konstellation  $h_l^T \in H^R$  zufällig gewählt. Für diesen Hashwert wird in der Menge aller möglichen Transformationen  $\mathcal{H}_l$  nach einer gültigen Transformation gesucht. Anschließend wird die Suche mit dem nächsten, zufällig gewählten  $h_l^T$  fortgesetzt. Es wird solange gesucht, bis entweder fünf gültige Bildpositionen gefunden oder bis 2000 Konstellationen ohne Erfolg überprüft wurden. Falls fünf gültige Transformationen gefunden wurden, wird mit Hilfe der Zellpositionen und Zellkernpositionen eine finale Position für das Testbild  $I^T$  berechnet.

## 6.1. Geometrisches Hashing

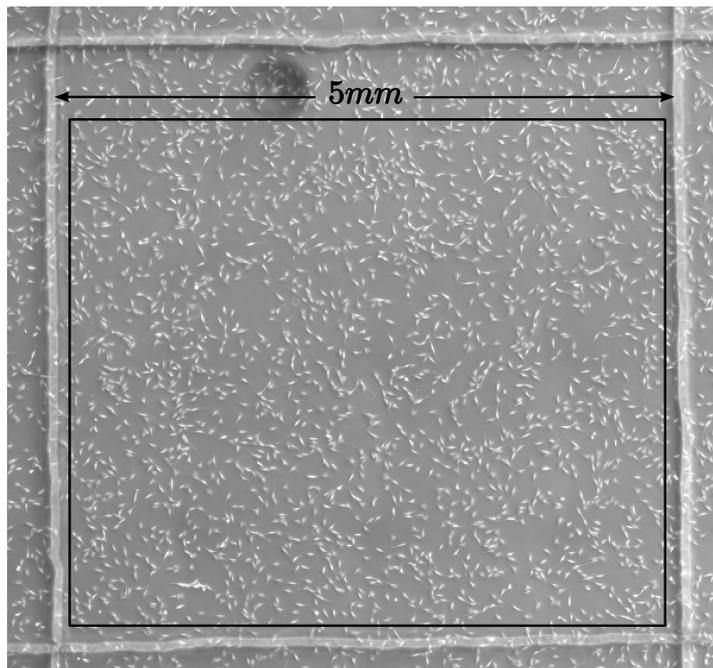


**Abbildung 6.8.:** Schematischer Ablauf der Positionssuche. Mit den Mittelpunkten der Zellkerne und Zellen werden die Hashtabellen  $H^R$  für das Referenzbild und  $H^T$  für das Testbild erstellt. Die Wahl der geometrischen Figur und die Wahl der zulässigen Länge ( $d_{min}, d_{max}$ ) hat entscheidenden Einfluss auf die Größe der Hashtabellen und somit auf die Laufzeit der Suche. Die Parameter  $\sigma_F^2, \theta_K$  werden für den Bayes-Test aus Gl. 6.12 benötigt und geben die Verschiebung der Zellkerne an. Werden fünf gültige Bildpositionen gefunden, wird eine finale Transformation ermittelt. Hierzu werden die Mittelpunkte der Zellen und Zellkerne der gültigen Hashwerte bzw. Konstellationen verwendet, um eine lineare Transformation für das Testbild zu bestimmen, die die richtige Bildposition im Referenzbild angibt (vgl. Mikroskopbild links unten).

## 6.2. Das Petrischalen-Modell

Das Suchschema in Abb. 6.8 zeigt, dass für die Positionssuche die datenabhängigen Parameter  $\sigma_F^2$ ,  $d_{max}$  und  $\theta_K$  bestimmt werden müssen, die den Versatz der Zellkerne, die Länge der verwendeten Konstellationen und den Schwellwert des Bayes-Tests angeben. Zuerst werden die Referenzaufnahmen aus Abschnitt 2.3.2 genutzt, um diese Parameter in dem vorliegenden *in vitro*-Zellszenario zu analysieren. Die verwendeten Referenzbilder zeigen einen großen Bereich einer Zellpopulation, der eine manuelle Markierung von  $5 \times 5 \text{mm}^2$  enthält, der sowohl im Phasenkontrast als auch im Fluoreszenzkontrast mit gefärbten Zellkernen abfotografiert wurde (vgl. Abb. 6.9). Mit den resultierenden Mosaikbildern ist es möglich, die exakte Verschiebung der Zellkerne zu bestimmen und das Rauschen bzw. die Varianz  $\sigma^2$  der Verteilung dieser Differenzen zu bestimmen und so den Parameter  $\sigma_F^2$  anzugeben.

Aufbauend auf diesem Rauschen wird ein künstliches Modell vorgestellt, mit dem die Schwerpunkte der Zellen bzw. Zellkerne simuliert werden. Mit diesen Daten wird das genaue Verhalten der Hashfunktionen abhängig von der Konstellationsgröße  $d_{max}$  analysiert. Anschließend wird das Verhalten des Bayes-Faktors  $K$  bzw. des Bayes-Tests aus Gl. (6.12) analysiert.



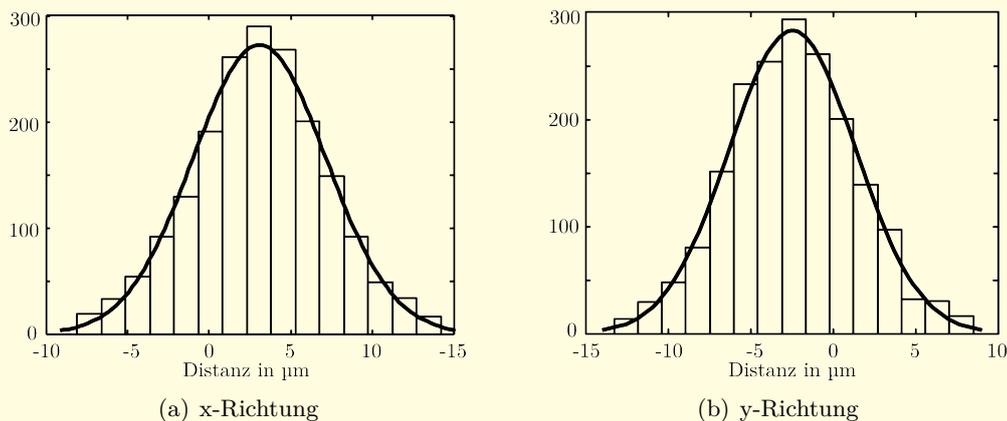
**Abbildung 6.9.:** Das Rauschmodell wurde als Versatz der Zellkerne in den Mosaikaufnahmen der Phasenkontrast- und Fluoreszenzaufnahmen aus Abb. 2.7 bestimmt. Um eine möglichst fehlerfreie Zell- und Zellkerndetektion zu erreichen, wurde ein Ausschnitt gewählt, der keine Artefakte wie die manuelle Markierung oder die Luftblase oben zeigt.

### Rauschmodell der Zellbilder (Petri-Rauschen)

Der entscheidende Parameter für das vorgestellte Suchverfahren ist der Versatz der Zell- und Zellkernpositionen, der als *in vitro*-Rauschmodell bezeichnet wird. Nur wenn die exakte Verteilung der Differenzen bekannt ist, kann hieraus der Parameter  $\sigma_F^2$  abgeleitet werden, der entscheidend für den Bayes-Test aus Gl. (6.12) ist.

Um diese Verschiebung zu bestimmen, wurde eine manuell markierte Petrischale mit Stammzellen (rPSC) besiedelt und im Phasenkontrast und Fluoreszenzkanal (DAPI-gefärbte Zellkerne) abfotografiert. Die resultierenden Mosaikbilder zeigen einen großen Bereich von  $5 \times 5 \text{ mm}^2$ , der mit einer Bildregistrierung übereinander gelegt wurde<sup>a</sup>.

Anschließend wurde die Bildsegmentierung aus Abschnitt 2.4 verwendet, um die Zellen bzw. Kerne in den zusammengesetzten Bildern zu detektieren. Die Schwerpunkte der segmentierten Masken wurden genutzt, um die Differenzen der Zellen zu ihren Zellkernen zu berechnen. So konnten die Differenzen von insgesamt 1800 Zellkernen berechnet werden, die einer Normalverteilung mit  $\mu = 3.1 \mu\text{m}$ ,  $\sigma^2 = 16.7 \mu\text{m}$  in x-Richtung (Abb. 6.10(a)) und  $\mu = -2.5 \mu\text{m}$ ,  $\sigma^2 = 14.8 \mu\text{m}$  in y-Richtung (Abb. 6.10(b)) entsprechen, die durch einen  $\chi^2$ -Test bestätigt wurden.



**Abbildung 6.10.:** Das Rauschmodell beschreibt die Verteilung der Distanzen zwischen dem Schwerpunkt einer Zelle und ihrem Kern.

<sup>a</sup>Die Bilder sind in Abschnitt 2.3.2 beschrieben und in Abb. 2.7 zu finden.

#### 6.2.1. Das *in vitro*-Rauschmodell

Das *in vitro*-Rauschmodell beschreibt die Differenz zwischen einer Zelle und ihrem Zellkern. Um die exakten Parameter in dem betrachteten Petrischalen-Szenario zu bestimmen, wurden die Mosaikaufnahmen verwendet, die in Abschnitt 2.3.2 beschrieben sind. Die resultierenden Bilder zeigen denselben Ausschnitt einer Zellpopulation unter

einem Langzeitmikroskop (Phasenkontrast- und Schräglichtaufnahmen) und als Fluoreszenzaufnahme mit DAPI-gefärbten Zellkernen. Die Aufnahmen enthalten eine manuelle Markierung der Größe  $5 \times 5mm^2$  die es ermöglichte, unabhängig von den Zellen dieselbe Positionen wiederzufinden.

### Registrierung und Vorverarbeitung der Testbilder

Um den exakten Versatz der Zellen und Zellkerne bestimmen zu können, wurden die Bilder in einem ersten Schritt übereinander gelegt. Hierzu wurde eine Bildregistrierung mit manuell gewählten Kontrollpunkten genutzt, um eine entsprechende Transformation zu bestimmen. Um alle störenden Artefakte zu entfernen, wurde anschließend ein verkleinerter Bildausschnitt gewählt, der in Abb. 6.9 für die Schräglichtaufnahmen gezeigt wird. Die resultierenden Bilder haben eine Größe von  $7131 \times 6134px$  was einer Größe von  $4.85 \times 4.17mm^2$  entspricht. Im Folgenden wird der Ausschnitt des Phasenkontrastbildes als  $I^{PH}$  bezeichnet und der Ausschnitt des IF-Bildes als  $I^{IF}$ .

### Detektion der Zellen und Zellkerne

Die Zelldetektion im Phasenkontrastbild  $I^{PH}$  und die Zellkerndetektion in der Fluoreszenzaufnahme  $I^{IF}$  erfolgte in beiden Bildern mit der globalen Schwellwertsegmentierung aus Abschnitt 2.4.2, wobei der Schwellwert mit der Methode von Otsu berechnet wurde (Otsu [84]). Die Segmentierung der Kerne und Zellen wurde manuell überprüft und die resultierenden Detektionsraten sind in Tab. 6.1 zusammengefasst. Die Erkennungsrate der DAPI-gefärbten Zellkerne war erwartungsgemäß hoch, da die Fluoreszenzbilder einen sehr guten Kontrast aufweisen. Die Erkennungsrate in den Schräglichtbildern liegt dagegen bei einer Genauigkeit von 0.9139 und einer Sensitivität von 0.8440, und damit leicht unter der Detektionsrate der Referenzdaten (vgl. Abb. 3.9).

### Approximation des Rauschens

Anhand der manuell markierten Daten konnte die Verschiebung  $\sigma^2$  der Zellkerne und Zellen bestimmt werden. Um das  $\sigma_F^2$  in dem Bayes-Test aus Gl.(6.4) zu bestimmen, muss aber beachtet werden, dass dieser Parameter nicht den Versatz der Zellkerne selber beschreibt, sondern den Versatz, mit dem die erstellte Transformation die Zellkerne in das Referenzbild projiziert. Das bedeutet, dass durch das Rauschen die zugrunde

		richtig	falsch	nicht	$\frac{RP}{RP+FP}$	$\frac{RP}{RP+FN}$
	Zellen	detektiert	detektiert	detektiert	(Genauigkeit)	(Sensitivität)
$I^{IF}$	2275	2272	1	3	0.9996	0.9987
$I^{PH}$	2275	1920	181	355	0.9139	0.8440

**Tabelle 6.1.:** Detektionsgenauigkeit der Zell- und Zellkerndetektion in Aufnahmen zur Bestimmung des Rauschmodells.

liegende Transformation mit einem systematischen Fehler geschätzt wird, d.h. je größer das Rauschen  $\sigma^2$  ist, desto ungenauer wird die rekonstruierte Transformation. Für die Validierung der Transformationen bedeutet das wiederum, dass das verwendete  $\sigma_F^2$  des Bayes-Tests größer gewählt werden sollte als das echte Rauschen  $\sigma^2$ : Im schlechtesten Fall wird eine Transformation nochmals mit einem Rauschen der Varianz  $\sigma^2$  verzerrt. Daher wird der Parameter  $\sigma_F^2$  über den Zusammenhang  $\sigma_F = 2\sigma$  bestimmt.

Für die betrachteten Referenzaufnahmen konnte die Varianz  $\sigma^2$  exakt bestimmt werden, was für ein allgemeines Zellbild nicht möglich ist. Es ist aber möglich, die Verschiebung der Zellkerne in einem beliebigen Zellbild über die durchschnittliche Fläche  $f_A$  der Zellen mit

$$\sigma_F = \sqrt{\frac{f_A}{\pi}} \quad (6.13)$$

zu bestimmen. In dieser Approximation wird das  $\sigma_F$  als der Radius  $r$  einer kreisrunden Zelle geschätzt. Bei dieser Abschätzung wird ausgenutzt, dass 95.54% aller Datenpunkte von normalverteilten Daten im Intervall  $[-2\sigma, 2\sigma]$  liegen. Somit entspricht der Radius  $r \approx 2\sigma$ , d.h. es gilt  $\sigma_F = r \approx 2\sigma$ , was eine gute Abschätzung für das  $\sigma_F$  aus Gl. (6.13) liefert.

### 6.2.2. Eigenschaften der Hashfunktionen

Aufbauend auf dem nun bekannten Rauschmodell wird ein künstliches Modell erzeugt, um verschiedene Eigenschaften der Hashfunktionen zu untersuchen. Diese artifiziellen Daten erlauben es, die genaue Stabilität der Hashwerte zu berechnen und den Einfluss der verschiedenen Parameter zu untersuchen. Im Folgenden wird die Größe der Hashstabellen abhängig von der maximalen Länge  $d_{max}$  betrachtet und es wird mit  $d_{min} = 0$  gerechnet.

#### Ein künstliches Rauschmodell

Es wird ein künstliches Modell verwendet, um den Einfluss des Rauschens bzw. der Zellkernverschiebung zu untersuchen. Hierzu wird ein Bild mit einer beliebigen Fläche  $S$  und Rauschen  $\sigma^2$  simuliert, in dem  $M^R$  zufällig verteilte Zellen bzw. Punkte generiert werden. Dieses erste Bild liefert die Zellen des Referenzbildes, die für die Erstellung der Hashtabellen  $H^R$  genutzt werden. Anschließend werden die Punkte mit  $\sigma^2$  verrauscht, um die Bildpunkte des Testbildes zu erzeugen. Hierbei wird die Zuordnung der Zellen gespeichert, d.h. es ist bekannt, welche Zelle  $c_m^T$  aus dem Testbild von welcher Zelle  $c_n^R$  aus dem Referenzbild erzeugt wurde. So ist es möglich, den genauen Einfluss des Rauschens zu analysieren und anzugeben, welche und wie viele Hashwerte durch das Rauschen verloren gehen.

Für die Berechnungen wurden teils die künstlichen, teils die echten Daten aus Abschnitt 6.2.1 verwendet. Es wird immer angegeben, welches Modell genutzt wurde. Falls das künstliche Modell Verwendung fand, werden die Parameter angegeben, mit denen die Daten simuliert wurden.

### Einfluss von $d_{max}$ auf die Größe der Hashtabelle

Die maximal erlaubte Länge  $d_{max}$  bzw. die Größe der Hashtabellen ist ein kritischer Parameter für die Laufzeit der Suche. Zum einen sollte  $d_{max}$  nicht zu groß gewählt werden, da die Anzahl an Konstellationen kombinatorisch wächst. Andererseits sollte  $d_{max}$  nicht zu klein gewählt werden, da die Auswirkung des Rauschens stärker wird, je kleiner die Länge einer Konstellation ist (Wolfson and Rigoutsos [120]). Der Parameter  $d_{max}$  ist daher entscheidend für die Suchzeit des Algorithmus bzw. für die Erstellung der Hashtabellen  $H^R$  und  $H^T$ . In Abb. 6.11 ist die Größe der Hashtabellen und die Laufzeit angegeben, die für die Erstellung der Referenzhashtabelle in dem Bild  $I^F$  benötigt wurde. Die beiden wichtigsten Informationen dieser Abbildung lassen sich wie folgt zusammenfassen:

1. Die Anzahl der Hashwerte, die für Konstellationen aus drei, vier und fünf Zellen gefunden werden, steigt unterschiedlich schnell an. Um diesen Anstieg mit  $d_{max}$  abzuschätzen, wird die Anzahl  $D$  der Zellen betrachtet, die in einem Kreis mit Durchmesser  $d_{max}$  liegen. Die Anzahl der Drei-, Vier- und Fünfecke, die eine maximale Länge  $d_{max}$  haben, kann mit dem Binomialkoeffizienten  $\binom{D}{3}$ ,  $\binom{D}{4}$ ,  $\binom{D}{5}$  abgeschätzt werden. Da die Anzahl der Zellen bei einer angenommenen Gleichverteilung proportional zur Fläche wächst, gilt  $D \sim \pi \left(\frac{d_{max}}{2}\right)^2$ . Somit kann die Anzahl der Drei-, Vier- und Fünfecke mit  $\binom{d_{max}^2}{3}$ ,  $\binom{d_{max}^2}{4}$ ,  $\binom{d_{max}^2}{5}$  abgeschätzt werden, d.h. die Anzahl der Konstellationen wächst kombinatorisch.
2. In Abb. 6.11(a) ist gezeigt, dass es eine Länge  $\tilde{d}_{max}$  gibt, mit der ungefähr gleich viele Drei-, Vier- und Fünfecke gefunden werden können. Da bei der Wahl von  $d_{max} = \tilde{d}_{max}$  die Hashtabellen  $H^R$  für Drei-, Vier- und Fünfecke dieselbe Größe haben, wird dieser Wert im Folgenden verwendet, um die verschiedenen Hashfunktionen zu vergleichen. Da die exakte Größe einer Hashtabelle nicht genau bestimmt werden kann, wird die Approximation

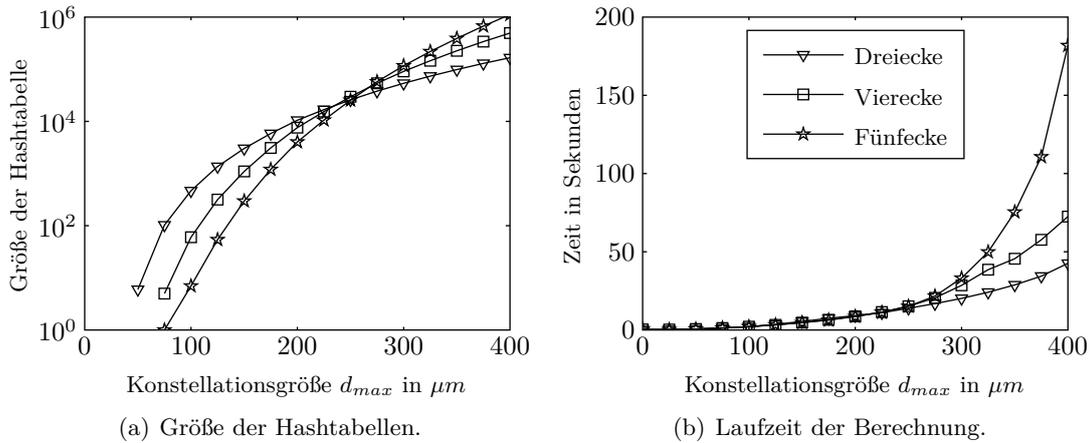
$$\tilde{d}_{max} \approx \sqrt{\frac{16S}{\pi M}} \quad (6.14)$$

genutzt, um  $\tilde{d}_{max}$  abzuschätzen. Hierbei beschreibt  $S$  die Fläche des Bildes und  $M$  die Zellzahl im Bild. Der entsprechende Nachweis ist im Anhang in Gl. (A.2) zu finden.

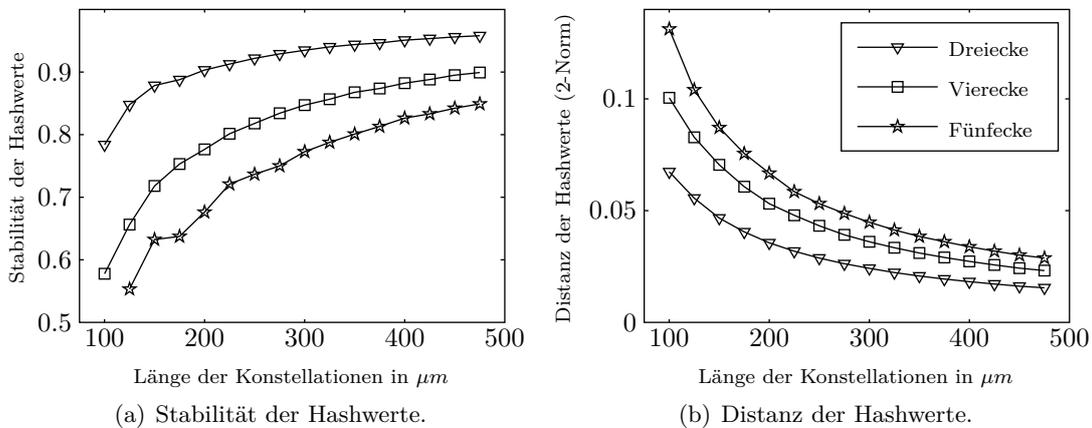
Der Wert von  $\tilde{d}_{max}$  aus Gl. (6.14) gibt den Wert an, bei dem dieselbe Anzahl an Dreiecken und Fünfecken mit der Länge  $\tilde{d}_{max}$  in einem Bild gefunden wird. Da die Hashtabelle auch Konstellationen von kleinerer Länge enthält, liefert Gl. (6.14) einen zu niedrigen Wert: In den Testbildern gilt  $\tilde{d}_{max} = 213\mu m$ , während in Abb. 6.11(a) gleichgroße Hashtabellen bei ca.  $240\mu m$  gefunden werden.

Dennoch liefert  $\tilde{d}_{max}$  eine gute Abschätzung für den Wert, mit dem dieselbe Rechenzeit für die Erstellung der Hashtabellen benötigt wird. Bei größeren Werten für  $d_{max}$  wächst der Speicherbedarf und die Rechenzeit für Drei-, Vier- und Fünfecke unterschiedlich schnell an (vgl. Abb.6.11(b)). Es empfiehlt sich daher, bei einem  $d_{max}$  Wert mit  $d_{max} \gg \tilde{d}_{max}$  aufgrund der benötigten Rechenzeit und Speicherkapazität für die Berechnung der Hashtabellen Dreiecke zu benutzen.

## 6.2. Das Petrischalen-Modell



**Abbildung 6.11.:** Die Größe der Hashtabellen ist für die Referenzaufnahmen aus Abschnitt 6.2.1 abhängig von der Konstellationsgröße  $d_{max}$  angegeben (a). Die logarithmische Darstellung wurde gewählt, um den Wert deutlicher darzustellen, an dem gleich viele Konstellationen für drei, vier oder fünf Zellen berechnet werden. In (b) ist gezeigt, dass die Laufzeit für die Berechnung der Konstellationen für Drei-, Vier- und Fünfecke kombinatorisch steigt.



**Abbildung 6.12.:** Die Stabilität und Distanz der Hashwerte wurde mit 50 simulierten Test- und Referenzbildern bestimmt. Die verwendete Bildgröße, Zellzahl und das Rauschen entspricht den Werten der Referenzaufnahmen (Bildgröße  $S = 5 \times 5 mm^2$ , Zellzahl  $M^R = 2000$ , Rauschen  $\sigma^2 = 16 \mu m^2$ ). Die Stabilität in (a) gibt den Prozentsatz an Konstellationen an, die sowohl in den Test- als auch in den Referenzdaten enthalten sind. Die Distanz in (b) gibt die 2-Norm  $\|h_l^T - h_k^R\|_2$  an, wobei  $h_l^T$  und  $h_k^R$  dieselben Konstellationen im Test- und Referenzbild beschreiben.

### Stabilität der Hashwerte für Drei-, Vier- und Fünfecke

Das größte Problem der beschriebenen Bildsuche ist der Versatz der Zellkerne, der in Abb.6.4 gezeigt ist und in dem vorherigen Kapitel 6.2.1 als Rauschmodell mit dem Parameter  $\sigma^2$  bestimmt wurde. Im besten Fall führt dieses Rauschen dazu, dass sich die Hashwerte derselben Konstellation von Zellen und Zellkernen (leicht) unterscheiden. Dieser Fehler wird mit der Distanz von zwei Hashwerten gemessen. Im schlechtesten Fall ist es aber möglich, dass es zu einem gesuchten Vieleck keine gültige Konstellation gibt. Eine Ursache können Fehlsegmentierungen sein, z.B. ein nicht detektierter Zellkern. Es ist aber auch möglich, dass durch das Rauschen eine gültige Konstellation verloren geht. Um das zu veranschaulichen, wird noch einmal die Abb. 6.5 betrachtet, in der das lokale Koordinatensystem der Hashwerte gezeigt ist.

Ein Hashwert eines Dreiecks ist z.B. nur dann eindeutig bzw. gültig, wenn die dritte Zelle näher an der ersten Zelle  $A$  liegt, als an der zweiten Zelle  $B$ . Ist dies nicht der Fall, werden die Zellen  $A$  und  $B$  vertauscht. Liegt nun die dritte Zelle genau zwischen den beiden ersten Zellen, kann es passieren, dass der Schwerpunkt der Zelle näher an Zelle  $A$  liegt, während der Schwerpunkt des Zellkerns näher an Zelle  $B$  liegt. In diesem Szenario ist es nicht mehr möglich, die Konstellation anhand der Hashwerte wiederzufinden: Auch wenn die Hashwerte in diesem Beispiel sehr ähnlich sind, so beschreiben sie doch eine gespiegelte Transformation, da die Zellen  $A, B$  vertauscht sind. Um den Effekt der Verschiebung der Zellkerne und den Einfluss dieser Fehler zu analysieren, werden zwei Maße definiert, mit denen die Stabilität und die Distanz der Hashwerte quantifiziert werden:

- Die **Stabilität**  $E_S$  gibt den Prozentsatz der Hashwerte  $h_l^T \in H^T$  an, zu denen es einen Hashwert  $h_k^R \in H^R$  aus dem Referenzbild  $I^R$  gibt, der dieselbe Konstellation beschreibt.
- Die **Distanz**  $E_D$  ist ein Maß für die Entfernung von zwei Hashwerten  $h_l^T$  und  $h_k^R$ , die dieselbe Konstellation im Test- und Referenzbild beschreiben. Sie wird mit  $E_D = \|h_l^T - h_k^R\|_2$  angegeben.

In Abb. 6.12 ist gezeigt, dass die Stabilität und das Rauschen, d.h. die Unähnlichkeit der Hashwerte, direkt von der Konstellationsgröße  $d_{max}$  abhängen: Je größer die Vielecke sind, desto robuster sind ihre Hashwerte. Diese Werte hängen wiederum von der Anzahl der verwendeten Zellen ab, d.h. ein Dreieck ist robuster als ein Fünfeck. Dieses Resultat war zu erwarten, da die Auswirkungen des Rauschens direkt mit der Größe der betrachteten Konstellationen korreliert (Wolfson and Rigoutsos [120]).

Die wichtigste Aussage der Abbildung ist jedoch, dass bei dem gegebenen Rauschen  $\sigma^2$  für alle verwendeten Figuren mehr als 50% der Hashwerte wiedergefunden werden können. Für die Suche bedeutet das, dass mindestens jedes zweite Vieleck wiedergefunden werden kann, falls genügend Transformationen überprüft werden. Die entscheidende Frage ist nun, wie viele Transformationen überprüft werden müssen, bis eine korrekte Bildposition gefunden wird. Diese Frage wird zum Teil durch die Distanz  $E_D$  in Abb. 6.12(b) beantwortet: Hier ist die gemittelte Entfernung der Hashwerte angegeben, die einen Wert für die Suchweite  $\epsilon$  aus Gl. (6.2) liefert. Wird das  $\epsilon$  entsprechend gewählt, liegen mindestens 50% der gesuchten Hashwerte in dem Suchradius.

### 6.2.3. Eigenschaften des Bayes-Tests

Um zu entscheiden, ob eine gegebene Transformation  $\mathcal{H}$  die richtige Bildposition zeigt, wird der Bayes-Test aus Gl.(6.12) verwendet, um mit Hilfe der Wahrscheinlichkeiten der Vordergrund- und Hintergrundmodelle  $F$  und  $B$  den Bayes-Faktor  $K$  zu berechnen. Liegt  $K$  über einem vorgegebenen Schwellwert  $\theta_K$ , wird die Transformation, die von  $\mathcal{H}$  beschrieben wird, als korrekt akzeptiert, d.h. die richtige Bildposition wurde gefunden. Der Bayes-Faktor  $K$  und der Schwellwert  $\theta_K$  werden daher direkt von den Werten und Parametern der Modelle  $F, B$  beeinflusst. Der Wert von  $K$  hängt folglich direkt von den Zellzahlen  $M^T, M^R$  der Test- bzw. Referenzbilder, von der Fläche des Testbildes  $I^T$  und dem verwendeten Rauschen  $\sigma_F^2$  ab. Um nun das Verhalten von  $K$  zu analysieren, werden die folgenden Auswertungen betrachtet:

1. Um das Verhalten des Bayes-Tests abzuschätzen, wird der maximale Wert des Bayes-Faktors  $K$  betrachtet, der abhängig von dem Rauschen  $\sigma_F^2$ , der Fläche  $S$  des Testbildes  $I^T$  und der Anzahl der Zellen  $M^T$  als

$$\log(K_{max}) = M^T \log\left(\frac{S}{2\pi\sigma_F M^T}\right) \quad (6.15)$$

bestimmt werden kann. Dieses  $K_{max}$  liefert eine obere Grenze für den Bayes-Faktor  $K$ , da er davon ausgeht, dass die Schwerpunkte der Zellkerne mit den Schwerpunkten der Zellen übereinstimmen. Die Herleitung für  $K_{max}$  ist im Anhang in Gl. (A.4) gegeben.

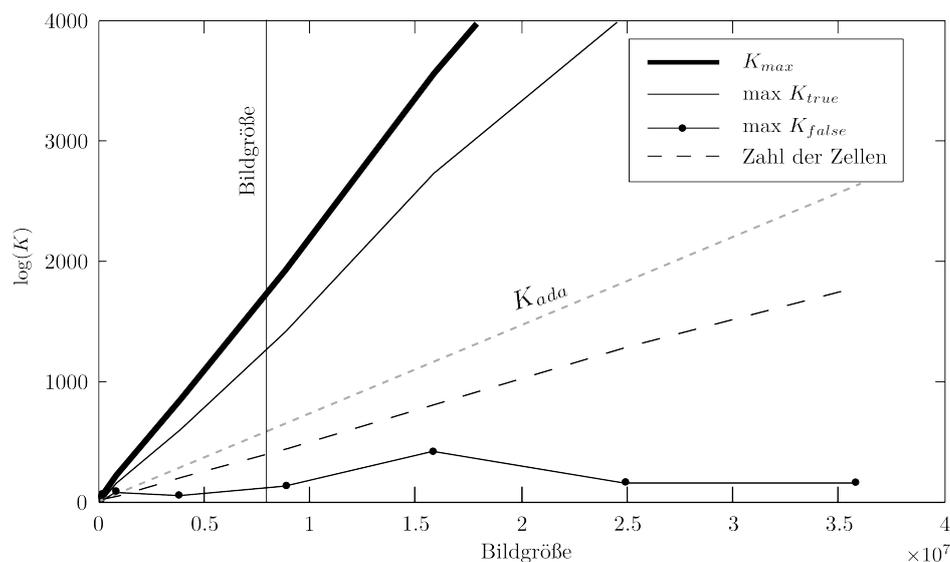
2. Es wurde das künstliche Rauschmodell aus Abschnitt 6.2.2 genutzt, um den Bayes-Faktor  $K$  aus Gl. (6.10) für Testbilder mit unterschiedlicher Fläche  $S$  und verschiedenen Zellzahlen  $M^T$  zu berechnen. In den simulierten Daten konnte unabhängig von dem Bayes-Test angegeben werden, welche Transformation die richtige Bildposition beschreibt und welche nicht, d.h. die größten  $K$ -Werte aller richtigen und aller falschen  $T_{k,l}$  konnten bestimmt werden.

Die Werte für die verschiedenen Bayes-Faktoren sind in Abb. 6.13 gezeigt. Hier ist zu erkennen, dass es keine statische Schwelle  $\theta_K$  für den Bayes-Faktor gibt, ab dem eine Transformation  $T_{k,l}$  als korrekt akzeptiert werden sollte.

Die wichtigste Aussage ist aber, dass es zwischen dem maximalen Bayes-Faktor der falschen und der richtigen Transformationen einen sicheren Korridor gibt, in dem nur die Bayes-Faktoren von gültigen Transformationen zu finden sind. Da die Fläche  $S$  und die Anzahl der Zellen  $M^T$  des gesuchten Testbildes  $I^T$  bekannt sind, bevor der Schwellwert  $\theta_K$  festgelegt wird, kann dieses *a-priori* Wissen genutzt werden, um einen adaptiven Wert für  $\theta_K$  zu bestimmen.

### 6.2.4. Vergleich der geometrischen Figuren

Die bisherigen Ergebnissen haben gezeigt, dass mit den Hashfunktionen für drei, vier und fünf Zellen stabile Hashwerte  $h_i^T$  berechnet werden können. Um zu entscheiden, welche Vielecke genutzt werden sollten, wird die Auswirkung der jeweiligen geometrischen Figur auf die gesamte Suche betrachtet.



**Abbildung 6.13.:** Das Verhalten des Bayes-Faktors  $K$  wurde mit künstlich generierten Testdaten analysiert (Bildgröße  $5 \times 5 \text{ mm}^2$ , Zellzahl  $M^R = 2000$ ,  $\sigma^2 = 16 \mu\text{m}^2$ ). In diesen Daten ist bekannt, welche Transformationen richtig sind, so dass der Bayes-Faktor  $K$  aus Gl. (6.10) für alle richtigen und falschen Transformationen berechnet werden konnte.  $K_{max}$  gibt den maximalen Wert an,  $\max K_{true}$  und  $\max K_{false}$  geben die maximalen Werte aller gültigen und aller ungültigen Transformationen an. Der eingezeichnete, adaptive Schwellwert wurde als  $K_{ada} = \frac{1}{3} K_{max}$  bestimmt. Die Werte wurden für sechs verschiedene Bildgrößen berechnet und es wurde jeweils das Maximum des Bayes-Faktors für die Suche nach zehn zufällig gewählten Bildausschnitten angegeben, wobei im Vordergrundmodell  $F$  eine Varianz von  $\sigma_F^2 = 64 \mu\text{m}^2 = (2\sigma)^2$  verwendet wurde.

Um die Suchzeit zu optimieren, sollte die Hashfunktion ausgewählt werden, mit der insgesamt die wenigsten Transformationen überprüft werden müssen. Um diese Anzahl abschätzen zu können, kann die Stabilität der Hashwerte aus Abb. 6.12 genutzt werden: Hier kann ein Suchradius angegeben werden, in dem mindestens 50% aller gültigen Hashwerte gefunden werden. Die entscheidende Frage ist nun, wie viele  $T_{k,l}$  in diesem Suchradius liegen. Diese Anzahl wird von zwei Faktoren bestimmt:

1. Die maximale Länge einer Konstellation  $d_{max}$  und die Zellzahl  $M$  bestimmen die Tabellengröße und damit auch die Zahl ähnlicher Hashwerte, die in einem Suchradius  $\epsilon$  liegen.
2. Die Dimensionalität der Hashwerte ist entscheidend für die Verteilungsdichte im Suchraum: Bei gleicher Tabellengröße ist ein sechsdimensionaler Suchraum wesentlich spärlicher besetzt als ein zweidimensionaler.

Um den Einfluss der Dimensionalität auf die Anzahl ähnlicher Hashwerte zu untersuchen, wird der Bereich betrachtet, der von einer Suche mit Radius  $r$  abgedeckt wird, d.h. es wird das Volumen einer Kugel mit  $n$ -Dimensionen betrachtet, das mit  $r^n$

wächst. Mit den gegebenen zwei-, vier- und sechsdimensionalen Hashwerten für Drei-, Vier- und Fünfecke kann die Anzahl der Hashwerte wie folgt abgeschätzt werden: Bei einer angenommenen Gleichverteilung der  $h_i^T$  im Suchraum muss davon ausgegangen werden, dass in einem Suchbereich, in dem  $N$  Fünfecke gefunden werden, eine Anzahl von  $N^3$  Vierecken und  $N^5$  Dreiecken gefunden werden. Um dieses Resultat für echte Daten zu überprüfen, wurde die Anzahl für Drei-, Vier- und Fünfecke in Hashtabellen mit  $d_{max} = \tilde{d}_{max}$  berechnet und die Anzahl der ähnlichen Hashwerte in einem Suchradius von  $\epsilon = 0.1$  bestimmt. Hier zeigt sich, dass durchschnittlich 4 Fünfecke, 91 Vierecke und 2037 Dreiecke gefunden wurden, was die obige Annahme bestätigt. Die genauen Daten sind in Abb. A.1 im Anhang angegeben.

### Optimierte Suche in einer diskretisierten Hashtabelle

In Tab. 6.2 ist die gesamte Suchzeit mit einer diskretisierten Hashtabelle und einer normalen Hashtabelle verglichen. Insgesamt zeigt sich, dass die Suche in einer diskretisierten Hashtabelle am performantesten durchgeführt werden kann. Das gilt besonders für hohe Dimensionen, d.h. für eine Suche nach Konstellationen mit vier oder fünf Zellen. Für den zweidimensionalen Suchraum der Dreiecke zeigt sich dagegen ein anderer Effekt: Hier dauert die Suche mit einer Diskretisierung am längsten. Diese lange Suchzeit ist ein direkter Einfluss der Diskretisierung selbst. Hierbei werden die Hashwerte in diskrete, gleichgroße Bereiche eingeteilt, wie in Abschnitt 6.1.4 beschrieben. Liegt ein Hashwert nah an der Grenze, kann der gesuchte Wert durch den Einfluss des Rauschens in den nächsten Bereich verrutschen und wird somit nicht mehr gefunden.

Geometrische Figur	Diskretisierung		Normale Suche	
	$H^R$	gesamt	$H^R$	gesamt
Dreiecke (2D)	8.8	717.5	8.1	494.4
Vierecke (4D)	13.7	79.3	12.6	135.1
Fünfecke (6D)	16.6	<b>48.4</b>	16.0	172.1

**Tabelle 6.2.:** Die gesamte Laufzeit der Bildsuche ist für verschiedene Datenstrukturen und Hashfunktionen in Sekunden angegeben. Die Berechnungen wurden mit künstlich generierten Daten durchgeführt (Referenzbild:  $5 \times 5mm^2$ , 4000 Zellen,  $\sigma^2 = 16\mu m^2$ ). In diesen Daten wurden 20 Suchbilder mit einer zufällig gewählten Position der Größe  $2 \times 2mm^2$  ausgeschnitten, deren Position anschließend bestimmt wurde. Die gesamte Suchzeit beinhaltet die Berechnung der Referenzhashtabelle  $H^R$ , den Aufbau der benötigten Diskretisierung und die zwanzigfache Suche. Die Suche wurde gestoppt, sobald 5 richtige Transformationen gefunden wurden. Die Zeit für die Erstellung der Referenzhashtabelle  $H^R$  und den Aufbau der benötigten Datenstrukturen ist separat angegeben.

### 6.3. Ergebnisse

Um die geometrische Positionssuche mit echten Zellbildern zu evaluieren, wurden in dem folgenden Abschnitt zwei Datensätze genutzt, die unterschiedliche Zellpopulationen zeigen. Als erster Datensatz wurden die Aufnahmen der rPSC-Population aus Abschnitt 2.3.2 verwendet, mit dem das Rauschmodell in Abschnitt 6.2.1 erstellt wurde. Um die Grenzen und die Robustheit der Algorithmen zu zeigen, wurde die Suche mit der automatisierten Zelldetektion und mit einer simulierten, fehlerhaften Zelldetektion durchgeführt. Es zeigte sich, dass die Bildpositionen auch noch erfolgreich gefunden wurden, wenn nur 50% der Zellen in dem Testbild betrachtet wurden.

Der zweite Datensatz zeigt die erste funktionale Fluoreszenzfärbung, in der ein Langzeitexperiment und eine Fluoreszenzanalyse kombiniert wurden. Hierzu wurde eine humane, Schweißdrüsen-abgeleitete Stammzellpopulation mehrere Tage in einem Langzeitexperiment beobachtet. Anschließend wurden die Zellen fixiert und das Protein Cytokeratin-19 und die Aktinfilamente gefärbt. Von den gefärbten Zellen wurde eine Mosaikaufnahme erstellt, die mit den letzten Aufnahmen der Langzeitbeobachtungen kombiniert wurde.

#### 6.3.1. Suche in den Referenzbildern

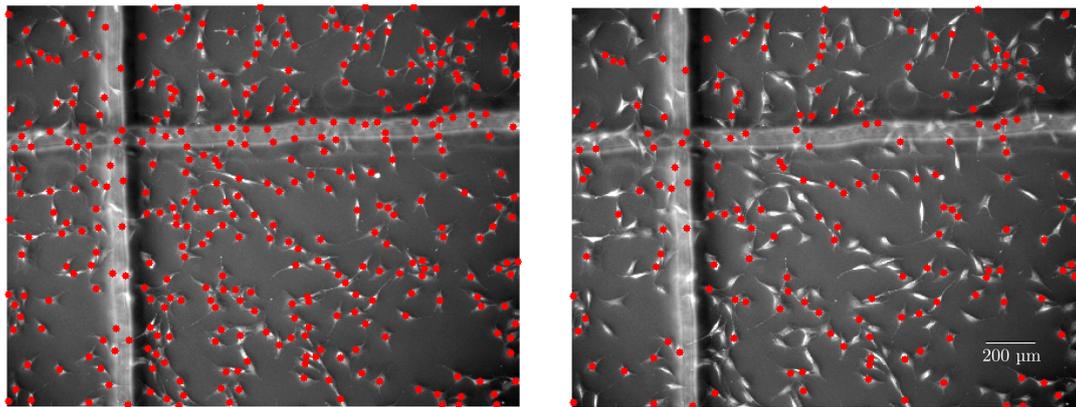
Der erste Datensatz aus Abschnitt 2.3.2 zeigt eine Zellpopulation von rPSC, die auf einer manuell markierten Petrischale ausgesät wurden (vgl. Abb. 6.9). In der gegebenen Suche wurden die Positionen der zwanzig Schräglichtaufnahmen in dem großen Mosaikbild aus Abb. 2.7, welches den DAPI-Kontrast zeigt, bestimmt. In Tab. 6.3 sind die Eigenschaften der verwendeten Bilder aufgelistet, die für die Suche verwendet wurden.

Die gesamte Positionssuche wurde mit Hashfunktionen für Drei-, Vier- und Fünfecke ausgeführt, wobei eine Diskretisierung der Hashtabelle  $H^R$  genutzt wurde (vgl. Abschnitt 6.1.4). Für die Zelldetektion wurde dieselbe Segmentierung verwendet, mit der das Rauschmodell bestimmt wurde (globale Schwellwertsegmentierung, wobei der Schwellwert mit der Methode von Otsu bestimmt wurde (Otsu [84])). Die Genauigkeit und Sensitivität der Zelldetektion lag bei 0.91 bzw. 0.84 (vgl. Tab. 6.1). Aufbauend auf dieser Zelldetektion wurden die folgenden zwei Positionssuchen durchgeführt:

1. In der ersten Auswertung wurden die Schwerpunkte der Zellen direkt aus den Zellmasken berechnet (vgl. Abb. 6.14(a)). Die Größe der Hashtabellen wurde mit einem  $d_{max} = \tilde{d}_{max} = 213.2\mu m$  nach Gl. (6.14) bestimmt, d.h. es wurde die Länge der Konstellationen gewählt, bei der dieselbe Anzahl an Drei- und Fünfecken in den Testbildern gefunden wird. Für den Bayes-Test aus Gl.(6.12) wurde mit dem Wert  $K_{max}$  aus Gl. (6.15) ein adaptiver Schwellwert  $\theta_K = \frac{K_{max}}{3}$  gewählt. Das Rauschen des Vordergrundmodells aus Gl. (6.6) wurde mit  $\sigma_F = 2\sigma = 8\mu m$  wie in Abschnitt 6.2 beschrieben gewählt.
2. In einer zweiten Auswertung wurde eine fehlerhafte Zelldetektion simuliert, indem 50% der gefundenen Zellen zufällig entfernt wurden (vgl. Abb. 6.14(b)). Um die geringe Zellzahl in den Testbildern auszugleichen, wurde die maximale Länge der Hashwerte auf  $d_{max} = 319.8\mu m$  erhöht. Es wurde ein adaptiver Schwellwert mit  $\theta_K = \frac{K_{max}}{4}$  und ein Rauschmodell mit  $\sigma_F = 2\sigma = 8\mu m$  gewählt.

### 6.3. Ergebnisse

Die vorgestellten Algorithmen waren in der Lage, die exakte Position aller zwanzig Referenzaufnahmen zu bestimmen. Um die Genauigkeit der Rekonstruktion zu evaluieren, ist für jedes Bild die Differenz der sichtbaren Zellen und Zellkerne berechnet worden, nachdem die Testbilder  $I^T$  mit der gefundenen Transformation in den Raum des Referenzbildes projiziert wurde. Die mittlere Verschiebung und die Varianz dieser Differenzen sind in Tab. 6.4 angegeben. Eine vollständige Übersicht für die Positionsbestimmung mit den fehlerhaften und fehlerfreien Daten für Drei-, Vier- und Fünfecke ist im Anhang in Tab. A.3 gegeben.



(a) Alle detektierten Zellen.

(b) Fehlerhafte Segmentierung.

**Abbildung 6.14.:** Das Bild zeigt die erste Position der Schräglichtaufnahmen, mit denen das Rauschmodell erstellt wurde. In (a) sind die Mittelpunkte aller detektierten Zellen als rote Punkte markiert, während in (b) eine schlechte Segmentierung simuliert wurde, indem zufällig 50% der Zellkerne entfernt wurden. Das geometrische Hashverfahren war in der Lage, die exakte Position für alle 20 Testbilder zu finden, selbst wenn die Hälfte aller Zellen nicht betrachtet wurde.

	Schräglichtbilder	Fluoreszenzbilder	Mosaikbild
Bilder	20 Schräglichtbilder (Testbilder)	80 DAPI Aufnahmen	1 Mosaikbild (Referenzbild)
Größe	$1376 \times 1036$	$1388 \times 1040$	$10126 \times 9466px$
Auflösung	$1.6\mu m/px$	$0.68\mu m/px$	$0.68\mu m/px$
Zellen	$\approx 100 - 180$	$\approx 20 - 40$	2275

**Tabelle 6.3.:** Die Referenzaufnahmen aus Abschnitt 2.3.2 wurden verwendet, um die geometrischen Hashverfahren zur Positionsbestimmung von Zellbildern an echten Daten zu evaluieren. Das Mosaikbild wurde mit dem *Image Composite Editor* erstellt (vgl. Abschnitt 2.3.2).

Position	überprüfte Konstellationen	$\sigma_F$ in $\mu m$		Mittelwert in $\mu m$	
		$x$	$y$	$x$	$y$
1	270	7.48	8.70	3.29	-3.28
2	291	6.52	6.98	-0.69	3.11
3	125	7.51	5.95	5.93	2.22
4	1164	16.95	19.18	-10.77	14.43
5	138	10.47	9.77	-4.01	6.89
6	157	6.66	7.12	-0.08	-0.82
7	70	7.75	5.89	-2.02	-3.01
8	187	7.73	7.18	-0.46	-1.91
9	154	6.09	6.76	-0.13	2.17
10	383	6.80	6.71	-0.40	-0.29
11	101	6.31	7.01	-0.34	-1.44
12	408	6.21	6.40	-0.33	1.91
13	473	6.16	7.44	-1.17	1.75
14	177	10.02	7.99	-11.54	-1.17
15	107	7.28	6.95	-0.52	-2.05
16	203	6.72	6.75	2.02	-4.26
17	1653	12.15	13.97	9.04	2.57
18	656	9.07	11.55	-0.19	-5.13
19	348	9.75	10.65	-2.50	6.44
20	911	6.58	8.01	0.47	1.54

**Tabelle 6.4.:** Die Ergebnisse der Positionssuche ist für die Referenzbilder als mittlere Verschiebung  $\mu$  und Standardabweichung  $\sigma$  der Zellkerne in  $x$ - und  $y$ -Richtung angegeben, die mit Hilfe der gefundenen Transformation bestimmt wurden. Die gezeigten Verschiebungen resultieren aus einer Suche, in der Fünfecke genutzt wurden.

Die Ergebnisse in Tab. A.3 zeigen, dass die Suche der Positionen der Referenzaufnahmen mit einer hohen Genauigkeit finden konnte. Die berechneten Transformationen wurden anschließend genutzt, um ein Mosaikbild der Schräglichtbilder zu erstellen, das dem exakten Ausschnitt der Referenzbilder der DAPI-Aufnahmen entspricht (vgl. in Abb. A.4 im Anhang).

Die Positionssuche mit den fehlerhaften Daten konnte ebenfalls alle Aufnahme-positionen wiederfinden. Allerdings zeigt sich hier die Grenze der Algorithmen, da zwei Bilder, die in den Ecken lagen, nicht immer gefunden wurden, wenn Hashfunktionen für Vier- und Fünfecken genutzt wurden. Der Grund ist, dass diese Bilder

- nur zu einem Drittel mit den Fluoreszenz-Mosaikbildern überlappen. Da hierbei nur die Hälfte der Zellen betrachtet werden, gibt es weniger Vier- und Fünfecke als gültige Dreiecke.
- die manuelle Markierung zeigen, was zu einer erhöhten Fehlsegmentierung führt (vgl. Abb. 6.4).

Es zeigt sich aber, dass erst in diesem extremen, konstruierten Szenario die Positionen nicht mehr bestimmt werden können.

### 6.3.2. Die erste funktionale Immunfluoreszenzfärbung

Die bisher analysierten Datensätze zeigen eine Zelllinie des Rattenpankreas, die aufgrund ihrer guten Verfügbarkeit und einfachen Kultivierung genutzt worden sind. Um die Übertragbarkeit auf weitere Zellen und die Kombination mit einer Fluoreszenzanalyse zu zeigen, wird in einem abschließend vorgestellten Experiment eine humane Zellpopulation analysiert.

Hierzu wurde eine Population von humanen, Schweißdrüsen-abgeleiteten Stammzellen in einem Langzeitexperiment an sechs Positionen über 90 Stunden beobachtet. Unmittelbar nach der letzten Aufnahme erfolgte eine Fixierung und eine Fluoreszenzanalyse der Zellen, wobei das Protein Cytokeratin-19 (CK19, roter Kanal) und die Aktinfilamente ( $\alpha$  *smooth muscle actin* – SMA, grüner Kanal) gefärbt wurden. Die entsprechenden Fluoreszenzaufnahmen sind in Abb. 6.17 gezeigt. Anschließend wurden 360 einzelne Aufnahmen im grünen, roten und blauen Fluoreszenzkanal erstellt und zu einem großen Mosaikbild der Größe von  $10128 \times 8425$  bzw.  $8.7 \times 7.3 \text{mm}^2$  zusammengefügt. Die Bildverarbeitung erfolgte analog zu den verwendeten Methoden aus Abschnitt 2.3.2. Die Details der verwendeten Bilder sind in Tab. 6.5 gegeben.

Der erste Schritt der Suche bestand darin, die Zellen und die Zellkerne in den Bilddaten zu detektieren. Hierzu wurde auf die Segmentierung aus Abschnitt 2.4 zurückgegriffen, d.h. es wurde die globale Schwellwertsegmentierung genutzt, die einen Schwellwert nach der Methode von Otsu berechnet. Die Mittelpunkte der detektierten Zellkerne aus dem Fluoreszenzbild (DAPI-Kanal) wurden anschließend genutzt, um die Hashwerte für das Referenzbild zu berechnen; analog dazu wurden die Hashtabellen für die sechs Aufnahmepositionen erstellt. Mit der Fläche der segmentierten Zellen im Testbild  $I^T$  wurde mit der Approximation aus Gl. (6.13) eine Varianz von  $\sigma_F^2 = 140 \mu\text{m}^2$  bestimmt. Als geometrische Figuren wurden Fünfecke gewählt und die maximale Länge der Figuren wurde mit  $d_{max} = \tilde{d}_{max} = 226 \mu\text{m}$  bestimmt, wobei die Zelldichte der Mosaikbilder genutzt wurde, um nach Gl. (6.14) das  $\tilde{d}_{max}$  zu bestimmen.

	Phasenkontrastbilder	Fluoreszenzbilder	Mosaikbild
Bilder	6 Schräglichtbilder (Testbilder)	360 DAPI Aufnahmen	1 Mosaikbild (Referenzbild)
Größe	$1376 \times 1036$	$1388 \times 1040$	$10128 \times 8425$
Auflösung	$1.6 \mu\text{m}/\text{px}$	$0.68 \mu\text{m}/\text{px}$	$1.36 \mu\text{m}/\text{px}$
Zellen	253 – 393	$\approx 20 - 50$	11422

**Tabelle 6.5.:** Bilddaten der funktionalen Immunfluoreszenzfärbung. Das Mosaikbild wurde mit dem *Image Composite Editor* (ICE Version 1.4.4.0, Microsoft Research, Redmond, USA) zusammengesetzt und um den Faktor zwei skaliert.

### Die verwendeten Parameter

Aufbauend auf den vorgestellten Ergebnissen liefert die folgende Liste einen Überblick über die verwendeten Parameter, die für eine robuste Positionsbestimmung eingesetzt wurden. Die Werte lassen sich direkt aus den Zellbildern ableiten und sind als Empfehlung zu verstehen, wie für ein gegebenes Zellbild optimale Parameter berechnet werden können.

1. Als geometrische Figuren wurden **Fünfecke** verwendet, da hier im Vergleich zu Drei- oder Vierecken die wenigsten Transformationskandidaten überprüft werden, bis die richtige Position gefunden wird.
2. Die Größe der Hashtabellen wird begrenzt, indem mit  $d_{max} = \tilde{d}_{max}$  aus Gl. (6.14) eine **maximale Länge der Konstellationen** angegeben wird. Hierbei beschreibt  $\tilde{d}_{max}$  die Länge, mit der in den gegebenen Bildern dieselbe Anzahl an Dreiecken und Fünfecken gefunden wird. Es hat sich gezeigt, dass dieser Wert einen guten Kompromiss für die Größe der Hashtabelle  $H^R$  liefert.
3. Der maximale Wert des Bayes-Faktors  $K_{max}$  lässt sich mit der Gl. (6.15) explizit für jedes gegebene Zellbild berechnen. Mit diesem  $K_{max}$  kann ein **adaptiver Schwellwert** als  $\theta_{ada} = \frac{K_{max}}{3}$  bestimmt werden. Ist die gegebene Zelldetektion allerdings sehr fehlerhaft, ist die Abschätzung für  $K_{max}$  nicht mehr gültig und es sollte mit  $\theta_{ada} = \frac{K_{max}}{4}$  ein niedrigerer, adaptiver Schwellwert gewählt werden.
4. Die **Varianz**  $\sigma_F^2$  des Vordergrundmodells  $F$  ist entscheidend für den Bayes-Faktor bzw. für den adaptiven Schwellwert  $\theta_{ada}$ . Der Wert wurde als  $\sigma_F = 2\sigma$  über die Varianz  $\sigma^2$  der Zellkernverschiebung berechnet, die mit Hilfe der Referenzaufnahmen exakt ermittelt werden konnte. Falls das exakte Rauschmodell, d.h. die Varianz  $\sigma^2$  nicht bekannt ist, kann  $\sigma_F$  mit der durchschnittlichen Fläche  $f_A$  der Zellen als  $\sigma_F = \sqrt{\pi f_A} \approx 2\sigma$  geschätzt werden (vgl. Gl. (6.13)).
5. In der gegebenen Suche wurden maximal 2000 **Konstellationen** überprüft, um eine gültige Position zu finden. Wird keine gültige Transformation gefunden, so liegt das Zellbild mit sehr hoher Wahrscheinlichkeit nicht in dem gegebenen Mosaikbild.

### Ergebnisse der Suche

Von den sechs Aufnahmepositionen der Langzeitbeobachtung wurden fünf Bilder wiedergefunden. Da das dritte Bild aber nur zur Hälfte in dem großen Fluoreszenzbild liegt, wird davon ausgegangen, dass die sechste Aufnahmeposition nicht in dem Mosaikbild enthalten ist. In Tab. 6.6 sind die Details der sechs Bilder angegeben.

In Abb. 6.17 ist die zweite Aufnahmeposition mit den überlagerten Fluoreszenzkanälen zu sehen. Das Fluoreszenzbild in Abb. 6.17(b) ist anhand der berechneten Transformation aus dem großen Mosaikbild ausgeschnitten worden. Beide Bilder zeigen, dass mit den fünf gefundenen Fünfecken eine sehr genaue Transformation erstellt werden konnte. Die Bilder aller fünf Positionen sind in Abb. 6.15 mit den überlagerten DAPI-Kernen gezeigt.

Position	1	2	3	4	5	6
Gefunden	ja	ja	ja	ja	ja	nein
Zellzahl $M^T$ im Testbild	270	361	253	257	301	393
Suchzeit in Sekunden	3.7	5.0	9.6	17.6	10.2	77.4
überprüfte Konstellationen	82	93	224	443	278	2000
Schwellwert $\theta_{ada}$	304	351	301	295	344	406

**Tabelle 6.6.:** Ergebnisse der Bildsuche. Die möglichen Transformationen wurden in zufälliger Reihenfolge überprüft und die Suche wurde fortgesetzt, bis fünf gültige  $T_{k,l}$  gefunden wurden (adaptiver Schwellwert  $\theta_{ada}$ , maximal 2000 Transformationskandidaten wurden überprüft). Die Suchzeit fasst die Berechnung der Hashtabellen  $H^T$  und der Validierung zusammen. Die gesamte Suchzeit inklusive der Berechnung der Hashtabelle  $H^R$  betrug 11 Minuten und 30 Sekunden. Die gesamte Auswertung wurde auf einem Intel® Core™ i7 mit 8GB RAM durchgeführt, es wurde eine parallele Berechnung mit acht Prozessen genutzt.

### 6.3.3. Zusammenfassung

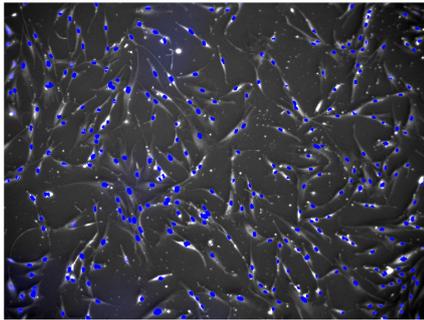
Mit der FFA wurde eine neue Methode vorgestellt, in der ein geometrisches Hashverfahren genutzt wird, um Zellbilder, die mit verschiedenen Kontrastmechanismen erstellt wurden, zu kombinieren. In der gezeigten Anwendung werden die Zellen in gegebenen Schräglichtaufnahmen in einem großen Mosaikbild wiedergefunden, wenn die DAPI-gefärbten Zellkerne sichtbar sind. So ist es möglich, das nicht-invasive Zelltracking, das in dieser Arbeit vorgestellt wurde, um eine Analyse der Proteinexpression zu erweitern.

Zwar gibt es bereits weitere Ansätze, in denen das Verhalten und die Differenzierung von Stammzellen in Langzeitexperimenten beobachtet wurde. In den ausgewerteten Experimenten wurden jedoch Reportersysteme eingesetzt, die auf einer genetischen Modifikation der Zelle bzw. der Zell-DNA basieren. Mit denen es nicht möglich war, eine automatisierte Auswertung durchzuführen (Eilken et al. [33]). In diesen Experimenten konnten daher nur wenige Zellen beobachtet werden, die zudem noch manuell ausgewertet werden mussten.

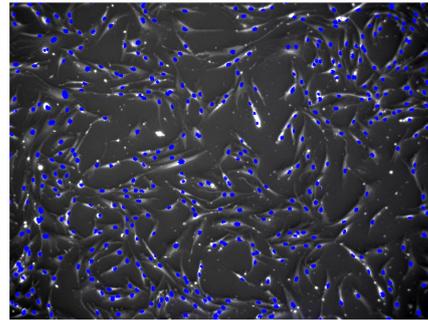
Weiterhin liefert die Positionssuche der FFA einen generellen Ansatz, um (Zell-) Bilder zu kombinieren, in denen dieselben Objekte in unterschiedlichen Bildmodalitäten gezeigt sind. In diesem Kontext der Bildverarbeitung liefert die folgende Liste verschiedene weitere Anwendungen und Optimierungen:

- Da die vorgestellten Algorithmen die Position der Zellen bzw. Objekte mit einer sehr hohen Genauigkeit bestimmen, kann diese Methode weiterhin genutzt werden, um z.B. automatisiert Kontrollpunkte für eine Bildregistrierung zu extrahieren.
- Die vorgestellten Algorithmen sind in der Lage, die Position der Testbilder mit einer hohen Genauigkeit zu bestimmen, d.h. es werden keine falschen Transformationen als gültig erkannt. Anhand der Ausgabe der Algorithmen ist es aber leicht möglich, eine Validierung der finalen Transformation einzuführen und so ein direktes Maß für die Güte der gefundenen Positionen zu erhalten. Diese Validierung kann anhand der fünf gefundenen  $T_{k,l}$  oder der approximierten Genauigkeit der Zelldetektion, die als  $\nu$  aus Gl. (6.8) bekannt ist, durchgeführt werden.
- Die kritische Operation für die Laufzeit der Suche ist die Erstellung der Hashtabellen  $H^R$ , die abhängig von der Größe des Übersichtsbildes bis zu 90% der gesamten Laufzeit ausmacht. Um die Berechnungen weiter zu beschleunigen, könnte eine parallele Implementierung unabhängig von MATLAB eingesetzt werden.

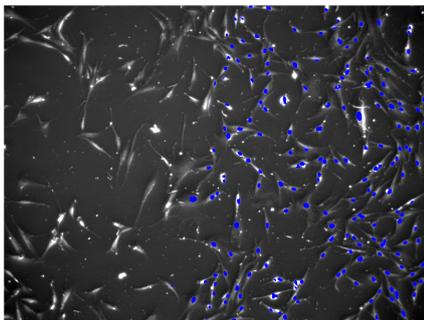
Insgesamt zeigt sich, dass mit der FFA ein genereller Algorithmus vorgestellt wurde, um Zellbilder unterschiedlicher Modalitäten zu kombinieren. Es wurde eine direkte Anwendung gezeigt, in dem die zellulären Genealogien um eine Analyse der Proteinexpression erweitert wurden. Somit kann die Differenzierung der Zellen in einem dynamischen und zeitlichen Kontext interpretiert werden, was direkt zu einem tieferen Verständnis der regulierenden Mechanismen der beobachteten *in vitro*-Zellkulturen führen kann.



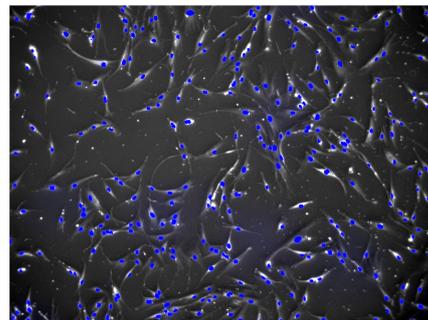
(a) Position 1.



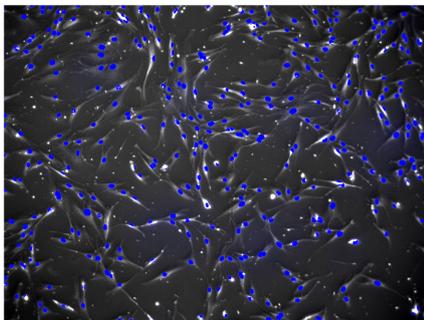
(b) Position 2.



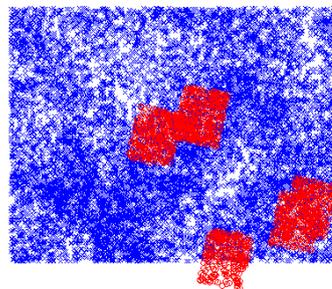
(c) Position 3.



(d) Position 4.

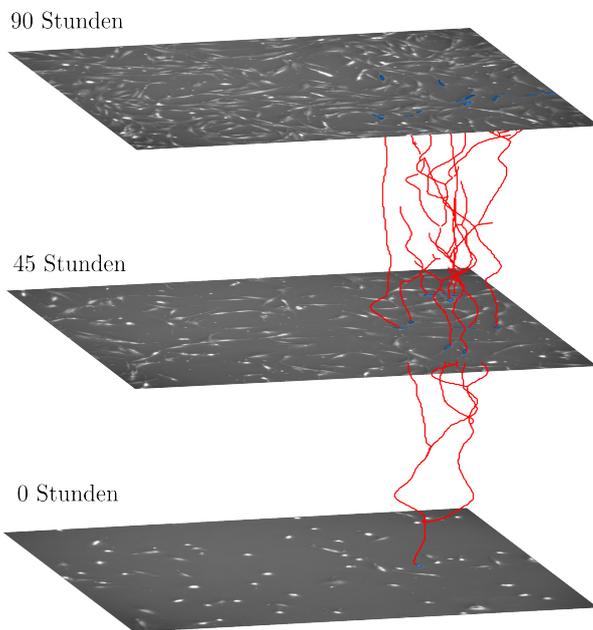
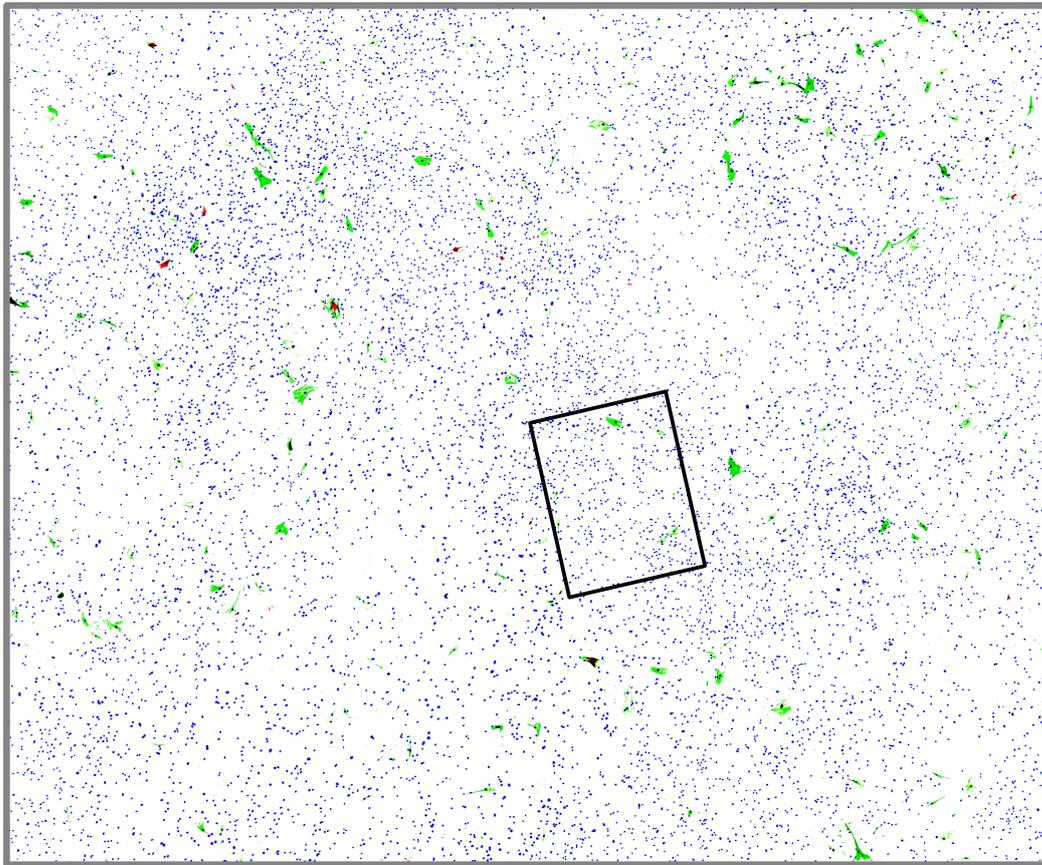


(e) Position 5.

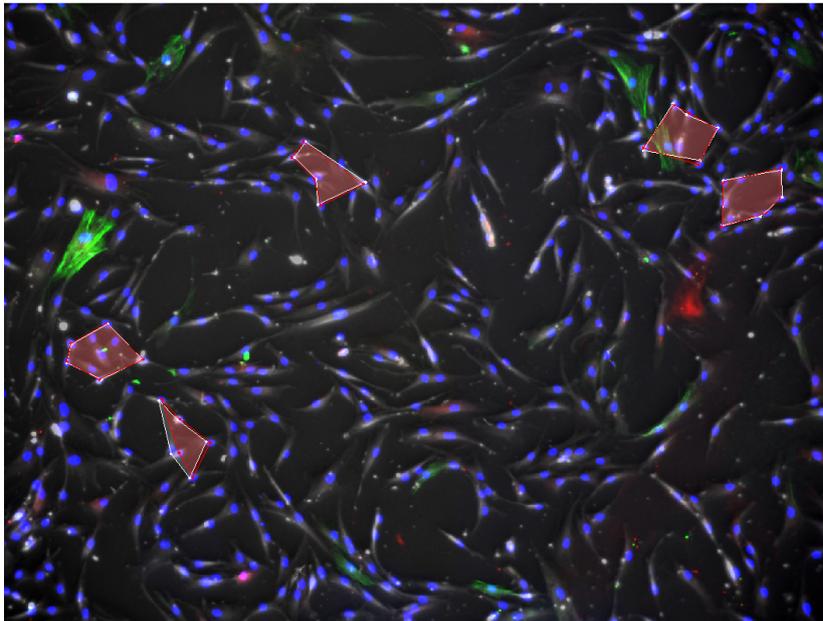


(f) Übersicht der gefundenen Bilder.

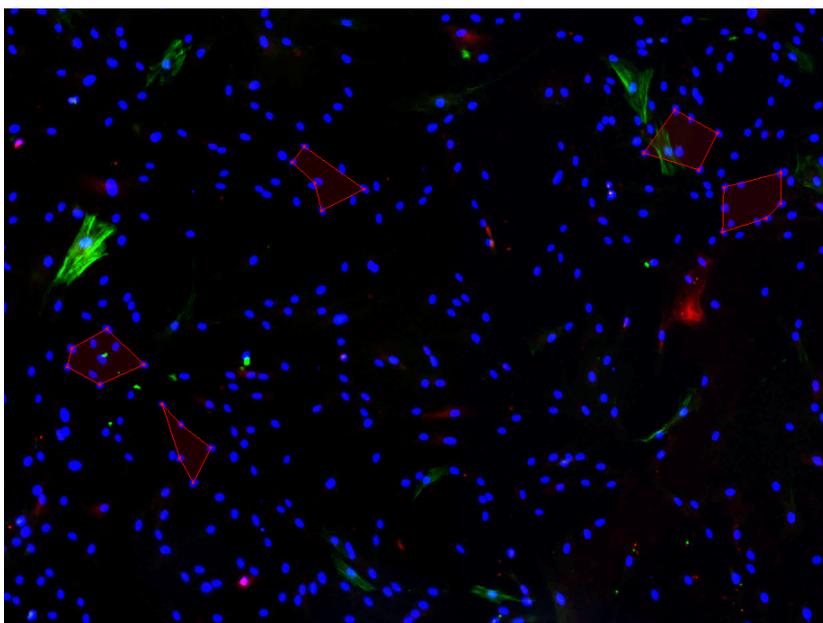
**Abbildung 6.15.:** Fünf der sechs Bilder konnten in dem Mosaikbild wiedergefunden werden. Die dritte Aufnahmeposition liegt nur halb in dem Bereich der Mosaikaufnahme. Daher ist es wahrscheinlich, dass das sechste Bild, das nicht wiedergefunden wurde, außerhalb der aufgenommenen Fläche des Mosaikbildes lag. Das Bild in (f) zeigt alle Positionen der DAPI-Kerne als blaue Kreuze und die Position der Zellen aus den einzelnen Bildern als rote Kreise.



**Abbildung 6.16.:** Kombiniertes Mosaikbild, in dem die drei Fluoreszenzaufnahmen im roten (CK-19), grünen ( $\alpha$ -SMA) und blauen (DAPI) Kanal überlagert sind. Zur besseren Darstellung ist der schwarze Hintergrund invertiert worden. Der Stammbaum unten links zeigt die größte automatisch extrahierte Zellfamilie.



(a) Kombiniertes Bild.



(b) Zusammengesetztes Fluoreszenzbild.

**Abbildung 6.17.:** In (a) ist die zweite von sechs Bildpositionen gezeigt, die in einem Langzeitexperiment beobachtet wurden. Die Bilddaten sind mit der Fluoreszenzfärbung überlagert, die im Anschluss an die Beobachtung erstellt wurde. Der Ausschnitt in (b) wurde mit Hilfe der Transformation kombiniert, die mit der geometrischen Suche berechnet wurde. Die Färbungen zeigen  $\alpha$ -SMA im grünen, CK19 im roten und die DAPI gefärbten Zellkerne im blauen Kanal. Die Fünfecke zeigen die gefundenen Konstellationen des Schräglichtbildes (weiß) und des Fluoreszenzbildes (rot).

mainmat

# 7

## Diskussion und Ausblick

In meiner Dissertation wurden verschiedene Neuerungen für die Analyse von *in vitro*-kultivierten Zellen mittels Langzeitbeobachtung umgesetzt. Im Bereich des Zelltracking umfassen sie zwei wichtige Aspekte: Die Validierung und die Evaluation. Als Validierung wird die unabhängige Analyse bzw. Identifizierung von möglichen Tracking-Fehlern verstanden. Diese ermöglicht es, die Genauigkeit der erzielten Ergebnisse deutlich zu verbessern und Fehler zweiter Art (falsch positive) zu vermeiden. Die Evaluation bezeichnet dagegen die systematische Analyse und Fehlerquantifizierung, mit der eine Zelldetektion oder ein Zelltracking-Algorithmus ausgeführt wird. Nur wenn beide Konzepte genutzt werden, d.h. wenn die Algorithmen evaluiert und die Ergebnisse validiert werden, wird eine automatisierte Auswertung verlässliche Ergebnisse liefern.

Weiterhin habe ich das Anwendungsspektrum der entwickelten Methoden untersucht und sukzessive ausgebaut. Mit der funktionalen Fluoreszenzanalyse aus Kapitel 6 wurde ein neues Konzept entwickelt, um das nicht-invasive Zelltracking mit einer Fluoreszenzanalyse zu kombinieren. So kann die genealogische Information automatisiert in einem biologischen Kontext interpretiert werden. Mit dieser Auswertung kann z.B. das Differenzierungspotential einer Zellpopulation auf Einzelzell-Ebene quantifiziert werden. Eine anschließende Auswertung liegt in dem weitgefassten Bereich der Systembiologie. Sie erlaubt eine holistische Analyse, in der die Ergebnisse der verschiedenen Auswertungen in einem gemeinsamen, interdisziplinären Kontext interpretiert werden.

### **Eine Parametrisierte Zellkultur mittels Validiertem Zelltracking**

Die Hauptaussage des dritten Kapitels ist, dass eine automatisierte Auswertung der Langzeitexperimente nur durchgeführt werden kann, wenn die extrahierten Daten mit einer Validierung aufbereitet werden. Dies wird sofort ersichtlich, wenn die starke Fehlerfortpflanzung in den genealogischen Bäumen betrachtet wird (vgl. Abschnitt 3.1.1). Weiterhin kann ein einzelner Fehler, wie z.B. ein verwechselter Zellpfad, dazu führen, dass eine gesamte Genealogie in einem falschen Kontext interpretiert wird (vgl. Abb. 3.2). Daher besteht die wichtigste Aufgabe der automatisierten Auswertung in der Minimierung der Fehler zweiter Art (falsch positive), d.h. in der Datenvalidierung.

Bei der durchgeführten Validierung nimmt die Mitosedetektion eine bedeutende Rolle ein: Nur wenn die Zellteilungen mit einer hohen Sicherheit gefunden werden, ist eine Extraktion der gesuchten Verwandtschaften bzw. der Konnektivität der Zellpfade mit einer hohen Genauigkeit möglich. In Kapitel 4 wurde eine unabhängige Mitosedetektion implementiert, um die einzelnen, vollständigen Zellpfade zu verknüpfen. Diese Unabhängigkeit wurde erreicht, indem die Zellen mit zwei Kontrastmechanismen aufgenommen

wurden: Der erste Kontrast wird für die Zelldetektion, der zweite für die Mitosedetektion genutzt. Die Auswirkung der Validierung und der unabhängigen Mitosedetektion sind in Abb. 4.16 für zwei verschiedene Zelltracking-Ansätze gezeigt. Die Ergebnisse verdeutlichen, dass diese generellen Konzepte unabhängig von der verwendeten Methode angewendet werden können und zu sehr genauen Ergebnissen führen (Genauigkeit  $\frac{RP}{RP+FP} = 0.96$ , Sensitivität  $\frac{RP}{RP+FN} = 0.66^1$ , vgl. Abb. 3.2). Für die Rekonstruktion der Genealogien bedeutet dies, dass 66% der Zellpfade bereits erfolgreich gefunden wurden. Hierauf aufbauend ist es leicht, vollständige Stammbäume zu erstellen: Dieselben Hilfsmittel, d.h. die Validierung und die grafischen Hilfsprogramme aus Abb. 3.7(a), können genutzt werden, um die fehlerhaften Pfade gezielt zu identifizieren und die wenigen Fehler zu korrigieren. Entscheidend ist hier, dass bereits  $\frac{2}{3}$  der gesamten Pfade bekannt sind und nur noch die verbleibenden 30% richtig verknüpft werden müssen. Somit können die genealogischen Informationen in kurzer Zeit extrahiert werden.

Eine mögliche Fehlerquelle sind systematisch übersehene Zellpfade, die von der konservativen Validierung entfernt werden. Zwar zeigen die bisher analysierten Daten, dass diese Fehler nicht auftraten (vgl. Abb. 3.16). Es ist aber möglich, dass Zellpfade mit gemeinsamen Eigenschaften wie z.B. einer sehr kurzen Dauer des Zellzyklus, übersehen werden. Jedoch finden sich diese Zellen in den Tracking-Ergebnissen als fehlerhafte Pfade und können somit leicht korrigiert werden. Der einzig denkbare, kritische Fall ist, wenn eine Zellen in keinem Bild, in dem sie aufgenommen wurde, von der Detektion erkannt werden kann. Nur in diesem sehr unwahrscheinlichen Fall wird die Zelle nicht von der Validierung betrachtet. Dennoch sollte berücksichtigt werden, dass dies eine Quelle für potentielle, systematische Fehler darstellen könnte.

Eine weitere, direkte Anwendung des validierten Zelltracking ist die Parametrisierung der Zellkultur, d.h. die Quantifizierung verschiedener Eigenschaften der Zellpopulation. So können Proliferationskurven, Verteilung und Änderung der Zellgröße oder der Grad der Konfluenz mit einer sehr hohen Genauigkeit nicht-invasiv bestimmt werden. Während diese Parameter nur eine Zelldetektion voraussetzen, bietet das validierte Zelltracking auch Maße, die mit keiner weiteren Methode extrahiert werden können: Die Migration der Zellen und die Verteilung der Lebenszeiten bzw. der Zellzyklusdauer der Zellen. Die statistische Verteilung dieser Parameter kann z.B. genutzt werden, um die Heterogenität der beobachteten Zellpopulation nicht invasiv zu beschreiben. Direkte Anwendungsbeispiele sind mit der Vitalitätsuntersuchung in Abschnitt 5.1 und der Migrationsstudie aus Abschnitt 5.2 gegeben.

Insgesamt kann die Parametrisierung als zwangsläufiges Ergebnis des Zelltracking angesehen und als logische Erweiterung der konventionellen Zellkultur betrachtet werden.

## Evaluation, Referenzdaten und Fehlermaße

Wie bereits in der Einleitung erwähnt, wird für eine systematische Evaluierung und Fehleranalyse von Zelltracking-Algorithmen ein Benchmark- oder Referenzdatensatz benötigt. In der Literatur wurden bisher verschiedene Datensätze genutzt, um eine Evaluation durchzuführen. So wurde z.B. auf (wenige) manuell erstellte Stammbäume

<sup>1</sup>FP = falsch positive, RP = richtig positive, FN = falsch negative

---

zurückgegriffen (Al-Kofahi et al. [9], Chen et al. [20], Yang et al. [122], Zhou et al. [125]), die zum Teil nur einen Ausschnitt bzw. nur einen kleinen Anteil der beobachteten Zellpopulation zeigen (Degerman et al. [28], Li et al. [66]). Ein weiterer Nachteil dieser Daten liegt darin, dass sie nicht (frei) verfügbar sind, aus künstlich generierten Daten bestehen oder das gefärbte Zellen beobachtet wurden. Viele der Daten enthalten nur sehr wenige Zellen, so dass eine Extraktion von genealogischen Informationen nicht zufriedenstellend evaluiert werden kann. Eine Übersicht verschiedener, verwendeter Referenzdaten ist im Anhang in Tab. A.4 gegeben.

Die Referenzdaten, die im Rahmen dieser Arbeit erstellt wurden, sind dagegen frei verfügbar und enthalten zwei Bildserien, in denen insgesamt mehr als 6000 unterscheidbare Zellen beobachtet wurden. Diese setzen sich aus  $> 320.000$  einzelnen Zellkonturen in 608 Bildern zusammen, die insgesamt  $> 2500$  Mitosen beinhalten. Weiterhin enthalten die Daten zum Teil doppelte Kontrastaufnahmen, so dass z.B. Mitosedetektionen verglichen werden können, die unterschiedliche Ansätze nutzen (vgl. Kapitel 4 und Zusammenfassung auf Seite 81).

Die Diversität der bisher verwendeten Referenzdaten findet sich ebenso in den genutzten Fehlermaßen wieder, die oftmals nur schwer zu vergleichen sind. Zusammen mit den Referenzdaten CeTReS wurden daher einheitliche Fehlermaße eingeführt, mit denen die Zelldetektion und vor allem das Zelltracking evaluiert und verglichen werden kann. Von besonderer Bedeutung ist hierbei das Fehlermaß aus Abschnitt 3.2.2, das einen Zelltracking-Algorithmus anhand der korrekt rekonstruierten vollständigen Pfade evaluiert (vgl. auch Abb. 4.16). Diese Pfade beschreiben den räumlich-zeitlichen Weg einer Zelle von der ersten Mitose bis zu ihrer eigenen Mitose und dienen als Bausteine, aus denen die gesuchten Stammbäume rekonstruiert werden. Die Genauigkeit und die Sensitivität, mit der diese Pfade von einem Zelltracking-Algorithmus extrahiert werden, liefern das erste, universelle Maß für die automatisierte Extraktion von genealogischen Informationen aus Bildserien.

## **Ausblick 1: Basistechnologie Zelltracking**

Mit den von mir entwickelten Algorithmen wurde die Grundlage für eine Parametrisierte Zellkultur geschaffen. Diese erlauben es erstmalig, die aufgenommenen Bildserien verlässlich und automatisiert in biologisch relevante Parameter und Zahlen umzuwandeln.

Aufgrund der hohen Anschaffungs- und Unterhaltungskosten stellt ein Langzeitmikroskop aber bisher in den meisten Zellkultur-Laboratorien eine Ausnahme da. Bevor sich diese Technik jedoch in allen Laboren als Standard etablieren kann, ist es notwendig, die erforderliche Apparatur zugänglicher zu machen. Eine Möglichkeit wäre es, das komplexe Langzeitmikroskop durch eine verkleinerte Apparatur zu ersetzen, die in einem Brutschrank Platz findet. Erste Ansätze, um dies zu erreichen, nutzen z.B. linsenlose Systeme, in denen die Zellen fast unmittelbar auf dem bildgebenden Sensor kultiviert werden (Zheng et al. [124]). Eine weitere Möglichkeit könnte die Oberflächenkontrastmikroskopie liefern, die in Kapitel 5 genutzt wurde. Hier wird der Kontrast auf der verwendeten Oberfläche erzeugt, d.h. es wird keine komplexe Optik benötigt.

Diese Entwicklungen zeigen, dass der Weg zu einer einfachen und breiten Anwendung des (validierten) Zelltracking bereits eingeschlagen ist. Daher ist es nicht eine Frage ob,

sondern wann diese Technik zu der Standardausrüstung eines modernen Zellkulturlabors gehören wird.

### **Ausblick 2: Mechanismen der *in vitro*-Zellkultur**

In der klinischen Anwendung sind adulte Stammzellen eine vielversprechende Quelle für autologe Zelltherapien oder liefern die Grundlage für eine personalisierte Medizin (Hankowski et al. [44]). Es ist allerdings noch nicht möglich, die Zellen in der benötigten Menge *in vitro* herzustellen oder die Qualität der gezüchteten Zellen nicht-invasiv zu beurteilen. Der Grund hierfür liegt in der *in vitro*-Kultur selbst: *in vivo*, d.h. im Körper, befinden sich adulte Stammzellen in einer streng regulierten Umgebung, während sie *in vitro* unter künstlichen Bedingungen gehalten werden. Diese Kultur wird von vielen Parametern wie dem verwendeten Nährmedium und verschiedenen Zusätzen wie z.B. fetalem Kälberserum (FKS) beeinflusst. Zu diesen äußeren Einflüssen zählen ebenso die Zelldichte, die Temperatur oder das Material der verwendeten Kulturflasche. Mit diesen Methoden ist es möglich, proliferierende Stammzellpopulationen reproduzierbar zu generieren. Die regulierenden Prozesse dieser Zellen sind allerdings noch nicht verstanden, d.h. es ist nicht bekannt, welche Faktoren welchen Einfluss auf Zellen ausüben: Es ist z.B. noch nicht hinreichend untersucht, welchen Einfluss die äußeren, extrinsischen Faktoren haben und wie wichtig die intrinsischen Faktoren sind, die z.B. durch die genetische Abstammung einer Zelle definiert werden.

Mit dem Zelltracking steht nun eine weitere Methode zur Verfügung, mit der diese Prozesse beobachtet werden können. Hier ist es möglich, neben den Verwandtschaften der Zellen auch die Zell-Zell-Kontakte zu extrahieren, die einen starken Einfluss auf die Differenzierung einer Stammzelle haben können. Eine erste Auswertung der aufgenommenen Daten zeigt z.B. ein sehr synchrones oder symmetrisches Teilungsverhalten der rPSC-Zellen (vgl. Abb. 3.17(c)). Zusammen mit der Fähigkeit zur spontanen Differenzierung (Gorjup et al. [38], Kruse et al. [62]) liefern diese Zellen einen guten Ausgangspunkt, um die Mechanismen, d.h. die Differenzierung und die asymmetrische Teilung, zu studieren.

Mit dem validierten Zelltracking ist es möglich, die intrinsischen und extrinsischen Einflüsse auf das Teilungs- und Differenzierungsmuster der Zellen automatisiert zu untersuchen. Der Ausgangspunkt für diese weiterführenden Experimente und Auswertungen ist die funktionale Fluoreszenzfärbung, bzw. das implementierte geometrische Hashverfahren, aus Kapitel 6. Sie erlaubt es, die intrinsischen Faktoren in Form der genetischen Abstammung einer Zelle zu extrahieren. Weiterhin werden die extrinsischen Faktoren um einen Informationsaustausch in Form von Zell-Zell-Kontakten und dem Bewegungsmuster der Zellen erweitert.

Diese erweiterte Analyse *in vitro*-kultivierter Zellen ist mit dem validierten Zelltracking und der funktionalen Fluoreszenzfärbung realisiert worden. Die Kombination der intrinsischen und extrinsischen Merkmale liefert die Grundlage für viele, weiterführende Modelle, mit denen der komplexe Charakter der Zellen beschrieben werden kann. Zusammengefasst ermöglichen es die Methoden, die in dieser Dissertation vorgestellt wurden, dem Verständnis der grundlegenden Mechanismen und des Verhaltens der *in vitro*-Zellkultur einen Schritt näher zu kommen.

---

## **Zum Schluss...**

dieser Arbeit möchte ich noch Prof. Amir Madany Mamlouk und Dr. Daniel H. Rapoport meinen Dank aussprechen. Von und mit Euch habe ich viel gelernt. Mein weiterer Dank gilt meinen Eltern, die mir diesen Weg geebnet haben. Liebe, die nicht mit Dankbarkeit verwechselt werden sollte, möchte ich meiner Frau Meike und unseren Kindern Elisa und Tammo aussprechen. Ihr habt mir gezeigt, was wirklich wichtig ist. Ein letzter Dank an Prof. Charli Kruse und die Graduiertenschule (DFG GSC 235/1). Bei dem einen forschte ich, die andere bezahlte mich. Danke!



# Abbildungsverzeichnis

1.1.	Trend der Publikationen zum Thema <i>cell tracking</i> . . . . .	3
1.2.	Übersicht der Kapitel. . . . .	6
2.1.	Experimentelle Übersicht. . . . .	11
2.2.	Langzeitmikroskop Olympus ix81. . . . .	15
2.3.	Die verwendeten Kontrasteinstellungen. . . . .	15
2.4.	Schema der Langzeitexperimente. . . . .	16
2.5.	rPSC. . . . .	18
2.6.	Zeitreihen der rPSC. . . . .	19
2.7.	Mosaikaufnahmen einer markierten Petrischale. . . . .	21
2.8.	Größte rekonstruierte Genealogie. . . . .	22
2.9.	Zelllinie HDF . . . . .	23
2.10.	Bildserien der HDF. . . . .	24
2.11.	Zelllinie HeLa. . . . .	25
2.12.	Ablauf der Apoptose- und Nekrose-Experimente. . . . .	26
2.13.	Import der Bildserien. . . . .	29
2.14.	Datenstruktur der Experimentdaten. . . . .	29
2.15.	Bildsegmentierung zur Zelldetektion . . . . .	31
2.16.	Otsu Schwellwert . . . . .	32
2.17.	Morphologische Eigenschaften. . . . .	33
2.18.	Überlapptracking in den Mikroskopbildern. . . . .	35
2.19.	Pfadfragmente des initialen Zelltracking. . . . .	35
2.20.	Bezeichnung der Zellpfade. . . . .	37
3.1.	Fehlerfortpflanzung in genealogischen Bäumen. . . . .	40
3.2.	Fehleranfälligkeit von genealogischen Bäumen. . . . .	41
3.3.	Fehlerquellen in Zelltrackingszenarien. . . . .	43
3.4.	Konsistenz der Zellpfade. . . . .	46
3.5.	Validiertes Tracking. . . . .	47
3.6.	Zellverschmelzung in den Referenzdaten. . . . .	49
3.7.	Erstellen der validierten Referenzdaten. . . . .	50
3.8.	Definition der verwendeten Fehlermaße. . . . .	51
3.9.	Evaluation der Zelldetektion. . . . .	53
3.10.	Genauigkeit des Zelltracking (vollständige Pfade). . . . .	54
3.11.	Fehlermaße zellulärer Genealogien. . . . .	56
3.12.	Wachstumskurve. . . . .	58
3.13.	Zelldichte. . . . .	59
3.14.	Zellmigration und Zellbewegung. . . . .	60
3.15.	Bewegung der Zellen als MSD. . . . .	61
3.16.	Vergleich der Zellzykluszeiten verschiedener Experimente. . . . .	63

3.17. Darstellung einer zellulären Genealogie. . . . .	63
4.1. Eine Zellteilung in zwei Kontrasten. . . . .	65
4.2. Evaluierung der Klassifikatoren. . . . .	67
4.3. Support Vector Machine (SVM). . . . .	68
4.4. Testdaten der Mitosedetektion in den Schräglichtbildern. . . . .	69
4.5. Änderung der Zellmorphologie. . . . .	70
4.6. Parametrisierte Änderung der Zellmorphologie. . . . .	70
4.7. Genauigkeit der statischen Mitosedetektion. . . . .	74
4.8. Parameter des adaptiven Mitosedetektion. . . . .	75
4.9. AUC-Werte der SVM basierten Klassifikation. . . . .	76
4.10. Fehler in der Mitosedetektion. . . . .	77
4.11. Testdaten der Mitosedetektion in den Phasenkontrastbildern. . . . .	79
4.12. Auswertung der Mitosedetektion in den Phasenkontrastbildern. . . . .	81
4.13. Schlechte Mitosen, gute Nicht-Mitosen. . . . .	82
4.14. Eigenmitosen. . . . .	83
4.15. Tote Zellen. . . . .	84
4.16. Evaluation verschiedenerer Tracking-Algorithmen und Mitosedetektionen. . . . .	87
4.17. Beispiel einer Mitose aus der Arbeit von Huh et al. [48]. . . . .	88
5.1. Ablauf der Apoptose / Nekrose Experimente mit HeLa Zellen . . . . .	93
5.2. Induzierter Zelltod (Apoptose und Nekrose). . . . .	94
5.3. Definition der Migrationsrichtungen und Winkel. . . . .	97
5.4. Bewegungsrichtung einer mitotischen Zelle. . . . .	99
5.5. Oberflächenkontrastmikroskopie (SCM). . . . .	103
5.6. Zelldetektion in SCM Bildern . . . . .	105
5.7. Automatisierte Auswertung von Mehrkanalbildern. . . . .	107
6.1. Kombination von Genealogien und Immunfluoreszenzanalysen. . . . .	109
6.2. Funktionale Fluoreszenzanalyse. . . . .	111
6.3. Positionsbestimmung mit einem geometrischen Hashverfahren. . . . .	113
6.4. Versatz der Zellkerne. . . . .	115
6.5. Geometrische Hashfunktion. . . . .	117
6.6. Übersicht des geometrischen Hashverfahrens. . . . .	120
6.7. Wahrscheinlichkeitsdichte des Vorder- und Hintergrundmodells. . . . .	123
6.8. Geometrische Suche. . . . .	127
6.9. Mosaikaufnahme zur Bestimmung des Zellkernversatzes. . . . .	128
6.10. Rauschmodell der Mikroskopbilder. . . . .	129
6.11. Größe der Hashtabellen. . . . .	133
6.12. Stabilität der geometrischen Hashwerte. . . . .	133
6.13. K-Werte des Bayes-Test. . . . .	136
6.14. Detektion mit fehlerhafter Segmentierung. . . . .	139
6.15. Positionen der Bildsuche. . . . .	145
6.16. Mosaikbild der Fluoreszenzaufnahmen. . . . .	146
6.17. Kombination von Schräglicht- und Fluoreszenzaufnahmen. . . . .	147

## Tabellenverzeichnis

2.1. Ausgewertete Langzeitexperimente. . . . .	27
2.2. Klassifikation der Pfadfragmente. . . . .	36
3.1. Referenzdaten CeTReS.A und CeTReS.B. . . . .	49
3.2. Alternative Bezeichnung der Fehlermaße. . . . .	52
3.3. Evaluation der Zelltracking Algorithmen. . . . .	55
4.1. Klassifikationsergebnisse der morphologischen Mitosedetektion. . . . .	73
4.2. Vergleich verschiedener Mitosedetektionen (Datensatz CeTReS.B). . . . .	82
4.3. Ergebnisse verschiedener Mitosedetektionen und Tracking-Algorithmen. . . . .	86
5.1. Daten der Migrationsstudie. . . . .	99
5.2. Zelldetektion in skalierten Bilddaten. . . . .	105
6.1. Detektionsgenauigkeit in den Mosaikaufnahmen. . . . .	130
6.2. Suchzeiten für verschiedene Datenstrukturen und Hashfunktionen. . . . .	137
6.3. Bilddaten der Mosaikaufnahmen. . . . .	139
6.4. Suchresultate der Referenzaufnahmen. . . . .	140
6.5. Bilddaten der funktionalen Immunofluoreszenzfärbung. . . . .	141
6.6. Ergebnisse der geometrischen Suche. . . . .	143



## Abkürzungsverzeichnis

<b>ATP</b>	Adenosintriphosphat
<b>AUC</b>	Area Under the Curve
<b>CAU</b>	Christian-Albrechts-Universität zu Kiel
<b>CCD</b>	Charge-coupled Device
<b>CeTReS</b>	Cell Tracking Reference Set
<b>CK19</b>	Cytokeratin-19
<b>DAPI</b>	4,6-Diamidin-2-phenylindol
<b>DIC</b>	Differential Interference Contrast
<b>DNA</b>	Desoxyribonukleinsäure
<b>EM</b>	Eigenmitosen
<b>FACS</b>	Fluorescence Activated Cell Sorting
<b>FFA</b>	Functional Fluorescence Analysis
<b>FITC</b>	Fluoresceinisothiocyanat
<b>FKS</b>	Fetales Kälberserum
<b>FN</b>	Falsch Negativ
<b>FP</b>	Falsch Positiv
<b>GIMP</b>	GNU Image Manipulation Program
<b>HDF</b>	Human Dermal Fibroblast
<b>ICE</b>	Image Composite Editor
<b>MP</b>	MegaPixel
<b>MSD</b>	Mean Square Displacement
<b>NAS</b>	Network Attached Storage
<b>PCA</b>	Principal Component Analysis
<b>PCS</b>	Photonic Crystal Slab
<b>PS</b>	Phosphatidylserine
<b>RN</b>	Richtig Negativ
<b>ROC</b>	Receiver Operating Characteristic
<b>RP</b>	Richtig Positiv
<b>rPSC</b>	rat Pancreatic Stem Cell
<b>SCM</b>	Surface Contrast Microscopy
<b>SIFT</b>	Scale-Invariant Feature Transform
<b>SMA</b>	Smooth Muscle Actin
<b>STS</b>	Staurosporin
<b>SVM</b>	Support Vector Machine



## Eigene Veröffentlichungen

- [1] Becker, T. and Madany Mamlouk, A. (2011). Automated morphology-based mitosis detection in stem cell cultures using phase contrast microscopy. In Proceedings 2011 World Congress on Engineering and Technology vol. 7, pp. 523–526, IEEE Computer Society Press.
- [2] Becker, T. and Madany Mamlouk, A. (2012). Morphology-based features for adaptive mitosis detection of in vitro stem cell tracking data. *Methods of Information in Medicine* 51, 449–56.
- [3] Becker, T., Rapoport, D. and Madany Mamlouk, A. (2012). From timelapse-data to genealogic trees: using different contrast mechanisms to improve cell tracking. In Proceedings of the 9th IEEE International Symposium on Biomedical Imaging pp. 386–389, IEEE Computer Society Press, Barcelona, Spain.
- [4] Becker, T., Rapoport, D. H. and Madany Mamlouk, A. (2011). Adaptive mitosis detection in large in vitro stem cell populations using timelapse microscopy. In Proceedings Bildverarbeitung für die Medizin (BVM) Informatik aktuell pp. 49–53, Springer, Lübeck, Germany.
- [5] Becker, T., Schultz, S., Rapoport, D. H. and Madany Mamlouk, A. (2013). Combining phase contrast and immunofluorescence images using geometric hashing. In Proceedings of the 10th IEEE International Symposium on Biomedical Imaging pp. 894–899, IEEE Computer Society Press, San Francisco, USA.
- [6] Danner, S., Kremer, M., Petschnik, A., Nagel, S., Zhang, Z., Hopfner, U., Reckhenrich, A., Weber, C., Schenck, T., Becker, T., Kruse, C., Machens, H. and Egana, J. (2012). The use of human gland-derived stem cells for enhancing vascularization during dermal regeneration. *Journal of Investigative Dermatology* 6, 1707–1716.
- [7] Nazirizadeh, Y., Becker, T., Reverey, J., Selhuber-Unkel, C., Rapoport, D. H. and Gerken, M. (2012). Photonic crystal slabs for surface contrast enhancement in microscopy of transparent objects. *Optics Express* 20, 14451–14459.
- [8] Rapoport\*, D. H., Becker\*, T., Madany Mamlouk, A., Schicktanz, S. and Kruse, C. (2011). A novel validation algorithm allows for automated cell tracking and the extraction of biologically meaningful parameters. *PLoS ONE* 6, e27315.



## Literaturverzeichnis

- [9] Al-Kofahi, O., Radke, R. J., Goderie, S. K., Shen, Q., Temple, S. and Roysam, B. (2006). Automated cell lineage construction: A rapid method to analyze clonal development established with murine neural progenitor. *Cell Cycle* 5, 327–335.
- [10] Allen, R. D., David, G. B. and Nomarski, G. (1969). The Zeiss-Nomarski differential interference equipment for transmitted-light microscopy. *Zeitschrift für wissenschaftliche Mikroskopie und mikroskopische Technik* 69, 193–221.
- [11] Atienza, J. M., Zhu, J., Wang, X., Xu, X. and Abassi, Y. (2005). Dynamic monitoring of cell adhesion and spreading on microelectronic sensor arrays. *Journal of Biomolecular Screening* 10, 795–805.
- [12] Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.
- [13] Berg, H. C. (1993). *Random Walks in Biology*. Princeton University Press.
- [14] Braun, V., Azevedo, R. B. R., Gumbel, M., Agapow, P.-M., Leroi, A. M. and Meinzer, H.-P. (2003). ALES: Cell lineage analysis and mapping of developmental events. *Bioinformatics* 19, 851–858.
- [15] Burger, W. and Burge, M. J. (2008). *Digital Image Processing. An Algorithmic Introduction Using Java*. Springer.
- [16] Burt, P. J., Edward and Adelson, E. H. (1983). The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications* 31, 532–540.
- [17] Campbell, N. A., Reece, J. B. and Markl, J. (2003). *Biologie*. Spektrum Akademischer Verlag.
- [18] Chae, H. J., Kang, J. S., Byun, J. O., Han, K. S., Kim, D. U., Oh, S. M., Kim, H. M., Chae, S. W. and Kim, H. R. (2000). Molecular mechanism of staurosporine-induced apoptosis in osteoblasts. *Pharmacological research : Journal of the Italian Pharmacological Society* 42, 373–381.
- [19] Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing* 10, 266–277.
- [20] Chen, X., Zhou, X. and Wong, S. T. C. (2006). Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy. *IEEE Transaction on Biomedical Engineering* 53, 762–766.
- [21] Clarke, D. L., Johansson, C. B., Wilbertz, J., Veress, B., Nilsson, E., Karlström, H., Lendahl, U. and Frisén, J. (2000). Generalized potential of adult neural stem cells. *Science (New York, N.Y.)* 288, 1660–1663.

- 
- [22] Cohen, A. R., Gomes, F. L. A. F., Roysam, B. and Cayouette, M. (2010). Computational prediction of neural progenitor cell fates. *Nature Methods* 7, 213–218.
- [23] Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Journal of Machine Learning* 20, 273–297.
- [24] Danner, S., Kajahn, J., Geismann, C., Klink, E. and Kruse, C. (2007). Derivation of oocyte-like cells from a clonal pancreatic stem cell line. *Molecular human reproduction* 13, 11–20.
- [25] Davis, P. J., Kosmacek, E. A., Sun, Y., Ianzini, F. and Mackey, M. A. (2007). The large-scale digital cell analysis system: an open system for nonperturbing live cell imaging. *Journal of Microscopy* 228, 296–308.
- [26] Debeir, O., Milojevic, D., Leloup, T., Ham, P. V., Kiss, R. and Decaestecker, C. (2005a). Mitotic tree construction by computer in vitro cell tracking: a tool for proliferation and motility features extraction. In *Proceedings of the EUROCON 2005 - The International Conference on Computer as a Tool* pp. 951–954, IEEE Computer Society Press, Belgrade, Serbia and Montenegro.
- [27] Debeir, O., Van Ham, P., Kiss, R. and Decaestecker, C. (2005b). Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. *IEEE Transactions on Medical Imaging* 24, 697–711.
- [28] Degerman, J., Fajerson, J., Althoff, K., Thorlin, T., Rodriguez, J. H. and Gustavsson, T. (2006). A comparative study between level set and watershed image segmentation for tracking stem cells in time-lapse microscopy. In *Proceedings of the International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB)*.
- [29] Degerman, J., Thorlin, T., Fajerson, J., Althoff, K., Eriksson, P. S., Put, R. V. and Gustavsson, T. (2009). An automatic system for in vitro cell migration studies. *Journal of Microscopy* 233, 178–191.
- [30] Deng, L. (2012). The MNIST database of handwritten digit images for machine learning research. *Signal Processing Magazine, IEEE* 29, 141–142.
- [31] Dingli, D., Traulsen, A. and Michor, F. (2007). (A) Symmetric stem cell replication and cancer. *PLoS Computational Biology* 3, e53.
- [32] Egaña, J. T., Danner, S., Kremer, M., Rapoport, D. H., Lohmeyer, J. A., Dye, J. F., Hopfner, U., Lavandero, S., Kruse, C. and Machens, H.-G. (2009). The use of glandular-derived stem cells to improve vascularization in scaffold-mediated dermal regeneration. *Biomaterials* 30, 5918–5926.
- [33] Eilken, H. M., Nishikawa, S. and Schroeder, T. (2009). Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature* 457, 896–900.
- [34] Eisenbrand, G., Pool-Zobel, B., Baker, V., Balls, M., Blaauboer, B. J., Boobis, A., Carere, A., Kevekordes, S., Lhuguenot, J.-C., Pieters, R. and Kleiner, J. (2002). Methods of in vitro toxicology. *Food and Chemical Toxicology* 40, 193–236.

- [35] Fadeel, B. and Orrenius, S. (2005). Apoptosis: a basic biological phenomenon with wide-ranging implications in human disease. *Journal of Internal Medicine* 258, 479–517.
- [36] Glauche, I., Lorenz, R., Hasenclever, D. and Roeder, I. (2009). A novel view on stem cell development: analysing the shape of cellular genealogies. *Cell Proliferation* 42, 248–263.
- [37] Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall.
- [38] Gorjup, E., Danner, S., Rotter, N., Habermann, J., Brassat, U., Brummendorf, T. H., Wien, S., Meyerhans, A., Wollenberg, B., Kruse, C. and von Briesen, H. (2009). Glandular tissue from human pancreas and salivary gland yields similar stem cell populations. *European journal of Cell Biology* 88, 409–421.
- [39] Gott, J. R., Jurić, M., Schlegel, D., Hoyle, F., Vogeley, M., Tegmark, M., Bahcall, N. and Brinkmann, J. (2005). A map of the universe. *The Astrophysical Journal* 624, 463.
- [40] Groth, E. J. (1986). A pattern-matching algorithm for two-dimensional coordinate lists. *The Astronomical Journal* 91, 1244–1248.
- [41] Haanen, C. and Vermes, I. (1995). Apoptosis and inflammation. *Mediators of Inflammation* 4, 5–15.
- [42] Hadjidemetriou, S., Gabrielli, B., Pike, T., Stevens, F., Mele, K. and Vallotton, P. (2008). Detection and tracking of cell divisions in phase contrast video microscopy. In *Proceedings of the Third International Workshop on Microscopic Image Analysis with Applications in Biology*, New York.
- [43] Hand, A. J., Sun, T., Barber, D. C., Hose, D. R. and MacNeil, S. (2009). Automated tracking of migrating cells in phase-contrast video microscopy sequences using image registration. *Journal of Microscopy* 234, 62–79.
- [44] Hankowski, K. E., Hamazaki, T., Umezawa, A. and Terada, N. (2011). Induced pluripotent stem cells as a next-generation biomedical interface. *Laboratory investigation; a journal of technical methods and pathology* 91, 972–977.
- [45] Harder, N., Mora-Bermúdez, F., Godinez, W. J., Ellenberg, J., Eils, R. and Rohr, K. (2006). Automated analysis of the mitotic phases of human cells in 3D fluorescence microscopy image sequences. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* vol. 9, pp. 840–848, Springer.
- [46] Hasegawa, Y., Shimizu, T., Takahashi, N. and Okada, Y. (2012). The apoptotic volume decrease is an upstream event of MAP kinase activation during staurosporine-induced apoptosis in HeLa cells. *International Journal of Molecular Sciences* 13, 9363–9379.
- [47] Hayflick, L. and Moorhead, P. (1961). The serial cultivation of human diploid cell strains. *Experimental Cell Research* 25, 585–621.

- [48] Huh, S., Ker, D. F. E., Bise, R., Chen, M. and Kanade, T. (2011). Automated mitosis detection of stem cell populations in phase-contrast microscopy images. *IEEE Transactions on Medical Imaging* *30*, 586–596.
- [49] Huh, S., Ker, D. F. E., Su, H. and Kanade, T. (2012). Apoptosis detection for adherent cell populations in time-lapse phase-contrast microscopy images. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention* pp. 331–339, Springer.
- [50] Jackson, W. M., Nesti, L. J. and Tuan, R. S. (2012). Concise review: clinical translation of wound healing therapies based on mesenchymal stem cells. *Stem Cells Translational Medicine* *1*, 44–50.
- [51] Jeffreys, H. (1939). *Theory of Probability*. International series of monographs on physics, The Clarendon Press.
- [52] Johnson, S., Fan, S., Villeneuve, P., Joannopoulos, J. and Kolodziejski, L. (1999). Guided modes in photonic crystal slabs. *Physical Review B* *60*, 5751–5758.
- [53] Karaman, M., Herrgard, S., Treiber, D., Gallant, P., Atteridge, C., Campbell, B., Chan, K., Ciceri, P., Davis, M., Edeen, P., Faraoni, R., Floyd, M., Hunt, J., Lockhart, D., Milanov, Z., Morrison, M., Pallares, G., Patel, H., Pritchard, S., Wodicka, L. and Zarrinkar, P. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nature Biotechnology* *26*, 127–132.
- [54] Ke, N., Wang, X., Xu, X. and Abassi, Y. A. (2011). The xCELLigence system for real-time and label-free monitoring of cell viability. *Methods in Molecular Biology* *740*, 33–43.
- [55] Ker, D. F. E., Weiss, L. E., Junkers, S. N., Chen, M., Yin, Z., Sandbothe, M. F., Huh, S.-I., Eom, S., Bise, R., Osuna-Highley, E., Kanade, T. and Campbell, P. G. (2011). An engineered approach to stem cell culture: automating the decision process for real-time adaptive subculture of stem cells. *PLoS ONE* *6*, e27672.
- [56] Kerr, J. F., Wyllie, A. H. and Currie, A. R. (1972). Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British Journal of Cancer* *26*, 239–57.
- [57] Kessler, W. (2007). *Multivariate Datenanalyse*. John Wiley & Sons.
- [58] Kimble, J. and Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in *Caenorhabditis elegans*. *Dev Biol* *70*, 396–417.
- [59] Koopman, G., Reutelingsperger, C., Kuijten, G., Keehnen, R., Pals, S. and van Oers, M. (1994). Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood* *84*, 1415–1420.
- [60] Krause, D. S., Theise, N. D., Collector, M. I., Henegariu, O., Hwang, S., Gardner, R., Neutzel, S. and Sharkis, S. J. (2001). Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. *Cell* *105*, 369–377.

- [61] Krohn, A. J., Preis, E. and Prehn, J. H. (1998). Staurosporine-induced apoptosis of cultured rat hippocampal neurons involves caspase-1-like proteases as upstream initiators and increased production of superoxide as a main downstream effector. *The Journal of Neuroscience* *18*, 8186–8197.
- [62] Kruse, C., Kajahn, J., Petschnik, A. E., Maaß, A., Klink, E., Rapoport, D. H. and Wedel, T. (2006). Adult pancreatic stem/progenitor cells spontaneously differentiate in vitro into multiple cell lineages and form teratoma-like structures. *Annals of Anatomy* *188*, 503–517.
- [63] Lang, D., Hogg, D. W., Mierle, K., Blanton, M. and Roweis, S. (2010). Astrometry.net: Blind astrometric calibration of arbitrary astronomical images. *The Astronomical Journal* *137*, 1782–2800.
- [64] Lee, S. H., Lee, J. H. and Cho, K. H. (2011). Effects of human adipose-derived stem cells on cutaneous wound healing in nude mice. *Annals of Dermatology* *23*, 150–155.
- [65] Lewis, C. J. (2013). Stem cell application in acute burn care and reconstruction. *Journal of Wound Care* *22*, 7–8, 10, 12–6.
- [66] Li, K., Chen, M., Kanade, T., Miller, E., Weiss, L. and Campbell, P. (2008). Cell population tracking and lineage construction with spatiotemporal context. *Medical Image Analysis* *12*, 546–566.
- [67] Li, K. and Kanade, T. (2007). Cell Population Tracking and Lineage Construction Using Multiple-Model Dynamics Filters and Spatiotemporal Optimization. In *Proceedings of the 2nd International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB)*.
- [68] Li, K., Miller, E., Weiss, L., Campbell, P. and Kanade, T. (2006). Online Tracking of Migrating and Proliferating Cells Imaged with Phase-Contrast Microscopy. In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW)* pp. 65–72, IEEE Computer Society Press.
- [69] Li, K., Miller, E. D., Chen, M., Kanade, T., Weiss, L. E. and Campbell, P. G. (2008). Computer vision tracking of stemness. In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging* pp. 847–850, IEEE Computer Society Press, Paris, France.
- [70] Liebe, C. C. (1993). Pattern recognition of star constellations for spacecraft applications. *IEEE Aerospace and Electronic Systems Magazine* *8*, 31–39.
- [71] Lin, S., Fonteno, S., Weng, J.-H. and Talbot, P. (2010). Comparison of the toxicity of smoke from conventional and harm reduction cigarettes using human embryonic stem cells. *Toxicological Sciences* *118*, 202–212.
- [72] Linkert, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., Macdonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K. W. and Swedlow, J. R. (2010). Metadata matters: access to image data in the real world. *Journal of Cell Biology* *189*, 777–782.

- 
- [73] Liu, A.-A., Li, K. and Kanade, T. (2012). A semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image sequences of stem cell populations. *IEEE Transactions on Medical Imaging* *31*, 359–369.
- [74] Lo, C. M., Wang, H. B., Dembo, M. and Wang, Y. L. (2000). Cell movement is guided by the rigidity of the substrate. *Biophysical journal* *79*, 144–152.
- [75] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* *60*, 91–110.
- [76] Massoudi, A., Semenovich, D. and Sowmya, A. (2012). Cell tracking and mitosis detection using splitting flow networks in phase-contrast imaging. *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society 2012*, 5310–5313.
- [77] Mehrayin, M., Farmanzad, F., Mozafari, M., Vashae, D. and Tayebi, L. (2012). A critical stress model for cell motility. *Theoretical Biology and Medical Modelling* *9*, 49.
- [78] Meijering, E., Dzyubachyk, O. and Smal, I. (2012). Methods for cell and particle tracking. *Methods in Enzymology* *504*, 183–200.
- [79] Michael, S., Sorg, H., Peck, C.-T., Koch, L., Deiwick, A., Chichkov, B., Vogt, P. M. and Reimers, K. (2013). Tissue engineered skin substitutes created by laser-assisted bioprinting form skin-like structures in the dorsal skin fold chamber in mice. *PloS ONE* *8*, e57741.
- [80] Mitchison, T. J. and Cramer, L. P. (1996). Actin-based cell motility and cell locomotion. *Cell* *84*, 371–9.
- [81] Mosig, A., Jaeger, S., Wang, C., Nath, S., Ersoy, I. and shing Chen, S. (2009). Tracking cells in life cell imaging videos using topological alignments. *Algorithms for Molecular Biology* *4*, 10.
- [82] Mumford, D. and Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics* *42*, 577–685.
- [83] Nazirizadeh, Y., Müller, J., Geyer, U., Schelle, D., Kley, E.-B., Tünnermann, A., Lemmer, U. and Gerken, M. (2008). Optical characterization of photonic crystal slabs using orthogonally oriented polarization filters. *Optics Express* *16*, 7153.
- [84] Otsu, N. (1979). A threshold selection method from gray level histograms. *IEEE Transaction on Systems, Man and Cybernetics* *9*, 62–66.
- [85] Padfield, D., Rittscher, J. and Roysam, B. (2008). Spatio-temporal cell segmentation and tracking for automated screening. In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging* pp. 376–379, IEEE Computer Society Press, Paris, France.

- [86] Pal, A. and Bakos, G. A. (2006). Astrometry in wide-field surveys. *The Publications of the Astronomical Society of the Pacific* 118, 1474–1483.
- [87] Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2, 559 – 572.
- [88] Pécereaux, J., Zimmer, C. and Olivo-Marin, J.-C. (2006). Biophysical active contours for cell tracking: Tension and Bending. In *Proceedings of the IEEE Signal Processing Society pp. 1949–1952*, IEEE Signal Processing Society, Atlanta, GA, USA.
- [89] Petschnik, A. E., Klatte, J. E., Evers, L. H., Kruse, C., Paus, R. and Danner, S. (2010). Phenotypic indications that human sweat glands are a rich source of nestin-positive stem cell populations. *The British Journal of Dermatology* 162, 380–383.
- [90] Qian, H., Sheetz, M. P. and Elson, E. L. (1991). Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophysical Journal* 60, 910–921.
- [91] Rahbari, R., Sheahan, T., Modes, V., Collier, P., Macfarlane, C. and Badge, R. M. (2009). A novel L1 retrotransposon marker for HeLa cell line identification. *BioTechniques* 46, 277–284.
- [92] Rapoport, D. H., Danner, S. and Kruse, C. (2009). Glandular stem cells are a promising source for much more than beta-cell replacement. *Annals of Anatomy* 191, 62–69.
- [93] Ray, N. and Acton, S. (2002). Active contours for cell tracking. In *Proceedings Fifth IEEE Southwest Symposium on Image Analysis and Interpretation pp. 7–9*, IEEE Computer Society Press.
- [94] Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C. and Schroeder, T. (2009). Hematopoietic cytokines can instruct lineage choice. *Science* 325, 217–218.
- [95] Rigoutsos, I. and Hummel, R. (1995). A Bayesian approach to model matching with geometric hashing. *Computer Vision and Image Understanding* 62, 11–26.
- [96] Rigoutsos, I. and Wolfson, H. J. (1997). Geometric hashing. *IEEE Computing in Science and Engineering* 4, 10–21.
- [97] Roeder, I., Braesel, K., Lorenz, R., and Loeffler, M. (2007). Stem cell fate analysis revisited: interpretation of individual clone dynamics in the light of a new paradigm of stem cell organization. *Journal of Biomedicine and Biotechnology* 2007.
- [98] Scherer, W. F., Syverton, J. T. and Gey, G. O. (1953). Studies on the propagation in vitro of poliomyelitis viruses. *The Journal of Experimental Medicine* 97, 695–710.
- [99] Scherf, N., Ludborz, C., Thierbach, K., Glauche, I. and Roeder, I. (2013). A fluid registration approach to cell tracking. *Current Medical Imaging Reviews* 9, 102–112.

- 
- [100] Scherf, N., Ludborz, C., Thierbach, K., Kuska, J.-P., Braumann, U.-D., Scheibe, P., Pompe, T., Glauche, I. and Roeder, I. (2012). FluidTracks. In Proceedings Bildverarbeitung für die Medizin (BVM) Informatik aktuell pp. 57–62, Springer, Lübeck, Germany.
- [101] Schmitz, S. (2011). Der Experimentator: Zellkultur. Springer.
- [102] Schwartz, J. T. and Sharir, M. (1987). Identification of partially obscured objects in two and three dimensions by matching noisy characteristic. *The International Journal of Robotics Research* 6, 29–44.
- [103] Schölkopf, B. and Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.
- [104] Scudellari, M. (2009). A closer look at the single cell. *Nature Reports Stem Cells* 1.
- [105] Shen, H., Nelson, G., Kennedy, S., Nelson, D., Johnson, J., Spiller, D., White, M. R. H. and Kell, D. B. (2006a). Automatic tracking of biological cells and compartments using particle filters and active contours. *Chemometrics and Intelligent Laboratory Systems* 82, 276–282.
- [106] Shen, H., Nelson, G., Nelson, D. E., Kennedy, S., Spiller, D. G., Griffiths, T., Paton, N., Oliver, S. G., White, M. R. H. and Kell, D. B. (2006b). Automated tracking of gene expression in individual cells and cell compartments. *Journal of the Royal Society Interface* 3, 787–794.
- [107] Shin, J.-H., Kim, H.-w., Rhyu, I. J., Song, K.-J. and Kee, S.-H. (2012). Axin expression reduces staurosporine-induced mitochondria-mediated cell death in HeLa cells. *Experimental cell research* 318, 2022–2033.
- [108] Skardal, A., Mack, D., Kapetanovic, E., Atala, A., Jackson, J. D., Yoo, J. and Soker, S. (2012). Bioprinted amniotic fluid-derived stem cells accelerate healing of large skin wounds. *Stem Cells Translational Medicine* 1, 792–802.
- [109] Solly, K., Wang, X., Xu, X., Strulovici, B. and Zheng, W. (2004). Application of real-time cell electronic sensing (RT-CES) technology to cell-based assays. *ASSAY and Drug Development Technologies* 2, 363–372.
- [110] Stephens, D. (2005). Cell Imaging: Methods Express Series. Scion Publishing Ltd.
- [111] Sulston, J. E. and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* 56, 110–156.
- [112] Takagi, H., Sato, M. J., Yanagida, T. and Ueda, M. (2008). Functional analysis of spontaneous cell movement under different physiological conditions. *PLoS ONE* 3, e2648.

- [113] Takashima, A. (2001). Establishment of fibroblast cultures. *Current Protocols in Cell Biology Chapter 2*, 2.1.1–2.1.12.
- [114] Tamaoki, T. and Nakano, H. (1990). Potent and specific inhibitors of protein kinase C of microbial origin. *Nature Biotechnology* 8, 732–735.
- [115] Tokuhisa, S. and Kaneko, K. (2010). The time series image analysis of the HeLa cell using viscous fluid registration. In *Proceedings of the Computational Science and Its Applications*, (Taniar, D., Gervasi, O., Murgante, B., Pardede, E. and Apduhan, B. O., eds), vol. 6018, of *Lecture Notes in Computer Science* pp. 189–200, Springer.
- [116] Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *J. Cognitive Neuroscience* 3, 71–86.
- [117] Vermes, I., Haanen, C., Steffens-Nakken, H. and Reutelingsperger, C. (1995). A novel assay for apoptosis. Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V. *Journal of Immunological Methods* 184, 39–51.
- [118] Watt, F. M. and Jensen, K. B. (2009). Epidermal stem cell diversity and quiescence. *EMBO Molecular Medicine* 1, 260–267.
- [119] Wayne, R. (2009). *Light and Video Microscopy*. Academic Press.
- [120] Wolfson, H. J. and Rigoutsos, I. (1997). Geometric hashing: An overview. *IEEE Computing in Science and Engineering* 4, 10–21.
- [121] Yan, K., Verbeek, F. J., Devedec, S. L. and van de Water, B. (2009). Cell Tracking and Data Analysis of in vitro Tumour Cells from Time-Lapse Image Sequences. In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications (VISAPP)* vol. 1, pp. 281–286, INSTICC Press.
- [122] Yang, F., Mackey, M. A., Ianzini, F., Gallardo, G. and Sonka, M. (2005). Cell segmentation, tracking, and mitosis detection using temporal context. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* pp. 302–309, Springer.
- [123] Zernike, F. (1942). Phase contrast, a new method for the microscopic observation of transparent objects. *Physica* 9, 686–698.
- [124] Zheng, G., Lee, S. A., Antebi, Y., Elowitz, M. B. and Yang, C. (2011). The ePetri dish, an on-chip cell imaging platform based on subpixel perspective sweeping microscopy (SPSM). *Proceedings of the National Academy of Sciences of the United States of America* 108, 16889–94.
- [125] Zhou, X., Li, F., Yan, J. and Wong, S. T. C. (2009). A novel cell segmentation method and cell phase identification using markov model. *IEEE Transaction on Information Technology in Biomedicine* 13, 152–157.

- [126] Zuba-Surma, E. K., Kucia, M., Wu, W., Klich, I., Lillard, J. W., Ratajczak, J. and Ratajczak, M. Z. (2008). Very small embryonic-like stem cells are present in adult murine organs: ImageStream-based morphological analysis and distribution studies. *Cytometry. Part A : Journal of the International Society for Analytical Cytology* 73A, 1116–1127.

# A

## Anhang

**Listing A.1:** Python-Skript um die jährliche Anzahl Publikationen zu einem gegebenen Suchbegriff aus der Pubmed-Datenbank zu extrahieren (vgl. Abb. 1.1).

```
#!/usr/bin/env python # script to plot pubmed trends
# Tim Becker,05.2012

from Bio import Entrez
from scipy import *
from pylab import *
Entrez.email = ""
yBegin = 1970
yEnd = 2012

# function to extract the number of publications / the trend
# for a search term. for publication date, the pubmed search
# field pdat is used
def getTrends(yBegin,yEnd,searchTerm):
    # init count
    Count = zeros(len(range(yBegin,yEnd)))
    for iYear in range(yBegin,yEnd):
        iTerm = "%s %d[pdat]" % (searchTerm,iYear)
        handle = Entrez.esearch(db="pubmed",term=iTerm)
        record = Entrez.read(handle)
        Count[iYear-yBegin] = int(record["Count"])
        print "count in year %d: %s" %(iYear, record["Count"])
    return Count

# get normalized number of publication per year
count_all = getTrends(yBegin,yEnd,"")
C1 = getTrends(yBegin,yEnd," cell AND tracking AND")
C1_n = 100 * C1 / count_all
C2 = getTrends(yBegin,yEnd," cell tracking AND")
C2_n = 100 * C2 / count_all
```

$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_K$	$AUC$	$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_K$	$AUC$
1	0	0	0	0.835	0	1	0	0	0.596
0	0	1	0	0.808	0	0	0	1	0.686
1	1	1	1	0.759	1	1	0	0	0.817
1	0	1	0	0.784	1	0	0	1	0.752
0	1	1	0	0.806	0	1	0	1	0.694
0	0	1	1	0.805	1	1	1	0	0.775
1	1	0	1	0.758	1	0	1	1	0.761
0	1	1	1	0.806					
1	0	0	0	0.799	0	1	0	0	0.654
0	0	1	0	0.821	0	0	0	1	0.711
1	1	1	1	0.783	1	1	0	0	0.796
1	0	1	0	0.801	1	0	0	1	0.746
0	1	1	0	0.823	0	1	0	1	0.724
0	0	1	1	0.818	1	1	1	0	0.796
1	1	0	1	0.755	1	0	1	1	0.785
0	1	1	1	0.823					
1	0	0	0	0.759	0	1	0	0	0.683
0	0	1	0	0.835	0	0	0	1	0.783
1	1	1	1	0.814	1	1	0	0	0.763
1	0	1	0	0.824	1	0	0	1	0.785
0	1	1	0	0.832	0	1	0	1	0.781
0	0	1	1	0.837	1	1	1	0	0.817
1	1	0	1	0.786	1	0	1	1	0.820
0	1	1	1	0.836					

**Tabelle A.1.:** Genauigkeit der Mitosedetektion  $K_{3P-MFML}$  auf dem Datensatz CeTReS.A für alle Parameterkombinationen. Die drei Abschnitte liefern die Werte für die drei Wachstumsphasen  $0h - 10h$ ,  $10h - 40h$  und  $40h - Ende$ .

$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_K$	$AUC$	$\Delta f_A$	$\Delta f_H$	$\Delta f_L$	$\Delta f_K$	$AUC$
1	0	0	0	0.820	0	1	0	0	0.768
0	0	1	0	0.828	0	0	0	1	0.763
1	1	1	1	0.781	1	1	0	0	0.830
1	0	1	0	0.839	1	0	0	1	0.791
0	1	1	0	0.842	0	1	0	1	0.778
0	0	1	1	0.778	1	1	1	0	0.843
1	1	0	1	0.799	1	0	1	1	0.772
0	1	1	1	0.793					

**Tabelle A.2.:** Genauigkeit der Mitosedetektion  $K_{SFML}$  und  $K_{MFML}$  auf dem Datensatz CeTReS.B für alle Parameterkombinationen.

---

### Optimierung der Kontrasteinstellungen der Schräglichtbilder

Um ein möglichst optimales Bild der Zellen zu erhalten, bietet das Olympus ix81 eine sehr gute Plattform, da es über viele manuelle und motorisierte Einstellmöglichkeiten verfügt. Diese erlauben bzw. erleichtern es, die Phasenkontraste optimal auszurichten und die Höhe des Kondensors anzupassen. Die gewählten Mikroskopeinstellungen, die zu den Schräglichtaufnahmen führen, entsprechen keiner Standardkonfiguration und werden daher kurz erläutert: Ein optimaler Kontrast wird erreicht, wenn der Kondensor an der untersten Position steht, die möglich ist, d.h. dass er sich kurz über der Petrischale bzw. der Kulturflasche befindet. Hierbei ist es sehr wichtig, dass die Phasenkontraste bzw. Phasenkontrastobjektive optimal ausgerichtet sind; gegebenenfalls müssen sowohl die Kondensoreinstellung als auch die Phasenkontraste vor jedem Experiment angepasst und eingestellt werden. Insgesamt lassen sich die Bedingungen für einen optimalen Kontrast wie folgt zusammenfassen:

1. Die Phasenkontraste sind optimal ausgerichtet.
2. Der Kondensor ist so positioniert, dass er sich möglichst nah an den Zellen befindet (unterste Position, die möglich ist).
3. Die Beleuchtung bzw. die Belichtung ist so eingestellt, dass die Bilder im Vergleich zu den automatischen geählten Werten leicht überbelichtet sind.

Die so erstellten Zellbilder sind durch einen hohen Kontrast charakterisiert, wie in Abb. 2.3 auf Seite 15 gezeigt. Somit liefern sie eine sehr gute Grundlage für eine anschließende Bildsegmentierung und Zelldetektion.

## Fluides Zelltracking

Das fluide Tracking aus Scherf et al. [99, 100] kombiniert eine Bildsegmentierung, die eine level-set-Methode und eine aktive Kontur nutzt (Chan and Vese [19]), mit einer nicht-parametrisierten Bildregistrierung (Hand et al. [43], Tokuhisa and Kaneko [115]). Hierzu wird für zwei aufeinanderfolgende Bilder ein Richtungsfeld berechnet, das für jeden Pixel die Verschiebung seiner Position im nächsten Bild angibt. Mit diesem Feld kann nun ein Tracking durchgeführt werden, indem eine Zellkontur von dem ersten Bild in das nachfolgende Bild transformiert wird. Die Zellkontur wird wiederum als parametrisierte Kurve interpretiert und nach dem fluiden Modell optimiert werden.

Der Vorteil dieser Methode liegt darin, dass zwei generelle Ansätze des Zelltracking kombiniert werden, die im Folgenden noch einmal explizit angegeben werden:

1. Der erste Ansatz beschreibt ein Verfahren, in dem die Zellen zuerst in jedem Bild einzeln detektiert werden. Anschließend werden die Zellkonturen in einem zweiten Schritt über die Zeit verknüpft. Dieser Ansatz ist in dem Überlapptracking aus Abschnitt 2.4 implementiert. Die separaten Detektions- und Trackingschritte führen dazu, dass die Genauigkeit von den Fehlern der Segmentierung und

der Rekonstruktion über die Zeit abhängt. Die vorgestellte Fehleranalyse aus Abschnitt 3.1 erlaubt es zwar, die Fehler zuverlässig zu erkennen und anschließend aus der Auswertung auszuschließen oder den Fehler zu beheben. Das generelle Problem dieses Ansatzes liegt aber in der Segmentierung. Diese wird in jedem Bild einzeln ausgeführt, ohne dass die Informationen der Zelldetektion aus dem vorherigen Bild genutzt wird. Im Gegensatz hierzu nutzt das fluide Tracking einen globalen Ansatz, der die Informationen bzw. die Bewegung der Zellen des vorherigen Bildes nutzt.

2. In dem zweiten Ansatz wird eine aktive Kontur verwendet, um die Zellen zu finden bzw. ihren Weg zu rekonstruieren. Hierbei wird ein parametrisiertes, mathematisches Modell der Zellen genutzt, um die Bildsegmentierung durchzuführen. Dieses Modell erfüllt spezifische, gegebene Anforderungen an die Elastizität, Krümmung der Kontur etc. und ist an das Verhalten der Zellen adaptiert. Eine Zellkontur wird anschließend in das nachfolgende Bild übertragen und dient dort als initiale Kontur, die mit dem Modell an das zuvor berechnete Richtungsfeld angepasst wird.

Das fluide Tracking nutzt ein viskoelastisches Modell, um die Verformung und Deformation der Zellen zu beschreiben. Dieses erlaubt es, Zellen in unmittelbarer Nähe, deren Zellkonturen optisch verschmelzen, zu trennen und die Identität jeder einzelnen Zelle zu erhalten.

## Evaluation des fluiden Trackings

Das fluide Zelltracking wurde mit dem Zelltracking-Framework und den Referenzdaten CeTReS.A evaluiert, d.h. es wurde eine systematische Fehleranalyse durchgeführt. Hierzu wurde die Validierung und die Integritätsprüfung genutzt, die aufbauend auf dem genealogischen Modell aus Abschnitt 3.1.2 vorgestellt wurden. Mit der Fehlerklassifizierung aus Tab. 2.2 auf Seite 36 war es weiterhin möglich, gezielt alle fehlerhaften Zellpfade zu finden. Mit dieser Analyse wurden die folgenden systematischen Fehler identifiziert, die anschließend in der Software behoben werden konnten:

- In einigen Fällen wurde die Identität einer Zelle nicht auf die richtige, nachfolgende Zelle übertragen sondern auf zwei nachfolgende Zellen aufgespalten. Folglich hatten zwei Zellen dieselbe Bezeichnung bzw. eine Zelle bestand aus einer zweiteiligen Zellmaske. Dieser Fehler wurden von der Integritätsprüfung aus Abschnitt 3.1.3 gefunden.
- Wenn sich zwei Zellen aufeinander zu bewegen, tendierte eine der Zellmasken dazu, ihre Größe zu verringern. Zum Teil wurden die Zellkonturen immer kleiner und waren ein paar Zeitpunkte später nicht mehr vorhanden. Dieser Fehler führte zu vielen verlorenen Enden, die mit einem *lost end*-Flag aus Tab. 2.2 auf Seite 36 identifiziert wurden.

Die Auswirkung der Evaluation sind in Abb. 4.16 auf Seite 87 als Verbesserung der Genauigkeit des Zelltracking angegeben.

---

## Größe der Hashtabellen und Bestimmung von $\tilde{d}_{max}$

Es wird der Wert  $\tilde{d}_{max}$  gesucht, mit dem in einem gegebenen Bild der Größe  $S$  mit  $M$  Punkten dieselbe Anzahl an Fünfecken und Dreiecken mit derselben Länge  $\tilde{d}_{max}$  gebildet werden kann. Als Ausgangspunkt wird der Fall betrachtet, dass zwei Punkte  $A, B$  mit einer Entfernung von  $d_{max}$  gewählt sind und die Grundseite einer geometrischen Figur bilden. Um ein gültiges Dreieck zu bilden, muss ein weiterer Punkt gewählt werden, der in einem Kreis mit dem Durchmesser  $d_{max}$  liegt, in dem auch die beiden Punkte  $A, B$  liegen (vgl. Berechnung der Hashwerte in Abschnitt 6.1.1). Für ein Fünfeck müssen drei weitere Punkte gewählt werden, die ebenfalls in diesem Kreis liegen. Die gleiche Anzahl an Drei- und Fünfecken ist daher durch den Wert von  $d_{max}$  definiert, an dem  $\binom{D}{1} = \binom{D}{3}$  gilt.  $D$  gibt die Anzahl an Punkten in dem beschriebenen Kreis an. Der folgende Beweis liefert den Wert für  $\tilde{d}_{max}$ , bei dem genau diese Bedingung erfüllt ist.

**Beweis A.1.** *Es gilt*

$$\binom{D}{1} = \binom{D}{3} \Leftrightarrow \frac{D!}{1!(D-1)!} = \frac{D!}{3!(D-3)!} \Leftrightarrow 1 = \frac{(D-1)(D-2)}{6},$$

woraus folgt, dass

$$0 = D^2 - 3D - 4 \quad \Leftrightarrow \quad D \in \{-1, 4\}. \quad (\text{A.1})$$

Da  $D$  die Anzahl der Punkte im Inneren des Kreises mit Durchmesser  $d_{max} = \|B - A\|$  angibt, kann die Fläche abhängig von  $d_{max}$  mit  $D = \left(\frac{d_{max}}{2}\right)^2 \pi \frac{M}{S}$  abgeschätzt werden. Mit  $D = 4$  aus Gl. (A.1) folgt direkt

$$4 = \left(\frac{d_{max}}{2}\right)^2 \pi \frac{M}{S} \quad (\text{A.2})$$

$$\Leftrightarrow d_{max} = \sqrt{\frac{16S}{\pi M}}. \quad (\text{A.3})$$

□

## Der maximale Bayes-Faktor $K_{max}$

Mit dem *a-priori* Wissen der Zelldetektion, d.h. der Anzahl der Zellen im Testbild  $I^T$  und dem gegebenen  $\sigma_F^2$  kann ein maximaler Wert für den Bayes-Faktor bestimmt werden. Hierzu wird der Bayes-Faktor aus Gl. (6.10) unter der Annahme betrachtet, dass mit  $T_{k,l}$  eine richtige Transformation gegeben ist. Dann kann der maximale Wert über

$$K_{max} = \prod_{m=1}^{M^T} \frac{\frac{\nu}{S} + \frac{1-\nu}{M^R} \sum_{n=1}^{M^R} \mathcal{N}(c_m^T | c_n^R, \sigma_F^2)}{\frac{1}{S}}$$

bestimmt werden. Da  $\mathcal{H}$  die richtige Transformation beschreibt, wird jede Testzelle von genau einer Referenzzelle erzeugt und es gilt  $\nu = 0$ . Da nur die Referenzzellen überprüft werden, die in dem transformierten Bild  $I^T$  liegen, gilt  $M^R = M^T$  und der Term reduziert sich zu

$$K_{max} = \prod_{m=1}^{M^T} \frac{S}{M^T} \sum_{n=1}^{M^T} \mathcal{N}(c_m^T | c_n^R, \sigma_F^2).$$

Weiterhin gilt, dass jede Testzelle von genau einer Referenzzelle generiert wird, d.h. die Summe vereinfacht sich zu  $\sum_{m=1}^{M^T} \mathcal{N}(c_m^T | c_n^R, \sigma_F^2) = \frac{1}{2\pi\sigma_F}$  und  $K_{max}$  kann explizit als

$$K_{max} = \left( \frac{S}{2\pi\sigma_F M^T} \right)^{M^T}$$

angegeben werden. Da die Werte für  $K_{max}$  sehr groß werden, wird für die Berechnung die logarithmierte Darstellung

$$\log(K_{max}) = M^T \log \left( \frac{S}{2\pi\sigma_F M^T} \right). \quad (\text{A.4})$$

gewählt.

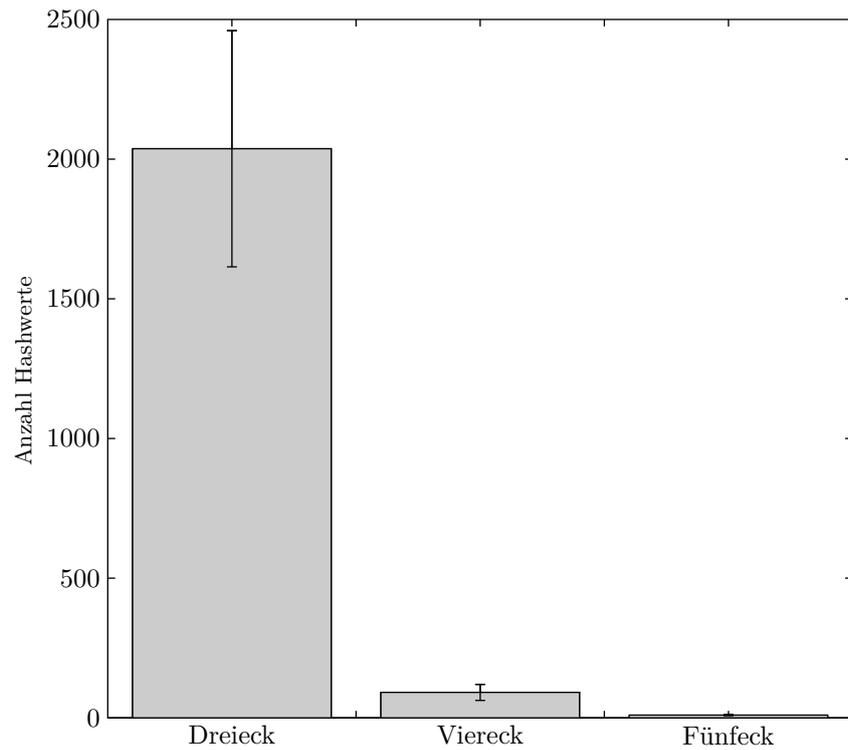
## Diskretisierte Hashtabellen

Im Folgenden wird eine Diskretisierung der Hashtabellen  $H^R, H^T$  beschrieben, wobei vorausgesetzt wird, dass die Hashwerte in jeder Dimension einen Wert zwischen null und eins annehmen, d.h. der Wertebereich der Hashwerte  $h_k^T$  liegt im offenen Intervall  $h_k^T \in (0, 1)^d$  (vgl. Abb. 6.5). Dieser Wertebereich wird in jeder Dimension in  $M$  Intervalle unterteilt, wobei im Folgenden  $M = 10$  betrachtet wird. Folglich werden die Daten in jeder Dimension in zehn Intervalle der Länge 0.1 unterteilt, die mit  $1, \dots, 10$  nummeriert werden. Für Drei-, Vier- und Fünfecke ergibt sich entsprechend eine Diskretisierung in  $10^2, 10^4$  und  $10^6$  Teilmengen. Jeder Hashwert kann nun mit der Hashfunktion

$$\text{index} = \text{round} \left( 10h_k^T + \frac{1}{2} \right) \quad (\text{A.5})$$

genau einem Index  $\{i_1, \dots, i_d\}$ ,  $i \in 1, \dots, 10$  zugeordnet werden. Beispielhaft wird der Hashwert  $h_k^T = [0.21, 0.01, 0.99, 0.57]$  betrachtet, der eine Konstellation aus vier Zellen beschreibt ( $d = 4$ ). Dieser wird auf die Indexmenge  $\{3, 1, 10, 6\}$  abgebildet.

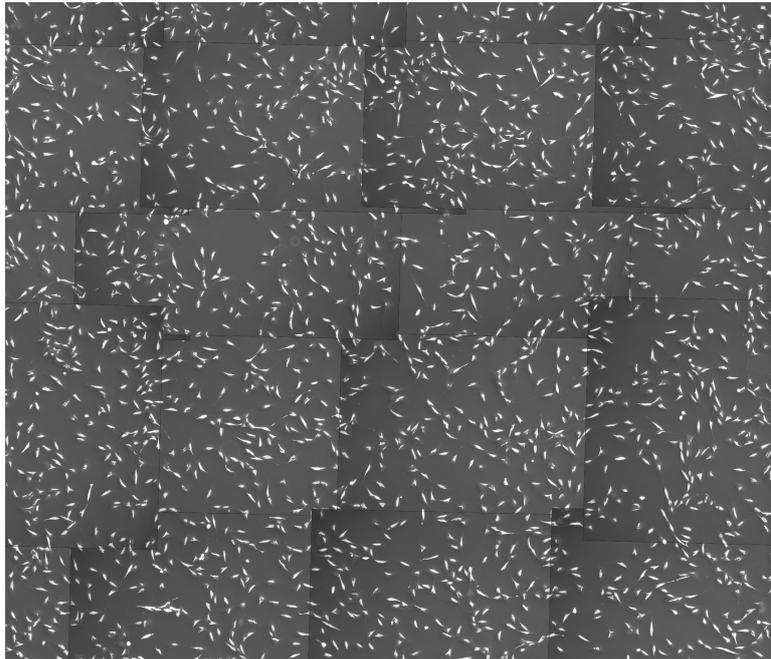
Die in Gl. (A.5) angegebene Funktion erlaubt eine schnelle und effiziente Diskretisierung der Hashtabelle  $H^R$  und reduziert die Laufzeit der Suche nach ähnlichen Konstellationen erheblich. Ein offensichtlicher Nachteil dieser Methode liegt jedoch in der Diskretisierung selbst. So ist es nicht möglich, ähnliche Hashwerte zu finden, die in unterschiedliche Intervalle eingeteilt werden. Wird also ein Hashwert  $[0.10, 0.10]$  in  $[0.10, 0.09]$  verwechselt, so ändert sich die Indexmenge von  $\{2, 2\}$  in  $\{2, 1\}$ .



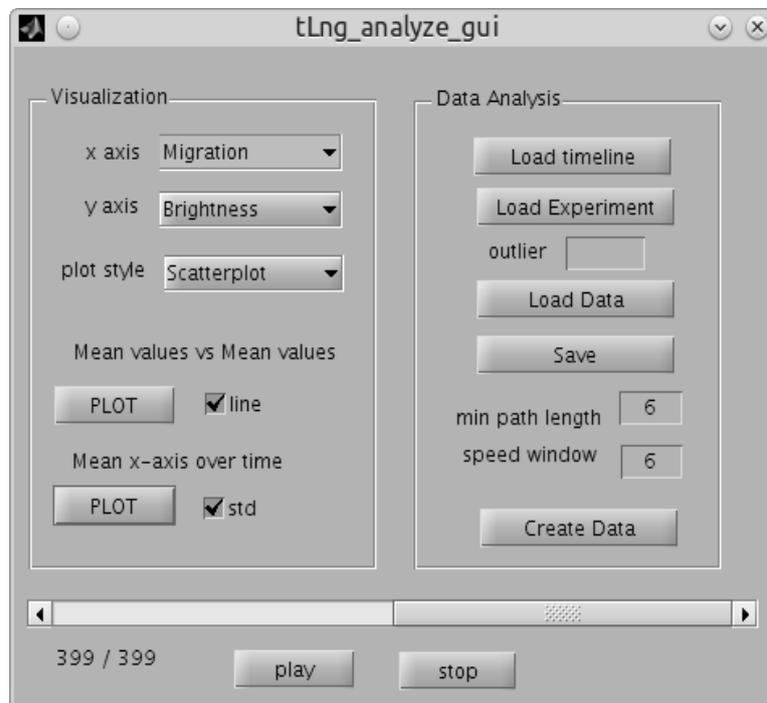
**Abbildung A.1.:** Die künstlich generierten Daten aus Abb. 6.12 wurden verwendet, um die Anzahl an ähnlichen Konstellationen für verschiedene geometrische Figuren zu berechnen (Bildgröße  $8 \times 8 \text{ mm}^2$ , 2000 Zellen,  $\sigma^2 = 16 \mu\text{m}^2$ ). Die Hashtabellen  $H^T, H^R$  wurden mit einer Konstellationsgröße von  $d_{max} = \tilde{d}_{max} = 403.7 \mu\text{m}$  erstellt. Für 4000 Hashwerte aus dem Testbild bzw.  $H^T$  wurde die Anzahl an ähnlichen Hashwerten in dem Referenzhashtabelle  $H^R$  bestimmt, d.h. für jeden Hashwert  $h_i^T \in H^T$  wurden alle Hashwerte aus  $h_k^R \in H^R$  mit einer Distanz  $\|h_i^T - h_k^R\|_2 < \epsilon = 0.1$  gesucht. Die Anzahl der ähnlichen Hashwerte ist im Balkendiagramm als Mittelwert und Standardabweichung angegeben. Die Mittelwerte lagen bei 2037, 91 und 4.

Position	Dreiecke		Vierecke		Fünfecke	
	normal	fehlerhaft	normal	fehlerhaft	normal	fehlerhaft
1	79.0	326.9	2.5	66.7	0.6	7.3
2	122.0	134.2	2.0	28.0	0.7	10.2
3	158.1	122.8	1.6	22.5	0.7	23.1
4	182.0	206.6	13.5	157.0	3.9	15.0
5	56.7	465.6	3.1	29.1	0.6	7.0
6	62.7	200.4	3.8	7.4	0.5	4.5
7	58.6	36.0	1.6	5.1	0.6	3.2
8	105.9	278.3	5.5	24.1	1.0	27.7
9	105.4	565.1	3.2	34.4	0.6	19.7
10	118.4	227.3	2.2	32.1	0.5	13.7
11	50.4	156.8	4.1	24.8	0.6	5.4
12	73.9	110.6	5.0	9.7	2.0	7.7
13	97.0	373.5	4.7	43.7	1.4	13.2
14	51.0	515.8	2.4	54.7	0.5	3.4
15	100.6	429.9	4.3	10.5	0.9	3.5
16	35.2	240.0	3.6	10.8	1.0	19.2
17	193.1	4989.7	27.7	310.3	4.3	154.0
18	100.5	423.8	6.3	67.0	0.9	34.8
19	47.0	406.0	2.4	93.2	1.0	5.2
20	348.2	2067.6	4.0	<b>81.5</b>	1.4	153.1

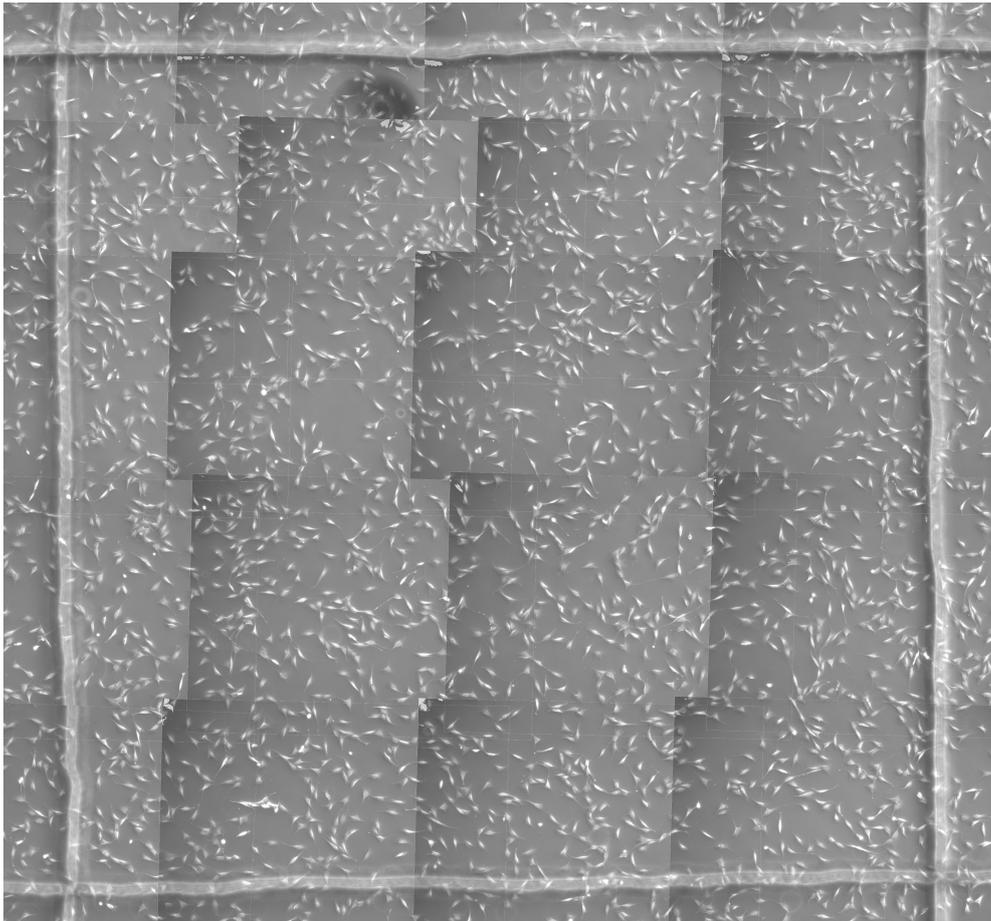
**Tabelle A.3.:** Suchzeit der Positionsbestimmung in Sekunden. Die Position der zwanzig Schräglichtbilder der Mosaikdaten aus Abschnitt 2.3.2 wurde mit den vorgestellten Algorithmen in dem Referenz-Mosaikbild ermittelt. In jedem Schräglichtbild wurden die Zellen detektiert und die Schwerpunkte der Zellmasken wurden als Zellposition genutzt ( $d_{max} = \tilde{d}_{max} = 213.2\mu$ ). Zusätzlich wurde die Positionssuche mit fehlerhaften Daten wiederholt, in denen 50% der Zellen entfernt wurden ( $d_{max} = \tilde{d}_{max} = 301.5\mu m$ ). Die Suche der fehlerhaften Bilddaten wurde mit einem niedrigeren Schwellwert berechnet ( $\theta_K = \frac{1}{4}k_{max}$ ) und benötigte ein Vielfaches an Suchzeit. Die Position von der zwanzigsten Aufnahme position wurde in der Suche mit Vierecken als geometrische Figur nicht wiedergefunden. Es wurde ein  $\sigma_F = 8\mu m$  bzw.  $\sigma_F^2 = 64$  gewählt.



**Abbildung A.2.:** Manuell erstelltes Mosaikbild der Schräglichtbilder. Die detektierten Zellen sind als weiße Konturen dargestellt.



**Abbildung A.3.:** Ein grafisches Hilfsprogramm erlaubt die Auswertung und Visualisierung einer parametrisierten Zellkultur.



**Abbildung A.4.:** Mit Hilfe der Positionssuche wurde aus den zwanzig Referenzbildern ein großes Mosaikbild zusammengesetzt. Hierzu wurden die transformierten Ausschnitte übereinander gelegt, so dass die Ränder der einzelnen Bilder noch als Artefakte erkennbar sind und ihre Position zeigen.

Arbeit	Zelltyp	Bilder	Intervall	Zellen	Mitosen	Bildserien
[9]	Maus, embryonal	2100	2 Min.	<200	<100	3
[42]	HeLa	41	6 Min.	-	51	5
[122]	U87-MG	3368	?	6654	26	50
				Zellkonturen		
[28]	Ratte Hippokampus	273	10 Min.	-	-	30
[81]	LSDCAS Daten	363	-	-	-	1
[66]	MG-63	150	4 Min.	-	-	4
[68]	MG-63	150	4 Min.	-	469	3
[121]	künstliche Objekte	300	-	100	1	10
[20]	HeLa (gefärbt)	100	15 min.	505	199	4
[85]	? (gefärbt)	110	15 Min.	6157	<20	4
				Zellkonturen		
[125]	HeLa (gefärbt)	190	15 Min.	300	-	1
[26]	U373	1080	4 Min.	1480	160	1
<b>CeTReS.A</b>	rPSC	209	15 Min.	4689	2019	1
<b>CeTReS.B</b>	(Ratte)	399	10 Min.	1431	573	1

**Tabelle A.4.:** Liste von Datensätzen, die in verschiedenen Publikationen für die Evaluation von Zelltrackingalgorithmen genutzt wurden. Quelle: Rapoport/Becker et al. [8], doi: 10.1371/journal.pone.0027315.s003.