

ALBERT-LUDWIGS-UNIVERSITÄT
FREIBURG
INSTITUT FÜR INFORMATIK

Lehrstuhl für Mustererkennung und Bildverarbeitung
Prof. Dr. Hans Burkhardt



Robuste invariante Merkmale

zu Analyse und Vergleich

von Chromosomenterritorien in Zellkernen

Diplomarbeit

Margret Keuper

16.8.2006 – 15.2.2007

Erklärung

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbstständig verfasst habe, keine anderen als die angegebenen Quellen/ Hilfsmittel verwendet habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe. Darüber hinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, bereits für eine andere Prüfung angefertigt wurde.

Freiburg, den 15.02.2007

15.08.2006

**Aufgabenstellung für die Diplomarbeit
von Frau Margret Keuper**

**Robuste invariante Merkmale zur Analyse und Vergleich von
Chromosomenterritorien in Zellkernen**

Chromosomen bilden bei einigen Organismen die Organisationsstruktur der DNA. Während die Chromosomen im kondensierten Zustand bei der Zellteilung einzeln sichtbar werden, ging man bis vor kurzem davon aus, dass sie im entspiralisierten Zustand als lange unstrukturierte „Fäden“ im Zellkern vorliegen. Durch spezielle Färbeverfahren ist es nun möglich, einzelne Chromosomen im Zellkern anzufärben, so dass sichtbar wird, dass jedes Chromosomen nur in einem bestimmten Territorium lokalisiert ist. Darüberhinaus wurden in der Arbeitsgruppe von Prof. Cremer am Biozentrum der LMU München diverse Zusammenhänge zwischen der morphologischen Struktur dieser dreidimensionalen Chromosomenterritorien und ihrer Funktion gefunden.

Aufgabe dieser Diplomarbeit soll es sein, anhand von konfokalen Aufnahmen dieser Chromosomenterritorien die verschiedenen Strukturen zu klassifizieren. In erster Linie soll dabei der Unterschied zwischen aktivierten und inaktivierten Chromosomen untersucht werden. Dazu liegen unter anderem folgende Datensätze vor:

- humane (weibliche) Zellkerne mit angefärbtem X-Chromosomen. In jedem Zellkern befindet sich ein aktives und ein inaktives X-Chromosom.
- humane Zellkerne mit angefärbtem Chromosom 8 (beide aktiv)
- genetisch veränderte Zellkerne von Mäusen, bei denen künstlich eines der beiden Chromosomen 11 inaktiviert wurde.

Die am LMB entwickelten grauwertbasierten Invarianten scheinen für diese Fragestellung prädestiniert zu sein, nur hat sich in der Vergangenheit oft gezeigt, dass sie sehr empfindlich auf Veränderungen in den Aufnahmebedingungen reagieren. Das zentrale Ziel dieser Arbeit ist es daher, robuste und reproduzierbare Merkmale zu finden, die trotz der verschiedenen zugrundeliegenden Daten von einem Experiment auf ein anderes übertragen werden können.

Die Arbeit soll in folgende Teilaufgaben gegliedert werden:

1. Literaturrecherche zum Thema
2. Analyse der Isotropie / Anisotropie der Texturen (inwieweit sind die 3D Informationen wirklich notwendig), z.B. mit co-occurrence Matrizen.
3. Untersuchungen mit vorhandenen Merkmals-Extraktions-Techniken, z.B.:

- Haralick-Texture features
 - local binary patterns
 - Autokorrelations-Analyse
 - Grauwertinvarianten mit relationalen Kernfunktionen
 - Grauwertinvarianten auf Gradientrichtungen
 - Phasendifferenz-basierte Grauwertinvarianten
 - etc.
4. Entwicklung von Verfahren zur Analyse der Robustheit und Übertragbarkeit der Features (z.B. bei Skalierung, Verrauschung, etc der Daten)
 5. evtl. Analyse durch Synthese der Texturen

Referent: Prof. Dr.-Ing. H. Burkhardt
Betreuer: Dipl.-Phys. O. Ronneberger
Ausgabedatum: 16.8.2006
Abgabedatum: 15.2.2007
Bearbeitungszeit: 6 Monate

Abstract

The DNA of eucaryotic cells is organized in chromosomes. In the nuclei, they can exist in different states of condensation. During proliferation, they are strongly condensed and therefore become visible under the light microscope. During interphase, they are decondensed and localized in distinct chromosome territories. The structure of these chromosome territories and their function are related in multiple ways.

The main issue of the present thesis consists in analysing the differences between the textures of active and inactivated interphase chromosomes with the help of confocal data. For this purpose, different gray-level based invariant features were tested. The calculation of gray-level co-occurrence matrices yielded good results. Features describing the relations of local extrema have proved to be a reliable means of describing the texture of the examined chromosomes as well. In both cases, the features could be successfully transferred from mouse to human chromosome X.

Zusammenfassung

Die DNA eukaryotischer Zellen ist in Chromosomen organisiert. Diese können in den Zellkernen in unterschiedlich kondensierter Form vorliegen. Während der Zellteilung sind sie stark kondensiert und werden so auch unter dem Lichtmikroskop sichtbar. Während der Interphase liegen sie decondensiert vor und sind in bestimmten Territorien lokalisiert. Zwischen der Struktur dieser Chromosomenterritorien und ihrer Funktion existieren diverse Zusammenhänge.

Die zentrale Aufgabe der vorliegenden Arbeit ist es, die Unterschiede zwischen den Texturen aktiver und inaktiver Chromosomen in der Interphase anhand konfokaler Aufnahmen zu untersuchen. Zu diesem Zweck wurden verschiedene grauwertbasierte, invariante Merkmale getestet. Durch die Berechnung von Grauwert Co-occurrence Matrizen konnten gute Ergebnisse erzielt werden. Auch Merkmale, die die Verhältnisse der lokalen Extrema zueinander beschreiben, haben sich als zuverlässige Mittel erwiesen, die Textur der betrachteten Chromosomen zu beschreiben. Die Übertragung der Merkmale vom Chromosom X der Maus auf das des Menschen war in beiden Fällen erfolgreich.

Inhaltsverzeichnis

Abkürzungsverzeichnis	11
1 Einleitung	13
2 Erstellen der Daten	15
2.1 Grundlagen	15
2.1.1 Differenzierung der Zelle	15
2.1.2 Natürliche Inaktivierung des Chromosom X	15
2.1.3 Chromosomen - Textur und Aktivierung	16
2.2 Verwendete Zellkerne	18
2.3 Präparation	18
2.3.1 Zellkulturen	18
2.3.2 FISH	19
2.4 Konfokale Laserscanning-Mikroskopie	21
2.5 Verwendete Daten	23
2.6 Manuelle Segmentierung	24
3 Vorverarbeitung	27
3.1 Segmentierung der Zellkerne	27
3.1.1 Schwellwertberechnung	28
3.1.2 Konturglättung	29
3.2 Segmentierung der Chromosomenterritorien	33
3.2.1 Segmentierung mittels Gradient-Magnitude	33
3.2.2 Segmentierung mittels A Posteriori Wahrscheinlichkeiten	35
3.3 Berechnung der Gradientenrichtungen	39
4 Invariante Merkmale	41
4.1 Erzeugung invarianter Merkmale	42
4.2 Relationale Kernfunktionen	43
4.3 Autokorrelation	45
4.4 Mutual Information	47
4.5 Co-occurrence Matrizen	49
4.5.1 Fuzzy Binning	51
4.5.2 Haralick Merkmale	53
4.6 Mean Depth-Width Ratio of Extrema	54
4.6.1 1D - Klassische Berechnung	54
4.6.2 2D und 3D - eine Erweiterung	55
4.7 Principle Component Analysis	58

5	Klassifikation	59
5.1	Der k-Nächste-Nachbarn Klassifikator	59
5.2	Support Vector Maschine	59
6	Ergebnisse	63
6.1	Mutual Information	63
6.2	Co-occurrence Matrizen	65
6.2.1	Gray-Level Co-occurrence Matrizen	66
6.2.2	Gradient-Direction Co-occurrence Matrizen	71
6.2.3	Co-occurrence Matrizen zwischen Grauwert und Gradientenrichtung, Grauwert und Gradientenlänge	74
6.2.4	Distanz-Grauwert und Distanz-Gradient-Direction Co-occurrence Ma- trizen	74
6.3	MDWRE	78
6.3.1	MDWRE-basierte Merkmale auf Grauwertbildern	78
6.3.2	MDWRE-basierte Merkmale auf Gradientenrichtungen	80
6.4	Principle Component Analysis	81
7	Schluss	83
	Literatur	87
A	Verwendete Software	93
A.1	Matlab	93
A.2	ImageJ	93
A.3	NetCDF	93
A.4	libSVMTL	93
B	Biologische Grundlagen und Hintergrund	94
B.1	Zellteilung	94
B.2	CT-IC Modell	94
C	Manuell und automatisch bestimmte Schwellwerte	97
D	Maximum Intensity Projektionen	99

Abkürzungsverzeichnis

CT	Chromosomenterritorium
DAPI	4',6'-Diamidino-2-Phenylindol
DNA	Desoxyribonukleinsäure
FISH	Fluoreszenz in situ Hybridisierung
FITC	Fluoreszencin-Isothiocyanat
IC	Interchromatin Compartment
HF-G0	menschliche Fibroblasten in G0-Phase
HF-S	menschliche Fibroblasten in S-Phase
Mbp	Mega-Basen-Paare
MEF	embryonale Mäuse-Fibroblasten
MES	embryonale Mäuse-Stammzellen
MESaa	embryonale Mäuse-Stammzellen, beide Chromosomen 11 aktiv
MESai	embryonale Mäuse-Stammzellen, ein Chromosom 11 aktiv, eines inaktiviert
MESar	embryonale Mäuse-Stammzellen, ein Chromosom 11 aktiv, eines reaktiviert
MDWRE	Mean Depth-Width Ratio of Extrema
MIP	Maximum Intensity Projektion
PSF	Point Spread Function
RKF	Relationale Kernfunktion
RNA	Ribonukleinsäure
Xist	X-inactive transcript

Kapitel 1

Einleitung

Die vollständige Entschlüsselung des menschlichen Genoms ist 2003 durch das 1990 ins Leben gerufene *Human Genome Project* geglückt, und doch sind die Fragen um die menschliche Genetik und die Rolle, die unsere Gene für unsere Entwicklung und nicht zuletzt für das Auftreten verschiedener Krankheiten spielen, noch lange nicht vollständig geklärt. Unser Verständnis davon, wie und durch welche Mechanismen Gene aktiviert oder unterdrückt werden und was genau ihre Funktion ist, bleibt weiterhin stark eingeschränkt.

Die Epigenetik bezeichnet nun Mechanismen, die bestimmte Muster der Genexpression und -funktion initiieren und erhalten, ohne dabei die Genomsequenz zu verändern - sie beschreibt also die Wechselwirkungen zwischen den Genen und ihrer Umgebung [7].¹ In diesem Kontext hat auch die Frage danach, wie sich die Architektur eines Zellkerns schematisch beschreiben lässt und ob diese von funktionaler Relevanz ist, an Bedeutung gewonnen. Über die Tatsache, dass jedes Chromosom während der Interphase² einen begrenzten Raum im Nukleus einnimmt, herrscht unter Spezialisten weitgehend Einigkeit. Dieser Raum eines Chromosoms wird als Chromosomenterritorium (CT) bezeichnet.

Einen zentralen Aspekt bei diesen Betrachtungen bildet der Zusammenhang zwischen der räumlichen Anordnung der Chromosomen und ihrer Aktivität. Unter anderem sind hier offenbar der Abstand der CTs zur Zellkernwand, die Form der CTs und der Grad ihrer Kompaktheit von Bedeutung.

Die Form der CTs wird hier im Weiteren nur am Rande betrachtet, der Schwerpunkt der vorliegenden Arbeit liegt vielmehr auf einer Untersuchung des Zusammenhangs zwischen Chromatinstruktur und Chromosomaktivität. Dazu wurden aktive und inaktive X-Chromosome in menschlichen Fibroblasten und Mäusefibroblasten ebenso wie aktive und künstlich inaktivierte Chromosome 11 in embryonalen Mäusestammzellen untersucht. Der Zusammenhang zwischen Struktur und Aktivität lässt sich mittels Methoden der Bildverarbeitung analysieren, da sich die Chromatinstruktur beziehungsweise -verteilung eines Chromosoms direkt auf die Textur in lasermikroskopischen Aufnahmen des Chromosoms auswirkt. Eine kompaktere Textur in dreidimensionalen Volumendaten weist also auf eine stärkere Kondensierung des Chromatins im entsprechenden Chromosom hin.

Klassische Methoden der Bildverarbeitung sind im Folgenden auf diese Fragestellung hin angepasst und erweitert worden. Zudem ist ein Merkmal generiert worden, das die lokalen Extrema der Grauwerte einer Aufnahme zueinander in Relation setzt. Als besonders effektiv bei der Erzeugung robuster Merkmale haben sich Co-occurrence Matrizen erwiesen, aber auch der auf lokalen Extrema basierte Ansatz führte zu guten Ergebnissen.

Im Folgenden werden zunächst grundlegend die Verfahren erläutert, mithilfe derer die betrachteten Daten erstellt wurden. Die Daten selbst werden vorgestellt. In Kapitel 3 folgt eine

¹Darunter fallen zum Beispiel DNA Methylierung, Histonmodifikationen und Chromatin Remodellierung.

²Zum Zellzyklus vergleiche Abschnitt B.1 im Anhang.

Darstellung der Vorverarbeitungsschritte, die für die Auswertung notwendig waren. Ein zentraler Aspekt war hier das Erstellen von Methoden zur Segmentierung. Die verwendeten Merkmale werden in Kapitel 4 theoretisch erläutert. Daraufhin werden kurz die zur Auswertung angewendeten Klassifikatoren vorgestellt. In Kapitel 6 findet sich schließlich eine Darstellung der Ergebnisse, die erzielt werden konnten.

Kapitel 2

Erstellen der Daten

2.1 Grundlagen

Der Nukleus einer Zelle enthält die Erbinformationen des Lebewesens in Form von DNA. Von dort kann sie abgelesen, also transkribiert werden. Die DNA ist mit Proteinen verknüpft und formt das Chromatin, ein Fadenwerk, das sich während der Zellteilung zu Chromosomen kondensiert. Eine Übersicht über den Zellteilungszyklus findet sich im Anhang B.1. Chromatin besteht also aus DNA, die mehr oder weniger stark dekondensiert ist.

Die Unterschiede im Dekondensierungsgrad des Chromatin sind sowohl am Beispiel von aktiven und inaktiven X-Chromosomen als auch in aktiven und künstlich inaktivierten Chromosomen 11 untersucht worden.

Im Folgenden werden die biologischen Grundlagen dargestellt, die für das Verständnis der vorliegenden Arbeit und das Zustandekommen der betrachteten Datensätze wichtig sind. Da sowohl differenzierte als auch undifferenzierte Zellen betrachtet wurden, wird zunächst eine Übersicht über die Differenzierung von Zellen im Allgemeinen gegeben. Es wird daraufhin kurz dargestellt, wie die Inaktivierung des weiblichen Chromosom X bei Säugetieren natürlicherweise abläuft. Schließlich wird der Zusammenhang zwischen Genreichtum und Chromatinstruktur von Chromosomen erläutert.

2.1.1 Differenzierung der Zelle

Nicht in allen Körperzellen eines Lebewesens werden alle Erbinformationen exprimiert, die in der DNA vorhanden sind. Vielmehr gibt es viele unterschiedliche Zelltypen, die unterschiedliche Funktionen haben. Sie bilden dementsprechend unterschiedliche RNA- und Proteinmoleküle - Gene werden hierzu gezielt ab- oder angeschaltet. Diese Differenzierung einer Zelle zeigt sich auch in ihrer Form. Fibroblasten (Bindegewebszellen) beispielsweise sind eher flach, während Stammzellen eher rund sind. Zellen, die unterschiedlich differenziert sind, werden als unterschiedliche Morphotypen bezeichnet. Im Gegensatz zu den differenzierten Zellen stehen Stammzellen. Sie sind sozusagen omnipotent, können sich also prinzipiell zu jedem beliebigen Zelltyp differenzieren.

2.1.2 Natürliche Inaktivierung des Chromosom X

In Zellkernen weiblicher Säugetiere sind in der Regel zwei X-Chromosomen enthalten, während männliche Zellkerne nur ein X-Chromosom und ein wesentlich gen-ärmeres Y-Chromosom haben. Um diesen Unterschied auszugleichen wird natürlicherweise eines der weiblichen X-Chromosomen ausgeschaltet, also inaktiviert.

Der Mechanismus, nach dem zufällig entweder das mütterliche oder das väterliche X-Chromosom

ausgeschaltet wird, läuft während der frühen post-implantationalen Embryogenese ab und ist irreversibel. Die Region im X-Chromosom, die die wichtigste Rolle bei der Steuerung dieser Inaktivierung spielt, wird als *X Inactivation Center* [7] bezeichnet. Sie enthält das *X-inactive transcript* (Xist) Gen, das einzige bekannte Gen, das ausschließlich vom inaktiven X-Chromosom transkribiert wird. [7] zufolge spielt Xist eine wichtige Rolle bei der Auslösung des Inaktivierungsprozesses, während es wahrscheinlich andere Faktoren sind, die dafür sorgen, dass das inaktive X-Chromosom auch inaktiv bleibt.

2.1.3 Chromosomen - Textur und Aktivierung

Wie oben erwähnt bestehen die Chromosomen aus Proteinen (Histone) und der DNA jedes Zellkerns. In den Keimzellen kommen die Chromosomen einfach (haploid) vor, in den übrigen Zellen doppelt (diploid). Zwei Chromosomen eines Chromosomenpaares werden als homologe Chromosomen bezeichnet - sie haben die gleiche Funktion. Innerhalb jeder Tier- und Pflanzenart ist die Anzahl der Chromosomenpaare konstant. Eine Übersicht über die Chromosomenzahlen einiger Tier und Pflanzenarten gibt Tabelle 2.1.

Pferdespulwurm	2	Champignon	8
Fruchtfliege	8	Erbse	14
Hausmaus	40	Mais	20
Mensch	46	Weizen	42
Schimpanse	48	Kartoffel	48
Hausschaf	54	Natternzunge	512
Hund	78		
Karpfen	104		

Tabelle 2.1: Chromosomenzahlen einiger Tier- und Pflanzenarten. Aus [15].

Menschliche Körperzellen enthalten 46 Chromosomen, zwei Paare von Autosomen¹ und ein Paar von Heterosomen, den geschlechtsbestimmenden Chromosomen X und Y.

Nicht alle Chromosomen enthalten gleichviel genetische Information und DNA. Beispielsweise liegen die menschlichen Chromosomen 18 und 19 in Bezug auf ihren DNA-Gehalt in ähnlicher Größenordnung: sie enthalten 77 beziehungsweise 60 Mbp (Mega-Basen-Paare). Dabei unterscheiden sie sich deutlich in ihrem Gen-gehalt: Chromosom 19 besteht hauptsächlich aus genreichem Chromatin (28 Gene/Mbp), während Chromosom 18 hauptsächlich aus genärmern Chromatin besteht (8 Gene/Mbp). (Vergleiche hierzu [8].)

In [8] konnte gezeigt werden, dass sich der Genreichtum eines Chromosoms auf dessen Positionierung im Zellkern auswirkt. Das genreichere Chromosom 19 tendiert dazu, sich eher nahe dem Zellkernzentrum zu befinden, während sich das genärmere Chromosom 18 eher in der Kernperipherie befindet. Je näher ein Chromosom am Kernzentrum liegt, desto leichter kann von ihm transkribiert werden. An der Zellkernhülle gibt es eine Schicht von inaktivem Chromatin. Gene, die dort liegen, können nicht transkribiert werden.

Die Gen-Dichte eines Chromosoms hat auch Einfluss auf dessen Chromatinstruktur. Chromatin genreicherer Chromosomen ist tendenziell weniger stark kondensiert als Chromatin genärmerer Chromosomen. Eine Begründung hierfür gibt das Chromosome Territory Interchromatin Compartment Modell (CT-IC Modell), das im Anhang B.2 zusammengefasst dargestellt ist. Dieses Modell zur Zellkernarchitektur wurde von [9] entworfen. Es modelliert den Zusammenhang zwischen Chromatinstruktur und Chromosomaktivität.

Die Gendichte eines Chromosoms ist für die vorliegende Untersuchung von Bedeutung, da sich

¹Chromosomen, die in männlichen und weiblichen Organismen gleich sind.

die Chromatinstruktur, genauer gesagt der Kondensierungsgrad eines Chromosoms direkt in seiner Textur äußert. Die These, die es zu belegen gilt ist ja: Die Textur von Chromatin ist *kompakter*, je stärker das Chromatin kondensiert ist, also je weniger von ihm transkribiert wird. Beim Vergleich der Chromatinstruktur zweier Chromosomen muss folglich nicht nur berücksichtigt werden, ob diese insgesamt aktiv oder inaktiviert sind, sondern auch, wie genreich sie sind. Genreiche Chromosomen, die durch bestimmte Mechanismen inaktiviert wurden, sind unter Umständen immer noch aktiver als genarme Chromosomen, die nicht inaktiviert wurden, da von genarmen Chromosomen ohnehin weniger Information transkribiert werden muss. Zusammengefasst gilt: Neben der Aktivität haben auch Gendichte und Chromosomgröße Einfluss auf die Chromatinstruktur eines Chromosoms [7].

Die vorliegende Arbeit untersucht nun grundsätzlich drei Arten von Chromosomen: das menschliche X-Chromosom, das X-Chromosom der Maus und das Chromosom 11 der Maus. Das menschliche X-Chromosom wie auch das X-Chromosom der Maus sind sehr genarm (7,8 respektive 7,9 Gene/Mbp), das Chromosom 11 der Maus dagegen ist sehr genreich (15,9 Gene/Mbp). Dabei enthält das X-Chromosom mehr DNA als das Chromosom 11. Vergleiche hierzu Tabelle 2.2.

Chromosom	Mbp	Gene	Gene/Mbp
11 (Maus)	121.8	1941	15.9
X (Maus)	163.7	1294	7.9
X (Mensch)	154.9	1209	7.8

Tabelle 2.2: Übersicht über den Gengehalt der betrachteten Chromosomen. Chromosom 11 ist das genreichste Chromosom der Maus, Chromosom X eines der genärmsten. Für eine vollständige Auflistung des Genoms der Maus vergleiche [16]. Die Angaben über das menschliche Genom stammen aus der Gen-Datenbank von Ensembl, Stand: 01.2007.

Chromosom 11 ist in etwa doppelt so genreich wie Chromosom X, wobei das Chromosom X des Menschen noch genärmer ist als das der Maus. Für die zu untersuchenden Texturen heißt das: Man erwartet, dass die Textur des aktiven Chromosom X dichter ist als die Textur des aktiven Chromosom 11. Die Textur des inaktiven Chromosom X sollte auch dichter sein als die des inaktivierten Chromosom 11. Jeweils sollte natürlich die Textur des inaktivierten Chromosoms dichter sein als die seines aktiven homologen Chromosoms.

Ob das inaktivierte Chromosom 11 stärker oder weniger stark kondensiert ist als das aktive X-Chromosom, bleibt zu zeigen.

2.2 Verwendete Zellkerne

Die verwendeten Daten wurden im Rahmen der Diplomarbeit von Corella Casas Delucchi [7] am Lehrstuhl für Anthropologie und Humangenetik der Ludwig-Maximilians-Universität München erstellt und uns zur Verfügung gestellt.

Insgesamt wurden für die Untersuchung sieben unterschiedliche Kulturen von Zellen generiert, von denen hier allerdings drei außer Acht gelassen werden müssen. Die Gründe hierfür werden in 2.3.1 näher erläutert.

Zusätzlich wurden zwei Datenmengen ausgewertet, die schon aus älteren Aufnahmen vorhanden waren. Hier werden die folgenden sechs Datenmengen betrachtet:

1. 39 embryonale Mäusefibroblasten (MEF), in denen das aktive und das inaktive X-Chromosom fluoreszent markiert sind,
2. 26 undifferenzierte embryonale Stammzellen von Mäusen (MES), in denen beide Chromosomen 11 - beide aktiv - fluoreszent markiert sind (MESaa),
3. 31 MES-Zellen, in denen beide Chromosomen 11 fluoreszent markiert sind, wobei eines künstlich inaktiviert wurde - das heißt: je eines aktiv und eines inaktiv (MESai) - und schließlich
4. 23 MES-Zellen, in denen beide Chromosomen 11 fluoreszent markiert sind, wobei eines künstlich inaktiviert und später reaktiviert wurde (MESar), das heißt prinzipiell: beide aktiv.

Zusätzlich:

5. 25 menschliche Fibroblasten in der G₀-Phase (HF-G₀)², in denen das aktive und das inaktive X-Chromosom fluoreszent markiert sind und
6. 25 menschliche Fibroblasten in der S-Phase (HF-S), in denen das aktive und das inaktive X-Chromosom fluoreszent markiert sind.

Wie die Zellen für die Aufnahme präpariert wurden, wird in Abschnitt 2.3 genauer beschrieben, mit der Aufnahme selbst befasst sich Abschnitt 2.4.

2.3 Präparation

2.3.1 Zellkulturen

Um in den Kulturen undifferenzierter embryonaler Mäusestammzellen ein Chromosom 11 auszuschalten, wurden diese genetisch manipuliert: in jedem Zellkern musste die Xist-Sequenz in eines der beiden Chromosomen 11 integriert werden. Ein Verfahren dazu wird in [25] beschrieben. Der Autor Anton Wutz stellte entsprechende Zellkulturen zur Aufnahme und Verwendung in [7] zur Verfügung (ES clone 36 Zellen). Im Rahmen dieser Arbeit wurden die Zellkulturen weiter nach dem in [25] beschriebenen Verfahren bearbeitet. Die Differenzierung der Zellen wird dabei durch die Zugabe von Retinoic Acid (R.Ac.) eingeleitet. Drei Zellkulturen wurden unter Bedingungen gezüchtet, unter denen keine Differenzierung stattfinden konnte (-R.Ac.), drei weiteren Zellkulturen embryonaler Mäusestammzellen wurde R.Ac. zugesetzt (+R.Ac.).

Leider sind im Nachhinein Fragen darüber aufgetreten, inwiefern die Differenzierung der embryonalen Stammzellen erfolgreich stattgefunden hat. Aufgrund dieser Unsicherheit bleiben

²Eine Übersicht über die verschiedenen Phasen des Zellteilungszyklus findet sich im Anhang B.1.

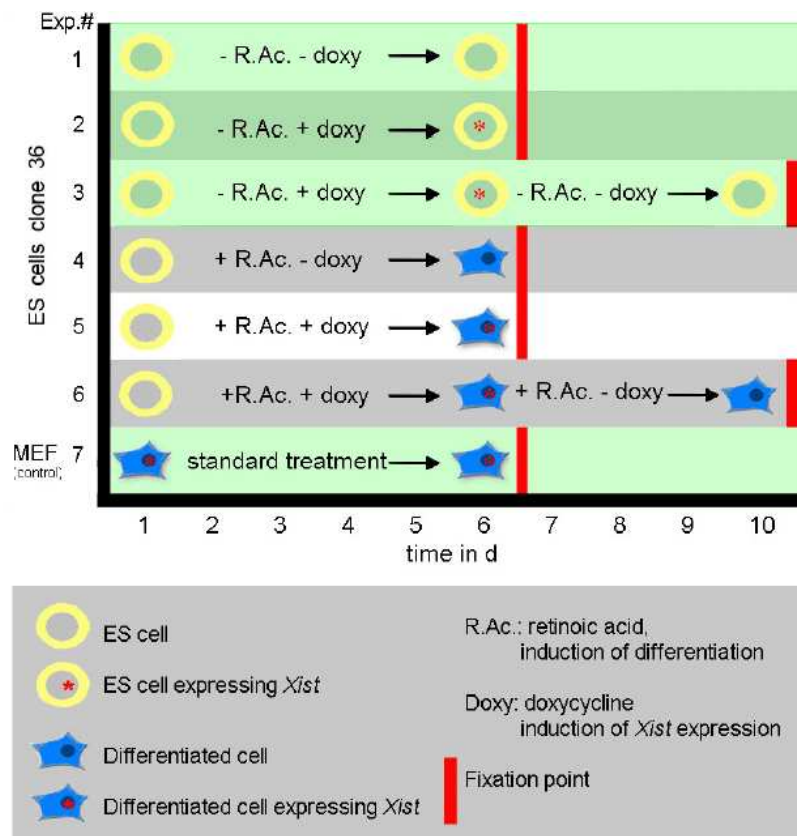


Abbildung 2.1: Zellkulturen der ES clone 36 Zellen und der MEF Zellen, aus [7]. In der vorliegenden Arbeit sind lediglich die Aufnahmen der Zellkulturen betrachtet worden, die grün unterlegt sind.

die entsprechenden Zellkulturen hier außer Acht. Vergleiche hierzu Abbildung 2.1.

Xist Inaktivierung der genetisch veränderten Chromosomen 11 wurde in jeweils zwei der drei betrachteten Zellkulturen durch Zugabe von Doxycycline erreicht. Die Zellkultur, in der keine Inaktivierung stattgefunden hat und eine der beiden Zellkulturen, die jeweils ein inaktiviertes Chromosom 11 enthalten, sind nach sechs Tagen fixiert worden. Der dritten Zellkultur wurde nach sechs Tagen das Xist-induzierende Doxycyclin entzogen. Bei undifferenzierten Zellen erwartet man [25] zufolge in diesem Fall, dass sich das betreffende Chromosom zumindest teilweise reaktiviert. Für die Aufnahmen von Chromosomen X in embryonalen Mäusefibroblasten wurde ein Standardverfahren genutzt, das gleichzeitig als Kontrolle für die Entwicklung der genetisch veränderten Zellkulturen fungierte. Vergleiche hierzu [7].

2.3.2 FISH

Zur Markierung der Chromosomen wurde *Fluoreszenz in situ Hybridisierung* (FISH) verwendet. Sie basiert auf komplementären Strukturen der DNA-Doppelhelix. Durch Denaturierungsmethoden zum Beispiel mittels Hitze oder hoher Salzkonzentration kann man die Doppelhelix in zwei Einzelstränge auftrennen. Dann verwendet man eine Sonde bestehend aus kurzen DNA-Strängen, die aufgrund ihrer Sequenz komplementär zu den gesuchten DNA-Sequenzen in der aufgetrennten DNA sind. Diese kurzen DNA-Stränge sind direkt oder indirekt markiert. Indirekt markierte Moleküle fluoreszieren nicht selbstständig, sie werden aber von anderen Mo-

lekülen erkannt, die ihrerseits an Fluorochrome gekoppelt sind.

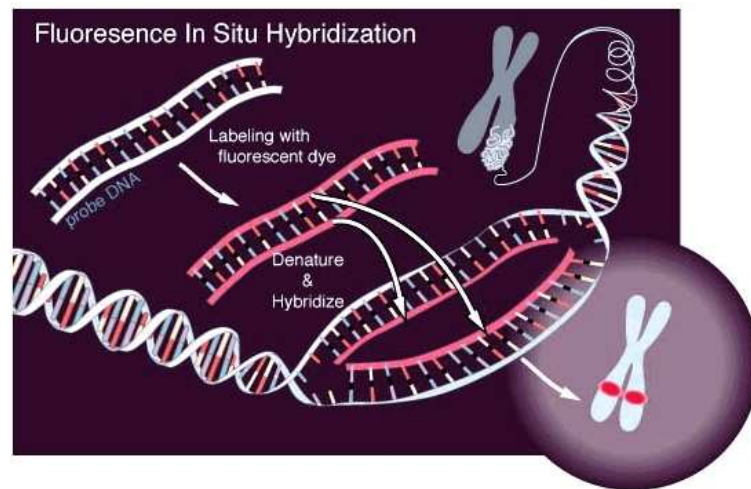


Abbildung 2.2: Schematische Darstellung der Fluoreszenz in situ Hybridisierung. Aus [27].

Zur Präparation der hier verwendeten Zellen sind nur indirekt markierte Sonden verwendet worden, nämlich Biotin zur Markierung der Chromosomen und Digoxigenin zur Markierung der Xist-Sequenz. Digoxigenin ist ein Antikörper.

Die Sonden werden ebenfalls denaturiert - also aufgespalten. Gibt man beides zusammen, so binden sich innerhalb von ein bis drei Tagen die Sonden an ihre komplementären Sequenzen im DNA-Strang. Diesen Vorgang nennt man Hybridisierung. Die Konzentration der Sonden muss dabei sehr hoch sein, damit die Wahrscheinlichkeit höher ist, dass sich die Sonden an die DNA der Chromosomen binden, als dass sich die DNA wieder zu einer Doppelhelix renaturiert.

Nach der Hybridisierung wäscht man stringent, so dass nicht- oder falschgebundene Sonden gewaschen werden. Der letzte Schritt ist die Detektion. Dazu gibt man die jeweiligen Moleküle - Chromosome-Paints - hinzu, die die Markierung der Sonde erkennen. Hier kann man mehrere Schichten verwenden, um das Signal zu verstärken. In [7] wurden zwei Schichten verwendet. Als erste Detektionsschicht fungierte Avidin-Alexa488; Avidin bindet sich an das Biotin und Alexa488 ist ein Fluorochrom. Die zweite Schicht bestand aus Goat-anti-Avidin-Fluoreszenzcin-Isothiocyant (FITC), das heißt aus Ziegen-Antikörpern, die sich an Avidin binden. Sie waren an FITC gekoppelt. FITC ist ebenfalls ein Fluorochrom; es fluoresziert grün. Zur Gegenfärbung wurde 4',6'-Diamidino-2-Phenylindol (DAPI) verwendet.

Aus der Detektion resultieren also fluoreszierende Moleküle, die indirekt an die Sequenz gebunden sind, für die man sich interessiert. Diese Sequenzen, zum Beispiel ganze Chromosomen, Chromosomenarme, oder einzelne Gene können dann mit einem konfokalen Laserscanning-Mikroskop sichtbar gemacht werden.

2.4 Konfokale Laserscanning-Mikroskopie

Im konfokalen Laserscanning-Mikroskop scannt ein Laser ein Objekt punktweise, das heißt, zu jedem Zeitpunkt wird der Laserstrahl auf genau einen Objektpunkt fokussiert. Damit ein Bild entstehen kann, müssen die gescannten Objekte fluoreszent leuchten. Objekte, die nicht über Eigenfluoreszenz verfügen, werden zum Beispiel mittels FISH Verfahren fluoreszent markiert. Jedes Fluoreszenzmolekül wird durch Licht einer spezifischen Frequenz angeregt. Mit dem kohärenten Licht eines Laserstrahls kann daher gezielt eine bestimmte Art von Fluoreszenzmolekülen angeregt werden.

Das durch eine Anregung an einem Objektpunkt entstehende Fluoreszenzsignal wird durch eine Lochblende in die Bildebene abgebildet, wodurch Signale, die nicht von diesem Objektpunkt stammen, ausgeblendet werden können. Automatisch werden auch gestreute Signale durch die Lochblende ausgeblendet. Dadurch wird der Kontrast insgesamt besser.

In der Bildebene wird das Signal mit einem lichtempfindlichen Detektor aufgenommen. Fluoreszenz unterschiedlicher Farben kann dabei getrennt und in unterschiedlichen Kanälen gleichzeitig aufgenommen werden.

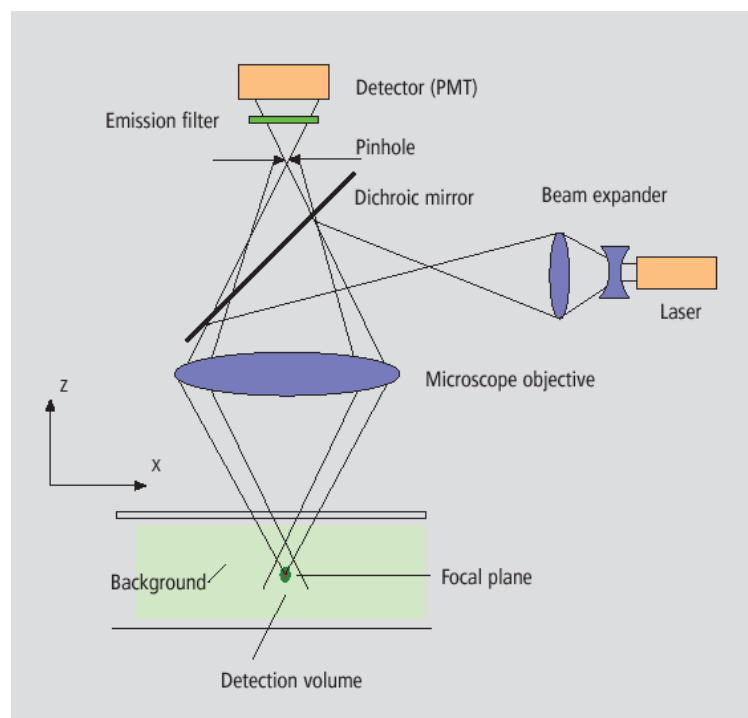


Abbildung 2.3: Schematische Darstellung eines konfokalen Laserscanning Mikroskops. Aus [24].

Durch dieses Verfahren ist es möglich bis zu $100\mu\text{m}$ dicke Objekte mit einer z -Auflösung von bis zu 500nm aufzunehmen. Wie dick eine aufgenommene Schicht letztendlich ist, hängt von dem verwendeten Objektiv und dem Durchmesser der Lochblende ab.

Indem das Objekt entlang der optischen Achse verschoben wird, können Aufnahmen von allen Schichten des Objekts gemacht werden, aus denen sich danach eine dreidimensionale Aufnahme des Objekts rekonstruieren lässt. Die Qualität dieser Aufnahme hängt nicht nur von der Dicke der einzelnen Schichten ab, sondern auch davon, wie weit aufeinanderfolgende Schichten voneinander entfernt sind. Dies wird bestimmt durch die *scanning rate*. Idealerweise gilt:

scanning rate in z -Richtung = $0.5 \times$ Schichtdicke.

Leider ist die Abbildung eines Objektpunktes nicht direkt ein Punkt, vielmehr ist das entstehende Bild stets um einem Punkt herum gestreut - eine Tatsache, die die Qualität einer Aufnahme deutlich verschlechtert. Das Bild eines Punktes wird mit einer *Point Spread Function* (PSF) beschrieben. Im idealen Fall, einem diffraktionsbeschränkten optischen System, sieht die PSF aus, wie in Abbildung 2.4 beschrieben. Aus dieser Darstellung wird deutlich,

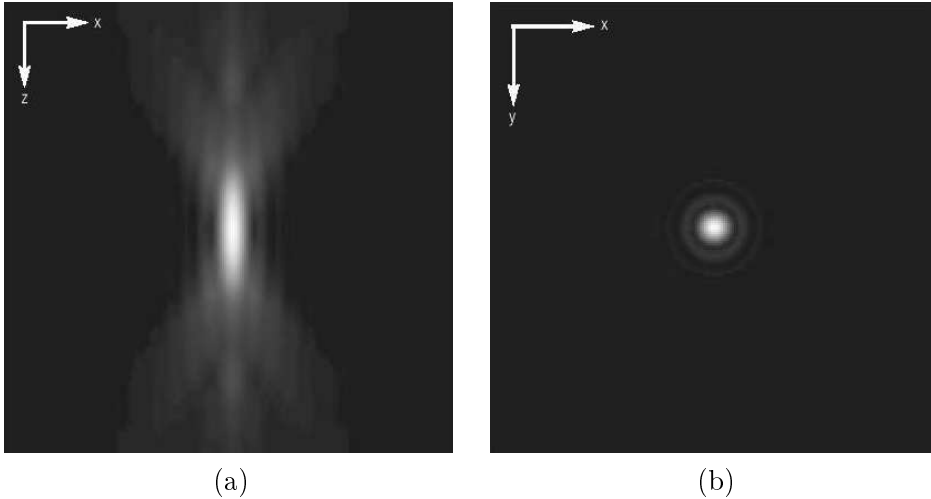


Abbildung 2.4: (a) Die ideale PSF in xz -Ansicht. (b) Die ideale PSF in xy -Ansicht. Aus [24].

dass die PSF das Signal in z -Richtung am stärksten verschmiert. Dieses Phänomen lässt sich auch in den in der vorliegenden Arbeit verwendeten Aufnahmen der CTs wiederfinden.

2.5 Verwendete Daten

Die verwendeten Daten sind mit zwei unterschiedlichen konfokalen Laserscanning-Mikroskopen aufgenommen worden, den Modellen Leica TCS-SP1 und Leica TCS-SP2. Das verwendete Objektiv ist in beiden Fällen ein 63x Plan-Achromat/1,4 Oil. Die Voxelgröße beträgt bei den jeweiligen Datenmengen:

MEF	$0,06 \times 0,06 \times 0,326\mu m$
MESaa	$0,05 \times 0,05 \times 0,326\mu m$
MESai	$0,05 \times 0,05 \times 0,326\mu m$ oder $0,05 \times 0,05 \times 0,285\mu m$
MESar	$0,05 \times 0,05 \times 0,326\mu m$
HF-G0	$0,06 \times 0,06 \times 0,285\mu m$
HF-S	$0,06 \times 0,06 \times 0,285\mu m$

Tabelle 2.3: Übersicht über die aufgenommenen Voxelgrößen in jeder der in Abschnitt 2.2 vorgestellten Datenmengen.

Diese Auflösung entspricht einer Aufnahmegröße von minimal $512 \times 512 \times 12$ Voxeln (flache Fibroblastenzellen) bis zu maximal $512 \times 512 \times 78$ Voxeln (rundliche Stammzellen). Zur weiteren Verarbeitung wurden alle Zellen mittels Interpolation auf eine Voxelgröße von $0,1 \times 0,1 \times 0,1\mu m$ skaliert. Von jedem Zellkern wurden drei Kanäle aufgenommen. Der erste ist der Gegenfärbungskanal (oder DAPI-Kanal), in welchem der gesamte Zellkern zu sehen ist. Der zweite zeigt die beiden Chromosomen, die untersucht werden sollen (FITC-Kanal). In einem dritten Kanal ist die Xist-Sequenz aufgenommen worden.

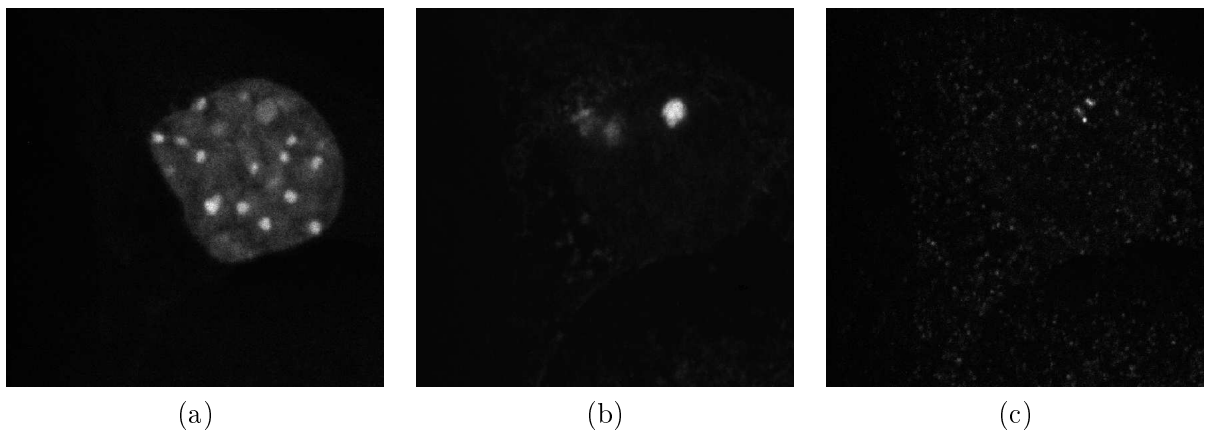


Abbildung 2.5: Jeweils eine Schicht aus dem (a) DAPI-Kanal, (b) dem FITC-Kanal und (c) dem Xist-Kanal der Aufnahme einer Mäusefibroblastenzelle. In der Xist-Aufnahme sieht man, dass alle Punkte mit hoher Intensität (Cluster oben rechts) klar einem der beiden CTs zugeordnet werden können. Dennoch gibt es viele Nebenmaxima, die eine automatische Zuordnung fragwürdig machen.

In diesem Kanal sind nur dort Punkte zu sehen, wo sich die Xist-Sequenz befindet. Bei den Aufnahmen von Fibroblasten hat man so eine *Ground Truth*, einen eindeutigen Marker, der anzeigt, welches der Chromosomen inaktiv ist. Bei den genetisch veränderten embryonalen Mäusestammzellen sind in jedem Zellkern zwei Xist-Sequenzen vorhanden: diejenige, die zu dem hier nicht dargestellten X-Chromosom gehört, und diejenige, die in das Chromosom 11 eingefügt wurde. In einem Teil der Aufnahmen (MESaa und MESar-Zellen) wurden Xist-Sequenz und CTs im selben Kanal aufgenommen, was eventuell zu Fehlinterpretationen der

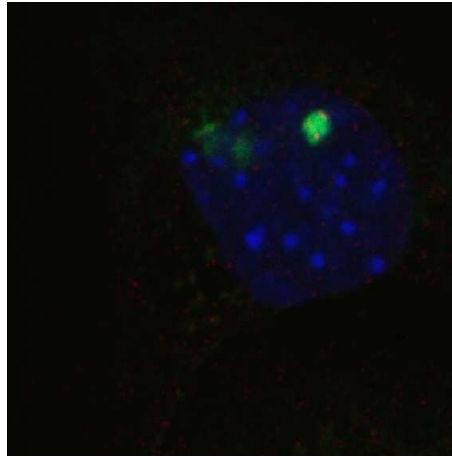


Abbildung 2.6: Die drei Einzelaufnahmen aus Abbildung 2.5 übereinandergelegt in unterschiedlichen Farbkanälen: *blau* entspricht dem DAPI-Kanal, *grün* dem FITC-Kanal und *rot* zeigt die Xist-Markierung.

Daten, speziell der Texturen der CTs führen kann.

Auch sind leider oft andere Punkte im Xist-Kanal zu sehen, die von Verunreinigungen der Zellkerne herrühren. Daher ist eine manuelle Auswertung einer automatischen in diesem Fall vorzuziehen.

2.6 Manuelle Segmentierung

Die Daten wurden im Rahmen der Arbeit [7] manuell vorsegmentiert. Mithilfe von ImageJ wurden die CTs sehr grob ausgeschnitten, so dass die Aufnahme des FITC-Kanals in zwei getrennten Datensätzen gespeichert werden konnte. Die segmentierten Aufnahmen enthalten einen großen Anteil Hintergrund, so dass die Chromosomen dennoch genauer vom Hintergrund segmentiert werden müssen, bevor sinnvoll Merkmale extrahiert werden können.

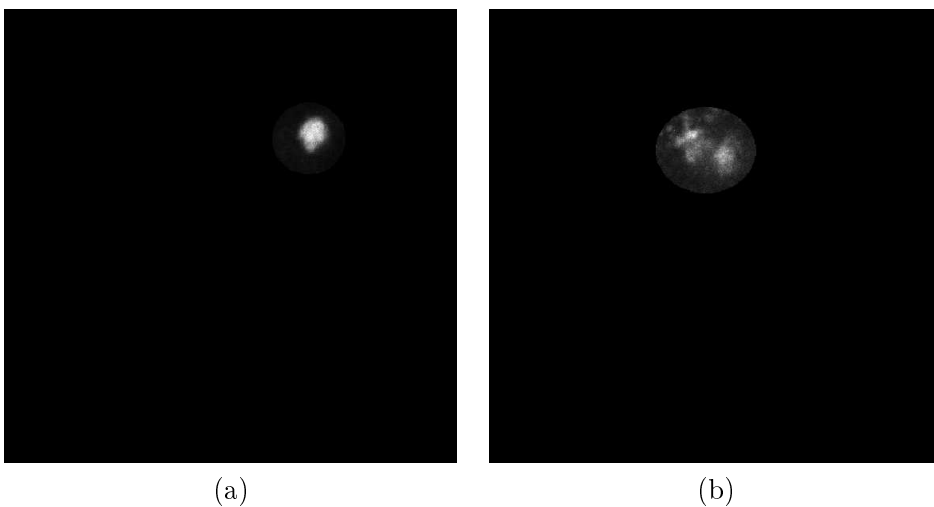


Abbildung 2.7: Die manuell segmentierte Version des Beispiels aus Abbildung 2.5(b). (a) Das inaktive X-Chromosom, (b) Das aktive X-Chromosom.

In [7] wurden zur genaueren Segmentierung zu jeder Aufnahme manuell eine Reihe von

Schwellwerten bestimmt: ein Schwellwert zur Segmentierung des Zellkerns, ein Schwellwert für jedes CT. Der Mittelwert der optimalen Schwellwerte für die einzelnen CTs wurde schließlich in [7] zur Evaluation der Daten benutzt. Diese Schwellwerte stehen hier zur Verfügung, um die Qualität der automatischen Segmentierungsverfahren für CTs zu messen.

Kapitel 3

Vorverarbeitung

Um die Daten für die Berechnung der invarianten Merkmale vorzubereiten, mussten verschiedene Vorverarbeitungsschritte durchgeführt werden. Zunächst wurden die Zellkerne vom Hintergrund segmentiert. Daraufhin wurden zwei unterschiedliche Verfahren zur Segmentierung der CTs entwickelt. Schließlich wurden zu den Aufnahmen der CTs die Richtungen der Gradienten bestimmt.

3.1 Segmentierung der Zellkerne

In dem Aufnahmekanal, in dem die CTs sichtbar sind (FITC-Kanal), ist klar zu erkennen, dass die Hintergrundhelligkeit vom Zellkern beeinflusst wird. Der Hintergrund hat in dem Bereich, der im Zellkern liegt, einen wesentlich höheren Grauwert, als in dem Bereich ausserhalb des Nukleus. Als Beispiel hierzu ist in Abbildung 3.1 jeweils dieselbe Schicht aus dem DAPI-Kanal und dem FITC-Kanal einer menschlichen Fibroblastenzelle abgebildet. Um die

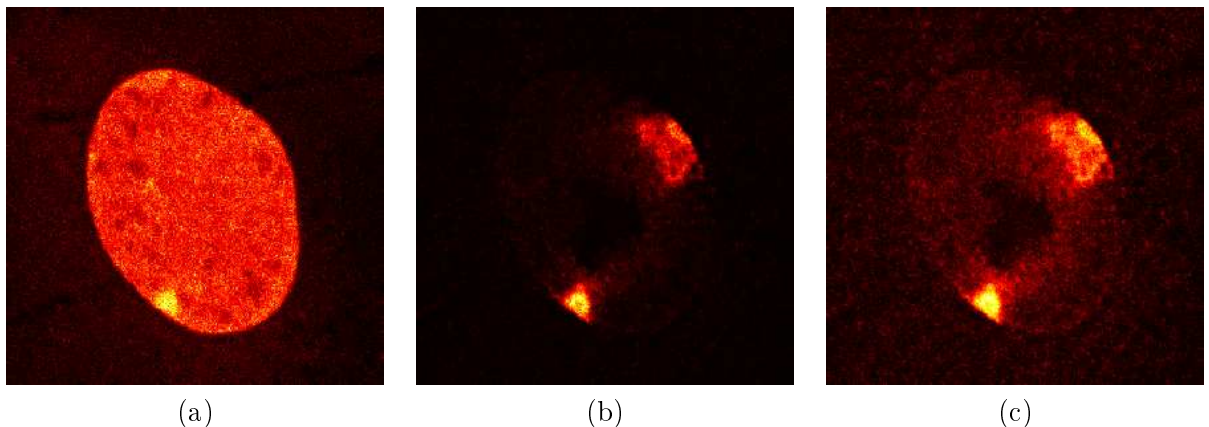


Abbildung 3.1: (a) Eine Schicht aus dem DAPI-Kanal eines menschlichen Fibroblasten Zellkerns. (b) Die entsprechende Schicht im FITC-Kanal. Man kann erkennen, dass die Grauwerte im Bereich des Zellkerns höher sind als außerhalb. (c) Dieselbe Aufnahme nach γ -Korrektur mit $\gamma=0.5$. So ist der Helligkeitsunterschied zwischen Zellkern und restlichem Hintergrund deutlicher zu erkennen.

Intensitätsunterschiede besser sichtbar zu machen, wurden die Grauwerte hier farblich auf einer Farbskala von schwarz (entspricht der Intensität Null) bis hellgelb (entsprechend der Intensität 255) dargestellt. Bei linearer Skalierung sind die Helligkeitsunterschiede zwar nur schwer zu erkennen, nach γ -Korrektur aber sieht man den Unterschied zwischen Zellkern und übrigem Hintergrund deutlich.

Um die Grundhelligkeit im Zellkern messen zu können ist es für die weitere Verarbeitung von Interesse, die Zellkerne vom Hintergrund zu segmentieren. Dies wurde im DAPI-Kanal vorgenommen.

3.1.1 Schwellwertberechnung

Die Methode, die hier zur Segmentierung der Zellkerne verwendet wurde, basiert auf Gradienten. Für eine dreidimensionale Funktion $F(x, y, z)$ lässt sich der Gradient folgendermaßen bestimmen:

$$\Delta(F) = \frac{\delta F}{\delta x} \mathbf{i} + \frac{\delta F}{\delta y} \mathbf{j} + \frac{\delta F}{\delta z} \mathbf{k}, \quad (3.1)$$

wobei $\frac{\delta F}{\delta x}$ bestimmt, wie stark sich die Funktionswerte in x -Richtung verändern, $\frac{\delta F}{\delta y}$ beschreibt die Änderung in y -Richtung, $\frac{\delta F}{\delta z}$ in z -Richtung. Die Gesamtlänge eines Gradienten an einer Stelle beschreibt, wie stark sich der Funktionswert an dieser Stelle ändert.

Dies lässt sich auch auf die dreidimensionalen Zellkerndatensätze anwenden: An den Stellen, an denen sich der Grauwert stark ändert, ist der Gradient relativ lang, eine geringe Grauwertänderung lässt sich an einem kurzen Gradienten messen. Lange Gradienten sind also ein Indiz dafür, dass an dieser Stelle einer Aufnahme die Kontur eines Objekts verläuft.

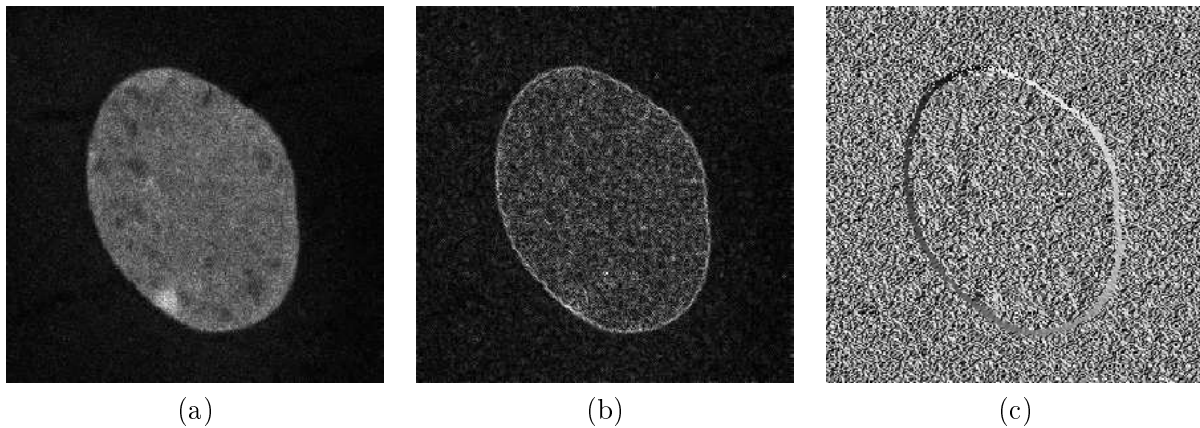


Abbildung 3.2: (a) Die Aufnahme aus Abbildung 3.1, hier als Grauwertdarstellung, (b) die Gradientenlängen dieser Aufnahme und (c) die Richtungen der Gradienten der Aufnahme.

Diese Tatsache nutzt man nun, um die Kontur eines Objektes mit einem globalen Schwellwert bestmöglich zu erfassen. Für eine Reihe von möglichen Schwellwerten τ wird die Kontur K berechnet, die bei Anwendung dieses Schwellwerts auf den Datensatz X entstehen würde:

$$K_t(\mathbf{x}) = \delta(X(\mathbf{x}) - \tau).$$

δ ist hierbei der *Dirac*-Impuls. Für jeden Schwellwert τ berechnet man dann, wie groß im Mittel die Gradienten sind, die sich auf der Kontur K befinden. Je größer dieses Mittel ist, desto stärker die Kante in der Originalaufnahme. Ist GM das $m \times n \times l$ große Array, in dem zu jedem Bildpunkt \mathbf{x} seine Gradientenlänge gespeichert ist, so ist der Schwellwert τ zu wählen, der folgenden Term maximiert:

$$\frac{1}{\int \delta(X(\mathbf{x}) - \tau) d\mathbf{x}} \int \delta(X(\mathbf{x}) - \tau) GM(\mathbf{x}) d\mathbf{x} \rightarrow \max \quad (3.2)$$

Wie sich dieser Term für Schwellwerte zwischen 1 und 200 entwickelt, ist in Abbildung 3.3 zu sehen.

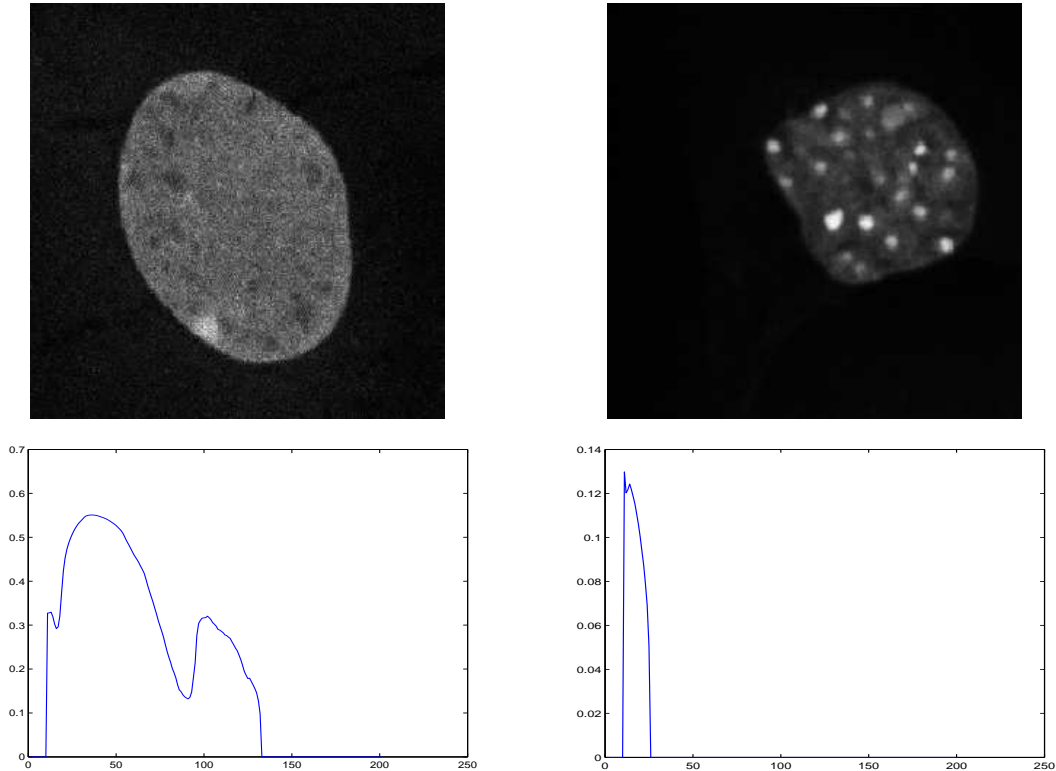


Abbildung 3.3: Jeweils eine Schicht aus dem DAPI-Kanal zweier Zellkerne - links ein Kern einer menschlichen, rechts einer Mäusefibroblastenzelle - und die Kurve, die sich bei Auswertung von Gleichung 3.2 auf den Zellkern ergibt. Der Kurvenverlauf in der menschlichen Zelle ist wesentlich breiter als in der Mäusezelle. Dies lässt darauf schließen, dass die Kontur sich bei Schwellwerten um den Wert 40 herum nur wenig verändert. Das Maximum erreicht die Kurve beim Schwellwert 35. In der Mäusezelle liegt der beste Schwellwert bei 10.

Mit dem richtigen Schwellwert lässt sich nun eine Maske erstellen, mit der diejenigen Bereiche ausgeblendet werden können, die außerhalb des Zellkerns liegen. Mithilfe dieser Maske lässt sich dann zum Beispiel die Hintergrundhelligkeit im Zellkern schätzen. Darauf wird in Abschnitt 3.2.2 genauer eingegangen werden.

In manchen der Aufnahmen sind mehrere, nah beieinander liegende Zellkerne dargestellt. Diese konnten durch das Segmentierungsverfahren nicht immer als zwei getrennte Objekte segmentiert werden. In den entsprechenden Fällen wurden die Zellkernmasken manuell nachkorrigiert.

3.1.2 Konturglättung

Im Gegensatz zu menschlichen Zellkernen weisen die Zellkerne von Mäusen dunkle und helle "Flecken" auf. Die dunkleren Flecken sind Nukleoli, weitgehend DNA-freier Raum, die hellen Flecken dagegen sind Chromozentren, das heißt Anhäufungen von pericentromerischem Chromatin - das Chromatin in der Nähe des Centrosom in jedem Chromosom. Ein Chromozentrum kann üblicherweise nicht einem bestimmten Chromosom zugeordnet werden, da es Chromatin

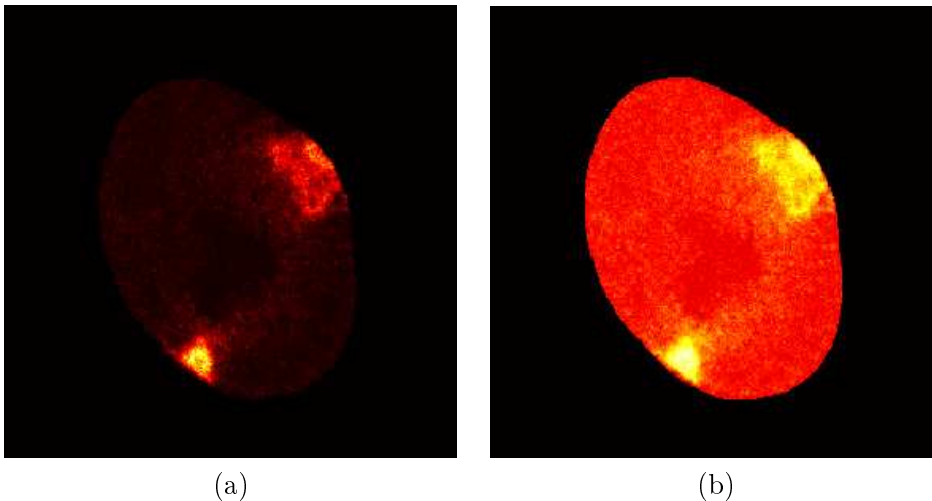


Abbildung 3.4: Die entstandene Maske kann genutzt werden, um im FITC-Kanal den Hintergrund auszublenden. Übrig bleiben die fluoreszenten Chromosomen und der Zellkernhintergrund. (a) zeigt die maskierte Aufnahme aus Abbildung 3.1(b). (b) Zur besseren Kenntlichmachung ist die Aufnahme hier nach γ -Korrektur mit $\gamma=0.3$ abgebildet.

aus verschiedenen Chromosomen aufweist. Es befindet sich auch nicht zwangsläufig in der Nähe des Chromosoms, das den größten Anteil an Chromatin beisteuert.

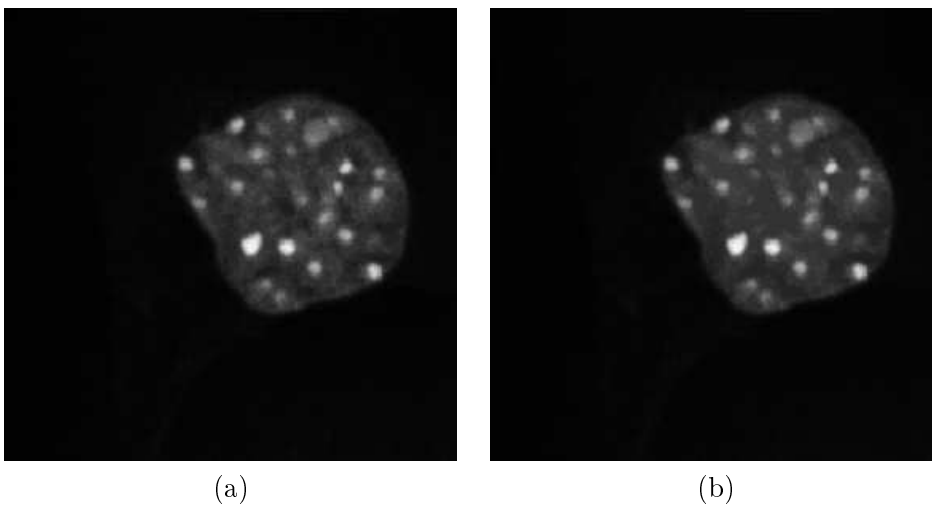


Abbildung 3.5: (a) Der Gegenfärbungskanal eines Mäusezellkerns. Es sind deutlich Nukleoli (dunkle Punkte) und Chromozentren (helle Punkte) zu sehen. (b) Der Gegenfärbungskanal des Mäusezellkerns nach der *fill*-Operation. Die Nukleoli sind nun aufgefüllt.

Die Nukleoli sind für die Segmentierung der Zellkerne meist kein größeres Problem, sie können mit einer einfachen morphologischen *fill*-Operation aufgefüllt werden. Diese *fill*-Operation setzt jeden Grauwert auf den Wert des minimalen Grauwerts der angrenzenden Pixel.

Die Chromozentren allerdings haben die Eigenschaft, die Form der Maske mehr oder weniger stark zu beeinflussen. Durch die PSF sind sie gerade in z -Richtung stark verzerrt, und beeinflussen durch ihre Helligkeit die durch Schwellwertsetzen bestimmte Kontur. Sie machen sich im segmentierten Bild als Ausbeulungen bemerkbar. Auch in xy -Richtung können sie die Form des segmentierten Kerns verfälschen. Gerade bei den flachen Fibroblasten allerdings

sind die deutlichen Auswölbungen in z -Richtung das größere Problem.

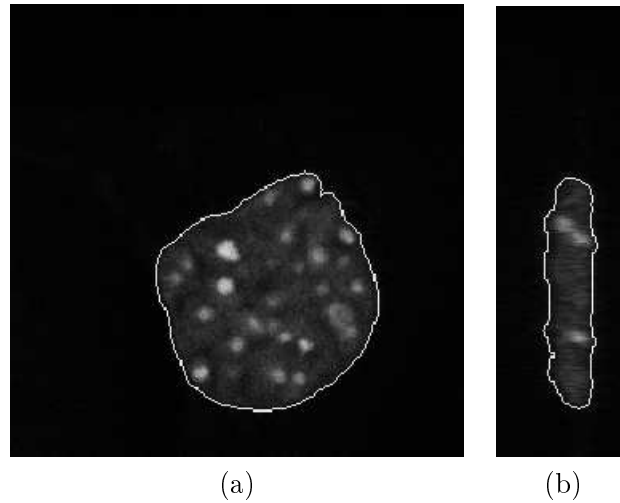


Abbildung 3.6: (a) Hier ist die Kontur eingezeichnet, an der das Verfahren den Kern ausschneidet. (b) Eine Seitenansicht der Aufnahme. Hier ist zu erkennen, wie Nukleoli und Chromozentren die Form des Nukleus verfälschen.

Auf der Suche nach einer Möglichkeit, diese Fehler zu korrigieren, hat sich eine sehr einfache Methode als praktikabel erwiesen. Die Masken, die sich aus dem oben beschriebenen Segmentierungsschritt ergeben, sind mit einem dreidimensionalen Gaußfilter iterativ geglättet und wieder binarisiert worden.

Das Gaußfilter ist ein lineares Tiefpassfilter.

$$f(x, y, z) = \frac{1}{2\pi^{N/2} |\Sigma|^{1/2}} \exp^{-1/2(x-\mu)^T \Sigma^{-1} (x-\mu)},$$

wobei Σ die Kovarianzmatrix ist, mit:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}.$$

Seine Form ist abhängig von der Standardabweichung σ in x -, y und z -Richtung. Bei jeder Iteration wird für jedes Voxel der Einflussbereich des Gaußfilters daraufhin ausgewertet, ob es größtenteils im Zellkern liegt, oder außerhalb. Liegt er größtenteils innerhalb des Zellkerns, so wird das betrachtete Voxel höchstwahrscheinlich auch zum Zellkern gehören. Durch die Binarisierung wird in jedem Schritt entschieden, welche Voxel eindeutig zu Vordergrund beziehungsweise Hintergrund gehören.

Damit starke Ausbeulungen nach und nach verschliffen werden, muss die Form des Gaußfilters bestmöglich der Form eines Zellkerns angepasst werden. Bei Fibroblasten heißt das, dass das Gaußfilter in z -Richtung eine wesentlich geringere Standardabweichung haben muss, als in x und y -Richtung, da Fibroblasten selbst ja auch sehr flach sind. Dadurch werden Auswölbungen in z -Richtung stärker ausgeglichen, als in x und y -Richtung. Bei den undifferenzierten Stammzellen ist dies schwieriger. Sie sind grundsätzlich runder, haben aber keine so leicht zu modellierende Form. Hier wurde also ein weniger flaches Gaußfilter gewählt, das allerdings insgesamt sehr klein ist, damit keine Artefakte entstehen. Tabelle 3.1 zeigt die verwendeten Standardabweichungen.

	σ_1	σ_2	σ_3
HF	0.15	0.15	0.03
MEF	0.15	0.15	0.03
MES-Zellen	0.12	0.12	0.07

Tabelle 3.1: Standardabweichungen (im μm), die zum Korrigieren der Zellkernkontur in den verschiedenen Zelltypen verwendet wurden.

Insgesamt wurden die binären Masken zehnmal iterativ mit dem Gaußfilter geglättet und wieder binarisiert. Abbildung 3.7 zeigt die Kontur, die nach der Korrektur der Maske entsteht. Grundsätzlich stellt diese Maske besser die Form des Nukleus dar. Leider fällt auf, dass in der xy -Ebene tatsächlich im Zellkern vorhandene Ausbeulungen auch leicht verschliffen werden. Die resultierenden Approximationen der tatsächlichen Kernkontur sind jedoch insgesamt hinreichend gut.

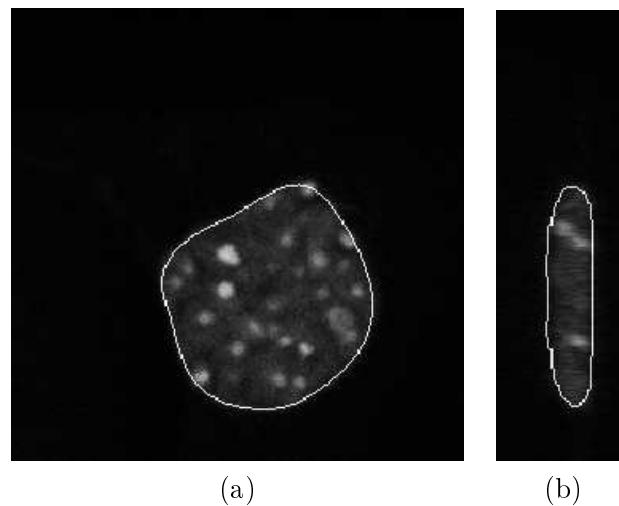


Abbildung 3.7: Die korrigierte Schnittlinie: (a) Frontalansicht (b) Seitenansicht.

Bei der Suche nach spezifischen Eigenschaften der unterschiedlichen CTs wurde auch deren Abstand von der Zellkernwand untersucht. Dies geschah insbesondere in den Untersuchungen in [7]. Dort wurden die Zellkerne zu diesem Zweck lediglich mit einem festen Schwellwert binarisiert. In der vorliegenden Arbeit wurden ähnliche Untersuchungen unter der Voraussetzung der verbesserten Approximation der Zellkernwand durchgeführt. Vergleiche hierzu Abschnitt 4.5.

3.2 Segmentierung der Chromosomenterritorien

Die Methode aus Abschnitt 3.1 hat sich bei der Anwendung auf die CTs als unpraktikabel erwiesen, da die von dieser Methode gewählten Schwellwerte hier stark variieren in Abhängigkeit davon, ob beide CTs in einem Zellkern gleichzeitig segmentiert werden, oder ein CT aus einer manuell vorsegmentierten Aufnahme segmentiert wird. Zwei weitere Verfahren, die an dieser Stelle getestet wurden und gute Ergebnisse erzielten, werden im Folgenden beschrieben.

3.2.1 Segmentierung mittels Gradient-Magnitude

Bei der Segmentierung mittels Gradient-Magnitude handelt es sich um ein Standardverfahren zur Segmentierung von Grauwertbildern. Prinzipiell sind die Grundlagen ähnlich wie in Abschnitt 3.1 - auch hier werden die Konturen der gesuchten Objekte durch die Gradientenlänge bestimmt. Abbildung 3.8 zeigt die Gradientenlängen zu Abbildung 3.1(b), einer Schicht aus dem FITC-Kanal eines Zellkerns.

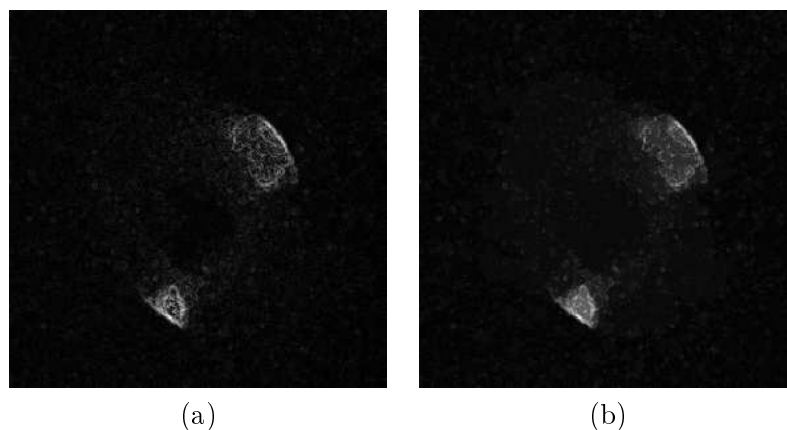


Abbildung 3.8: (a) Gradientenlängen zur Schicht aus 3.1(b). Man sieht, dass die Gradienten um die CTs herum deutlich länger sind als im restlichen Zellkern. Innerhalb eines CTs allerdings schwanken die Gradientenlängen stark. (b) Die Gradientenlängen nach *fill*-Operation. Nun sind die Gradientenlängen innerhalb der CTs insgesamt etwas höher als im restlichen Zellkern.

Man sieht, dass die Gradienten um die beiden CTs herum und auch innerhalb der CTs länger sind als im Hintergrund. Um möglichst im ganzen Inneren des CTs hohe Einträge im Gradient-Magnitude Array zu erhalten, wurde auch hier wieder eine dreidimensionale *fill*-Operation durchgeführt. Die Werte dieses Arrays jeder Aufnahme wurden daraufhin mit dem Min-Max-Algorithmus¹ auf das Intervall $[0,1]$ abgebildet. Binarisiert man diese Daten, kann man indirekt aus der Gradientenlänge eine Maske für die CTs erstellen.

Leider zeigt sich hier das Problem, dass der beste Schwellwert für die Binarisierung offenbar nicht für beide CTs der gleiche ist. Abbildung 3.9 zeigt die Ergebnisse der Binarisierung der Aufnahme aus Abbildung 3.8 (b) mit verschiedenen Schwellwerten. Das obere CT wird erst mit einem sehr geringen Schwellwert als zusammenhängende Region ausgeschnitten, für das untere CT könnte man auch einen wesentlich höheren Schwellwert wählen, bei dem weniger Hintergrundelemente mitausgeschnitten werden.

Das Resultat lässt sich mittels morphologischer Operationen deutlich verbessern. Es hat sich gezeigt, dass die aufeinanderfolgende Anwendung einer *Close*-Operation mit einem runden

¹Beim Min-Max Algorithmus subtrahiert man den minimalen vorkommenden Wert von allen Werten, so dass das Minimum Null beträgt. Dann wird dit dem maximalen vorkommenden Wert normiert.

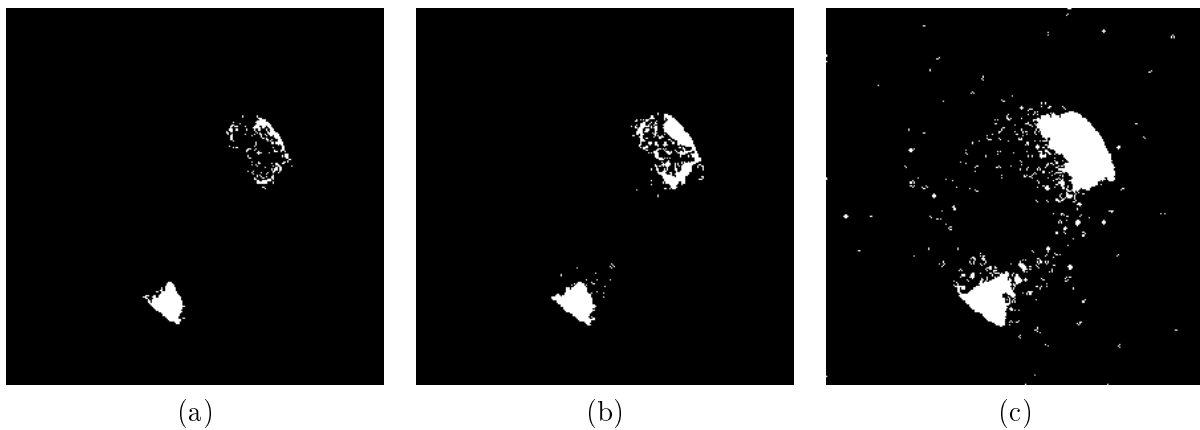


Abbildung 3.9: Binarisierung der Aufnahme aus Abbildung 3.8 (b) mit den Schwellwerten (a) 0.4 (b) 0.3 (c) 0.2. Je niedriger der Schwellwert gewählt wird, desto größer wird die Maske, das heisst ein desto größerer Anteil des Zellkerns wird dem CT zugerechnet.

Strukturelement der Größe 5×5 und einer *Open*-Operation mit einem runden Strukturelement der gleichen Größe meist zum gewünschten Ergebnis führt. Vergleiche hierzu Abbildung 3.10.

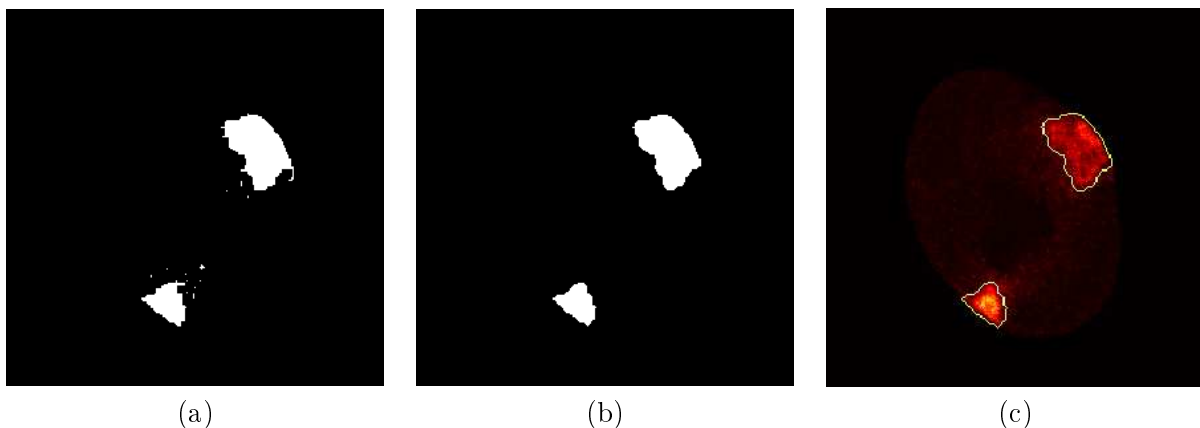


Abbildung 3.10: (a) Die Maske aus Abbildung 3.9(b) nach *Close*-Operation. Die Masken beider CTs haben nun keine Löcher mehr. (b) Die Maske aus (a) nach *Open*-Operation. Kleine Maskenfragmente, die nicht nahe genug an der restlichen Maske liegen, verschwinden hierdurch. (c) Die resultierende Kontur, eingezeichnet in die Aufnahme aus Abbildung 3.4 (a).

Nutzt man zusätzlich aus, dass von jedem CT auch eine manuell vorsegmentierte Aufnahme vorhanden ist, lassen sich mit dieser Methode Masken erstellen, die nach dem visuellen Eindruck gut passen. Objektiv lässt sich die Qualität dieser Masken allerdings schlecht messen, als einziger Vergleichswert die im Rahmen der Arbeit [7] manuell erstellten Grauwertschwellwerte herangezogen werden können. Daher wurde im Weiteren nach einem Verfahren gesucht, das im Gegensatz zu dem oben vorgestellten Verfahren einen Schwellwert im Grauwertbereich setzt.

3.2.2 Segmentierung mittels A Posteriori Wahrscheinlichkeiten

Die Grundlage für die hier verwendete Methode bildet das Bayes-Theorem:

$$P(\omega|\mathbf{x}) = \frac{p(\mathbf{x}|\omega)P(\omega)}{p(\mathbf{x})},$$

wobei $p(\mathbf{x}|\omega)$ die klassenspezifische Verteilungsdichte der Klasse ω und $P(\omega|\mathbf{x})$ die a-posteriori Wahrscheinlichkeitsdichte ist, also die Wahrscheinlichkeit dafür ist, dass das beobachtete Ereignis \mathbf{x} der Ereignisklasse ω angehört. Vergleiche hierzu [5].

Im Folgenden werden die in einer Aufnahme auftretenden Grauwerte genauso wie die Zugehörigkeit eines Voxels zur Klasse CT oder zur Klasse Hintergrund als mögliche, abhängige Ereignisse \mathbf{x} , ω_1 und ω_2 betrachtet.

Sofern man nun in der Lage ist, die klassenspezifische Verteilungsdichte $p(\mathbf{x}|\omega_i)$ zu schätzen, also zu schätzen, mit welcher Wahrscheinlichkeit ein Grauwert \mathbf{x} im Hintergrund oder im CT auftritt, und sofern die a-priori Wahrscheinlichkeiten von ω_1 und ω_2 bekannt sind, ist man in der Lage jeden Grauwert mit einer bestimmten Wahrscheinlichkeit dem Vordergrund oder dem Hintergrund zuzuordnen. Hierzu benutzt man den Bayes- oder MAP-Klassifikator:

$$\frac{p(\mathbf{x}|\omega_1)P(\omega_1)}{p(\mathbf{x})} \geq \frac{p(\mathbf{x}|\omega_2)P(\omega_2)}{p(\mathbf{x})}.$$

Dies ist gleichbedeutend mit der Entscheidungsfunktion

$$p(\mathbf{x}|\omega_1)P(\omega_1) \geq p(\mathbf{x}|\omega_2)P(\omega_2). \quad (3.3)$$

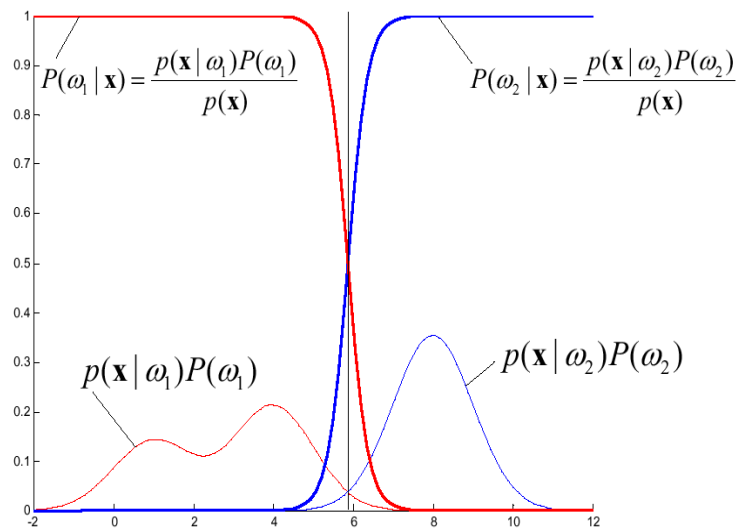


Abbildung 3.11: Schematische Darstellung der Bayes-Klassifikation aus [5].

Als Grundlage für die Schätzung der klassenspezifischen Wahrscheinlichkeitsdichten wurden die Grauwertistogramme der Aufnahmen verwendet. Diese geben an, wie häufig welcher Grauwert in einer Aufnahme vorkommt, sie können also auch als Wahrscheinlichkeitsverteilungen der Grauwerte in einer Aufnahme interpretiert werden.

Das in Abbildung 3.12 dargestellte Histogramm weist zwei deutliche Maxima auf, das globale Maximum liegt beim Grauwert 10. Dann fällt es recht schnell ab - um etwa ab dem Grauwert 20 flach auszulaufen. Die Grauwerte der Voxel, die zu den CTs gehören, sind deutlich heller als

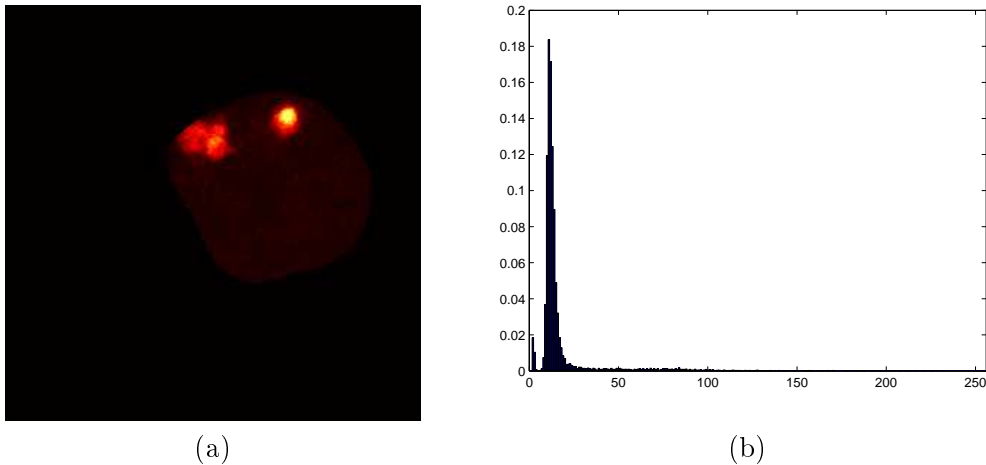


Abbildung 3.12: (a) Eine Schicht des FITC-Kanals der MEF-Zelle, deren DAPI-Kanal in Abbildung 3.5(a) dargestellt ist; der Hintergrund ist hier ausgeblendet. (b) Das Histogramm dieser Aufnahme. Die Hintergrundhelligkeit des Zellkerns nimmt den größten Teil der nach Maskierung verbleibenden Aufnahme ein. Das Histogramm weist daher ein deutliches Maximum auf, das den häufigsten Hintergrundgrauwert markiert.

die Hintergrundgrauwerte. Der obere, flache Teil des Histogramms stammt daher größtenteils von Voxeln aus den CTs.

Im Folgenden wird die Hintergrundgrauwertverteilung durch eine Poissonverteilung angenähert. Eine Zufallsgröße heißt *poissonverteilt*, wenn sie die abzählbar unendlich vielen möglichen Werte $0, 1, 2, \dots$ mit den Wahrscheinlichkeiten

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, (k = 0, 1, 2, \dots)$$

annimmt[6]. Die Poissonverteilung hat also nur einen Parameter λ . Sie lässt sich daher zwar nicht besonders gut an die jeweilige Grauwertverteilung anpassen, stellt andererseits allerdings eine sehr einfache Approximation der tatsächlichen Verteilung dar. Als λ -Wert der Poissonverteilungen wurde hier jeweils der Mittelwert der Grauwerte einer Aufnahme gewählt.

Zusätzlich stellt sich die Frage, wie die Poissonverteilung am Besten zu skalieren ist, das heißt, wieviele Grauwerte am ehesten einem Ereignis entsprechen. Es hat sich gezeigt, dass sich mit einer Skalierung von einem Grauwert je Ereignis brauchbare Ergebnisse erzielen lassen, leichte Veränderungen an dieser Skalierung beeinflussen das Ergebnis nur minimal.

Grundsätzlich kann man davon ausgehen, dass Grauwerte $x \leq \lambda$ zum Hintergrund und nicht zum CT gehören. Daher kann man aus dem Histogramm und der Poissonverteilung die Vordergrundgrauwertverteilung schätzen, indem man an jedem Punkt zwischen 1 und 256 die Poissonverteilungsdichtefunktion vom Histogramm subtrahiert. Dabei entstehende negative Werte werden ebenso wie Werte für $x \leq \lambda$ auf Null gesetzt. Abbildung 3.14 zeigt die entstehenden Wahrscheinlichkeitsdichtefunktionen für die Aufnahme aus Abbildung 3.12(a). Die a-priori Wahrscheinlichkeiten für die Zugehörigkeit eines Voxels zum Hintergrund oder zum CT entsprechen dem Anteil, den das CT am Volumen des Zellkerns einnimmt. Mithilfe der manuell bestimmten Schwellwerte wurde dies anhand einiger Beispielzellkerne aus der Datenbank getestet: im Mittel nahmen die CTs fünf Prozent des Zellkernvolumens ein. Daher wurde als a priori Wahrscheinlichkeit für die Zugehörigkeit eines Voxels zum CT 0.05 gewählt, für die Zugehörigkeit zum Hintergrund 0.95.

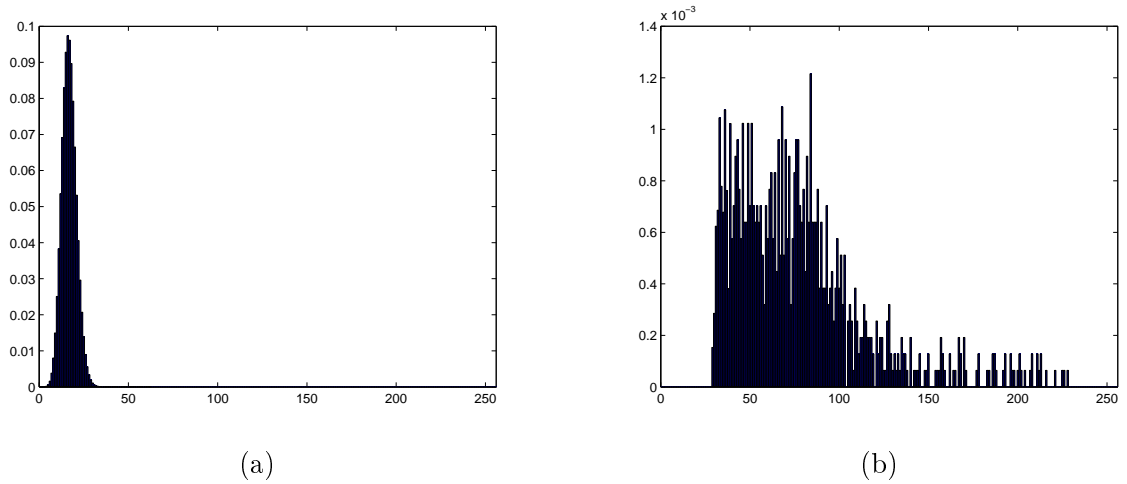


Abbildung 3.13: (a) Poissonverteilung mit $\lambda = 16$ (b) Das Histogramm aus 3.12 (b) nach Subtraktion der Poissonverteilung. Übrig bleiben die Grauwerte, die wahrscheinlich nur im Vordergrund auftreten. Das Histogramm gibt nun Aufschluss darüber, wie häufig welcher Grauwert im Vordergrund auftritt.

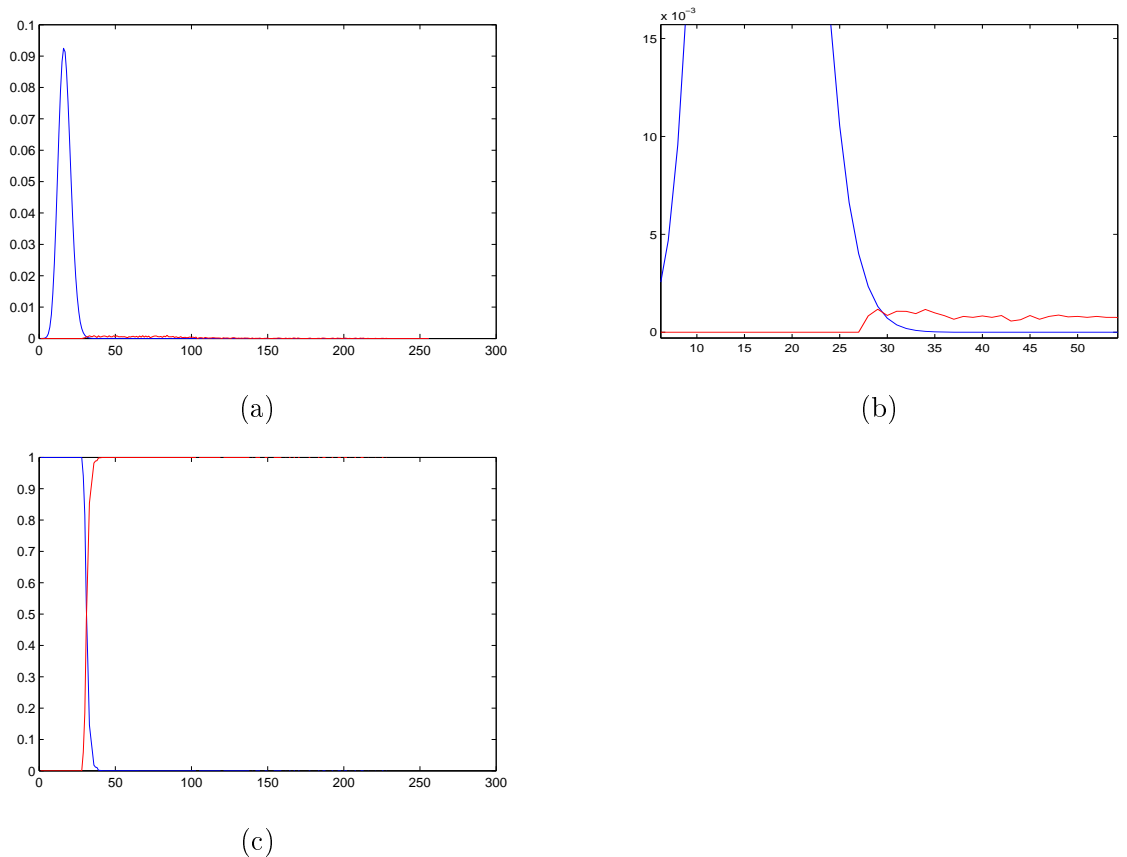


Abbildung 3.14: (a) Gleichung 3.3 ist hier für die sich aus dem Histogramm ergebenden Verteilungen dargestellt: blau: $p(x|Hintergrund)P(Hintergrund)$, rot: $p(x|CT)P(CT)$, (b) Um die sehr geringen Vordergrundwahrscheinlichkeiten sichtbar zu machen, ist hier eine Detailansicht von (a) dargestellt. (c) blau: $\frac{p(x|Hintergrund)P(Hintergrund)}{p(x)}$, rot: $\frac{p(x|CT)P(CT)}{p(x)}$.

Mittels Gleichung 3.3 lassen sich die in Abbildung 3.14 dargestellten Entscheidungsfunktionen erzeugen. Für jeden Grauwert zwischen 1 und 256 ist in Abbildung 3.14(b) aufgetragen, mit welcher Wahrscheinlichkeit er der Klasse Hintergrund (blau) oder Vordergrund (rot) zugeordnet werden kann. Die Schwellwerte, die sich aus diesem Verfahren ergeben sind für die MEF und MES Zellen im Anhang C aufgelistet und können dort mit den manuell bestimmten Schwellwerten verglichen werden.

Filtert man vor der Klassifikation die Aufnahme mit einem dreidimensionalen Tiefpassfilter, so steigt die Wahrscheinlichkeit, dass benachbarte Voxel der gleichen Klasse zugeordnet werden. Daher wurden die Aufnahmen vor der Klassifikation gaußgefiltert. Abbildung 3.15 zeigt die sich aus Abbildung 3.12(a) ergebenden Wahrscheinlichkeiten für jeden Punkt, dem CT anzugehören. Die Maske, die sich aus diesem Verfahren für das Beispiel aus Abbildung 3.1(b) ergibt, ist in Abbildung 3.16 genutzt worden, um den Hintergrund auszublenden.

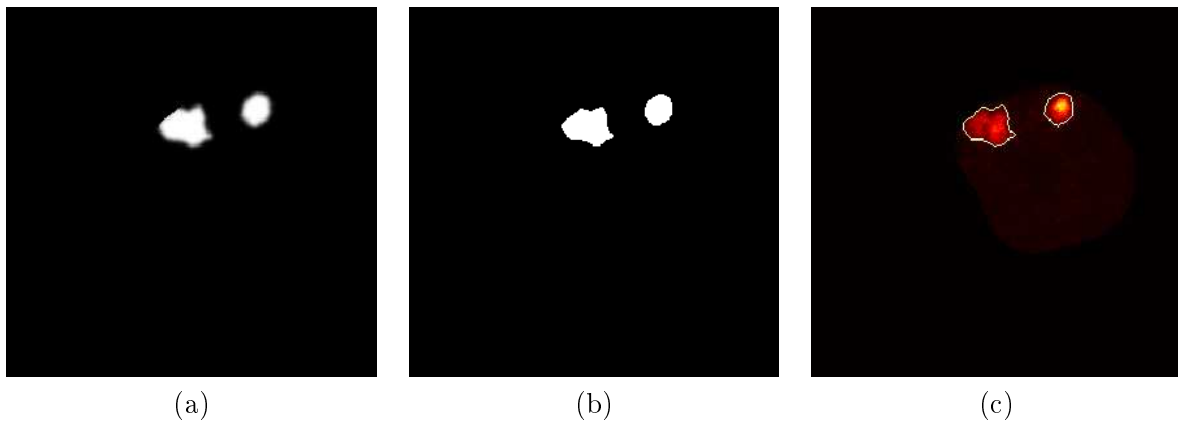


Abbildung 3.15: (a) Wahrscheinlichkeiten für die Zugehörigkeit jedes Pixels zur Klasse CT, (b) Maske nach MAP-Klassifikation (c) die Kontur der Maske aus (b) eingezeichnet in die Aufnahme aus 3.12(a).

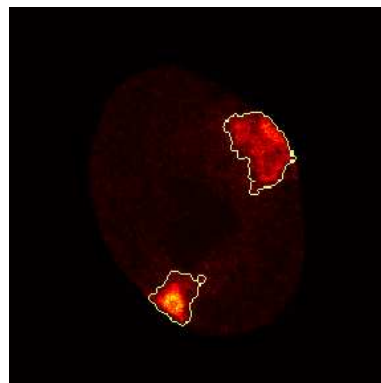


Abbildung 3.16: Die mittels MAP-Klassifikation erstellte Maske ist hier als Kontur in den Datensatz aus Abbildung 3.4 eingezeichnet.

3.3 Berechnung der Gradientenrichtungen

Auf der Suche nach Merkmalen, die vor allem gegenüber Grauwerttransformationen invariant sind, bietet sich eine Betrachtung der Gradientenrichtungen an - vergleiche hierzu Gleichung 3.1.

Im zweidimensionalen Raum können die Gradientenrichtungen einer Bildmatrix der Größe $m \times n$ als Matrix der gleichen Größe dargestellt werden, deren Einträge im Wertebereich $[0, 360]$ liegen. Die Werte entsprechen dem absoluten Winkel zur x -Achse, der Wert 0 ist daher gleichbedeutend mit dem Wert 360.

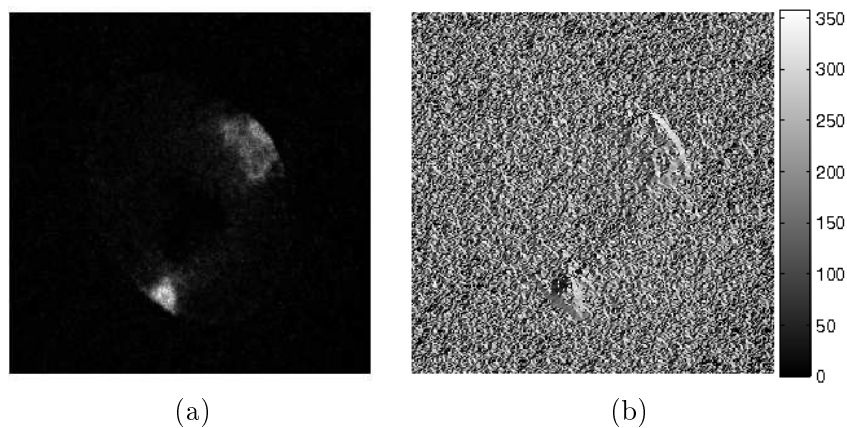


Abbildung 3.17: (a) Eine zweidimensionale Schicht aus einem Datensatz, (b) Gradientenrichtungen der zweidimensionalen Schicht.

Abbildung 3.17 zeigt eine Bildmatrix und die dazugehörigen Gradientenrichtungen. Der Wert 0 ist hier schwarz eingezeichnet, der Wert 360 weiß. Man sieht, dass die Gradientenrichtungen hier besonders im Hintergrund nicht sehr stabil sind. Leider ändern sie sich auch in den CTs sehr oft. Dies rührt daher, dass durch leichtes Rauschen gerade die Gradientenrichtungen besonders stark verfälscht werden. Daher müssen die Aufnahmen vor der Berechnung der Gradientenrichtungen geglättet werden. Hierzu wurde ein 5×5 Pixel großes Gaußfilter verwendet. In Abbildung 3.18 sieht man deutlich, dass die Gradientenrichtungen in den CTs durch diese Tiefpassfilterung stabiler werden.

Abbildung 3.19 zeigt die Aufnahme aus Abbildung 3.17 nach dreifacher Gaußfilterung. Hier sind nicht nur die Gradientenrichtungen im Vordergrund stabiler, sondern auch im Hintergrund ergeben sich Strukturen: unerwünschte Artefakte.

Um zu einem dreidimensionalen Datensatz der Größe $m \times n \times k$ eine Darstellung der Gradientenrichtung derselben Dimension zu erhalten, werden die dreidimensionalen Gradienten auf die zur Zellkernwand zeigenden Gradienten projiziert. Damit erhält man Matrixeinträge im Wertebereich $[-1,1]$, wobei 1 bedeutet, dass der Gradient in Richtung Zellkernwand zeigt, -1 dagegen, dass der Gradient von der Zellkernwand wegzeigt. Diese *relativen Gradientenrichtungen* sind im Gegensatz zu den absoluten auch rotationsinvariant.

Mithilfe der Zellkernmasken, die in Abschnitt 3.1 vorgestellt wurden, lässt sich diese Projektion leicht umsetzen. Aus den inversen Masken muss dazu zunächst eine Distanztransformation berechnet werden. Diese enthält nun für jedes Voxel des Zellkerns seinen Abstand zur Zellkernwand. Die Werte wachsen linear zum Zellkernzentrum hin. Berechnet man nun von dieser Distanztransformation die Gradienten, so erhält man Gradienten der Länge 1, die alle zur Zellkernwand zeigen. Das Skalarprodukt zwischen diesen Gradienten und den normierten Gradienten des Originaldatensatzes liefert das gewünschte Ergebnis.

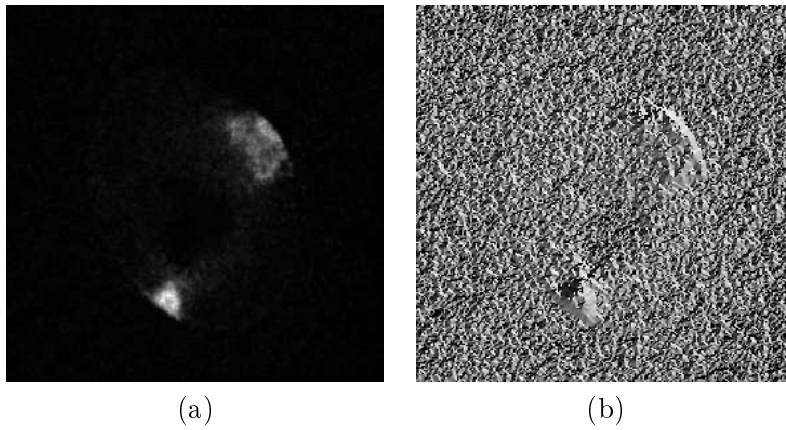


Abbildung 3.18: (a) Der gaußgefilterte Datensatz aus Abbildung 3.17, (b) Gradientenrichtungen des gaußgefilterten Datensatzes. Die Gradientenrichtungen um die CTs herum sind jetzt stabiler; im Hintergrund zeigen sie immer noch in willkürliche Richtungen.

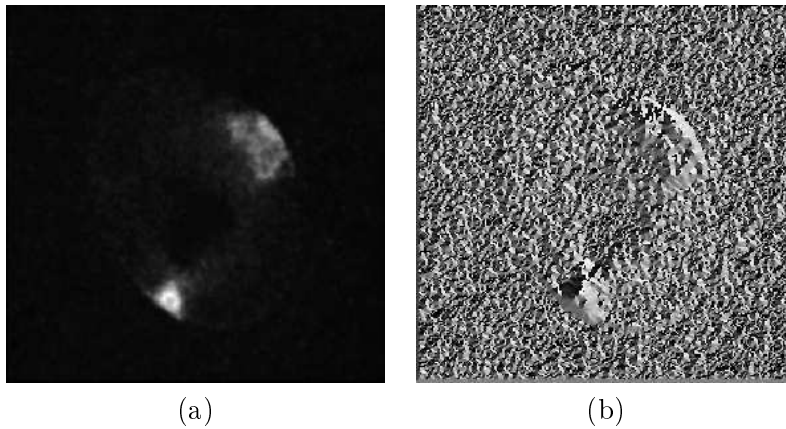


Abbildung 3.19: (a) Der Datensatz nach dreifacher Gaußfilterung und (b) die zugehörigen Gradientenrichtungen.

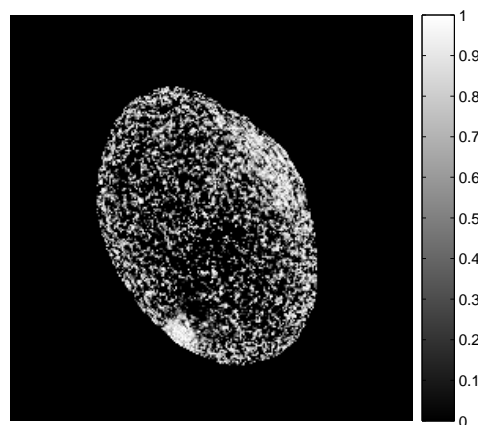


Abbildung 3.20: Die dem Beispiel aus Abbildung 3.19 (a) entsprechende Schicht der Projektion der dreidimensionalen Gradienten auf die Gradienten zur Zellkernwand. Die Werte sind hier auf das Intervall $[0, 1]$ normiert.

Kapitel 4

Invariante Merkmale

Ziel dieser Arbeit ist es, Merkmale zu generieren, die die CTs nicht nur individuell gut beschreiben; vielmehr sollen diese Merkmale möglichst spezifisch einen bestimmten Grad der *Aktivität* eines Chromosoms beschreiben. Sie sollen sich also nach Möglichkeit nicht nur gut von einer Zelle auf die andere übertragen lassen, sondern auch von einem Zelltyp auf den anderen und von einem Chromosom aufs andere.

Augenscheinlich geht es darum, die Struktur des Chromatin im jeweiligen CT zu beschreiben. Diese äußert sich in der Textur der CTs.

Mit einer genauen Definition des Begriffs Textur tut man sich in der Regel schwer. Qualitativ kann eine Textur beschrieben werden als mehr oder weniger rauh, grobkörnig, homogener oder weniger homogen, regelmäßig oder unregelmäßig, gerichtet oder ungerichtet, vergleiche hierzu [1] und [22]. Dementsprechend gibt es viele unterschiedliche Ansätze, Textur zu quantifizieren; man kann grundsätzlich zwischen statischen, geometrischen oder strukturellen oder modellbasierten Methoden unterscheiden (zu einer genaueren Einteilung siehe [22]). Der visuelle Eindruck lässt darauf schließen, dass die Textur des Chromatin in den CTs ungerichtet, also isotrop ist. Allerdings sind CTs meist nicht rund, sie haben manchmal eine flache, scheibenförmige, manchmal eine eher längliche Form. Es stellt sich die Frage, inwiefern dies eine Rolle spielt, wenn man die Textur nur in einer Richtung untersucht. Um dies zu untersuchen, wurde bei manchen der berechneten Texturmerkmale auch Rotationsinvarianz hergestellt. Die Ergebnisse dieser Merkmale wurden mit denen der nicht rotationsinvarianten Merkmale verglichen. So kann man gleichzeitig entscheiden, ob die Annahme der Isotropie der Textur selbst korrekt ist.

In diesem Kapitel wird zunächst dargestellt, wie invariante Merkmale zu zwei- und dreidimensionalen Datensätzen erstellt werden können. Dann werden diejenigen Merkmale vorgestellt, die zur Beschreibung der Textur des Chromatin berechnet wurden. Vor allem statistische Merkmale waren hier erfolgsversprechend.

4.1 Erzeugung invarianter Merkmale

Zur Erstellung invarianter Merkmale für n-dimensionale Datenmengen stehen verschiedene Methoden zur Verfügung. In [17] beispielsweise wird eine Methode vorgestellt, solche Merkmale zu berechnen. Dazu müssen lediglich zwei Dinge bekannt sein. Zum einen benötigt man mindestens ein Kriterium, mit welchem eine gesuchte Eigenschaft eines Datensatzes gemessen werden kann, zum anderen muss man wissen, welche mathematischen Transformationen die gesuchte Eigenschaft nicht beeinflussen. Für ein beliebiges n-dimensionales Datum $X = \{X(\mathbf{x}) | 0 \leq x_1 < N_1, \dots, 0 \leq x_n < N_n\}$ lassen sich dann mittels Haar-Integration über der endlichen, kompakten Transformationsgruppe G invariante Merkmale folgendermaßen berechnen:

$$T[f](X) := \int_G f(gX) dg \quad (4.1)$$

Hierbei ist g ein Element der Transformationsgruppe G , gX das mit g transformierte, n-dimensionale Datum X . f ist eine nichtlineare Kernfunktion. Um zum Beispiel im dreidimensionalen Raum rotations- und translationsinvariante Grauwertmerkmale zu erhalten, wird ein Haar-Integral über der Gruppe der dreidimensionalen Euklid'schen Transformationen gebildet.

$$T[f](X) := \int_{\mathbf{x} \in N} \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{\psi=0}^{2\pi} f(g(\mathbf{x}, \varphi, \theta)X) \sin(\theta) d\psi d\theta d\varphi d\mathbf{x} \quad (4.2)$$

Die Berechnung kann in zwei Schritte unterteilt werden: die Auswertung der Kernfunktion für jeden Voxelwert, und das Summieren über die dadurch entstandenen Zwischenwerte:

$$T[f](X) := \sum_{\mathbf{x} \in N} \sum_{r_1=0}^{R-1} \sum_{r_2=0}^{R-1} \sum_{r_3=0}^{R-1} f(g(\mathbf{x}, \varphi, \theta, \psi)X), \quad (4.3)$$

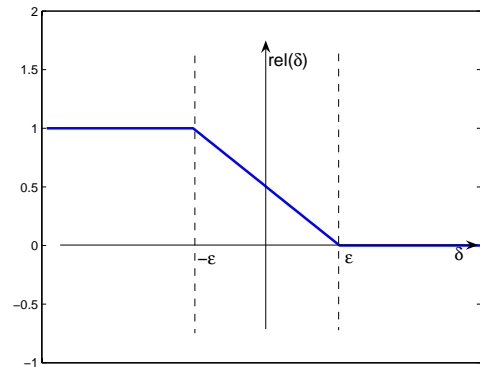
wobei $\varphi = 2\pi \frac{r_1}{R}$, $\theta = \pi \frac{r_2}{R}$ und $\psi = 2\pi \frac{r_3}{R}$.

4.2 Relationale Kernfunktionen

Relationale Kernfunktionen wurden in [19] erfolgreich zur Klassifikation von zweidimensionalen Texturen eingesetzt. Sie sind sowohl invariant gegenüber zweidimensionalen Euklid'schen Transformationen als auch gegenüber additiven Grauwerttransformationen. Die relationalen Kernfunktionen (RKF) wurden als eine Erweiterung der *local binary pattern* (LBP) entwickelt, die das Verhältnis zwischen dem Grauwert eines Pixels zu den Grauwerten der Pixel seiner 3x3-Nachbarschaft auf ein binäres Muster abbilden. Dieses resultierende Muster ist schon bei geringen Grauwertvariationen nicht stabil, was sich als entscheidender Nachteil der LBP erwiesen hat. Da die RKF die Differenz zwischen zwei Grauwerten auf eine reelle Zahl abbilden, bilden sie ein kontinuierliches Merkmal und weisen somit eine wesentlich höhere Stabilität auf.

Ein Beispiel für eine einfache RKF sieht für eine Breite $\epsilon \geq 0$ folgendermaßen aus.

$$rel(\delta) = \begin{cases} 1 & \text{für alle } \delta < -\epsilon \\ \frac{1}{2\epsilon}(\epsilon - \delta) & \text{für alle } -\epsilon \leq \delta \leq \epsilon \\ 0 & \text{für alle } \epsilon < \delta \end{cases}$$



Die RKF wird auf Kreisen ausgewertet, die sich auf verschiedenen Radien um einen Bildpunkt befinden. Diese Kreise werden mit $p_i = \{r_i, \phi_i\}$, $i \in \mathbb{N}$ definiert; $r_i \in \mathbb{R}_+$ ist hierbei der Radius und $\phi_i \in [0, 2\pi[$ die Phase eines Kreises i . Wie viele Grauwerte auf einem Kreis ausgewertet werden, legt die *sampling rate* fest.

In Abbildung 4.1 ist ein Beispiel für zwei Kreise $p_1 = \{1, \pi/8\}$ und $p_2 = \{2, 7\pi/8\}$ aufgeführt, wobei als *sampling rate* 16 gewählt wurde.

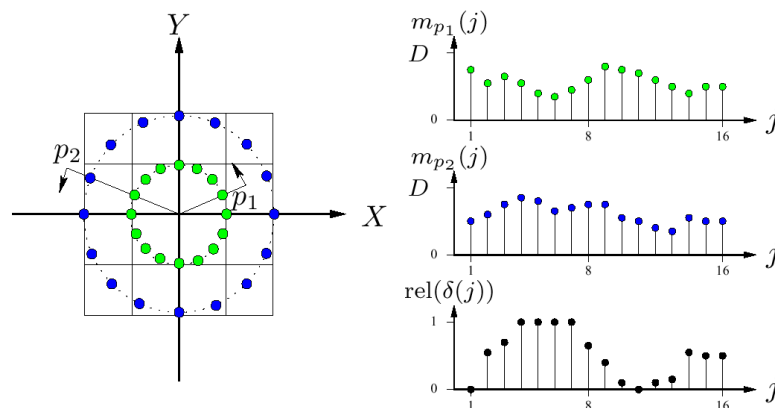


Abbildung 4.1: Relationale Kernfunktionen auf Beispielradien. Aus [19].

Auf der rechten Seite sind die Grauwerte, die sich auf jedem Kreis ergeben, abgebildet. Die Differenzen zwischen den korrespondierenden Punkten werden von der RKF auf das Intervall $[0,1]$ abgebildet. Dann wird mittels Integration über dem Kreis Rotationsinvarianz hergestellt. Für jeden Datensatz der Größe $m \times n \times l$ ergeben sich bei k verschiedenen Kreisen $k(k-1)/2$ Merkmalsmatrizen der Größe $m \times n \times l$. Durch Mittelung über die Merkmale an jedem Punkt x erhält man $k(k-1)/2$ globale Merkmale für den Datensatz X .

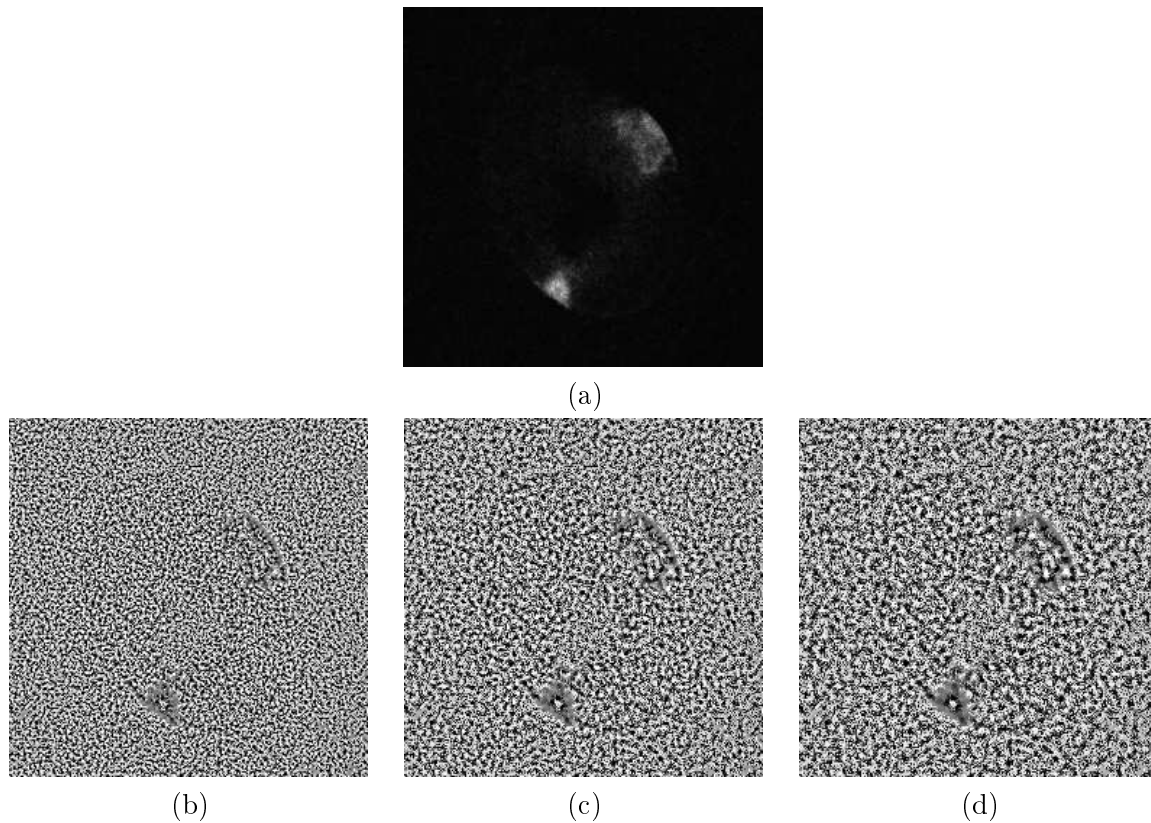


Abbildung 4.2: (a) Eine Schicht aus einem Element der Datenbank. (b)-(d) Merkmalsmatrizen, die sich bei Anwendung der RKF auf (a) ergeben. Der verwendete Schwellwert ϵ beträgt 1, die Kreise, auf denen die RKF ausgewertet wurden, sind: $p_1 = \{0,0\}$, $p_1 = \{1,0\}$ und $p_3 = \{2,0\}$.

Abbildung 4.2 zeigt eine Schicht aus einer 3D-Aufnahme von CTs aus dem Datensatz. Leider zeigen die sich bei Anwendung der RKF auf dieser Aufnahme ergebenden Merkmalsmatrizen, dass die RKF die Textur der CTs nicht angemessen beschreiben können. In sämtlichen betrachteten Radien gibt es in der Ergebnismatrix kaum noch Unterschiede zwischen Hintergrund und CTs. Die RKF gleichen die Unterschiede zwischen Signal und Hintergrundrauschen aus. Lediglich der Rand der CTs ist in den Matrizen noch zu erkennen.

Die RKF sind daher im Folgenden nicht weiter ausgewertet worden.

4.3 Autokorrelation

Grundsätzlich stellt sich bei der Untersuchung von Texturen die Frage, ob es möglich ist, eine Periodizität nachzuweisen, also zu zeigen, dass sich ein Muster in gewissen Abständen wiederholt. Eine solche Periodizität ließe sich mittels Autokorrelation messen.

Bei diesem Ansatz werden die dreidimensionalen Bilderstapel als Zufallsvariablen betrachtet. Für diese Zufallsvariablen wird dann mittels Autokorrelation berechnet, wie ähnlich sie bei unterschiedlichen Translationen m zu sich selbst sind. Die Autokorrelation versteht sich als Spezialfall der Kreuzkorrelation:

$$R_{xy}(m) = E\{x_{m+n}y_n^*\} = E\{x_n y_{m-n}^*\}$$

Hierbei sind x_n und y_n stationäre Zufallsprozesse mit $-\infty < n < \infty$. Sind x und y identisch, so entspricht dies:

$$R_x(m) = E\{x_{m+n}x_n^*\} = E\{x_n x_{m-n}^*\},$$

oder nach Normierung

$$R_x(m) = \frac{E\{x_{m+n}x_n^*\}}{\sigma_0\sigma_m}.$$

σ_0 ist die Standardabweichung zur Translation 0, σ_m die Standardabweichung zur Translation m . Für $m = 0$ hat die normierte Autokorrelationsfunktion stets den Wert 1. Weist ein Signal eine gewisse Periodizität auf, so ist auch die Autokorrelation periodisch mit der gleichen Periode; sie weist also in regelmäßigen Intervallen Werte nahe 1 auf. Die Autokorrelationsfunktion von weißem Rauschen hingegen entspricht einem *Dirac*-Impuls.

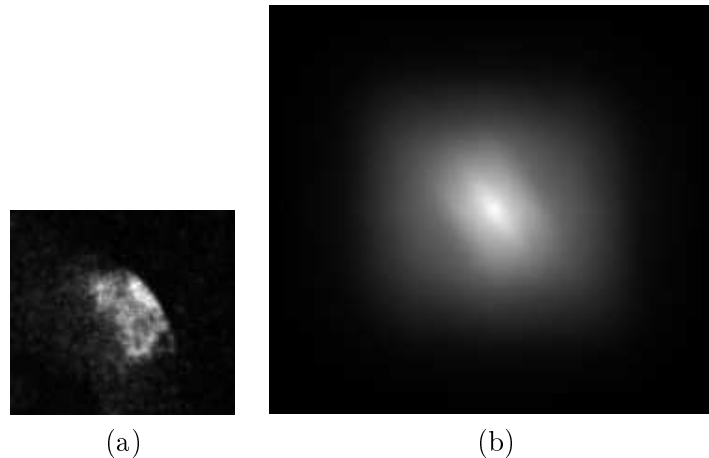


Abbildung 4.3: (a) Schichtaufnahme des CT eines aktiven X-Chromosoms, (b) Autokorrelation dieses CT.

Abbildung 4.3 zeigt eine Schichtaufnahme eines CT und die dazugehörige Autokorrelation. Man sieht, dass die Autokorrelation in einem ellipsenförmigen Bereich um die Translation 0 herum sehr hoch ist. Dies ist durch die Form des CT begründet: das Chromosom selbst ist relativ länglich. Allerdings weist die Autokorrelation keinerlei Periodizität auf. Offenbar kann sie in dieser Form nicht genutzt werden, um die Textur des CTs zu beschreiben.

Aus dem Datensatz aus Abbildung 4.3(a) wurde im Folgenden ein Fenster ausgeschnitten, das eine lokale Maximalstelle und ihre Umgebung enthält. Optisch hat man den Eindruck, dass sich das aktive CT aus mehreren solcher kleinen Auswölbungen zusammensetzt. Setzt sich eine Textur aus mehreren gleichen Fragmenten zusammen, so können diese als Texturelemente (Texel) betrachtet werden. Mittels Kreuzkorrelation zwischen einem Template eines solchen

Texel und der Textur kann dann ein spezifischer Abstand der Texel bestimmt werden. Ob eine solche Regelmäßigkeit in der Textur der CTs feststellbar ist, wurde im Folgenden untersucht. Hierzu wurde das ausgeschnittene Segment gaußgefiltert, um seine Form zu verallgemeinern. Es wurde daraufhin als Template für die Kreuzkorrelationsberechnung verwendet.

Abbildung 4.5 zeigt die Ergebnisse der Kreuzkorrelation des Templates mit dem der Origi-

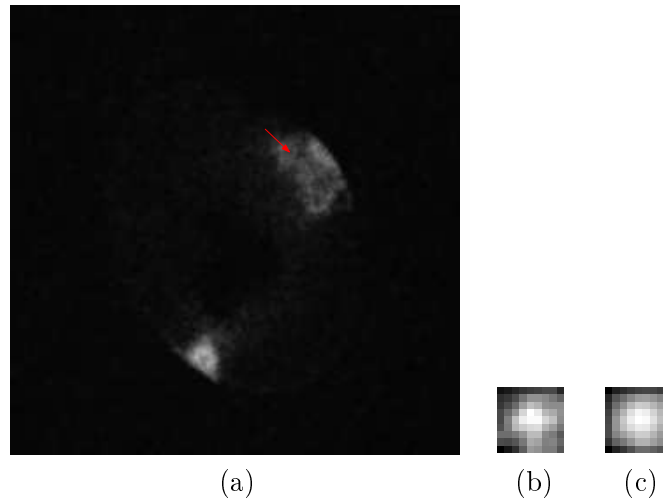


Abbildung 4.4: (a) Der rote Pfeil zeigt auf das ausgeschnittene lokale Maximum. (b) Das ausgeschnittene Fenster. (c) Das im Folgenden verwendete Template.

nalschicht des CT, mit dem CT des inaktiven Chromosoms aus der selben Aufnahme und mit einem Stück Hintergrundtextur der Aufnahme. Aus dieser Abbildung wird klar, dass auch

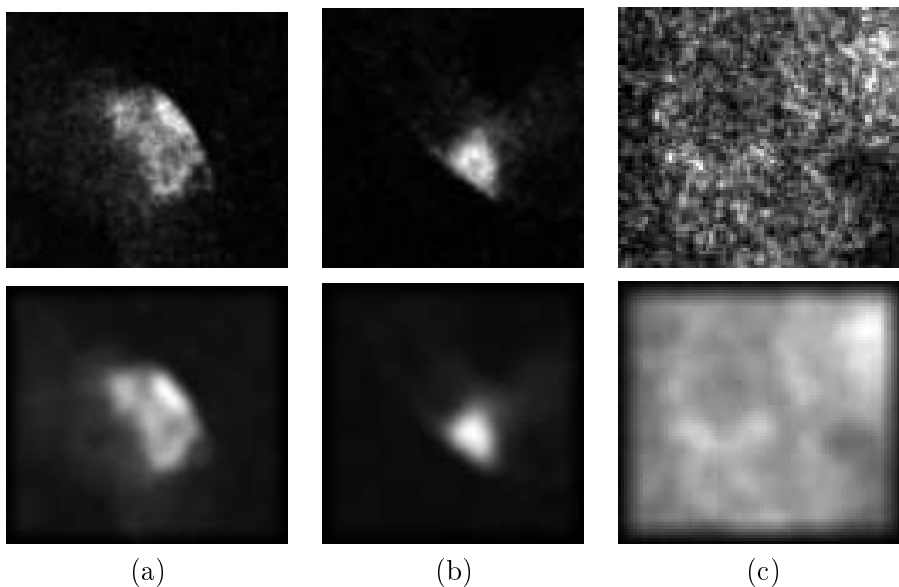


Abbildung 4.5: (a) Die Korrelation des Templates aus Abbildung 4.4 und dem aktiven CT - ihr Maximum liegt bei 0.78, (b) die Korrelation des Templates mit dem inaktiven CT - hier liegt das Maximum bei 0.82 und (c) die Korrelation mit einem Stück Hintergrundtextur, das Maximum beträgt 0.76.

durch Korrelation mit einem Texel-Template keine Regelmäßigkeit in der Textur feststellbar ist. Auch ist die Form des Templates nicht spezifisch für das aktive CT aus dem es entnommen ist. Die Korrelation des geglätteten Ausschnitts ist im inaktiven CT höher als im aktiven -

die Korrelationswerte im Hintergrund sind ebenfalls sehr hoch.

Ähnliche Untersuchungen wurden auch auf den Gradientenrichtungen der Aufnahmen durchgeführt und führten zu ebenso unbefriedigenden Ergebnissen. Die Autokorrelation wurde folglich als Untersuchungsmethode der Chromosomentextur verworfen.

4.4 Mutual Information

Die Mutual Information ist ein Maß für die gegenseitige Abhängigkeit zweier Zufallsvariablen A und B . Sie basiert auf der Entropie einer Zufallsvariablen A , die definiert ist als:

$$H(A) = \sum_{a \in A} p_A(a) \log \left(\frac{1}{p_A(a)} \right) = - \sum_{a \in A} p_A(a) \log(p_A(a)),$$

wobei $p_A(a)$ die Randverteilung von A ist. Die Verbundentropie von A und B ist dann gegeben durch:

$$H(A, B) = \sum_{a \in A, b \in B} p_{A,B}(a, b) \log \left(\frac{1}{p_{A,B}(a, b)} \right) = - \sum_{a \in A, b \in B} p_{A,B}(a, b) \log(p_{A,B}(a, b)),$$

wobei $p_B(b)$ die Randverteilung von B und $p_{A,B}(a, b)$ die Verbundverteilung von A und B ist. Die Mutual Information ist dann definiert als:

$$MI = \sum_{a \in A, b \in B} p_{A,B}(a, b) \log \left(\frac{p_{A,B}(a, b)}{p_A(a)p_B(b)} \right).$$

Nach einigen Umformungsschritten [13] lässt sich dieser Term als Summe von Entropien schreiben.

$$\begin{aligned} MI &= \sum_{a \in A, b \in B} p_{A,B}(a, b) \log \left(\frac{p_{A,B}(a, b)}{p_A(a)p_B(b)} \right) \\ &= \sum_{a \in A, b \in B} (p_{A,B}(a, b) \log(p_{A,B}(a, b)) \\ &\quad - p_{A,B}(a, b) \log(p_A(a)p_B(b))) \\ &= \sum_{a \in A, b \in B} p_{A,B}(a, b) \log(p_{A,B}(a, b)) \\ &\quad - \sum_{a \in A, b \in B} p_{A,B}(a, b) \log(p_A(a)) - \sum_{a \in A, b \in B} p_{A,B}(a, b) \log(p_B(b)) \\ &= \sum_{a \in A, b \in B} \log(p_{A,B}(a, b)) \\ &\quad - \sum_{a \in A} \log(p_A(a)) - \sum_{b \in B} \log(p_B(b)) \\ &= -H(A, B) + H(A) + H(B). \end{aligned}$$

Betrachtet man nun die Intensitäten in einem Datensatz X als Zufallsvariable, so lässt sich für eine gegebene Translation \mathbf{t} die Mutual Information zwischen diesem Datensatz und dem verschobenen Datensatz berechnen.

$$MI_{X, \mathbf{t}} = -H(X, X + \mathbf{t}) + H(X) + H(X + \mathbf{t}) = -H(X, X + \mathbf{t}) + 2H(X)$$

Sie gibt Auskunft darüber, mit welcher Sicherheit man von der Intensität eines Voxels im Datensatz auf die Intensität des um \mathbf{t} verschobenen Voxels schließen kann. Sie wird also

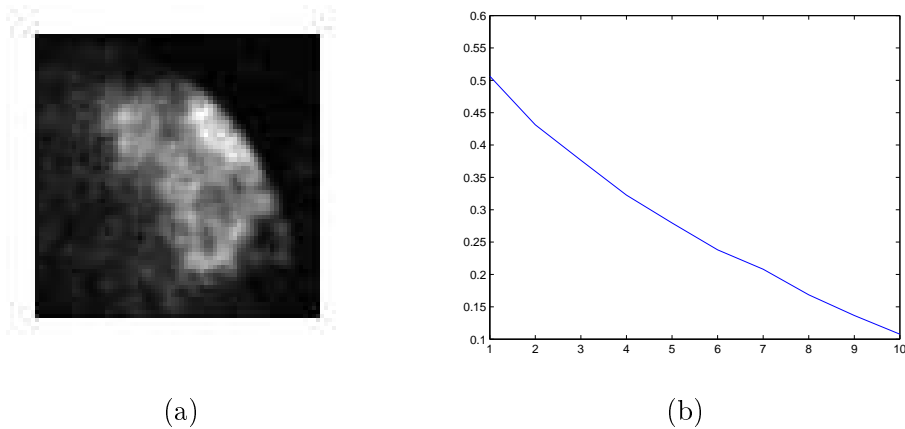


Abbildung 4.6: (a) Ein Ausschnitt aus der Aufnahme eines CTs (b) Die Mutual Information dieses Ausschnitts bei den vertikalen Translationen um 1 bis 10 Pixel. Zur Berechnung der Mutual Information wurden fuzzy histograms verwendet (vergleiche Abschnitt 4.5.1).

kleiner, je schwächer die Intensitätswerte "korreliert" sind. Bei völlig unkorrelierten Signalen beträgt sie Null.

Abbildung 4.6 zeigt deutlich, dass die Mutual Information des Datenbankeintrags geringer wird, je größer der Translationsvektor \mathbf{t} . Bei einem Abstand von zehn Pixel (das entspricht einem μm) ist die Mutual Information schon bei 0,1. Die Werte sind also nur noch sehr wenig voneinander abhängig. Da sich die Texturen aktiver und inaktiver Chromosomen in ihrer Dichte unterscheiden, kann man davon ausgehen, dass sich auch Unterschiede in ihrer Mutual Information zeigen werden. Es stellt sich die Frage, ob die Werte, die die Mutual Information annimmt, spezifisch für die Aktivität von Chromosomen sind.

Die Mutual Information ist translationsinvariant. Berechnet man sie für Translationsvektoren \mathbf{t}_i in allen Raumrichtungen und mittelt über diese, so erhält man auch Rotationsinvarianz. Da allerdings grundsätzlich von einer isotropen Textur ausgegangen wurde, wurde Rotationsinvarianz in diesem Fall nicht hergestellt.

Detailliertere Informationen über die Häufigkeit, mit der bestimmte Intensitäten gemeinsam auftreten, bietet die Betrachtung von Co-occurrence Matrizen. Dies wird im folgenden Abschnitt erläutert.

4.5 Co-occurrence Matrizen

Im Rahmen zahlreicher Studien haben sich Gray-Level Co-occurrence Matrizen als wertvolles Werkzeug erwiesen, um statistische Informationen aus Bildern zu extrahieren. [23] hat gezeigt, dass sich mittels genetischer Algorithmen optimierte Gray-Level Co-occurrence Matrizen dazu eignen, die Textur von Chromatin zu beschreiben. Ursprünglich wurden Co-occurrence Matrizen von Haralick et al.[12] vorgestellt, die sie allerdings nicht direkt für die Klassifikation von Texturen verwendeten. Sie berechneten wiederum Merkmale aus den Matrizen, um schnell Merkmale der Bilder zu berechnen. Dies wird genauer in Abschnitt 4.5.2 beschrieben. Für einen Datensatz X und einen Translationsvektor \mathbf{t} ist die Co-occurrence Matrix definiert als

$$C_{i,j}(X, \mathbf{t}) = \int_X \delta(X(\mathbf{x}) - i)\delta(X(\mathbf{x} + \mathbf{t}) - j)d\mathbf{x}. \quad (4.4)$$

Durch einen Translationsvektor wird also ein Datensatz X gegen sich selbst verschoben. Dadurch entstehen korrespondierende Datensatzpunkte, die in der Co-occurrence Matrix gegeneinander aufgetragen werden. Ist die betrachtete Textur isotrop, dann ist es ausreichend, nur Translationen in einer Richtung zu untersuchen. Es ist dann nur nötig, die Weite der Translation t in diese Richtung, zum Beispiel der x -Richtung, zu bestimmen. Es ist $\mathbf{t} = (t, 0, 0)$. Folgendes Beispiel veranschaulicht prinzipiell die Vorgehensweise.

$$X = \begin{pmatrix} 1 & 4 & 2 & 1 & 9 & 5 & 2 & 3 \\ 3 & 6 & 7 & 0 & 5 & 1 & 1 & 4 \\ 2 & 4 & 2 & 5 & 9 & 9 & 0 & 3 \\ 1 & 4 & 2 & 2 & 9 & 1 & 1 & 3 \\ 3 & 4 & 1 & 0 & 9 & 4 & 2 & 6 \\ 5 & 7 & 0 & 1 & 8 & 5 & 2 & 6 \\ 7 & 2 & 6 & 0 & 8 & 1 & 2 & 3 \end{pmatrix}$$

$$C(X, 1) = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 2 & 2 & 2 & 0 & 3 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 1 & 1 & 3 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

X ist ein Datensatz, der Werte zwischen 0 und 9 enthält. C ist die Co-occurrence Matrix von X für eine Verschiebung um 1 in x -Richtung. Das Wertepaar 4,2 beispielsweise kommt in X 4mal vor, dementsprechend steht in $C(X, 1)(4, 2)$ der Wert 4.

Abbildung 4.7 zeigt Gray-Level Co-occurrence Matrizen, die sich für eine CT-Aufnahme aus der Datenmenge ergeben. Man sieht, dass die Werte stärker gestreut sind, je größer die Translation t ist. Für $t = 0$ liegen alle Werte auf der Hauptachse.

Möchte man nun invariante Merkmale generieren, so setzt man die Co-occurrence Matrix als Kernfunktion in Gleichung 4.1 ein. Rotationsinvariante Merkmale lassen sich dann folgendermaßen generieren:

$$IC_{i,j}(X, \mathbf{t}) := \int_{|\mathbf{t}|=q} \int_X \delta(X(\mathbf{x}) - i)\delta(X(\mathbf{x} + \mathbf{t}) - j)d\mathbf{x}d\mathbf{t}, \quad (4.5)$$

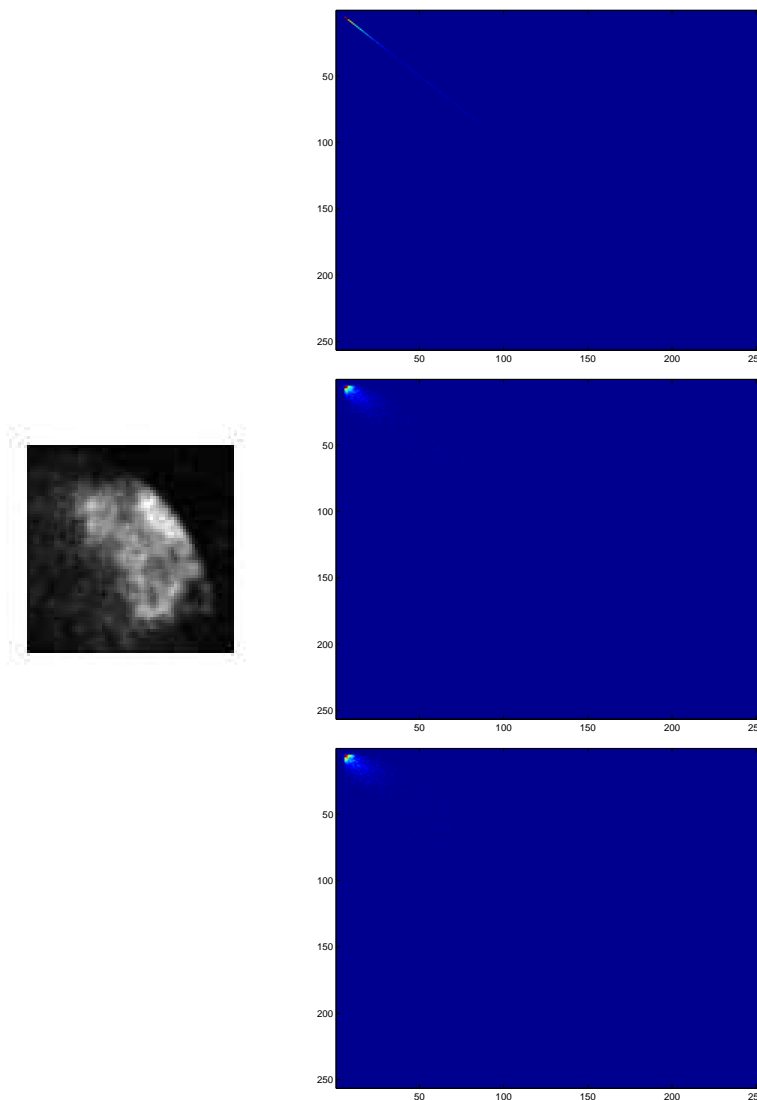


Abbildung 4.7: Zu dem CT aus Abbildung 4.6 sind hier die Co-occurrence Matrizen für $t=0$, 0.2 und $0.4\mu\text{m}$ abgebildet. Für $t=0$ entspricht die Co-occurrence Matrix dem Histogramm. Je größer die Translation t ist, desto stärker sind die Werte gestreut.

wobei q die konstante Länge des Translationsvektors ist. Vergleiche hierzu auch [14]. Im Laufe meiner Untersuchungen hat sich gezeigt, dass es sinnvoll ist, die Einträge der Co-occurrence Matrix so zu gewichten, dass Punkte, die mit größerer Wahrscheinlichkeit im Zentrum der CTs liegen, einen größeren Einfluss auf das Ergebnis haben. Die einfachste Möglichkeit hierzu bietet eine Gewichtung mit den Grauwerten der betrachteten Bildpunkte. Je höher die Grauwerte sind, desto wahrscheinlicher liegen die Punkte im CT.

$$C_{i,j}(X, t) = ij \int_{\mathbf{X}} \delta(X(\mathbf{x}) - i) \delta(X(\mathbf{x} + \mathbf{t}) - j) d\mathbf{x} \quad (4.6)$$

Auch die in Abschnitt 3.2.2 berechneten Wahrscheinlichkeiten für die Zugehörigkeit zum CT sowie natürlich die berechneten Masken wurden als Gewichte für die Co-occurrence Matrizen getestet und ausgewertet. Zusätzlich wurde auch untersucht, wie gut die Gradientenlänge an einer Stelle als Gewichtung funktioniert - wie aussagekräftig also die Gradientenlänge ist bei der Frage, welche Punkte zum CT gehören und welche nicht.

Grundsätzlich wurden Co-occurrence Matrizen hier nicht nur von Grauwertdatensätzen berechnet. Um Invarianz gegenüber Grauwerttransformationen zu erhalten, wurden für jeden Datensatz auch Co-occurrence Matrizen seiner relativen Gradientenrichtungen erstellt. Relative Gradientenrichtung und Grauwert wurden gegeneinander aufgetragen, genauso wie Gradientenlänge und Grauwert.

Es wurden auch Matrizen berechnet, in denen die Grauwerte beziehungsweise Gradientenrichtungen an jedem Punkt gegen dessen Distanz zur Zellkernwand aufgetragen wurden.

Schließlich wurden ganz analog zu den zweidimensionalen Co-occurrence Matrizen dreidimensionale Co-occurrence Matrizen generiert, in denen Gradientenlänge, relative Gradientenrichtungen und Distanz zur Zellkernwand gegeneinander aufgetragen wurden.

Alle Matrizen wurden vor der Auswertung so normiert, dass sich alle Einträge zu 1 aufsummieren.

Damit nicht geringe Wertverschiebungen zu starken Verschiebungen in der Matrix führen, damit also alle Merkmale besser übertragbar sind, ist bei der Berechnung ein Fuzzy Binning verwendet worden. Dies wird im folgenden Abschnitt (4.5.1) vorgestellt. Zusätzlich wurden Haralick-Merkmale von den Co-occurrence Matrizen berechnet. Diese werden in Abschnitt 4.5.2 genauer beschrieben.

4.5.1 Fuzzy Binning

1D - Fuzzy Histograms

Das klassische Histogramm eines Datensatzes X kann definiert werden als:

$$hist_X(i) = \int_X \delta(X(\mathbf{x}) - i),$$

wobei δ der *Dirac*-Impuls ist. Die Länge des Histogramms entspricht dabei der Größe N des Wertebereichs von X , sofern X diskrete Werte hat. Ist dies nicht der Fall, so müssen die in einem Bin zusammengefassten Werte quantisiert werden. Dies bietet sich allerdings in jedem Fall an, damit einerseits die Rechen- und Speicherintensivität verringert wird, andererseits sehr ähnliche Werte zum gleichen Bin gezählt werden können. Die Gewichtungsfunktion h sei hier definiert als:

$$h_m(i) = \begin{cases} 1 & \text{für } i \in [(m - 1/2) \cdot q, (m + 1/2)q] \\ 0 & \text{sonst,} \end{cases}$$

wobei N/q Bins der Größe q entstehen. Dann kann das Histogramm geschrieben werden als:

$$hist_m = \sum_{\mathbf{x} \in X} h_m(X(\mathbf{x})),$$

Dieses klassische Histogramm hat einen entscheidenden Nachteil: es ist diskontinuierlich. Eine beliebig kleine Intensitätsänderung in der Nähe einer Bingenze kann zur Zuordnung des Werts zum falschen Bin führen. [21] stellt daher ein Histogramm vor, das anstelle einer Rechtecksfunktion eine kontinuierliche Quantisierungsfunktion benutzt. Die so entstehenden Histogramme werden als *Fuzzy Histograms* bezeichnet. Vergleiche hierzu Abbildung 4.8.

Wichtig ist hierbei, dass die Summe über alle h_m 1 betragen muss, damit sich die Beiträge eines Eintrags $X(\mathbf{x})$ zu 1 summieren.

$$\int_{-\infty}^{\infty} h_m(i) dm = 1 \quad \forall i$$

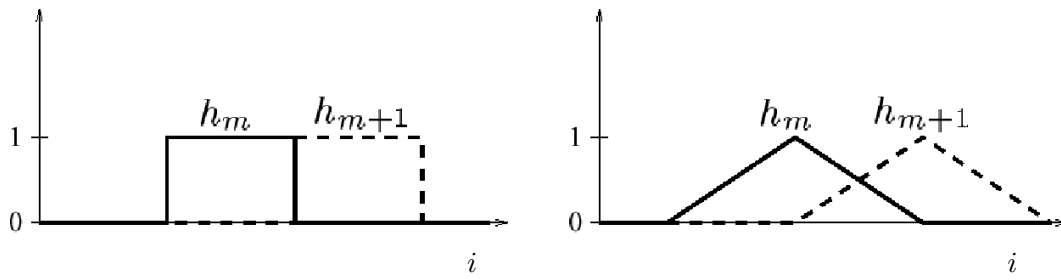


Abbildung 4.8: Klassisches Histogramm (links) und Fuzzy Histogramm (rechts). Aus [21].

Eine einfache Gewichtungsfunktion, die diese Bedingung erfüllt, ist die Dreiecksfunktion:

$$h_m(i) = \begin{cases} 1 - |m - i/q| & \text{für } i \in [(m-1) \cdot q, (m+1) \cdot q] \\ 0 & \text{sonst.} \end{cases}$$

Es wird also im Histogramm zum m -ten Bin 1 addiert, wenn $i = m \cdot q$. Dann liegt der Grauwert i genau im Zentrum des Bins. Für alle i , die zwischen $m \cdot q$ und $(m+1) \cdot q$ liegen, wird zum m -ten Bin $1 - |m \cdot q - i|/q$, zum $(m+1)$ -ten Bin $|m \cdot q - i|/q$ addiert.

2D - Fuzzy Binning

Das eindimensionale Fuzzy Binning kann direkt auf den zweidimensionalen Fall übertragen werden. Die Gewichtungsfunktion entspricht dann nicht mehr einer Dreiecksfunktion, sondern einer Pyramide.

$$h_{k,l}(i,j) = \begin{cases} (1 - |k - i/q|)(1 - |l - j/q|) & \text{für } i \in [m - q, m + q] \\ & \text{und } j \in [n - q, n + q] \\ 0 & \text{sonst,} \end{cases}$$

wobei $k = m/q$ und $l = n/q$. So lassen sich zum Beispiel die zweidimensionalen Co-occurrence Matrizen kontinuierlich quantisieren; sie können so auch direkt miteinander verglichen werden. Nach Fuzzy Binning mit Bingröße 8 entsteht aus der Co-occurrence Matrix aus Abbildung 4.7(mitte) die in Abbildung 4.9 dargestellte.

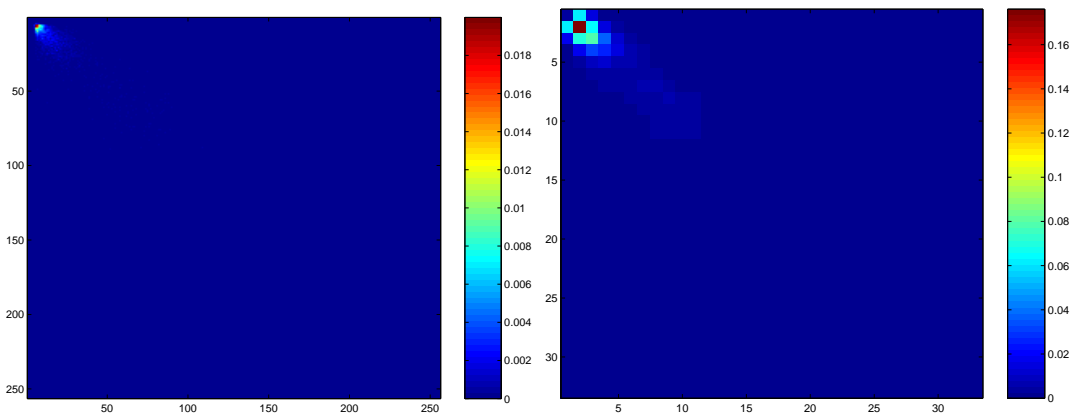


Abbildung 4.9: Klassische und Fuzzy Co-occurrence Matrix für den Datensatz aus Abbildung 4.7 mit $t=0.2\mu\text{m}$. Die 256 Bins in der klassischen Co-occurrence Matrix sind auf 32 Bins in der Fuzzy Matrix reduziert.

4.5.2 Haralick Merkmale

Schon sehr früh wurden Gray-Level Co-occurrence Matrizen zur Analyse von Texturen eingeführt. Bereits 1973 veröffentlichten Haralick et al. eine Studie zur Auswertung von Luftaufnahmen [12]. Hier war die Berechnung der Merkmale in zwei Schritte unterteilt: Die Berechnung der Co-occurrence Matrix aus der Originalaufnahme und die Berechnung verschiedener Merkmale aus dieser Matrix. Auch im Folgenden haben sich einige der von ihm vorgestellten Merkmale bewährt. Tabelle 4.5.2 gibt eine Übersicht über die Merkmale, die hier im Weiteren verwendet wurden.

<i>Energie</i> (angular second moment)	$f = \sum_i \sum_j (C(i, j))^2$
<i>Kontrast</i>	$f = \sum_i (i - j)^2 C(i, j)$
<i>Korrelation</i>	$f = \frac{\sum_i \sum_j (i - \mu)(j - \mu) C(i, j)}{\sigma^2},$
<i>Varianz</i>	$f = \frac{1}{2} \sum_i \sum_j ((i - \mu)^2 C(i, j) + (j - \mu)^2 C(i, j))$
<i>Entropy</i>	$f = - \sum_i \sum_j C(i, j) \log C(i, j)$
<i>Homogenität</i>	$f = \sum_i \sum_j \frac{C(i, j)}{1 + i - j }$
<i>Clustertendenz</i>	$f = \sum_i \sum_j (i + j - 2\mu)^k C(i, j)$

Tabelle 4.1: Verwendete Haralick-Merkmale. μ ist dabei der Mittelwert, σ die Standardabweichung der Co-occurrence Matrix.

Die *Energie* eines Datensatzes ist höher, je mehr Intensitätspaare in der Co-occurrence Matrix gehäuft vorkommen; sie misst die Anzahl an sich wiederholenden Paaren. Das Merkmal *Kontrast* misst den lokalen Kontrast eines Datensatzes. Dieser ist gering, wenn die benachbarten Grauwerte im Datensatz ähnlich sind.

Die *Korrelation* ist die Korrelation zweier Pixel beziehungsweise Voxel in den Voxelpaaren. Je stärker die gegeneinander aufgetragenen Daten korreliert sind, desto höher ist die *Korrelation*. Die *Varianz* misst den Streuungsgrad der Werte im Datensatz. Sie ist größer, je stärker die Werte gestreut sind.

Die *Entropy* ist höher, je willkürlicher die Werte im Datensatz verteilt sind.

Die *Homogenität* ist höher, je ähnlicher sich die Intensitäten der korrespondierenden Voxel sind, je homogener also die Werte im Datensatz sind. Ist die *Homogenität* hoch, so ist der *Kontrast* gering.

Die *Clustertendenz* schließlich ist groß, wenn korrespondierende Voxel mit ähnlichen oder gleichen Intensitätswerten bei bestimmten Intensitäten besonders häufig vorkommen, wenn sie also auf bestimmte Intensitäten clustern.

4.6 Mean Depth-Width Ratio of Extrema

4.6.1 1D - Klassische Berechnung

Der Mean Depth-Width Ratio of Extrema (MDWRE) wurde in [11] erfolgreich angewendet, um aus Grauwertbildern von Zellkernmengen diejenigen Zellkerne zu identifizieren, die sich in Mitose und Interphase befanden. Das Merkmal ist prinzipiell eine Erweiterung der *Relativen Extrema-Dichte*, die in [18] vorgestellt wurde. Beim MDWRE werden die Grauwertbilder dreidimensional als Oberfläche oder Relief interpretiert.

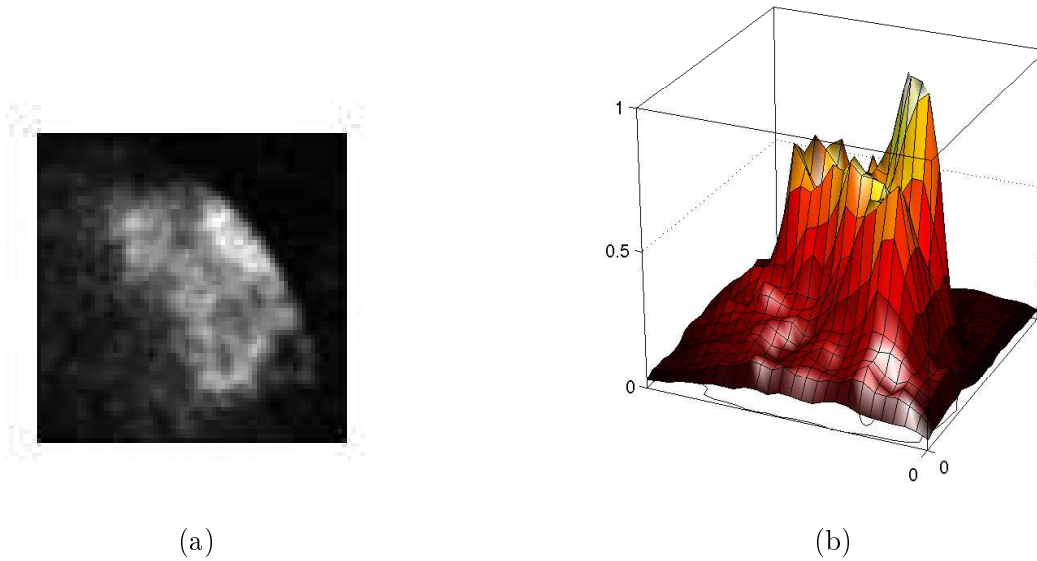


Abbildung 4.10: (a) Eine Grauwertschicht eines CTs und (b) das dazugehörige dreidimensionale Grauwertprofil.

Dieses Relief wird daraufhin untersucht, wie sich die verschiedenen Ausbeulungen und Eindellungen zueinander verhalten. Im klassischen Ansatz wird dies nur eindimensional ausgewertet, das heißt, es werden die Zeilen jedes Datensatzes einzeln betrachtet. Für jede Datensatzzeile werden im Relief Maxima und Minima bestimmt.

Jedes dieser Extrema wird dann zu benachbarten Extrema ins Verhältnis gesetzt. Dazu werden Grauwertdifferenz (Tiefe) und Distanz betrachtet. Der MDWRE ist definiert als:

$$\text{MDWRE} = 1/N \sum_{i=1}^N f_i \text{pwd}_i,$$

wobei

$$\text{pwd}_i = \frac{(V_{A_i} + V_{B_i})/2}{H_{A_i} + H_{B_i}}$$

und

$$f_i = \begin{cases} 1 & \text{falls } \frac{\min(V_{A_i}, V_{B_i})}{\max(V_{A_i}, V_{B_i})} > \tau \\ 0 & \text{sonst.} \end{cases}$$

A_i verbindet hierbei das $i - 1$ -te Maximum mit dem i -ten Minimum, B_i das i -te Minimum mit dem i -ten Maximum. V_{A_i} ist die Projektion von A_i auf die y -Achse. Es misst die Grauwertdifferenz zwischen Maximum und dem Minimum. H_{A_i} ist die Projektion auf die x -Achse;

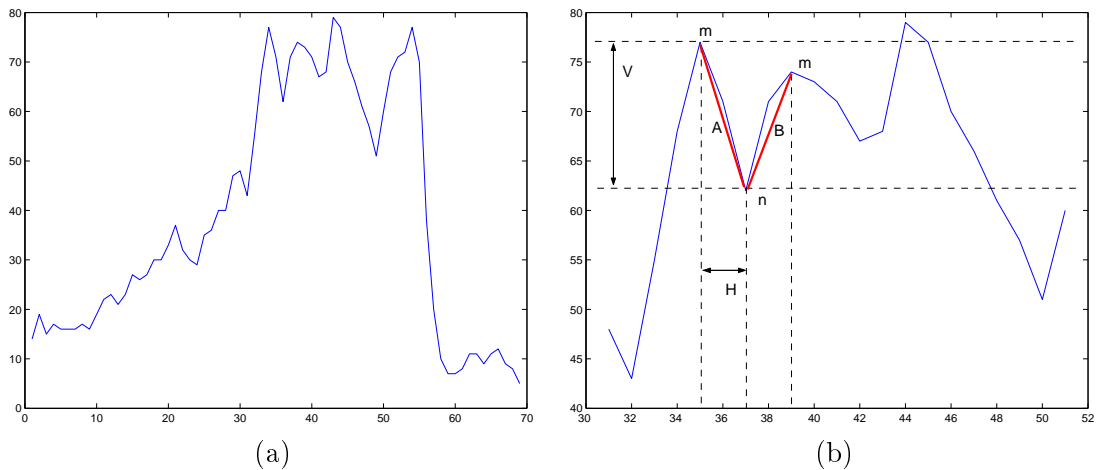


Abbildung 4.11: (a) Das zweidimensionale Grauwertprofil einer Zeile der CT-Schicht aus Abbildung 4.10 (b) eine Detailansicht aus diesem Profil zur Illustration des MDWRE. V ist die vertikale Projektion von A , H die horizontale Projektion von A .

dies entspricht dem Abstand zwischen Maximum und Minimum. Die Funktion f_i sorgt dafür, dass nur diejenigen Extrempaare in die Berechnungen einfließen, deren Differenz größer ist als ein bestimmter Schwellwert τ . So werden zu schwache Ausbeulungen und Sattelpunkte nicht berücksichtigt.

Um den MDWRE für einen kompletten Datensatz zu berechnen, wird über die MDWRE-Werte der einzelnen Zeilen des Datensatzes gemittelt. Grundsätzlich ist der MDWRE invariant gegenüber Translationen, nicht aber gegenüber Rotationen. Auch gegenüber Grauwerttransformationen ist der MDWRE nicht invariant. Nimmt allerdings die Helligkeit zu einem Bildbereich hin monoton zu, so kann dies durch den MDWRE ausgeglichen werden. Zumindest ist der MDWRE also robust gegenüber solchen Grauwerttransformationen.

4.6.2 2D und 3D - eine Erweiterung

Sind die Maxima und Minima im Grauwertprofil eines Datensatzes bestimmt, so lassen sich auch Merkmale aus dem zwei- oder dreidimensionalen Datensatz berechnen, die an den MDWRE angelehnt sind.

In dieser Arbeit wurden sechs weitere Merkmale für eine zweidimensionale Schicht jedes Datensatzes und für jeden dreidimensionalen Datensatz X generiert. Für die zweidimensionale Auswertung musste eine Schicht aus jeder segmentierten Aufnahme ausgewählt werden, die das abgebildete Chromosom möglichst gut darstellt - sie sollte möglichst in dem Teil des CTs liegen, in dem die Intensitäten am größten sind. Es wurde also in jedem Chromosom die Schicht benutzt, die im Mittel die höchsten Intensitäten aufweist.

Abbildung 4.12 verdeutlicht die Unterschiede zum eindimensionalen Fall. Zu jedem Maximum gibt es nicht mehr genau zwei korrespondierende Minima. Allerdings kann zu jedem Maximum das Minimum mit der geringsten Distanz bestimmt werden. Umgekehrt kann zu jedem Minimum ein korrespondierendes Maximum bestimmt werden.

Sei k die Anzahl der bestimmten lokalen Maxima an den Stellen \mathbf{m}_i eines Datensatzes X und l die Anzahl der bestimmten lokalen Minima an den Stellen \mathbf{n}_i . Für eine Maximalstelle \mathbf{m} sei \mathbf{n}_m die Position des Minimums, das zu \mathbf{m} die geringste Distanz hat. Die berechneten MDWRE-basierten Merkmale sind in Tabelle 4.2 aufgeführt.

Alle in Tabelle 4.2 vorgestellten Merkmale sind rotationsinvariant. Die Merkmale E_1 und E_2 messen das Verhältnis der Grauwerte der Extrema zum global maximalen Grauwert. Die Wer-

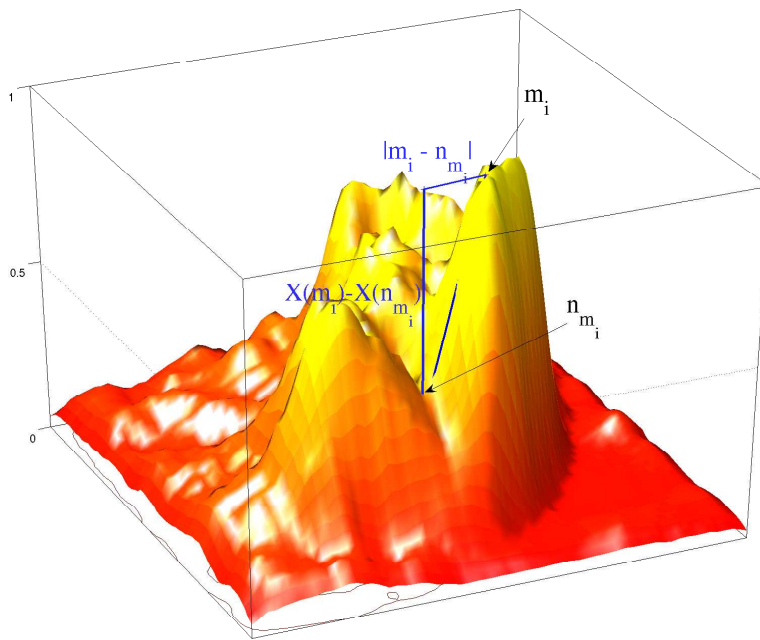


Abbildung 4.12: In der Reliefdarstellung des in Abbildung 4.10(a) dargestellten Beispiels wurden ein lokales Maximum \mathbf{m}_i und das dazugehörige Minimum mit der geringsten Distanz $\mathbf{n}_{\mathbf{m}_i}$ eingezeichnet. Die Höhendifferenz zwischen den Extrema in der Reliefdarstellung entspricht der Grauwertdifferenz der Datensatzpunkte.

$E_1 = \frac{\sum(X(\mathbf{m}_i))}{\max(X)}$	Durchschnittliches Intensitätsverhältnis der Maxima zum globalen Maximum
$E_2 = \frac{\sum(X(\mathbf{n}_i))}{\max(X)}$	Durchschnittliches Intensitätsverhältnis der Minima zum globalen Maximum
$E_3 = \frac{\sum_{i=1}^k \sum_{j=1}^k \mathbf{m}_i - \mathbf{m}_j \cdot X(\mathbf{m}_i)X(\mathbf{m}_j)}{\max(X) \cdot k^2}$	Durchschnittliche Distanz zwischen den Maxima gewichtet mit Grauwerten
$E_4 = \frac{\sum_{i=1}^l \sum_{j=1}^l \mathbf{n}_i - \mathbf{n}_j \cdot X(\mathbf{n}_i)X(\mathbf{n}_j)}{\max(X) \cdot l^2}$	Durchschnittliche Distanz zwischen den Minima gewichtet mit Grauwerten
$E_5 = \frac{1}{k} \sum_{i=1}^k \frac{X(\mathbf{m}_i) - X(\mathbf{n}_{\mathbf{m}_i})}{ \mathbf{m}_i - \mathbf{n}_{\mathbf{m}_i} }$	Durchschnittliches Verhältnis von Grauwertdifferenz zwischen Maxima und jeweils nächstem Minimum und Distanz zwischen Maximum und Minimum
$E_6 = \frac{1}{l} \sum_{i=1}^l \frac{X(\mathbf{m}_{\mathbf{n}_i}) - X(\mathbf{n}_i)}{ \mathbf{m}_{\mathbf{n}_i} - \mathbf{n}_i }$	Durchschnittliches Verhältnis von Grauwertdifferenz zwischen Minima und jeweils nächstem Maximum und Distanz zwischen Minimum und Maximum

Tabelle 4.2: Zwei- und dreidimensionale Merkmale als Erweiterung des MDWRE.

te von E_1 und E_2 sind größer, je mehr Extrema es gibt, die deren Grauwerte ähnlich dem des globalen Maximums sind. Ist das globale Maximum deutlich höher als alle Nebenmaxima, so sind die resultierenden Werte sehr klein. Die Merkmale E_3 und E_4 betrachten die durchschnittliche Distanz zwischen den Extrema. Sie sind dementsprechend nicht skalierungsinvariant. Die Extrema werden hier mit ihren Intensitätswerten gewichtet, da die Grauwerte insgesamt in den CTs höher sind als außerhalb. So kann man gewährleisten, dass lokale Extrema, die am Rand des CTs liegen, weniger stark in das Merkmal eingehen. Durch die Normierung mit dem maximalen Grauwert wird bei den Merkmalen E_1 - E_4 zumindest Invarianz gegenüber additiven Grauwertänderungen hergestellt. Merkmale E_5 und E_6 sind am stärksten verwandt mit dem klassischen MDWRE. Hier wird für jedes Maximum (beziehungsweise Minimum) betrachtet, welche Grauwertdifferenz zu dem Minimum (beziehungsweise Maximum) besteht, das zu ihm die geringste Distanz hat. Diese Differenz wird zur Distanz ins Verhältnis gesetzt. Die Merkmale sind also ebenso invariant gegenüber additiven Grauwerttransformationen. Berechnet man diese MDWRE-basierten Merkmale nicht direkt auf den Intensitätsaufnahmen sondern auf den Gradientenrichtungen beziehungsweise auf den Projektionen der Gradienten in Richtung Zellkernwand (vergleiche Abschnitt 3.3), so sind diese invariant gegenüber Grauwerttransformationen jeder Art.

4.7 Principle Component Analysis

Bei der Untersuchung der Topologie des Zellkerns wird am Lehrstuhl für Anthropologie und Humangenetik der Ludwig-Maximilians-Universität München der Ansatz verfolgt, die Form der CTs mittels Principle Component Analysis zu beschreiben. Grundsätzlich wird hierbei versucht, Ellipsoide so zu bestimmen, dass sie möglichst gut der Form des CTs entsprechen. Die Principle Components oder Hauptachsen sind dabei die Achsen, auf denen diese Ellipsoide liegen.

Sie lassen sich mithilfe der Kovarianzmatrix bestimmen. Für einen dreidimensionalen Datensatz X ist diese gegeben durch:

$$C(X) = \begin{pmatrix} E((X_1 - \mu_1)(X_1 - \mu_1)) & E((X_1 - \mu_1)(X_2 - \mu_2)) & E((X_1 - \mu_1)(X_3 - \mu_3)) \\ E((X_2 - \mu_2)(X_1 - \mu_1)) & E((X_2 - \mu_2)(X_2 - \mu_2)) & E((X_2 - \mu_2)(X_3 - \mu_3)) \\ E((X_3 - \mu_3)(X_1 - \mu_1)) & E((X_3 - \mu_3)(X_2 - \mu_2)) & E((X_3 - \mu_3)(X_3 - \mu_3)) \end{pmatrix},$$

wobei E der Erwartungswert einer Zufallsvariable ist, μ_i der Mittelwert von X in Dimension i . Die Eigenvektoren dieser Kovarianzmatrix entsprechen den gesuchten Hauptachsen, die Eigenwerte der Ausdehnung des dargestellten Objekts in Richtung der jeweiligen Achse.

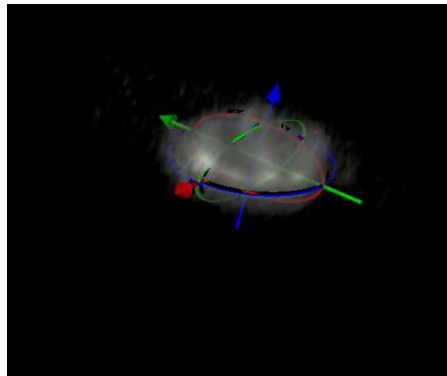


Abbildung 4.13: Eine dreidimensionale Darstellung eines aktiven X-Chromosoms mit seinen Principle Components und den Ellipsoiden, die sich aus ihnen ergeben.

Vor der Berechnung der Hauptachsen der CTs wurde in den Aufnahmen der Hintergrund ausmaskiert. Abbildung 4.13 zeigt die bestimmten Ellipsoide für ein CT eines aktiven Chromosoms. Als aussagekräftig für die Aktivität des Chromosoms betrachtet man dessen Form, die aktiven Chromosomen sind eher länglich, die inaktiven eher scheibenartig, und dessen Ausrichtung zu Aufnahmeebene (in den Fällen, in denen sie der Hauptebene des Zellkerns entspricht). Die These ist hier, dass die aktiven Chromosomen eher parallel zur optischen Ebene liegen, während die inaktiven Chromosomen keine Vorzugsrichtung haben.

Kapitel 5

Klassifikation

5.1 Der k-Nächste-Nachbarn Klassifikator

Der k-Nächste-Nachbarn (k-NN) Klassifikator ist ein einfaches Verfahren zur Klassifikation von Objekten anhand ihrer Merkmale. Alle n Merkmale eines Objekts werden als Punkte im n -dimensionalen Raum dargestellt. Über eine Metrik ist ein Distanzmaß für diesen Raum vorhanden. Für Punkte $x, y \in \mathbb{R}^n$ kann man hier zum Beispiel das Euklid'sche Distanzmaß

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$$

wählen.

Mit einer gelabelten Menge von Trainingsdaten kann ein Punkt x so klassifiziert werden. Die k Trainingspunkte werden bestimmt, die den geringsten Abstand zu dem zu klassifizierenden Punkt haben. Dieser wird schließlich der Klasse zugeordnet, zu der die meisten seiner k nächsten Nachbarn gehören.

5.2 Support Vector Maschine

Die *Support Vector Maschine* ist ein optimaler Klassifikator, der auf einem statistischen, lerntheoretischen Ansatz beruht. Aus einer gelabelten Lernstichprobe von l Beobachtungen soll ein Modell erstellt werden, welches die Wahrscheinlichkeit der Fehlklassifikation von Testdaten minimiert. Zur Lösung dieser *Structural Risk Minimization* liegt ein Ansatz von Vapnik und Chervonenkis vor. Er wird in [2] genauer erläutert.

Die VC-Dimension einer Menge von Funktionen ist hierbei von zentraler Bedeutung. Mit ihr bezeichnet man "die maximale Anzahl von Trainingspunkten, welche durch diese Klasse von Funktionen in allen möglichen Zuordnungen separiert werden können"[2]. Bei der *Structural Risk Minimization* wird die VC-Dimension der Funktion, die die Klassen trennt, minimiert. Im einfachen Fall eines Zweiklassenproblems wird in einem höherdimensionalen Merkmalsraum eine Hyperebene $\langle w, x \rangle + b = 0$ gefunden, die beide Klassen voneinander trennt. Die Klassifikation eines Objekts x geschieht dann über

$$f(x) = \text{sgn}(\langle w, x \rangle + b).$$

Je größer der Abstand zwischen den Klassen ist, also zwischen den jeweiligen Elementen der beiden Klassen, die am nächsten zur jeweiligen anderen Klasse liegen, desto sicherer ist die Klassifikation eines Objekts, sofern konstante Intraklassenstreuung angenommen werden kann. Der Abstand dieser nächsten Objekte zur Trennebene bestimmt den Rand der Trennebene. Je breiter der Rand der Hyperebene ist, umso geringer ist ihre VC-Dimension.

Die Hyperebene ist also durch diejenigen Trainingsdaten genau bestimmt, die auf ihrem Rand liegen. Sie heißen *Support Vektoren* (SV). Wird die Distanz der trennenden Hyperebene w zu den Supportvektoren x_i maximiert, so entspricht dies dem Minimieren von $\|w\|^2$.

$$\begin{aligned} \langle w, w \rangle &\implies \min, \\ \text{wobei } \forall i \quad y_i(\langle w, x_i \rangle + b) &\geq 1 \end{aligned} \tag{5.1}$$

Probleme, welche in ihrem ursprünglichen Merkmalsraum F nicht linear separierbar sind, können mit einer nichtlinearen Abbildung ψ in einen höherdimensionalen Merkmalsraum F' abgebildet werden, so dass sie dort linear trennbar sind. Die Klassifikation wird dann über

$$f(x) = \text{sgn}(\langle w, \psi(x) \rangle + b)$$

vorgenommen.

Nimmt man beispielsweise an, ein Problem sei im ursprünglichen Merkmalsraum mit einem Polynom separierbar. Ist dieses Polynom im n -dimensionalen Merkmalsraum F p -dimensional, so hat der resultierende Merkmalsraum F' $O(n^p)$ Dimensionen. In dieser Dimensionalität ist die explizite Berechnung des Skalarprodukts sehr rechenaufwändig, weshalb man auf eine implizite Berechnung mithilfe von Kernfunktionen K ausweicht.

$$K(x_i, x_j) = \langle \psi(x_i), \psi(x_j) \rangle$$

Verbreitete Kernfunktionen sind hierbei:

$$\begin{aligned} \text{Polynomkerne} \quad K(x_i, x_j) &= (\langle x_i, x_j \rangle + 1)^d \\ \text{Gaußkerne} \quad K(x_i, x_j) &= \exp(-\|x_i - x_j\|^2 / (2\sigma^2)) \\ \text{Sigmoidkerne} \quad K(x_i, x_j) &= \tanh(\kappa \langle x_i, x_j \rangle + \theta) \end{aligned}$$

Das Optimierungsproblem aus (5.1) wird folglich ersetzt durch das dazu Wolf-duale Optimierungsproblem des Maximierens von

$$L'(\alpha) = \sum_{i=1}^l \alpha_i - \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle,$$

$$\text{wobei } \sum_{i=1}^l y_i \alpha_i = 0 \quad \text{und} \quad 0 \leq \alpha_i \leq c.$$

Dabei ist c der *cost*-Wert, der Wert, mit dem eine falsche Klassifikation in der Trainingsmenge bestraft wird. Ist c zu groß, kann dies zu einem Overfitting-Effekt führen. Im ursprünglichen Merkmalsraum F wird die Entscheidungsfunktion berechnet durch

$$f(x) = \text{sgn}\left(\sum_{x_i \in SV} \alpha_i y_i \langle x_i, x \rangle + b\right).$$

Die trennende Hyperebene ist dabei bestimmt durch

$$\sum_{x_i \in SV} \alpha_i y_i \langle x_i, x \rangle + b = 0.$$

Wenn das betrachtete Problem kontinuierlich ist, ist man eventuell nicht an einer binären Entscheidung interessiert. Vielmehr ist es dann von Interesse zu wissen, wie stark ein Objekt

die Eigenschaften einer Klasse aufweist. Die Support Vector Maschine liefert diese Information in Form des *decision values* dv :

$$dv(\mathbf{x}) = \sum_{x_i \in SV} \alpha_i y_i \langle x_i, \mathbf{x} \rangle + b$$

Je weiter der Punkt \mathbf{x} von der Hyperebene entfernt ist, desto größer ist der Betrag seines *decision values*. Damit ist auch die Sicherheit größer, mit der er seiner Klasse zugeordnet wird.

Es gibt zwei Ansätze, die Zweiklassen-SVM auf Probleme mit mehreren Klassen zu erweitern: den *one versus one* und den *one versus rest* Ansatz. Bei *one versus one* werden jeweils nur zwei Klassen betrachtet und für diese eine trennende Hyperebene bestimmt. Insgesamt werden dabei also für n Klassen $n(n-1)/2$ Hyperebenen festgelegt. Im *one versus rest* Ansatz wird in jedem Schritt eine Klasse herausgenommen - alle übrigen Klassen werden als eine betrachtet. Für die zwei sich in jedem Schritt dadurch ergebenden Klassen kann dann eine Hyperebene zur Trennung bestimmt werden. Hier werden also nur n Hyperebenen bestimmt.

Kapitel 6

Ergebnisse

Bei der Auswertung der verschiedenen Merkmale auf den betrachteten Datensätzen sind mehrere Aspekte von Interesse. Grundsätzlich stellt sich zunächst die Frage, wie gut das jeweilige Merkmal die Textur für einen einzelnen Zelltyp beschreibt, wie gut es also dazu geeignet ist, aktiveres von inaktiverem Chromatin des gleichen Chromosoms und im gleichen Zelltyp zu trennen.

Weitaus interessanter ist allerdings, inwiefern Merkmale von einem Zelltyp auf den anderen und von einem Chromosom auf das andere übertragen werden können. Daher wurde jeweils getestet, welche Resultate sich ergeben, wenn man HF- und MEF-Zellen nur unter Kenntnis der jeweils anderen Menge klassifiziert. In diesen beiden Zellmengen werden jeweils aktive und inaktive X-Chromosomen betrachtet.

Auch wurde für jedes Merkmal getestet, ob es sich von Chromosom X auf Chromosom 11 und damit zugleich von Fibroblasten auf undifferenzierte Stammzellen übertragen lässt.

Grundsätzlich wurde dabei zunächst davon ausgegangen, dass alle Chromosomen entweder der Klasse *aktiv* oder der Klasse *inaktiv* zuzuordnen sind. Bei den folgenden Klassifikationsergebnissen ist zu beachten, dass sich bei den MESar Zellen, bei denen ja ein Chromosom 11 nach der Inaktivierung wieder reaktiviert wurde, dieses Chromosom in einem Zwischenzustand zwischen aktiv und inaktiv befindet. Bei der Auswertung wurde davon ausgegangen, dass es in allen Fällen inaktiv ist. Dies führt in vielen Fällen dazu, dass die Klassifikationsergebnisse für diesen Zelltyp in Prozent sehr schlecht sind. Die Vorgehensweise hat allerdings den Vorteil, dass nur so getestet werden kann, wie aufgrund der Kenntnis dieser MESar-Zellen die MESaa-Zellen klassifiziert werden. So kann man einen Eindruck davon gewinnen, ob sich die reaktivierten Chromosomen in ihrer Textur von den aktiven unterscheiden.

Beim Betrachten der Daten gewinnt man den Eindruck, dass ein Großteil der Information über die Aktivität schon im mittleren Grauwert der CTs enthalten ist. Die Ergebnisse, die bei Klassifikation der Datensätze mit dem mittleren Grauwert erzielt werden können, sind allerdings nicht überzeugend. Da schon eine Übertragung von HF-S auf HF-G0 nicht möglich ist und auch die Übertragbarkeit des Merkmals von Menschen- auf Mäusefibroblasten nicht besonders erfolgreich gelingt, ist hier der Versuch einer Übertragung von Chromosom X auf Chromosom 11 nicht aufgeführt. Lediglich innerhalb einer Aufnahmereihe scheint die Klassifikation mittels mittlerem Grauwert sinnvoll zu sein.

6.1 Mutual Information

Bei der Auswertung der Mutual Information gab es im Wesentlichen nur einen Parameter zu variieren, nämlich die Translation t mit der ein Datensatz gegen sich selbst verschoben wird. Für Translationen zwischen $t=0.1\mu\text{m}$ und $t=0.3\mu\text{m}$ in x - und in y -Richtung wurden für jeden Datensatz zwei Mutual-Information-Werte berechnet. Es hat sich gezeigt, dass diese Werte

MEF - MEF	69.2%
HF - HF	75%
MESai - MESai	58.6%
MESar - MESar, MESaa	41.7%
HF-G0 - HF-S	50%
HF-S - HF-G0	50%
MEF - HF	55.5%
HF - MEF	56.4%

Tabelle 6.1: Klassifikationsergebnisse mit mittlerem Grauwert. Es wurde der 10-NN Klassifikator verwendet.

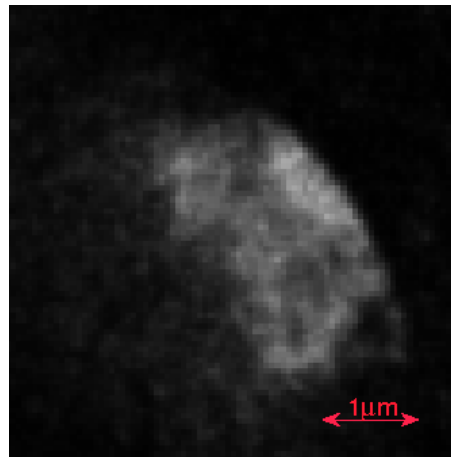


Abbildung 6.1: Aufnahme eines aktiven CTs. Um nicht an Stelle der Textur die Größe des CTs auszuwerten sollten keine Translationen von mehr als $0.3\mu\text{m}$ verwendet werden.

im Allgemeinen sehr ähnlich ausfallen. Vergleiche hierzu Abbildung 6.2.

Es stellt sich die Frage, ob es eine Translation t gibt, für die die Differenz zwischen Mutual Information von aktiven und inaktiven CTs besonders hoch ist. Tabelle 6.1 zeigt die sich ergebenden Klassifikationsergebnisse, sowie die Ergebnisse, die sich aus der Auswertung der Mutual Information auf Gradientenrichtungen ergeben. Zur Klassifikation wurde auch hier der 10-NN Klassifikator verwendet. Insgesamt sind die hier erreichten Ergebnisse nicht befriedigend. Innerhalb eines Zelltyps liegen die Klassifikationsergebnisse jeweils bestenfalls bei zwischen 60% und 70%. Eine Übertragung der Merkmale von einem Zelltyp auf den anderen ist nicht möglich, ebenso wenig eine Übertragung der Merkmale von Chromosom X auf Chromosom 11. Allerdings fällt auf, dass die Ergebnisse der Klassifikation der HF-G0 Zellen nur aufgrund der Kenntnis der HF-S Zellen ebenso wie umgekehrt die Klassifikation der HF-S Zellen aufgrund der Kenntnis der HF-G0 Zellen verhältnismäßig gut funktioniert. Grundsätzlich lässt man wohl bei der Betrachtung der Mutual Information zuviele Informationen außer Acht.

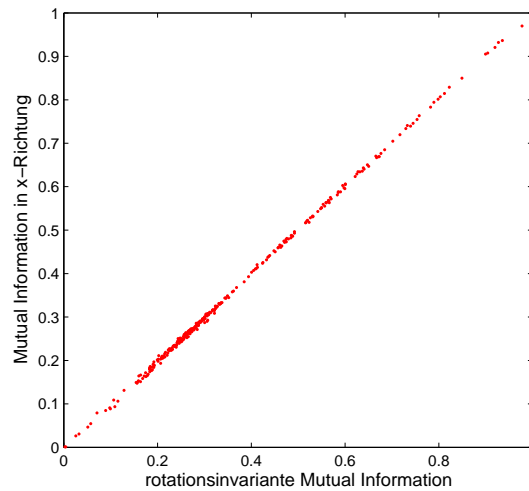


Abbildung 6.2: Die Mutual Information in x-Richtung ist hier gegen die rotationsinvariante Mutual Information aufgetragen. Man sieht, dass die Werte stark korreliert sind. Es genügt also vollkommen, die Mutual Information in nur einer Richtung zu betrachten.

	Grauwerte			Gradientenrichtungen
	$0.1\mu\text{m}$	$0.2\mu\text{m}$	$0.3\mu\text{m}$	$0.2\mu\text{m}$
MEF - MEF	69.2%	62.8%	61.5%	50%
HF - HF	62%	60%	58%	55%
MESai - MESai	53.4%	53.4%	55.1%	51.7%
MESar - MESar, MESaa	59.4%	62.5%	58.3%	48%
HF-G0 - HF-S	64%	62%	60%	62%
HF-S - HF-G0	50%	52%	52%	68%
MEF - HF	50%	50%	50%	53%
HF - MEF	50%	50%	50%	47.4%
MESai - MESaa	32%	38%	32%	36%
MESai - MESar	43.5%	45.7%	43.5%	47.8
HF,MEF - MES	48.7%	47.4%	49.4%	46.8%
HF,MEF - MESai	62.1%	53.4%	58.6%	55.2%
HF,MEF - MESaa	36%	44%	44%	36%
HF,MEF - MESar	45.7%	43.5%	43.5%	47.8%

Tabelle 6.2: Ergebnisse der Mutual Information der Grauwerte für $t=0.1\mu\text{m}$, $0.2\mu\text{m}$ und $0.3\mu\text{m}$ und der Gradientenrichtungen für $t=0.2\mu\text{m}$.

6.2 Co-occurrence Matrizen

Bei der Auswertung mehrdimensionaler Merkmale mit dem k-NN Klassifikator stellt sich zunächst die Frage, welche Norm man am Besten verwendet. Auf einer Teilmenge der Gesamtdaten bestehend aus 25 HF-S Zellen und 39 MEF-Zellen wurde für die Grauwert- und Gradientenrichtungs Co-occurrence Matrizen zunächst ausgewertet, ob die L1-Norm oder die L2-Norm besser geeignet ist. Hierzu wurden die Co-occurrence Matrizen zunächst nur auf einer Schicht jedes Datensatzes berechnet, auf der Schicht, in der der mittlere Grauwert maximal ist. Dann wurden alle aktiven und inaktiven CT-Aufnahmen gegeneinander klassifiziert, ohne dabei den Zelltyp zu berücksichtigen.

	Gradientenrichtung		Grauwert	
	Haralick Merkmale	ganze Matrix	Haralick Merkmale	ganze Matrix
L1	74.2%	55.4%	86.3%	79.8%
L2	73.4%	54.8%	84.6%	80.6%

Tabelle 6.3: Für eine Teilmenge der Daten (25 HF-S Zellen und 39 MEF Zellen) wurde ausgewertet, welche Norm für die Klassifikation besser geeignet ist. Die hier verwendeten Co-occurrence Matrizen sind nicht gewichtet, $t=0.2 \mu\text{m}$, Grauwerte und Gradientenrichtungen sind auf 32 Bins quantisiert.

Tabelle 6.3 zeigt, dass die Unterschiede zwischen L1- und L2-Norm nicht groß sind, in den betrachteten Fällen allerdings die L1-Norm etwas besser abschneidet. Im Folgenden wurde daher mit der L1-Norm gearbeitet.

6.2.1 Gray-Level Co-occurrence Matrizen

Die Wahl der Gewichtung

Eine der zentralen Fragestellungen ist, welche Gewichtung bei der Berechnung der Co-occurrence Matrizen am Besten die Zugehörigkeit der Voxel zum CT repräsentiert. Es wurden daher Co-occurrence Matrizen mit verschiedenen Gewichtungen berechnet. Vor der Auswertung wurden die Matrizen jeweils so normiert, dass sich die Einträge zu 1 aufsummieren lassen.

Der einfachste Ansatz ist hier die Gewichtung mit den Grauwerten der jeweiligen Bildpunkte. Tabelle 6.4 zeigt die Ergebnisse, die sich aus den Grauwert Co-occurrence Matrizen ergeben, wenn als Gewichtung Grauwerte (GWG-CM), die Wahrscheinlichkeiten für jeden Punkt zum CT zu gehören (GWP-CM), die daraus gebildeten Masken (GWM-CM) und die Gradientenlängen (GWGL-CM) verwendet werden.

	GWG-CM	GWP-CM	GWM-CM	GWGL-CM
MEF - MEF	82.05%	82.05%	80.8%	80.8%
HF - HF	98%	92%	91%	86%
MESai - MESai	57.9%	46.6%	44.8%	67.2%
MESar - MESar,MESaa	44.8%	49%	42.7%	35.4%
HF-G0 - HF-S	98%	94%	90%	96%
HF-S - HF-G0	98%	98%	100%	96%
MEF - HF	92%	70%	69%	69%
HF - MEF	76.9%	64.1%	64.1%	76.9%
MESai - MESaa	72%	0%	0%	64%
MESai - MESar	52.17%	50%	50%	47.8%
HF,MEF - MES	69.5%	54.5%	56.5	68.2%
HF,MEF - MESai	65.5%	55.2%	56.5%	56.9%
HF,MEF - MESaa	100%	60%	72%	100%
HF,MEF - MESar	50%	50%	54.3%	50%

Tabelle 6.4: Die Ergebnisse der nicht rotationsinvarianten Grauwert Co-occurrence Matrizen mit unterschiedlichen Gewichtungen bei $t=0.2\mu\text{m}$ und einer Quantisierung auf 32×32 Bins. Auch hier wurde der 10-NN Klassifikator verwendet.

Die hier erzielten Ergebnisse sind weitaus besser. Sie übersteigen für alle getesteten Gewich-

tungen die Ergebnisse, die mit dem mittleren Grauwert erzielt werden konnten. Für die Fibroblastenzellen führen die meisten Gewichtungen zu brauchbaren Ergebnissen, auch wenn Wahrscheinlichkeiten und Masken in den meisten Fällen etwas schlechter abschneiden als Gewichtung mit Grauwert oder Gradientenlänge. Bei den MES-Zellen versagt die Gewichtung mit Wahrscheinlichkeiten oder binärer Maske gänzlich. Dies könnte daran liegen, dass in diesen Zellen Xist-Sequenz und CTs in einem Kanal aufgenommen worden ist. Die Masken und Wahrscheinlichkeiten sind ja so bestimmt, dass jedem CT möglichst eine zusammenhängende Region zugeordnet werden kann. Die sehr hellen Punkte der Xist Sequenz des inaktiven X-Chromosoms allerdings gehören zu keinem der CT der Chromosomen 11 - wenn ihnen durch die Maske ein hohes Gewicht zugeordnet wird, verfälschen sie eventuell die Ergebnisse stark. Grauwertgewichtete Co-occurrence Matrizen führen insgesamt zu den besten Ergebnissen. Hier fallen auch eventuelle Störungen durch die Xist Sequenz weniger stark ins Gewicht, da diese nur an einzelnen Punkten sehr hell ist, die Nachbarpunkte recht dunkel sind und damit auch das Gesamtgewicht dieser Punkte in der Co-occurrence Matrix gering ist.

Mit grauwertgewichteten Co-occurrence Matrizen kann für die X-Chromosomen sehr genau unterschieden werden, ob sie aktiv oder inaktiv sind. Die Merkmale sind von Zellen in unterschiedlichen Phasen des Zellzyklus aufeinander übertragbar: für die HF-S und HF-G0 Zellen liegen die Klassifikationsraten bei allen Gewichtungen zwischen 90% und 100%. Beeindruckender ist die Tatsache, dass eine Übertragung von Mäusefibroblasten auf menschliche Fibroblastenzellen möglich ist. Trotz der Klassifikation mit dem sehr einfachen 10-NN Klassifikator liegen die Ergebnisse hier schon bei 92%. Eine umgekehrte Übertragung ist offenbar etwas schwieriger (76.9%), was wohl damit in Zusammenhang steht, dass sich die Chromosomen der Mäusefibroblasten schlechter trennen lassen. Innerhalb der Klasse der MEF-Zellen wurde ebenfalls nur eine Klassifikationsrate von 82% erreicht.

Die Übertragbarkeit vor Chromosom X auf Chromosom 11 ist eher fragwürdig. Bei der Klassifikation der Chromosomen der MESai-Zellen mit Hilfe der X-Chromosomen aus den Fibroblastenzellen liegen die Ergebnisse bei 65.5%. Die meisten der Chromosomen 11 wurden als aktiv klassifiziert. Dies könnte darin begründet liegen, dass sie grundsätzlich aktiver sind, als

MESai	Xa	Xi	MESaa	Xa	Xi	MESar	Xa	Xi
11a	29	0	11a	50	0	11a	23	0
11i	20	9				11r	23	0

Tabelle 6.5: Confusion Matrices zu Spalte 1, Zeilen 12, 13 und 14 aus Tabelle 6.4.

die X-Chromosomen - Chromosom 11 ist ja wesentlich genreicher.

Insgesamt fällt allerdings auf, dass die Ergebnisse, die für Chromosom 11 erzielt werden können, wesentlich schlechter sind, als für die X-Chromosomen. Selbst bei Vergleich der Merkmale der MESai Zellen untereinander wurden bestenfalls 67% der Chromosomen richtig klassifiziert. Dies gelang durch die Gewichtung mit der Gradientenlänge, eine Gewichtung die auf den Fibroblastendatensätzen zu eher schlechteren Ergebnissen führte.

Die Klassifikation der aktiven und reaktivierten Chromosomen der MESar-Zellen mit den MESai-Datensätzen zeigt Ergebnisse von um die 50%, die Klassifikationsraten der MESaa-Datensätze mit den MESar Daten ist ebenso zufällig. Dies lässt darauf schließen, dass die Textur der reaktivierten Chromosomen 11 mehr Ähnlichkeit zu den aktiven Chromosomen 11 aufweist als zu den inaktiven.

Isotropie der Textur

Im Folgenden wurde getestet, ob sich die Ergebnisse im Wesentlichen verbessern, wenn die Grauwert Co-occurrence Matrizen rotationsinvariant erstellt werden. Es wurden die grauwertgewichteten Matrizen berechnet, da diese im nicht rotationsinvarianten Fall die besten Ergebnisse lieferten. Zusätzlich ist hier getestet worden, bei welcher Distanz t zwischen den Punkten die charakteristischsten Informationen über die Textur zu finden sind. Die hier erziel-

	$t=0.1\mu\text{m}$	$t=0.2\mu\text{m}$	$t=0.3\mu\text{m}$
MEF - MEF	80.8%	80.8%	83.3%
HF - HF	98%	97%	97%
MESai - MESai	50%	50%	72.4%
MESar - MESar, MESaa	46.9%	50%	51%
HF-G0 - HF-S	98%	98%	98%
HF-S - HF-G0	98%	98%	98%
MEF - HF	93%	94%	97%
HF - MEF	79.5%	76.9%	82.1%
MESai - MESaa	50%	52%	76%
MESai - MESar	34.8%	30.4%	47.8%
HF,MEF - MES	66.2%	66.2%	71.4%
HF,MEF - MESai	50%	50%	63.8%
HF,MEF - MESaa	100%	100%	100%
HF,MEF - MESar	50%	50%	50%

Tabelle 6.6: Hier sind die Ergebnisse für rotationsinvariante grauwertgewichtete Co-occurrence Matrizen für $t=0.1\mu\text{m}$, $t=0.2\mu\text{m}$ und $t=0.3\mu\text{m}$ aufgeführt. Die Matrizen sind auch hier auf 32×32 Bins quantisiert.

ten Ergebnisse sind sehr ähnlich zu denen, die mit nicht rotationsinvarianten Matrizen erreicht werden können. Die Annahme, die Textur der CTs sei weitgehend isotrop, ist also korrekt. In Abbildung 6.3 sind die unterschiedlichen Klassifikationsraten für rotationsinvariante und nicht rotationsinvariante grauwertgewichtete Co-occurrence Matrizen mit $t=0.2\mu\text{m}$ graphisch dargestellt.

Im oberen Teil der Abbildung sieht man, dass Rotationsinvarianz das Ergebnis bei der Klassifikation der Fibroblastenzellen nur minimal beeinflusst. Die Klassifikationsergebnisse bei den MES-Zellen sind insgesamt nicht befriedigend und werden durch rotationsinvariante Co-occurrence Matrizen noch schlechter. Eine mögliche Erklärung hierfür wäre, dass MES-Zellen runder sind als Fibroblastenzellen, und damit grundsätzlich mehr Information in z -Richtung enthalten ist. Da diese allerdings durch die PSF stark gestört ist, werden die Informationen in den Co-occurrence Matrizen verwaschen.

Die Wahl der korrekten Translation t

Betrachtet man die Entwicklung der Klassifikationsrate in Abhängigkeit von der Translation t , so erkennt man, dass diese bei $t=0.3\mu\text{m}$ besser ausfällt, als bei kleineren Translationen.

Abbildung 6.4 verdeutlicht graphisch die Unterschiede in den Klassifikationsraten mit Grauwert-Co-occurrence Matrizen für die unterschiedlichen verwendeten Translationen. An den aufgeführten Beispielen wird klar, dass die Unterschiede bei den Fibroblastenzellen marginal sind. Bei den MES-Zellen sind die erzielten Ergebnisse mit $t=0.3\mu\text{m}$ deutlich besser als bei kleineren Translationen t .

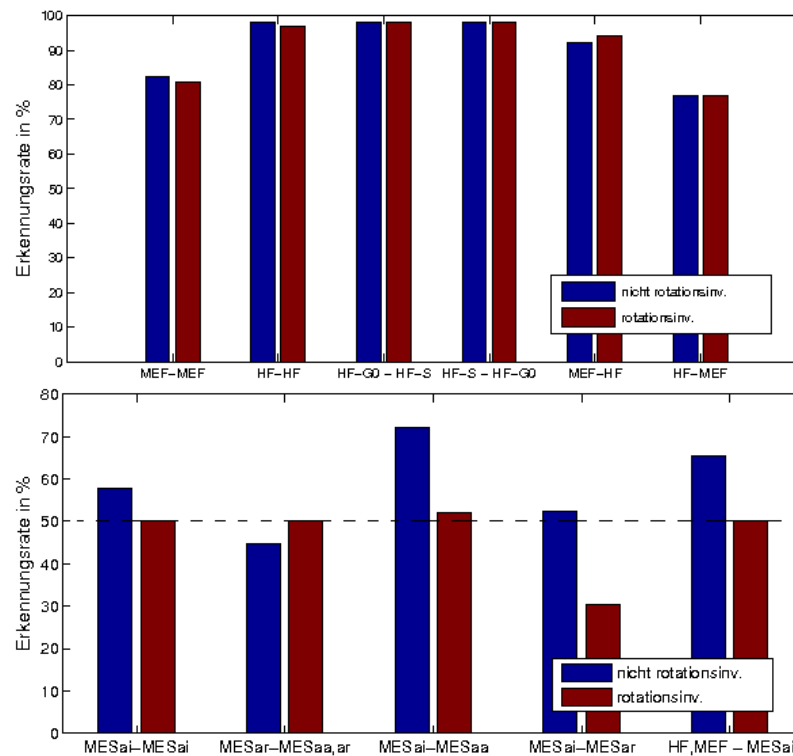


Abbildung 6.3: Die unterschiedlichen Klassifikationsergebnisse bei rotationsinvarianten und nicht rotationsinvarianten Grauwert Co-occurrence Matrizen. (oben) Die Ergebnisse auf Fibroblastenzellen, (unten) auf MES-Zellen.

Abbildung 6.5 zeigt die mittleren grauwertgewichteten Grauwert Co-occurrence Matrizen für X-Chromosomen aus HF-Zellen und MEF-Zellen sowie für das aktive und das künstlich inaktivierte Chromosom 11 aus MES-Zellen. Man sieht deutlich, dass die mittleren Co-occurrence Matrizen für das aktive Chromosom X von HF und MEF-Zellen sehr ähnlich sind (Abbildung 6.5(a) und (c)), genauso wie die Co-occurrence Matrizen inaktiver X-Chromosomen sehr ähnlich sind (Abbildung 6.5(b) und (d)). Die Grauwerte in CT-Aufnahmen aktiver X-Chromosomen sind offenbar stärker korreliert als die Grauwerte der inaktiven X-Chromosomen. Dies gilt besonders im hohen Grauwertbereich. Für Chromosom 11 trifft dies offenbar nicht zu; selbst im hohen Grauwertbereich sind hier die Werte der mittleren Co-occurrence Matrix noch stark um die Hauptachse konzentriert.

In Abbildung 6.6 sind die mittleren Grauwert Co-occurrence Matrizen für die aktiven und reaktivierten Chromosomen 11 aus den MESaa und MESar Datensätzen dargestellt. Die drei Matrizen sind sehr ähnlich. Dies lässt darauf schließen, dass die Textur der reaktivierten Chromosomen der der aktiven entspricht. Beim Vergleich der Matrizen aus Abbildung 6.6 mit der Matrix aus 6.5(e) fällt auf, dass letztere Matrix wesentlich weniger gestreute Werte enthält. Grundsätzlich allerdings sollten die Texturen der Datenmengen sehr ähnlich sein, handelt es sich doch in jedem Fall um aktive Chromosomen 11 in MES-Zellen.

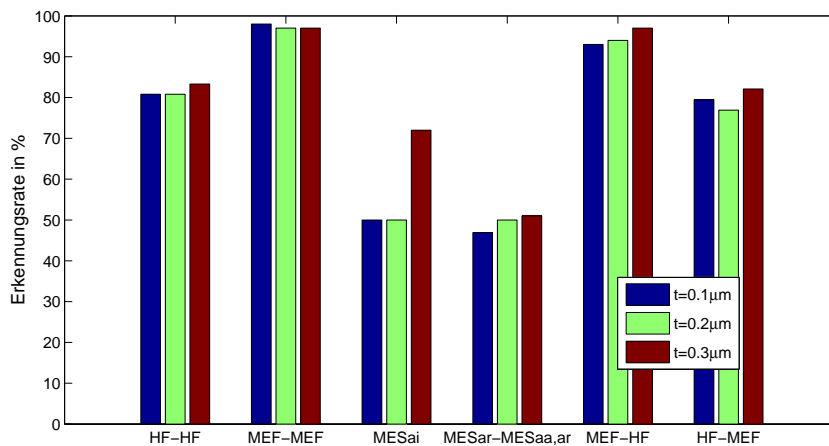


Abbildung 6.4: Unterschiede der Klassifikationsraten der Datenmengen mit grauwertgewichteten Grauwert- Co-occurrence Matrizen für $t=0.1\mu\text{m}$, $t=0.2\mu\text{m}$ und $t=0.3\mu\text{m}$.

Haralick Merkmale - Unabhängigkeit vom absoluten Grauwert

Schließlich wurden Haralick Merkmale aus den nicht rotationsinvarianten grauwertgewichteten Co-occurrence Matrizen ausgewertet. Auch hier hat sich gezeigt, dass die Merkmale in x -Richtung ähnlich wie die Mutual Information sehr stark mit denen in y -Richtung korreliert sind. Dies bestätigt erneut die Annahme der Isotropie der Textur. Auch bei der Klassifikation mit Haralick Merkmalen wurde der 10-NN Klassifikator verwendet. Die Ergebnisse dieser Auswertung sind in Tabelle 6.7 aufgeführt. Sie sind jeweils um wenige Prozent besser als die mithilfe der vollständigen Matrizen erzielten Ergebnisse. Dabei ist zu beachten, dass die Haralick Merkmale vollkommen unabhängig vom absoluten Grauwert der Aufnahme sind.

Bei Übertragung der Merkmale von Mäuse- auf menschliche Fibroblasten können hier 96% Genauigkeit erreicht werden, bei der Übertragung von Menschen- auf Mäusefibroblasten immerhin 82.1% bei $t=0.3\mu\text{m}$. Ebenso sind die Ergebnisse für Chromosom 11 hier besser, selbst wenn eine Übertragung von Chromosom X auf Chromosom 11 auch mit Haralick Merkmalen nicht gelingt. Eine Klassifikationsrate von mehr als 62.1% kann bei Klassifikation der CTs der MESai-Zellen auf Grundlage der Kenntnis der Chromosomen X aus den Fibroblastenzellen nicht erreicht werden.

Auswertung mittels Support Vector Maschine

Für die grauwertgewichteten nicht rotationsinvarianten Co-occurrence Matrizen ist schließlich eine Auswertung mit der Support Vektor Maschine durchgeführt worden. Die Ergebnisse sind hier erwartungsgemäß besser als unter Verwendung des einfachen 10-NN Klassifikators.

Grundsätzlich bleiben jedoch die gleichen Phänomene zu beobachten: Die Ergebnisse sind sehr gut im Falle der X-Chromosomen aus den Fibroblastenzellen. Eine Übertragung zwischen menschlichen Zellen und Mäusezellen kann sehr erfolgreich durchgeführt werden. Nach Training mit den Aufnahmen der MEF-Zellen können die menschlichen X-Chromosomen mit 98 prozentiger Genauigkeit klassifiziert werden. Für Chromosom 11 allerdings bleiben die Ergebnisse weiterhin weitgehend unbefriedigend. Immerhin kann eine Erkennungsrate von fast 70% bei der Klassifikation der Chromosomen 11 aus dem MESai-Zellen erreicht werden, wenn mit den X-Chromosomen aus den Fibroblastenzellen trainiert wird.

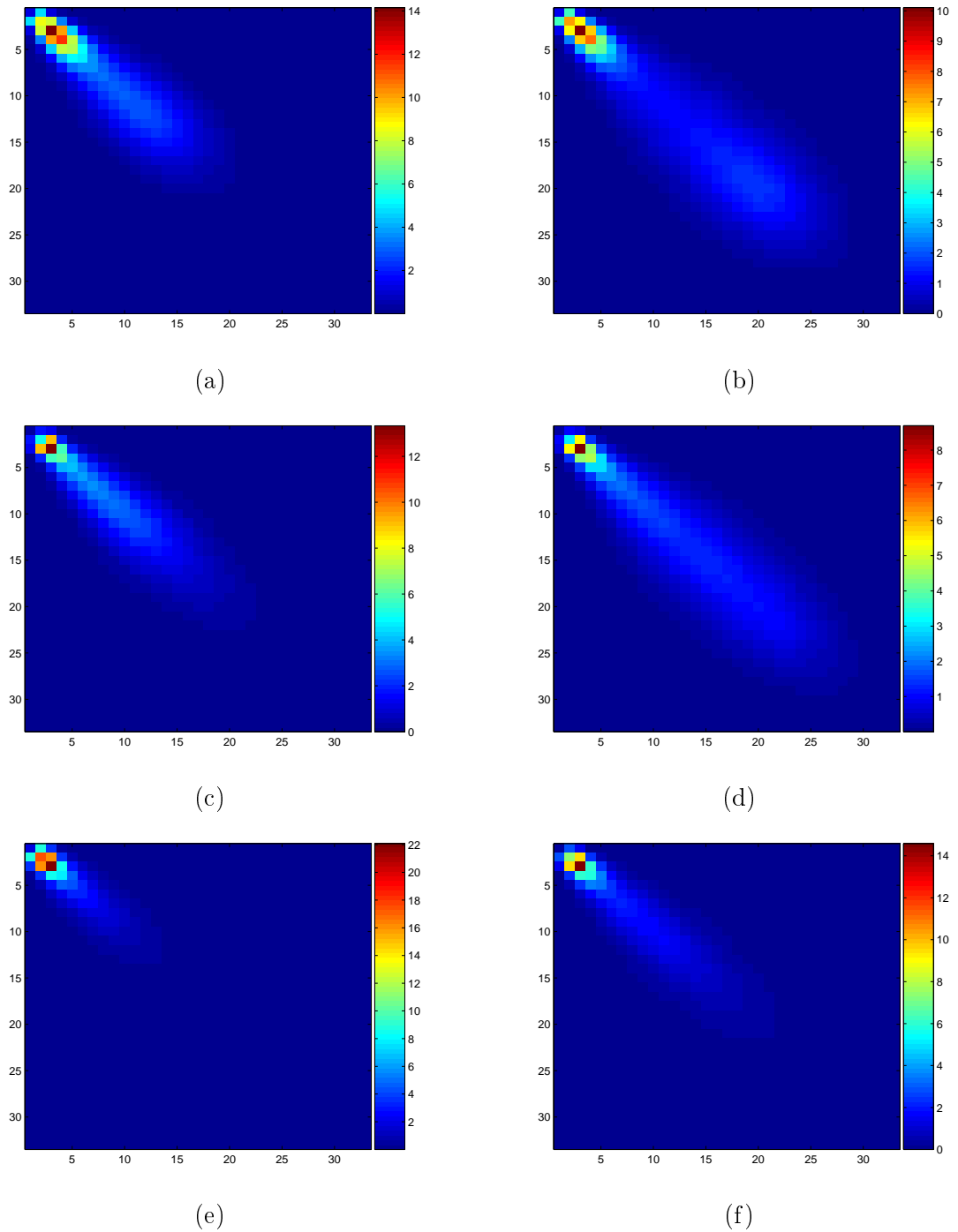


Abbildung 6.5: Mittlere grauwertgewichtete Grauwert-Co-occurrence Matrizen. (a) Aktive X-Chromosomen der HF-Zellen, (b) inaktive X-Chromosomen der HF-Zellen, (c) aktive X-Chromosomen der MEF-Zellen, (d) inaktive X-Chromosomen der MEF-Zellen, (e) aktive Chromosomen 11 der MES-Zellen (MESai) und (f) inaktivierte Chromosomen 11 der MES-Zellen.

6.2.2 Gradient-Direction Co-occurrence Matrizen

Die Co-occurrence Matrizen der relativen Gradientenrichtungen wurden für die Parameter berechnet, die sich auf den Grauwerten bewährt haben. Als Translation wurde folglich $t=0.3\mu\text{m}$

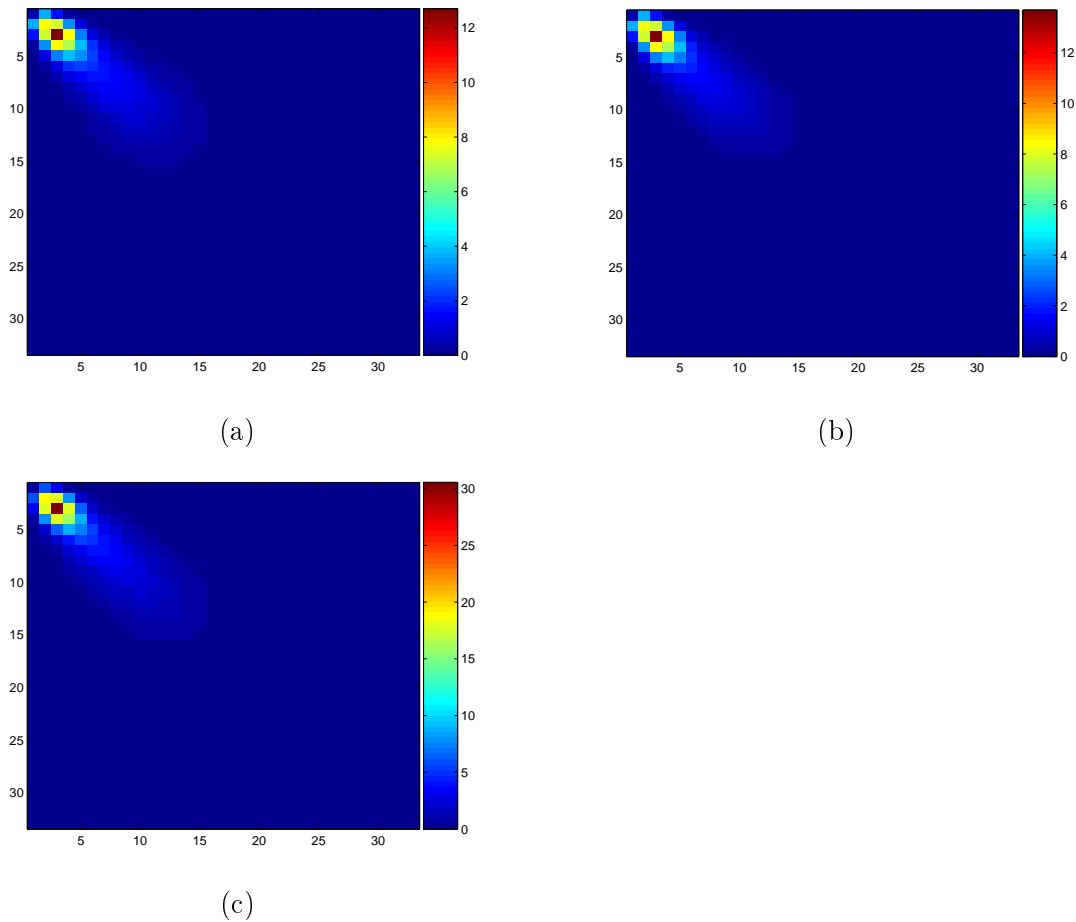


Abbildung 6.6: Mittlere grauwertgewichtete Grauwert-Co-occurrence Matrizen. (a) Der aktiven Chromosomen 11 der MESar-Zellen, (b) der reaktivierten Chromosomen 11 der MESar-Zellen und (c) der aktiven Chromosomen 11 der MESar-Zellen. Diese drei Co-occurrence Matrizen sind sehr ähnlich. Für die Matrizen in (a) und (c) war dies auch zu erwarten. Die Ähnlichkeit der Matrix aus (b) zu den Matrize aus (a) und (c) lässt darauf schließen, dass die Textur der reaktivierten Chromosomen der der aktiven entspricht.

gewählt, die Gewichtung wurde mit Grauwerten vorgenommen. Da der Wertebereich der Gradientenprojektionen in Richtung Zellkernwand allerdings zwischen -1 und $+1$ liegt und damit recht klein ist, wurde eine kleinere Quantisierung gewählt. Die Werte wurden auf 16×16 Bins quantisiert. Die erzielten Ergebnisse sind in Tabelle 6.9 aufgeführt. Sie bleiben in den meisten Fällen deutlich hinter den Ergebnissen zurück, die mit Co-occurrence Matrizen auf den Grauwerten erzielt wurden.

Gerade im Fall der X-Chromosomen aus den Fibroblastenzellen sind die Ergebnisse deutlich schlechter. Bei der Klassifikation der aktiven und inaktivierten Chromomen 11 sind hier allerdings die Ergebnisse etwas besser als bei der Auswertung der Grauwert Co-occurrence Matrizen. Hier können bei der Klassifikation der Chromosomen aus den MESai-Zellen Ergebnisse von fast 80% erreicht werden. Eine Übertragung der Merkmale von Chromosom X auf Chromosom 11 gelingt auch hier nicht. Offenbar lässt sich die Textur von Chromosom 11 besser durch die Gradientenrichtung beschreiben, als die von Chromosom X. Dies könnte im Wesentlichen daran liegen, dass die Aufnahmen von Chromosom 11 aus Stammzellen stammen. Diese sind rund wohingegen die Fibroblasten, in denen Chromosom X aufgenommen wurde, sehr flach sind. Da hier die Projektion der Gradientenrichtung auf die Gradienten in

	t=0.1 μ m	t=0.2 μ m	t=0.3 μ m
MEF - MEF	80.8%	80.8%	80.8%
HF - HF	96%	96%	97%
MESai - MESai	70.7%	70.7	72.4%
MESar - MESar,MESaa	46.9%	41.7%	41.7%
HF-G0 - HF-S	98%	98%	98%
HF-S - HF-G0	98%	98%	98%
MEF - HF	96%	96%	96%
HF - MEF	80.8%	80.8%	82.1%
MESai - MESaa	70%	72%	82%
MESai - MESar	34.8%	34.8%	45.6%
HF,MEF - MES	70.8%	70.8%	70.8%
HF,MEF - MESai	62.1%	62.1%	62.1%
HF,MEF - MESaa	100%	100%	100%
HF,MEF - MESar	50%	50%	50%

Tabelle 6.7: Auswertung der Haralick Merkmale aus den nicht rotationsinvarianten grauwertgewichteten Co-occurrence Matrizen mit dem 10-NN Klassifikator.

	SVM
HF-G0 - HF-S	98%
HF-S - HF-G0	98%
MEF - HF	98%
HF - MEF	83.3%
MESai - MESaa	68%
MESai - MESar	50%
HF,MEF - MES	69.5%
HF,MEF - MESai	58.6%
HF,MEF - MESaa	100%
HF,MEF - MESar	50%

Tabelle 6.8: SVM-Klassifikation der nicht rotationinvarianten grauwertgewichteten Co-occurrence Matrizen mit t=0.3 μ m. Verwendet wurde eine lineare Kernfunktion.

Richtung Zellkernwand betrachtet werden, spielt dieser Unterschied eine große Rolle. Dies lässt einen Vergleich der Projektionen der Gradientenrichtungen zwischen diesen beiden Zelltypen fragwürdig werden. Auch zwischen der Dicke der menschlichen Fibroblasten und der Mäusefibroblasten gibt es Unterschiede. Die menschlichen Zellen sind am flachsten, hier haben also die Gradienten in Richtung Zellkernwand klare Vorzugsrichtungen nach unten und nach oben. Auch die Schwierigkeiten der Übertragbarkeit der gradientenbasierten Merkmale zwischen Mäuse- und Menschenzellen könnten in der Form der Zellkerne begründet liegen. Es fällt auf, dass die Erkennungsraten insgesamt schlechter werden, wenn man nicht die gesamten Matrizen, sondern nur die daraus berechneten Haralick Merkmale betrachtet. Für die auf Grauwerten berechneten Matrizen war dies umgekehrt. Dort war es so, dass durch Berechnung der Haralick Merkmale eine Invarianz gegenüber dem absoluten Grauwert erreicht wurde; in den Co-occurrence Matrizen der Gradientenrichtungen liegt von vorne herein Invarianz gegenüber Grauwerttransformationen vor. Eine weitere Reduktion der Informationen ist hier also nicht nötig und sogar eher von Nachteil.

	GradWG-CM	GradWG-Haralick M.
MEF - MEF	74.4%	62.8%
HF - HF	82%	64%
MESai - MESai	79.3%	69%
MESar - MESar, MESaa	50%	38.5%
HF-G0 - HF-S	80%	56%
HF-S - HF-G0	74%	56%
MEF - HF	60%	51%
HF - MEF	73.1%	55.1%
MESai - MESaa	70%	48%
MESai - MESar	54.3%	56.5%
HF, MEF - MES	70.1%	60.4%
HF, MEF - MESai	60.3%	58.6%
HF, MEF - MESaa	100%	82%
HF, MEF - MESar	50%	39.1%

Tabelle 6.9: Erkennungsraten bei Klassifikation der CTs mit Co-occurrence Matrizen auf Gradientenrichtungen unter Verwendung des 10-NN Klassifikators. Die Matrixeinträge wurden vor der Normierung mit Grauwerten gewichtet. Die Matrizen wurden für die Translation $t=0.3\mu\text{m}$ ausgewertet.

6.2.3 Co-occurrence Matrizen zwischen Grauwert und Gradientenrichtung, Grauwert und Gradientenlänge

Im Folgenden wurde getestet, ob sich die Ergebnisse verbessern lassen, indem man Informationen aus den Grauwerten der Aufnahmen und den relativen Gradientenrichtungen verbindet. Es wurden hierzu Matrizen berechnet, die die Co-occurrence von Grauwerten und relativen Gradientenrichtungen einer Aufnahme messen (GradDG-CM). Ebenso wurden Matrizen betrachtet, in denen die Grauwerte einer Aufnahme gegen die Gradientenlänge an der entsprechenden Stelle aufgetragen wurden (GradLG-CM). Die Ergebnisse, die hiermit erreicht werden konnten, sind in Tabelle 6.10 aufgeführt. Es fällt zunächst auf, dass die Ergebnisse bei Betrachtung der Gradientenlänge in Kombination mit dem Grauwert zu wesentlich besseren Ergebnissen führt, als die Betrachtung der relativen Gradientenrichtung. Die Klassifikationsergebnisse innerhalb einer Datenmenge sind hier relativ gut. Bei den CTs der HF und der MEF-Zellen können sie zwar nicht mit denen der reinen Grauwert Co-occurrence Matrizen konkurrieren, für die MESai Datenmenge allerdings sind die hier erreichten 81% wesentlich besser, als die mittels Grauwert Co-occurrence Matrizen im besten Fall erreichten 72.4%.

6.2.4 Distanz-Grauwert und Distanz-Gradient-Direction Co-occurrence Matrizen

In Anlehnung an den Ansatz aus [7] sind Haralick Merkmale auf Co-occurrence Matrizen ausgewertet worden, die die für jeden Punkt seine Distanz zur Zellkernwand zum seinem Grauwert beziehungsweise zu seiner relativen Gradientenrichtung ins Verhältnis setzen. Die These, die in [7] belegt wird, besagt, dass das inaktivierte Chromosom eine sehr kompakte Form annimmt und sich an einer Position nah an der Zellkernwand aber nicht in Kontakt mit der Zellkernwand befindet. Das aktive Chromosom hingegen berührt mit den inaktiven Genen die Zellkernwand, ragt aber mit den Teilen, die transkribiert werden in den Kern hinein.

In Abbildung 6.7 wird der Ansatz beispielhaft dargestellt. Für das aktive und das inaktive X-Chromosom eines HF-S Datensatzes sind die Co-occurrence Matrizen zwischen Grauwert

	GradDG-CM	GradLG-CM
MEF - MEF	76.9%	83.3%
HF - HF	89%	90%
MESai - MESai	69%	81%
MESar - MESar,MESaa	45.8%	41.7%
HF-G0 - HF-S	74%	86%
HF-S - HF-G0	88%	90%
MEF - HF	70%	74%
HF - MEF	64.1%	73.1%
MESai - MESaa	56%	40%
MESai - MESar	43.5%	41.3%
HF,MEF - MES	65.6%	72.1%
HF,MEF - MESai	60.3%	69%
HF,MEF - MESaa	92%	90%
HF,MEF - MESar	43.5%	56.5%

Tabelle 6.10: Co-occurrence von Gradientenrichtung und Grauwert (GradDG) und von Gradientenlänge und Grauwert (GradLG), jeweils quantisiert auf 16×16 Bins. Auch hier wurde der 10-NN Klassifikator verwendet.

und Distanz zur Zellkernwand dargestellt. Die Grauwerte der Aufnahmen sind hier vor der Berechnung der Co-occurrence Matrizen so normiert worden, dass der minimale Grauwert 0 beträgt, der maximale 255. An den Matrizen sieht man, dass in der Tat hohe Grauwerte im aktiven X-Chromosom in unterschiedlichen Distanzen zur Kernwand auftreten, für das inaktive Chromosom sind die meisten Grauwerte in einer konstanten Distanz zur Zellkernwand zu finden. Die Ergebnisse, die insgesamt mit Haralick Merkmalen mit diesen Co-occurrence Matrizen erzielt werden konnten, sind in Tabelle 6.11 aufgeführt. Auch hier wurden verschiedene

	DGWG-CM	DGWP-CM	DGWM-CM
MEF - MEF	69.2%	64.1%	64.1%
HF - HF	55%	80%	84%
MESai - MESai	65.5%	58.6%	63.8%
MESar - MESar,MESaa	40.6%	38.5%	35.4%
HF-G0 - HF-S	52%	50%	50%
HF-S - HF-G0	68%	84%	92%
MEF - HF	50%	74%	74%
HF - MEF	51.3%	60.3%	62.8%
MESai - MESaa	58%	72%	70%
MESai - MESar	47.8%	52.2%	50%
HF,MEF - MES	68.2%	68.8%	70.1%
HF,MEF - MESai	63.8%	56.9%	62.1%
HF,MEF - MESaa	90%	98%	96%
HF,MEF - MESar	50%	52.2%	56.5%

Tabelle 6.11: Haralick Merkmale zur Co-occurrence von Distanz zur Zellkernwand und Grauwert. Die Merkmale sind für die Gewichtung mit Grauwert (DGWG-CM), die Gewichtung mit Wahrscheinlichkeiten (DGWP-CM) und Ausmaskieren des Hintergrunds (DGWM-CM) berechnet worden.

Gewichtungen getestet. Die Maskierung der Chromosomen scheint hier am besten geeignet.

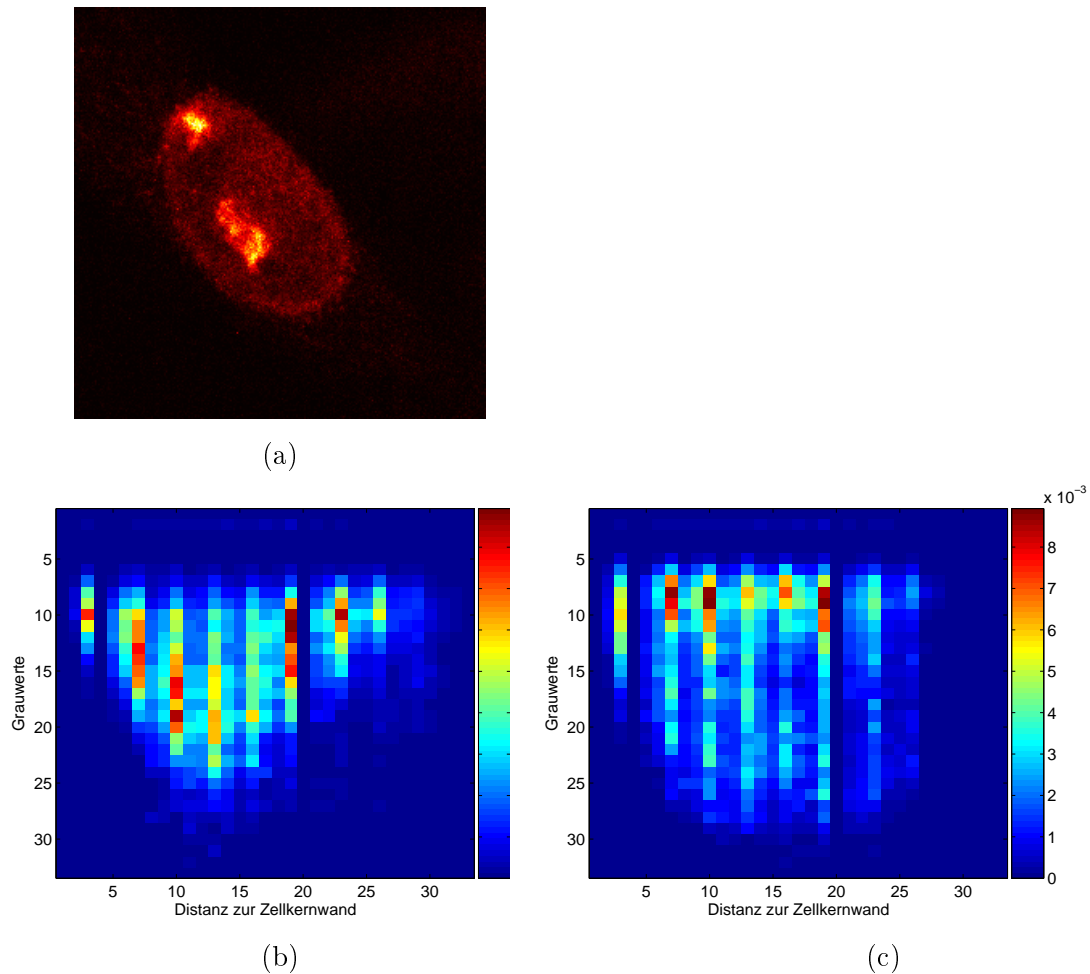


Abbildung 6.7: Die Abbildung zeigt (a) eine Schicht aus dem FITC-Kanal einer HF-Zelle. Man erkennt am oberen Ende des Kerns das kompaktere, inaktive X-Chromosom, etwa im Zentrum das wesentlich weniger kompakte, aktive X-Chromosom. (b) Die Co-occurrence Matrix des aktiven X-Chromosoms, in der Grauwert gegen Distanz zur Zellkernwand aufgetragen ist, (c) die Co-occurrence Matrix des inaktiven X-Chromosoms. Die vertikalen Streifen, die in den Matrizen sichtbar sind rühren vom Binning der Distanzwerte. Dies hat keine negativen Effekte, da nicht die Matrizen selbst sondern die aus ihnen berechneten Haralick Merkmale zur Klassifikation verwendet wurden.

Leider bleiben die Ergebnisse allerdings auch hier in jeder Hinsicht hinter den mit reinen Grauwert Co-occurrence Matrizen erzielten Ergebnissen zurück. Bei der Übertragung von X-Chromosomen menschlicher Fibroblasten auf Chromosomen der Maus können keine guten Klassifikationsergebnisse erzielt werden, die Übertragung von Chromosom 11 auf Chromosom X gelingt ebenfalls nicht. Innerhalb eines Zelltyps immerhin scheint man aufgrund der Co-occurrence Matrizen eine Aussage über die Aktivität eines CTs treffen zu können.

Tabelle 6.12 zeigt die Ergebnisse für die Co-occurrence von relativer Gradientenrichtung und Distanz zur Zellkernwand. Auch hier sind die Ergebnisse für keine der Gewichtungen überzeugend, wenn auch in manchen Fällen geringfügig besser als die Ergebnisse aus Grauwert und Distanz zur Zellkernwand.

Um mehr Informationen in den Co-occurrence Matrizen abzubilden wurde im Folgenden eine dreidimensionale Form der Co-occurrence Matrizen generiert und getestet. Es wurden relati-

	DGradWG-CM	DGradWP-CM	DGradWM-CM
MEF - MEF	53.8%	71.8%	73.1%
HF - HF	63%	78%	78%
MESai - MESai	56.9%	51.7%	51.7%
MESar - MESar,MESaa	52.1%	59.4%	62.5%
HF-G0 - HF-S	66%	78%	80%
HF-S - HF-G0	68%	74%	74%
MEF - HF	51%	58%	58%
HF - MEF	50%	61.5%	60.3%
MESai - MESaa	22%	40%	44%
MESai - MESar	45.7%	67.4%	69.6%
HF,MEF - MES	66.9%	60.4%	61.7%
HF,MEF - MESai	51.7%	55.2%	55.2%
HF,MEF - MESaa	100%	78%	84%
HF,MEF - MESar	50%	47.8%	45.7%

Tabelle 6.12: Haralick Merkmale zur Co-occurrence von Distanz zur Zellkernwand und Gradientenrichtung.

ve Gradientenrichtung, Gradientenlänge und Distanz zur Zellkernwand in eine mit $10 \times 10 \times 10$ Bins quantisierte Co-occurrence Matrix eingetragen. Für die Gewichtung mit Grauwerten und einfache Ausmaskierung des Hintergrunds sind die erzielten Ergebnisse in Tabelle 6.13 aufgelistet. Bei Verwendung von Masken können für die menschlichen X-Chromosomen relativ gute

	GGD-WG	GGD-M
MEF - MEF	61%	62.8%
HF - HF	77%	82%
MESai - MESai	72.4%	69%
MESar - MESar,MESaa	63.5%	44.8%
HF-G0 - HF-S	68%	92%
HF-S - HF-G0	78%	84%
MEF - HF	59%	76%
HF - MEF	62.8%	78.2%
MESai - MESaa	24%	52%
MESai - MESar	54.3%	45.7%
HF,MEF - MES	66.9%	58.4%
HF,MEF - MESai	63.8%	67.2
HF,MEF - MESaa	94%	46%
HF,MEF - MESar	41.3%	60.9%

Tabelle 6.13: Klassifikationsergebnisse der dreidimensionalen Co-occurrence Matrizen von Gradientenrichtung, Gradientenlänge und Distanz zur Zellkernwand. Die Werte wurden auf $10 \times 10 \times 10$ Bins quantisiert.

Ergebnisse erreicht werden. Die Übertragbarkeit der Merkmale von HF-G0 auf HF-S Zellen und umgekehrt ist gegeben, allerdings können die Merkmale nicht von MEF- auf HF-Zellen übertragen werden, was wohl in der unterschiedlichen Form der Zellkerntypen begründet liegt. Die Klassifikationsergebnisse auf Chromosm 11 sind ebenfalls nicht überzeugend, ein Übertragung der Merkmale von Chromosom X auf Chromosom 11 ist nicht möglich.

6.3 MDWRE

Die MDWRE-basierten Merkmale wurden sowohl auf Grauwertbildern als auch auf den Gradientenrichtungen berechnet. Für jeden Datensatz wurden die Merkmale ein- zwei- und dreidimensional erstellt und ausgewertet.

6.3.1 MDWRE-basierte Merkmale auf Grauwertbildern

Die erzielten Ergebnisse sind in Tabelle 6.14 dargestellt. Man sieht, dass die ein- und zweidi-

	1D	2D	3D
MEF - MEF	78.2%	83.33%	55.1%
HF - HF	76%	91%	66%
MESai - MESai	69%	75.9%	43.1%
MESar - MESar, MESaa	57.3%	56.3%	67.1%
HF-G0 - HF-S	80%	90%	52%
HF-S - HF-G0	80%	86%	54%
MEF - HF	80%	91%	56%
HF - MEF	79.5%	82%	47.4%
MESai - MESaa	28%	58%	62%
MESai - MESar	60.9%	54.3%	50%
HF, MEF - MES	59.1%	68.2%	42.9%
HF, MEF - MESai	63.8%	67.2%	41.4%
HF, MEF - MESaa	62%	88%	34%
HF, MEF - MESar	50%	47.8%	54.3%

Tabelle 6.14: Extrema-Merkmale auf maskierten Aufnahmen.

mensionale Auswertung deutlich bessere Ergebnisse liefert als die dreidimensionale. Die Ergebnisse der zweidimensionalen Auswertung sind für die X-Chromosomen nur wenig schlechter als die Ergebnisse der Grauwert Co-occurrence Matrizen, auf Chromosom 11 sind sie sogar etwas besser. Abbildung 6.8 zeigt graphisch die unterschiedlichen Erkennungsraten, die jeweils mit den beiden Verfahren erreicht werden konnten. Bei der Übertragbarkeit von Mäusefibroblastenzellen auf menschliche Fibroblasten sind die MDWRE-basierten Merkmale den Co-occurrence Matrizen offenbar unterlegen, erzielen aber immer noch sehr gute Ergebnisse. Die Übertragbarkeit von Chromosom X auf Chromosom 11 ist bei den MDWRE-basierten Merkmalen besser.

Die Ergebnisse, die mittels Support Vektor Maschine auf den MDWRE-basierten Merkmalen erzielt werden konnten, zeigen hier ähnliche Phänomene. Sie sind in Tabelle 6.15 aufgelistet. Die Ergebnisse auf den X-Chromosomen sind sehr gut, wenn auch etwas schlechter als die Ergebnisse der Co-occurrence Matrizen. Bei der Übertragbarkeit auf Chromosom 11 allerdings sind die MDWRE-basierten Merkmale etwas besser.

Dass die Ergebnisse bei der Klassifikation der Chromosomen 11 aus den MESaa und MESar Datensätzen mit den MESai Datensätzen relativ schlecht funktioniert, könnte an einem Unterschied in der Aufnahme liegen: Bei den MESaa und MESar Datensätzen wurden die Chromosomen und die Xist-Sequenz im gleichen Kanal aufgenommen. In der Aufnahme der Xist-Sequenz sind meist neben dem tatsächlichen Maximum an der Stelle, an der sich diese Sequenz befindet, sehr viele Nebenmaxima zu sehen. Dass diese gerade die MDWRE-basierten Merkmale, die ja auf der Berechnung der lokalen Extrema beruhen, besonders stören, ist klar.

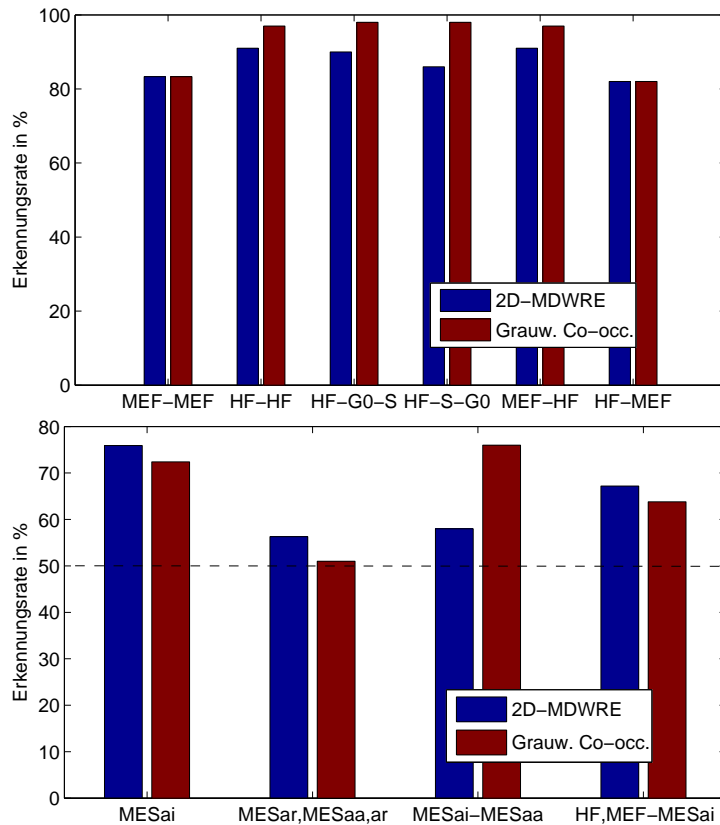


Abbildung 6.8: Vergleich der Erkennungsraten, die mit den zweidimensionalen MDWRE-basierten Merkmalen erzielt wurden, mit denen der grauwertgewichteten Grauwert Co-occurrence Matrizen. Die obere Graphik zeigt die Unterschiede in der Klassifikation der X-Chromosomen. Die MDWRE-basierten Merkmale schneiden jeweils etwas schlechter ab. Die Untere Graphik zeigt die Ergebnisse auf Chromosom 11 beziehungsweise bei der Übertragung von Chromosom X auf Chromosom 11. Innerhalb der MESai-Zellen schneidet der MDWRE besser ab als die Co-occurrence Matrizen, und auch bei der Übertragung von Chromosom X auf Chromosom 11 ist der MDWRE etwas besser. Bei der Klassifikation der Chromosomen 11 aus den MESaa-Zellen allerdings ist die Klassifikation mit den Grauwert Co-occurrence Matrizen besser.

	2D SVM
HF-G0 - HF-S	94%
HF-S - HF-G0	84%
MEF - HF	91%
HF - MEF	87.2%
MESai - MESaa	72%
MESai - MESar	37%
HF,MEF - MES	68.9%
HF,MEF - MESai	67.2%
HF,MEF - MESaa	88%
HF,MEF - MESar	50%

Tabelle 6.15: SVM-Klassifikation der MDWRE-basierten-Merkmale auf maskierten Grauwert-Aufnahmen. Verwendet wurde eine RBF-Kernel Funktion mit $\gamma = 0.001$.

6.3.2 MDWRE-basierte Merkmale auf Gradientenrichtungen

Tabelle 6.16 zeigt die Ergebnisse, die mit den MDWRE-basierten Merkmalen auf den Projektionen der Gradienten in Richtung Zellkernwand erzielt werden konnten. Auch hier ist die zweidimensionale Auswertung besser als die ein- und dreidimensionale. Allerdings sind die Ergebnisse in den meisten Fällen denen, die auf Grauwertbildern erzielt werden konnten, auch hier unterlegen. Grundsätzlich fällt auch hier auf, dass dies in allen Fällen für Chromosom X

	1D	2D	3D
MEF - MEF	78.2%	69.2%	62.8%
HF - HF	76%	81%	76%
MESai - MESai	69%	75.9%	43.1%
MESar - MESar,MESaa	57.3%	72.9%	67.7%
HF-G0 - HF-S	80%	78%	76%
HF-S - HF-G0	84%	78%	88%
MEF - HF	80%	75%	56%
HF - MEF	79.5%	60.3%	62.8%
MESai - MESaa	28%	38%	62%
MESai - MESar	60.9%	32.6%	50%
HF,MEF - MES	59.1%	67.5%	42.9%
HF,MEF - MESai	63.8%	72.4%	41.4%
HF,MEF - MESaa	62%	82%	34%
HF,MEF - MESar	50%	45.7%	54.3%

Tabelle 6.16: Extrema-Merkmale auf maskierten Gradientenrichtungen.

gilt, die Ergebnisse für Chromosom 11 allerdings oft auf den relativen Gradientenrichtungen genauso gut wenn nicht sogar etwas besser ausfallen als auf Grauwerten. Allerdings sind auch auf Gradientenrichtungen die Ergebnisse für Chromosom 11 meist nicht besonders gut. Immerhin liegt hier die Erkennungsrate bei Übertragung von Chromosom X auf Chromosom 11 aus den MESai-Zellen bei 72.4%. Bei der Klassifikation der Chromosomen 11 aus den MESaa Datensätzen mit den aktiven und reaktivierten Chromosomen aus den MESar Datensätzen kann hier eine Klassifikationsrate von 72% erreicht werden. Daraus lässt sich schließen, dass sich mithilfe der MDWRE-basierten Merkmale ein Unterschied in der Textur aktiver und reaktiver Chromosomen feststellen lässt. Die Texturen sind also nicht in jeder Hinsicht gleich.

6.4 Principle Component Analysis

Als Letztes wurde getestet, wie aussagekräftig der Winkel der Hauptkomponenten auf den betrachteten Datensätzen ist. Dazu wurden für die Principle Components der Fibroblastenzellen die Winkel zur optischen Ebene bestimmt. Für die MES-Zellen war diese Aswertung schwer möglich, da dort die Hauptebene des Zellkerns nicht eindeutig definiert ist; die Stammzellen sind ja eher rund. Die Ergebnisse, die für die Fibroblastenzellen erzielt wurden, sind in Tabelle 6.17 dargestellt. Die Ergebnisse zeigen, dass man aus der Ausrichtung der CTs zur optischen

	PCA
MEF - MEF	64%
HF - HF	70%
HF-G0 - HF-S	67%
HF-S - HF-G0	64%
MEF - HF	65%
HF - MEF	63.5%

Tabelle 6.17: Ergebnisse der Hauptachsenwinkel auf den Fibroblasten.

Ebene Informationen darüber gewinnen kann, wie aktiv das Chromosom ist. Eine genauere Approximation der Form der CTs würde die Merkmale sicher merklich verbessern. Gerade die Form der dekondensierten, aktiven Chromosomen lässt sich nicht immer brauchbar durch Ellipsoide annähern. Eine detailliertere Beschreibung der Form ließe sich zum Beispiel mittels Spherical Harmonic Descriptors erstellen. Bisläng allerdings ist die Ausrichtung der CTs als Kriterium bei der Unterscheidung zwischen aktiven und inaktiven Chromosomen nicht so sicher wie der Kondensierungsgrad des Chromatin. Die Merkmale, die aus der Textur gewonnen werden können, sind wesentlich zuverlässiger.

Kapitel 7

Schluss

Im Rahmen dieser Arbeit konnte gezeigt werden, dass sich aufgrund der Textur dekondensierten Chromatins gute Merkmale für den Grad der Aktivität von Chromosomen erzeugen lassen. Texturmerkmale auf der Basis von Grauwert Co-occurrence Matrizen zeigten sich als sehr zuverlässige Merkmale, ebenso wie sich MDWRE-basierte Merkmale sehr erfolgreich zur Beschreibung der betrachteten Texturen einsetzen ließen. Die Arbeit mit relativen Gradientenrichtungen hat sich aufgrund der zelltypspezifischen Zellkernform als weniger zuverlässig erwiesen. Mittels Haralick Merkmalen konnte jedoch gezeigt werden, dass auch im Fall der Grauwert Co-occurrence Matrizen die Textur selbst und nicht etwa der absolute Grauwert die entscheidenden Informationen über den Kondensierungsgrad des Chromatin enthält.

Es hat sich gezeigt, dass die vorgestellten MDWRE-basierten Merkmale genauso wie Grauwert Co-occurrence Matrizen wesentlich bessere Ergebnisse erzielen, als es bislang mithilfe von Merkmalen möglich ist, die auf der CT-Form beruhen.

Im Folgenden bleiben noch einige Aufgaben zu bewältigen. Eine detailliertere Beschreibung der Form der CTs ist wünschenswert, nicht nur um konkret die hier erzielten Klassifikationsergebnisse zu verbessern, sondern um insgesamt die Topologie des Zellkerns besser beschreiben zu können.

Auch ist es bislang noch nicht gelungen Merkmale zu finden, die sich ähnlich robust von Chromosom X auf Chromosom 11 übertragen lassen, wie es bei der Übertragung vom Chromosom X der Maus auf das des Menschen gelingt.

Die Arbeitsgruppe von Prof. Cremer am Biozentrum der LMU München hat in Aussicht gestellt, zusätzliche Datenmengen zu erstellen, damit gerade die Form der CTs auch anderer Chromosomen und anderer Zelltypen genauer untersucht werden kann. So lässt sich hoffentlich mit größerer Sicherheit entscheiden, welche der beobachteten Phänomene auf die Struktur der untersuchten Chromosomen selbst und welche auf die Eigenschaften des Zellkerns zurückzuführen sind. Auch wie gut die in der vorliegenden Arbeit generierten Merkmale von einem Chromosom aufs andere übertragen werden können, könnte mit zusätzlichen Datenmengen besser untersucht werden. Interessant wäre hier ein Vergleich mit einem Chromosom, dessen Gen-Gehalt zwischen dem des Chromosom X und des Chromosom 11 der Maus liegt.

Literaturverzeichnis

- [1] Aksoy, S.; Haralick, R.M.: Using Texture in Image Similarity and Retrieval. In: Pietikainen, M.; Bunke, H.(Hrsg.): Texture Analysis and Machine Vision, World Scientific, 2000.
- [2] Burkhardt, H.: Grundlagen der Bilderzeugung und Bildanalyse. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [3] Burkhardt, H.: Algorithmen zur Digitalen Bildverarbeitung. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [4] Burkhardt, H.: Bildverarbeitungspraktikum I. Veranstaltung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [5] Burkhardt, H.: Grundlagen der Bilderzeugung und Bildanalyse. Vorlesung am Institut für Informatik. Albert-Ludwigs-Universität Freiburg.
- [6] Bronstein, I.N.; Semendjajew, K.A.: Taschenbuch der Mathematik, 19., völlig überarbeitete Auflage, Hrsg: Grosche, G.; Ziegler, V., Verlag Harri Deutsch, Thun und Frankfurt/Main, 1980.
- [7] Casas Delucchi, C. S.: Structural and Positional Changes Accompanying Xist-Mediated Whole Chromosome Inactivation. Diplomarbeit, Department II, Anthropologie und Humangenetik der Ludwig-Maximilians-Universität München. München, 2006.
- [8] Cremer, M.; Küpper, K.; Wagler, B.; Wizelman, L.; Hase, J. v., Weiland, Y.; Kreja, L.; Diebold, J.; Speicher, M. R.; Cremer, T.: Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. In: The Journal of Cell Biology, 162/5, 2003. S.809-820.
- [9] Cremer, T.; Cremer, C.: Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells. In: Nature Reviews, Genetics, 2, 4/2001, S.292-301.
- [10] Cremer, T.; Kreth, K.; Koester, H.; Fink, R. H.; Heintzmann, R.; Cremer, M.; Solovei, I.; Zink, D.; Cremer, C.: Chromosome Territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. In: Critical Review in Eukaryotic Gene Expression, 10, 2/2000, S.179-212.
- [11] Corkidi, G.; Vega, L.; Márquez, J.; Rojas, E.; Ostrosky-Wegman, P.: Roughness feature of metaphase chromosome spreads and nuclei for automated cell proliferation analysis, In: Medical and biological engineering and computing, 36(6), 1998, S.679-85.
- [12] Haralick, R.M.; Shammugam, K.; Dinstein, I.: Textural Features for Image Classification. In: Transactions on System, Man, and Cybernetics, Vol. SMC 3,(6), 11/1973, S.610-621.

- [13] Kuijper, A: Mutual Information Aspects of Scale Space Images. In: Visualization, Imaging and Image Processing, 396, 2003.
- [14] Kurani, A.S.; Xu, D.-H.; Furst, J and Raicu, D.S.: Co-occurrence Matrices for Volumetric Data, In: Computer Graphics and Imaging,(426), 2004.
- [15] Linder Biologie, Lehrbuch für die Oberstufe, 20., neubearbeitete Auflage, Bayrhuber, H.; Kull, U.; Bäßler, U.; Danzer, A., Gesamtband. Schroedel Schulbuchverlag GmbH, Hannover, 1998.
- [16] Mayer, R.: Anordnung und Struktur von Chromosomenterritorien in Mauszellen: Zelltypspezifische Unterschiede und Gemeinsamkeiten, Dissertation an der Fakultät für Biologie, Ludwig-Maximilians-Universität München, 2/2006.
- [17] Ronneberger, O.; Fehr, J.; Burkhardt, H.: Voxel-Wise Gray Scale Invariants for Simultaneous Segmentation and Classification, Albert-Ludwigs-Universität Freiburg, Institut für Informatik, Lehrstuhl für Mustererkennung und Bildverarbeitung, Internal Report, April, 2005.
- [18] Rosenfeld, A.; Troy, E.: Visual Texture Analysis. Technical Report, University of Maryland, College Park, 1970, S.70-116.
- [19] Schael, M.: Invariant Texture Classification using Group Averaging with Relational Kernel Functions. In: Texture 2002 the 2nd international workshop on texture analysis and synthesis, S.129-134, 2002.
- [20] Schulz-Mirbach, H.: Invariant features for gray scale images. In G. Sagerer, S. Posch, and F. Kummert, Hrsg., 17. DAGM - Symposium "Mustererkennung", S. 1-14, Bielefeld, 1995. Reihe Informatik aktuell, Springer.
- [21] Siggelkow, Sven; Burkhardt, Hans: Improvement of Histogram-Based Image Retrieval and Classification. In: Proceedings of the International Conference on Pattern Recognition, Quebec, Canada, Vol.3, 9/2002, S. 367-370.
- [22] Tuceryan, M.; Jain, A.K.: Texture Analysis. In: Chen,C.H.; Pau,L.F.; Wang,P.S.P.(Grs.):The Handbook of Pattern Recognition and Computer Vision (2nd Edition), World Scientific Publishing Company, River Edge, NJ, 1998, S. 207-248.
- [23] Walker, R.F.; Jackway, P.T.; Longstaff, D.: Genetic Algorithm Optimisation of Adaptive Multi-Scale GLCM Features. In: International Journal of Pattern Recognition and Artificial Intelligence, World Scientific Publishing Company,10/2002.
- [24] Wilhelm, S.; Gröbler, B.; Gluch. M.; Heinz, H.: Confocal Laser Scanning Microscopy. Carl Zeiss Jena GmbH, 2003.
- [25] Wutz, A.; Jaenisch, R.:A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. In: Molecular Cell 5, 4/2000, S.605-705.
- [26] Gray's Anatomy of the human Body. 20. Auflage, erstveröffentlicht 1918, online edition. (<http://www.bartleby.com/107/>).
- [27] www.genome.gov/10000206.
- [28] <http://lmb.informatik.uni-freiburg.de/lmbsoft/libsvm/mtl/>.

Abbildungsverzeichnis

2.1	Zellkulturen der ES clone 36 Zellen und der MEF Zellen, aus [7]. In der vorliegenden Arbeit sind lediglich die Aufnahmen der Zellkulturen betrachtet worden, die grün unterlegt sind.	19
2.2	Schematische Darstellung der Fluoreszenz in situ Hybridisierung. Aus [27]. . .	20
2.3	Schematische Darstellung eines konfokalen Laserscanning Mikroskops. Aus [24].	21
2.4	(a) Die ideale PSF in xz-Ansicht. (b) Die ideale PSF in xy-Ansicht. Aus [24].	22
2.5	Jeweils eine Schicht aus dem (a) DAPI-Kanal, (b) dem FITC-Kanal und (c) dem Xist-Kanal der Aufnahme einer Mäusefibroblastenzelle. In der Xist-Aufnahme sieht man, dass alle Punkte mit hoher Intensität (Cluster oben rechts) klar einem der beiden CTs zugeordnet werden können. Dennoch gibt es viele Nebenmaxima, die eine automatische Zuordnung fragwürdig machen.	23
2.6	Die drei Einzelaufnahmen aus Abbildung 2.5 übereinandergelegt in unterschiedlichen Farbkanälen: <i>blau</i> entspricht dem DAPI-Kanal, <i>grün</i> dem FITC-Kanal und <i>rot</i> zeigt die Xist-Markierung.	24
2.7	Die manuell segmentierte Version des Beispiels aus Abbildung 2.5(b). (a) Das inaktive X-Chromosom, (b) Das aktive X-Chromosom.	24
3.1	(a) Eine Schicht aus dem DAPI-Kanal eines menschlichen Fibroblasten Zellkerns. (b) Die entsprechende Schicht im FITC-Kanal. Man kann erkennen, dass die Grauwerte im Bereich des Zellkerns höher sind als außerhalb. (c) Dieselbe Aufnahme nach γ -Korrektur mit $\gamma=0.5$. So ist der Helligkeitsunterschied zwischen Zellkern und restlichem Hintergrund deutlicher zu erkennen.	27
3.2	(a) Die Aufnahme aus Abbildung 3.1, hier als Grauwertdarstellung, (b) die Gradientenlängen dieser Aufnahme und (c) die Richtungen der Gradienten der Aufnahme.	28
3.3	Jeweils eine Schicht aus dem DAPI-Kanal zweier Zellkerne - links ein Kern einer menschlichen, rechts einer Mäusefibroblastenzelle - und die Kurve, die sich bei Auswertung von Gleichung 3.2 auf den Zellkern ergibt. Der Kurvenverlauf in der menschlichen Zelle ist wesentlich breiter als in der Mäusezelle. Dies lässt darauf schließen, dass die Kontur sich bei Schwellwerten um den Wert 40 herum nur wenig verändert. Das Maximum erreicht die Kurve beim Schwellwert 35. In der Mäusezelle liegt der beste Schwellwert bei 10.	29
3.4	Die entstandene Maske kann genutzt werden, um im FITC-Kanal den Hintergrund auszublenden. Übrig bleiben die fluoreszenten Chromosomen und der Zellkernhintergrund. (a) zeigt die maskierte Aufnahme aus Abbildung 3.1(b). (b) Zur besseren Kenntlichmachung ist die Aufnahme hier nach γ -Korrektur mit $\gamma=0.3$ abgebildet.	30

3.5	(a) Der Gegenfärbungskanal eines Mäusezellkerns. Es sind deutlich Nukleoli (dunkle Punkte) und Chromozentren (helle Punkte) zu sehen. (b) Der Gegenfärbungskanal des Mäusezellkerns nach der <i>fill</i> -Operation. Die Nukleoli sind nun aufgefüllt.	30
3.6	(a) Hier ist die Kontur eingezeichnet, an der das Verfahren den Kern ausschneidet. (b) Eine Seitenansicht der Aufnahme. Hier ist zu erkennen, wie Nukleoli und Chromozentren die Form des Nukleus verfälschen.	31
3.7	Die korrigierte Schnittlinie: (a) Frontalansicht (b) Seitenansicht.	32
3.8	(a) Gradientenlängen zur Schicht aus 3.1(b). Man sieht, dass die Gradienten um die CTs herum deutlich länger sind als im restlichen Zellkern. Innerhalb eines CTs allerdings schwanken die Gradientenlängen stark. (b) Die Gradientenlängen nach <i>fill</i> -Operation. Nun sind die Gradientenlängen innerhalb der CTs insgesamt etwas höher als im restlichen Zellkern.	33
3.9	Binarisierung der Aufnahme aus Abbildung 3.8 (b) mit den Schwellwerten (a) 0.4 (b) 0.3 (c) 0.2. Je niedriger der Schwellwert gewählt wird, desto größer wird die Maske, das heisst ein desto größerer Anteil des Zellkerns wird dem CT zugerechnet.	34
3.10	(a) Die Maske aus Abbildung 3.9(b) nach <i>Close</i> -Operation. Die Masken beider CTs haben nun keine Löcher mehr. (b) Die Maske aus (a) nach <i>Open</i> -Operation. Kleine Maskenfragmente, die nicht nahe genug an der restlichen Maske liegen, verschwinden hierdurch. (c) Die resultierende Kontur, eingezeichnet in die Aufnahme aus Abbildung 3.4 (a).	34
3.11	Schematische Darstellung der Bayes-Klassifikation aus [5].	35
3.12	(a) Eine Schicht des FITC-Kanals der MEF-Zelle, deren DAPI-Kanal in Abbildung 3.5(a) dargestellt ist; der Hintergrund ist hier ausgeblendet. (b) Das Histogramm dieser Aufnahme. Die Hintergrundhelligkeit des Zellkerns nimmt den größten Teil der nach Maskierung verbleibenden Aufnahme ein. Das Histogramm weist daher ein deutliches Maximum auf, das den häufigsten Hintergrundgrauwert markiert.	36
3.13	(a) Poissonverteilung mit $\lambda = 16$ (b) Das Histogramm aus 3.12 (b) nach Subtraktion der Poissonverteilung. Übrig bleiben die Grauwerte, die wahrscheinlich nur im Vordergrund auftreten. Das Histogramm gibt nun Aufschluss darüber, wie häufig welcher Grauwert im Vordergrund auftritt.	37
3.14	(a) Gleichung 3.3 ist hier für die sich aus dem Histogramm ergebenden Verteilungen dargestellt: blau: $p(x Hintergrund)P(Hintergrund)$, rot: $p(x CT)P(CT)$, (b) Um die sehr geringen Vordergrundwahrscheinlichkeiten sichtbar zu machen, ist hier eine Detailansicht von (a) dargestellt. (c) blau: $\frac{p(x Hintergrund)P(Hintergrund)}{p(x)}$, rot: $\frac{p(x CT)P(CT)}{p(x)}$	37
3.15	(a) Wahrscheinlichkeiten für die Zugehörigkeit jedes Pixels zur Klasse CT, (b) Maske nach MAP-Klassifikation (c) die Kontur der Maske aus (b) eingezeichnet in die Aufnahme aus 3.12(a).	38
3.16	Die mittels MAP-Klassifikation erstellte Maske ist hier als Kontur in den Datensatz aus Abbildung 3.4 eingezeichnet.	38
3.17	(a) Eine zweidimensionale Schicht aus einem Datensatz, (b) Gradientenrichtungen der zweidimensionalen Schicht.	39
3.18	(a) Der gaußgefilterte Datensatz aus Abbildung 3.17, (b) Gradientenrichtungen des gaußgefilterten Datensatzes. Die Gradientenrichtungen um die CTs herum sind jetzt stabiler; im Hintergrund zeigen sie immer noch in willkürliche Richtungen.	40

3.19	(a) Der Datensatz nach dreifacher Gaußfilterung und (b) die zugehörigen Gradientenrichtungen.	40
3.20	Die dem Beispiel aus Abbildung 3.19 (a) entsprechende Schicht der Projektion der dreidimensionalen Gradienten auf die Gradienten zur Zellkernwand. Die Werte sind hier auf das Intervall $[0, 1]$ normiert.	40
4.1	Relationale Kernfunktionen auf Beispielradien. Aus [19].	43
4.2	(a) Eine Schicht aus einem Element der Datenbank. (b)-(d) Merkmalsmatrizen, die sich bei Anwendung der RKF auf (a) ergeben. Der verwendete Schwellwert ϵ beträgt 1, die Kreise, auf denen die RKF ausgewertet wurden, sind: $p_1 = \{0, 0\}$, $p_1 = \{1, 0\}$ und $p_3 = \{2, 0\}$	44
4.3	(a) Schichtaufnahme des CT eines aktiven X-Chromosoms, (b) Autokorrelation dieses CT.	45
4.4	(a) Der rote Pfeil zeigt auf das ausgeschnittene lokale Maximum. (b) Das ausgeschnittene Fenster. (c) Das im Folgenden verwendete Template.	46
4.5	(a) Die Korrelation des Templates aus Abbildung 4.4 und dem aktiven CT - ihr Maximum liegt bei 0.78, (b) die Korrelation des Templates mit dem inaktiven CT - hier liegt das Maximum bei 0.82 und (c) die Korrelation mit einem Stück Hintergrundtextur, das Maximum beträgt 0.76.	46
4.6	(a) Ein Ausschnitt aus der Aufnahme eines CTs (b) Die Mutual Information dieses Ausschnitts bei den vertikalen Translationen um 1 bis 10 Pixel. Zur Berechnung der Mutual Information wurden fuzzy histograms verwendet (vergleiche Abschnitt 4.5.1).	48
4.7	Zu dem CT aus Abbildung 4.6 sind hier die Co-occurrence Matrizen für $t=0, 0.2$ und $0.4\mu\text{m}$ abgebildet. Für $t=0$ entspricht die Co-occurrence Matrix dem Histogramm. Je größer die Translation t ist, desto stärker sind die Werte gestreut.	50
4.8	Klassisches Histogramm (links) und Fuzzy Histogramm (rechts). Aus [21].	52
4.9	Klassische und Fuzzy Co-occurrence Matrix für den Datensatz aus Abbildung 4.7 mit $t=0.2\mu\text{m}$. Die 256 Bins in der klassischen Co-occurrence Matrix sind auf 32 Bins in der Fuzzy Matrix reduziert.	52
4.10	(a) Eine Grauwertschicht eines CTs und (b) das dazugehörige dreidimensionale Grauwertprofil.	54
4.11	(a) Das zweidimensionale Grauwertprofil einer Zeile der CT-Schicht aus Abbildung 4.10 (b) eine Detailansicht aus diesem Profil zur Illustration des MDWRE. V ist die vertikale Projektion von A , H die horizontale Projektion von A	55
4.12	In der Reliefdarstellung des in Abbildung 4.10(a) dargestellten Beispiels wurden ein lokales Maximum \mathbf{m}_i und das dazugehörige Minimum mit der geringsten Distanz $\mathbf{n}_{\mathbf{m}_i}$ eingezeichnet. Die Höhendifferenz zwischen den Extrema in der Reliefdarstellung entspricht der Grauwertdifferenz der Datensatzpunkte.	56
4.13	Eine dreidimensionale Darstellung eines aktiven X-Chromosoms mit seinen Principle Components und den Ellipsoiden, die sich aus ihnen ergeben.	58
6.1	Aufnahme eines aktiven CTs. Um nicht an Stelle der Textur die Größe des CTs auszuwerten sollten keine Translationen von mehr als $0.3\mu\text{m}$ verwendet werden.	64
6.2	Die Mutual Information in x-Richtung ist hier gegen die rotationsinvariante Mutual Information aufgetragen. Man sieht, dass die Werte stark korreliert sind. Es genügt also vollkommen, die Mutual Information in nur einer Richtung zu betrachten.	65

6.3	Die unterschiedlichen Klassifikationsergebnisse bei rotationsinvarianten und nicht rotationsinvarianten Grauwert Co-occurrence Matrizen. (oben) Die Ergebnisse auf Fibroblastenzellen, (unten) auf MES-Zellen.	69
6.4	Unterschiede der Klassifikationsraten der Datenmengen mit grauwertgewichteten Grauwert- Co-occurrence Matrizen für $t=0.1\mu\text{m}$, $t=0.2\mu\text{m}$ und $t=0.3\mu\text{m}$	70
6.5	Mittlere grauwertgewichtete Grauwert-Co-occurrence Matrizen. (a) Aktive X-Chromosomen der HF-Zellen, (b) inaktive X-Chromosomen der HF-Zellen, (c) aktive X-Chromosomen der MEF-Zellen, (d) inaktive X-Chromosomen der MEF-Zellen, (e) aktive Chromosomen 11 der MES-Zellen (MESai) und (f) inaktivierte Chromosomen 11 der MES-Zellen.	71
6.6	Mittlere grauwertgewichtete Grauwert-Co-occurrence Matrizen. (a) Der aktiven Chromosomen 11 der MESar-Zellen, (b) der reaktivierten Chromosomen 11 der MESar-Zellen und (c) der aktiven Chromosomen 11 der MESar-Zellen. Diese drei Co-occurrence Matrizen sind sehr ähnlich. Für die Matrizen in (a) und (c) war dies auch zu erwarten. Die Ähnlichkeit der Matrix aus (b) zu den Matrizen aus (a) und (c) lässt darauf schließen, dass die Textur der reaktivierten Chromosomen der der aktiven entspricht.	72
6.7	Die Abbildung zeigt (a) eine Schicht aus dem FITC-Kanal einer HF-Zelle. Man erkennt am oberen Ende des Kerns das kompaktere, inaktive X-Chromosom, etwa im Zentrum das wesentlich weniger kompakte, aktive X-Chromosom. (b) Die Co-occurrence Matrix des aktiven X-Chromosoms, in der Grauwert gegen Distanz zur Zellkernwand aufgetragen ist, (c) die Co-occurrence Matrix des inaktiven X-Chromosoms. Die vertikalen Streifen, die in den Matrizen sichtbar sind rühren vom Binning der Distanzwerte. Dies hat keine negativen Effekte, da nicht die Matrizen selbst sondern die aus ihnen berechneten Haralick Merkmale zur Klassifikation verwendet wurden.	76
6.8	Vergleich der Erkennungsraten, die mit den zweidimensionalen MDWRE-basierten Merkmalen erzielt wurden, mit denen der grauwertgewichteten Grauwert Co-occurrence Matrizen. Die obere Graphik zeigt die Unterschiede in der Klassifikation der X-Chromosomen. Die MDWRE-basierten Merkmale schneiden jeweils etwas schlechter ab. Die Untere Graphik zeigt die Ergebnisse auf Chromosom 11 beziehungsweise bei der Übertragung von Chromosom X auf Chromosom 11. Innerhalb der MESai-Zellen schneidet der MDWRE besser ab als die Co-occurrence Matrizen, und auch bei der Übertragung von Chromosom X auf Chromosom 11 ist der MDWRE etwas besser. Bei der Klassifikation der Chromosomen 11 aus den MESaa-Zellen allerdings ist die Klassifikation mit den Grauwert Co-occurrence Matrizen besser.	79
B.1	(I)Interphase, (II)Prophase, (III)Prometaphase, (IV)Metaphase, (V und IV)Anaphase, (VII)Telophase, (VIII)Cytokinese. Aus [26].	95
B.2	Ein Modell der funktionalen Zellkernarchitektur. Aus [9].	96
D.1	MIP des DAPI und FITC-Kanals der Fibroblasten.	99
D.2	MIP des DAPI und FITC-Kanals der MES Datensätze.	100

Tabellenverzeichnis

2.1	Chromosomenzahlen einiger Tier- und Pflanzenarten. Aus [15].	16
2.2	Übersicht über den Gengehalt der betrachteten Chromosomen. Chromosom 11 ist das genreichste Chromosom der Maus, Chromosom X eines der genärmsten. Für eine vollständige Auflistung des Genoms der Maus vergleiche [16]. Die Angaben über das menschliche Genom stammen aus der Gen-Datenbank von Ensembl, Stand: 01.2007.	17
2.3	Übersicht über die aufgenommenen Voxelgrößen in jeder der in Abschnitt 2.2 vorgestellten Datenmengen.	23
3.1	Standardabweichungen (im μm), die zum Korrigieren der Zellkernkontur in den verschiedenen Zelltypen verwendet wurden.	32
4.1	Verwendete Haralick-Merkmale. μ ist dabei der Mittelwert, σ die Standardabweichung der Co-occurrence Matrix.	53
4.2	Zwei- und dreidimensionale Merkmale als Erweiterung des MDWRE.	56
6.1	Klassifikationsergebnisse mit mittlerem Grauwert. Es wurde der 10-NN Klassifikator verwendet.	64
6.2	Ergebnisse der Mutual Information der Grauwerte für $t=0.1\mu\text{m}$, $0.2\mu\text{m}$ und $0.3\mu\text{m}$ und der Gradientenrichtungen für $t=0.2\mu\text{m}$	65
6.3	Für eine Teilmenge der Daten (25 HF-S Zellen und 39 MEF Zellen) wurde ausgewertet, welche Norm für die Klassifikation besser geeignet ist. Die hier verwendeten Co-occurrence Matrizen sind nicht gewichtet, $t=0.2\mu\text{m}$, Grauwerte und Gradientenrichtungen sind auf 32 Bins quantisiert.	66
6.4	Die Ergebnisse der nicht rotationsinvarianten Grauwert Co-occurrence Matrizen mit unterschiedlichen Gewichtungen bei $t=0.2\mu\text{m}$ und einer Quantisierung auf 32×32 Bins. Auch hier wurde der 10-NN Klassifikator verwendet.	66
6.5	Confusion Matrices zu Spalte 1, Zeilen 12, 13 und 14 aus Tabelle 6.4.	67
6.6	Hier sind die Ergebnisse für rotationsinvariante grauwertgewichtete Co-occurrence Matrizen für $t=0.1\mu\text{m}$, $t=0.2\mu\text{m}$ und $t=0.3\mu\text{m}$ aufgeführt. Die Matrizen sind auch hier auf 32×32 Bins quantisiert.	68
6.7	Auswertung der Haralick Merkmale aus den nicht rotationsinvarianten grauwertgewichteten Co-occurrence Matrizen mit dem 10-NN Klassifikator.	73
6.8	SVM-Klassifikation der nicht rotationinvarianten grauwertgewichteten Co-occurrence Matrizen mit $t=0.3\mu\text{m}$. Verwendet wurde eine lineare Kernfunktion.	73
6.9	Erkennungsraten bei Klassifikation der CTs mit Co-occurrence Matrizen auf Gradientenrichtungen unter Verwendung des 10-NN Klassifikators. Die Matrixeinträge wurden vor der Normierung mit Grauwerten gewichtet. Die Matrizen wurden für die Translation $t=0.3\mu\text{m}$ ausgewertet.	74

6.10	Co-occurrence von Gradientenrichtung und Grauwert (GradDG) und von Gradientenlänge und Grauwert (GradLG), jeweils quantisiert auf 16×16 Bins. Auch hier wurde der 10-NN Klassifikator verwendet.	75
6.11	Haralick Merkmale zur Co-occurrence von Distanz zur Zellkernwand und Grauwert. Die Merkmale sind für die Gewichtung mit Grauwert (DGWG-CM), die Gewichtung mit Wahrscheinlichkeiten (DGWP-CM) und Ausmaskieren des Hintergrunds (DGWM-CM) berechnet worden.	75
6.12	Haralick Merkmale zur Co-occurrence von Distanz zur Zellkernwand und Gradientenrichtung.	77
6.13	Klassifikationsergebnisse der dreidimensionalen Co-occurrence Matrizen von Gradientenrichtung, Gradientenlänge und Distanz zur Zellkernwand. Die Werte wurden auf $10 \times 10 \times 10$ Bins quantisiert.	77
6.14	Extrema-Merkmale auf maskierten Aufnahmen.	78
6.15	SVM-Klassifikation der MDWRE-basierten-Merkmale auf maskierten Grauwert-Aufnahmen. Verwendet wurde eine RBF-Kernel Funktion mit $\gamma = 0.001$	79
6.16	Extrema-Merkmale auf maskierten Gradientenrichtungen.	80
6.17	Ergebnisse der Hauptachsenwinkel auf den Fibroblasten.	81
C.1	Manuell und automatisch bestimmte Schwellwerte der MES Zellen in denen beide Chromosomen 11 aktiv sind, bzw. eines aktiv ist, das andere reaktiviert wurde.	97
C.2	Manuell und automatisch bestimmte Schwellwerte der MEF und der MES Zellen, in denen ein Chromosom 11 aktiv ist, eines inaktiviert wurde.	98

Anhang A

Verwendete Software

A.1 Matlab

Die angewendeten Methoden wurden in Matlab implementiert. Ebenso wurde Matlab zum Erstellen der Graphiken eingesetzt.

A.2 ImageJ

ImageJ ist ein Java basiertes Bildverarbeitungsprogramm. Es wurde in manchen Fällen dazu benutzt, die Masken der Zellkerne zu korrigieren, wenn mehrere Zellkerne in einer Aufnahme abgebildet waren.

A.3 NetCDF

Die verwendeten Daten wurden im NetCDF (network Common Data Form) Format gespeichert und verarbeitet. Dieses Datenformat ermöglicht es, alle Informationen zu einem Datensatz in einer Datei zu speichern.

A.4 libSVMTL

Zur Klassifikation mittels Support Vector Machine wurde die LibSVMTL verwendet - eine objektorientierte C++ Support Vector Machine Library, die auf der LibSVM von Chih-Jen Lin aufbaut. Sie wurde am Lehrstuhl für Mustererkennung und Bildverarbeitung an der Universität Freiburg entwickelt.

Anhang B

Biologische Grundlagen und Hintergrund

B.1 Zellteilung

Die Zellteilung aller Eukarioten, also aller Lebewesen, die einen deutlich vom Rest der Zelle abgegrenzten Zellkern haben, läuft prinzipiell auf die gleiche Art ab. Der Zellzyklus besteht grundsätzlich aus Interphase und Mitose. Die Mitose fasst hierbei die einzelnen Phasen der Zellkernteilung zusammen: Prophase, Prometaphase, Metaphase, Anaphase und Telophase. In der Prophase ziehen sich die Chromosomen zu zwei Schwesterchromatiden zusammen, sie sind durch das Centromer verbunden. Hier werden sie unter dem Lichtmikroskop sichtbar. Tierische Zellen enthalten in der Regel zwei Centriolen, die aus Mikrotubuli, also aus Molekülen des Proteins Tubulin bestehen. In der Prophase trennen sich diese Centriolen und bewegen sich zu den entgegengesetzten Enden der Zelle. An ihnen wird später der Spindelapparat aufgebaut. In der Prometaphase fragmentiert sich die Zellkernhülle; im Zellzentrum sammeln sich die Chromosomen. Sie werden in der Metaphase in der Äquatorialebene ausgerichtet, der Spindelapparat entsteht zwischen den Centriolen aus Spindelfasern und Mikrotubuli. Die beiden Chromatiden werden dann in der Anaphase durch den Spindelapparat auseinandergezogen, so dass jede Zellhälfte einen vollständigen Chromatidensatz erhält. Um diese bildet sich in der Telophase jeweils eine neue Zellkernhülle und die Chromosomen dekondensieren wieder. Daraufhin folgt in der Cytokinese die Zellteilung. Vergleiche hierzu Abbildung B.1.

Die Interphase fasst die G₀-Phase, die G₁-Phase, die S-Phase und die G₂-Phase des Zellzyklus zusammen. Zellen in der G₀-Phase teilen sich auf unbestimmte Zeit nicht. Sie können in die G₁-Phase übergehen, in der die Zelle wächst, allerdings keine Synthese von DNA oder Proteinen vorgenommen wird. In der Synthesephase wird dann die DNA verdoppelt: es entstehen zwei gleiche Chromatiden. Danach folgt eine weitere Wachstumsphase ohne Synthese von DNA, die G₂-Phase, auf die dann die Zellkernteilung, die Mitose folgt.

B.2 CT-IC Modell

[9] zufolge spielt die Architektur des Zellkerns eine wichtige Rolle bei der Frage, welche Gene eines Chromosoms exprimiert werden und welche nicht. In Zellkernen von Säugetieren sind die Chromosomen in bestimmte, diskrete Regionen eingeteilt; jedes Chromosom hat sein Chromosomenterritorium (CT). Nun scheint, nach [9], die Position eines Gens in einem solchen CT Einfluss auf die Genexpression zu haben, da das Gen nicht an jeder Stelle gleichen Zugang zur "Transkriptionsmaschinerie" des Zellkerns hat.

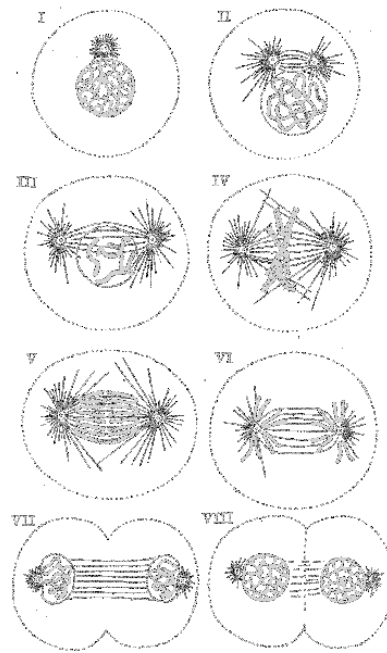


Abbildung B.1: (I)Interphase, (II)Prophase, (III)Prometaphase, (IV)Metaphase, (V und VI)Anaphase, (VII)Telophase, (VIII)Cytokinese. Aus [26].

Studien konnten zeigen, dass CTs eine komplex gefaltete Oberfläche aufweisen, die etwa der eines Schwammes gleicht (Vergleiche hierzu [7]). Chromatinfreie Zwischenräume erstrecken sich vom äußeren Rand der CTs ins Innere [10] und [7]. Daher plädieren die Autoren von [10] für ein Modell, das dieser Tatsache Rechnung trägt: das *Chromosome Territory - Interchromatin Compartment*-Modell (CT-IC Modell). Demzufolge sind Chromosomen in ihren CTs angeordnet, zwischen denen bestimmte Räume existieren, in denen die zur Replikation und Transkription notwendigen makromolekularen Strukturen vorhanden sind. Diese Räume heißen *Interchromatin Compartments* (IC). Dort, wo Chromatin an die IC angrenzt, befindet sich eine dritte wichtige Komponente des Modells, die *Perichromatin Region*, in welcher Transkription und DNA-Replikation ablaufen.

Abbildung B.2 zeigt eine Übersicht der Strukturmerkmale des Zellkerns, die das CT-IC Modell stützen. **a** zeigt eine Detailansicht der komplex gefalteten Oberfläche eines CTs: eine Schleife von aktiver und inaktiver DNA ragt in das IC hinein. **b** CTs sind in zwei Chromosom-Arm Gebiete, eines für den kurzen, eines für den langen Arm, und die Centromerische Region (*) unterteilt. Die Detailansicht zeigt (oben) aktiv transkribierte Gene (weiß) - sie liegen auf einer Schleife und haben keinen Kontakt zur Centromerischen Region. (unten) Liegen die gleichen Gene (schwarz) an dem centromerischen Chromatin, so werden sie inaktiviert. **c** Das Chromatin eines CTs ist an unterschiedlichen Stellen unterschiedlich dicht gepackt. Ist es locker (gelb), so ragt es in das IC hinein, dicht gepackte Chromatin hat keinen Kontakt zum IC. **d** Genarmes Chromatin (rot) liegt bevorzugt am Rand des Nukleus, während genreiches Chromatin (grün) eher in Kerninneren liegt. **e** Die Detailansicht zeigt komplexere Chromatinstrukturen aus einzelnen Chromatinfäden. Aktive Gene (weiße Punkte) liegen auf den Oberflächen der Strukturen. **f** Das IC (grün) enthält Komplexe, die für die Transkription, DNA-Replikation und Reparatur benötigt werden (orangene Punkte) und chromatinfreie Regionen. **g** zeigt Chromatingebiete (rot) und die ICs zwischen diesem Chromatin (grün). Aktive Gene (weiß) liegen außen an den Chromatingebieten, inaktive Gene (schwarz) sind von Chromatin umgeben.

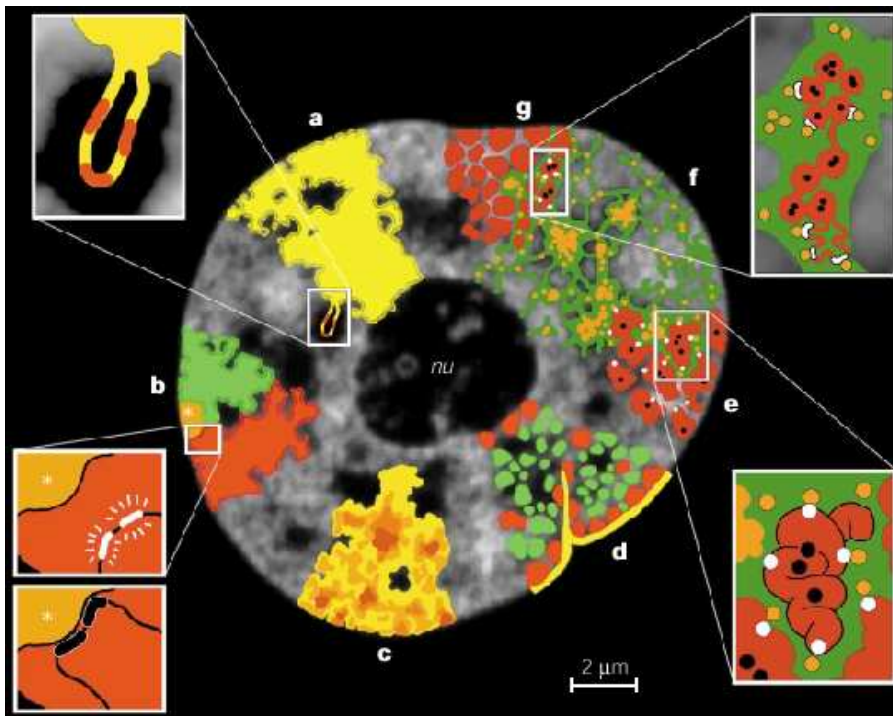


Abbildung B.2: Ein Modell der funktionalen Zellkernarchitektur. Aus [9].

Zusammenfassend gilt: Gene, die aktiv sind und daher transkribiert werden, liegen dem Modell zufolge in den äußeren Schichten des Chromatin, in der *Perichromatin Region*; Gene, die dauerhaft inaktiv sind, sind von dichtem Chromatin umgeben und haben daher keinen Zugang zu den für die Transkription notwendigen makromolekularen Strukturen. Je dichter das Chromatin eines Chromosoms gepackt ist, desto weniger seiner Gene können also transkribiert werden; folglich ist es zu einem geringeren Grad aktiv.

Anhang C

Manuell und automatisch bestimmte Schwellwerte

Tabelle C.1: Manuell und automatisch bestimmte Schwellwerte der MES Zellen in denen beide Chromosomen 11 aktiv sind, bzw. eines aktiv ist, das andere reaktiviert wurde.

Zelltyp	Zellkern	man. SW	autom. SW	Zelltyp	Zellkern	man. SW	autom. SW
MES	1	33	31	MES	1	32	31
beide Chr. 11 aktiv	2	24	25	1 Chr. 11 aktiv	2	37	32
	3	30	29	1 Chr. 11 reaktiviert	3	32	29
	4	27	25		4	29	26
	5	26	26		5	31	27
	6	27	23		6	-	29
	7	26	30		7	31	34
	8	35	28		8	29	29
	9	34	29		9	29	29
	10	30	30		10	33	31
	11	27	24		11	27	31
	12	40	38		12	-	27
	13	30	30		13	30	27
	14	24	24		14	29	30
	15	30	30		15	31	29
	16	28	27		16	32	30
	17	36	31		17	25	27
18	24	27	18		34	31	
19	34	32	19	-	31		
20	30	29	20	26	23		
21	30	22	21	31	17		
22	25	23	22	27	19		
23	27	27	23	29	24		
24	36	29					
25	30	24					
26	26	27					

Tabelle C.2: Manuell und automatisch bestimmte Schwellwerte der MEF und der MES Zellen, in denen ein Chromosom 11 aktiv ist, eines inaktiviert wurde.

Zelltyp	Zellkern	man. SW	autom. SW	Zelltyp	Zellkern	man. SW	autom. SW
MEF	1	29	29	MES	1	26	25
	2	32	34	1 Chr. 11 aktiv	2	29	25
	3	34	30	1 Chr. 11 inaktiv	3	42	26
	4	34	30		4	28	27
	5	37	34		5	50	44
	6	39	33		6	34	29
	7	42	36		7	23	21
	8	38	29		8	34	29
	9	33	30		9	34	32
	10	37	29		10	35	30
	11	40	34		11	25	22
	12	38	29		12	27	23
	13	39	33		13	32	28
	14	34	29		14	30	29
	15	36	35		15	37	34
	16	35	31		16	38	33
	17	35	28		17	25	22
	18	36	30		18	20	18
	19	33	32		19	22	18
	20	38	39		20	29	27
	21	-	37		21	19	17
	22	-	37		22	30	24
	23	43	34		23	34	29
	24	41	37		24	26	22
	25	42	36		25	25	19
	26	35	28		26	23	19
	27	40	30		27	25	23
	28	38	26		28	29	22
	29	38	28		29	23	23
	30	35	27		30	27	23
	31	28	24		31	31	28
	32	36	29				
	33	39	31				
	34	33	27				
	35	35	30				
	36	33	29				
	37	33	28				
	38	40	33				
	39	37	31				

Anhang D

Maximum Intensity Projektionen

Abbildung D.1: MIP des DAPI und FITC-Kanals der Fibroblasten.
HF-G0 HF-S MEF

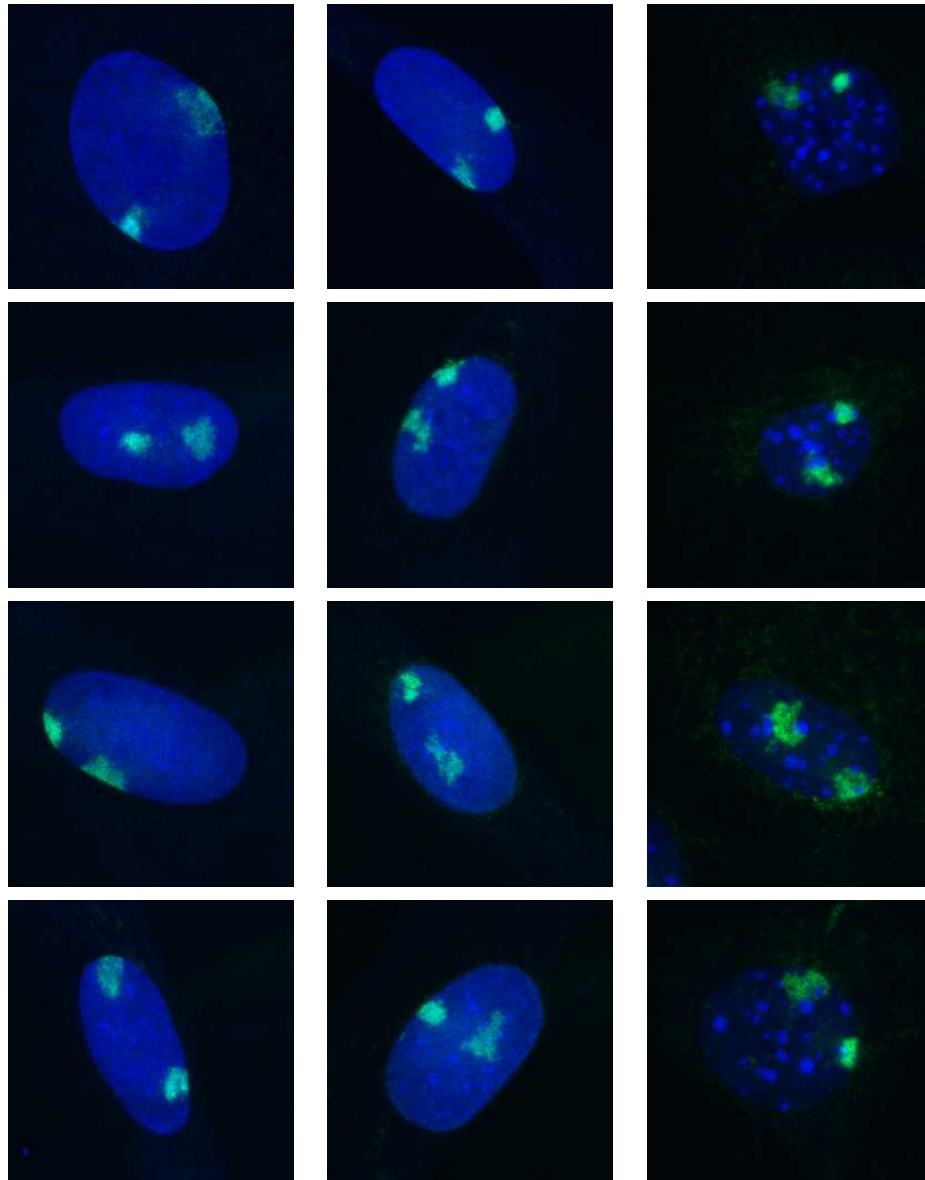


Abbildung D.2: MIP des DAPI und FITC-Kanals der MES Datensätze.
MESai MESaa MESar

