

INAUGURAL-DISSERTATION
zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen
Gesamtfakultät
der
Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Dipl.-Phys. **Kaan Saracoglu**
aus Kirkuk

Tag der mündl. Prüfung 7. November 2001

Bildanalyse von M-FISH

Gutachter:

Prof. Dr. Bernd Jähne

Prof. Dr. Dr. Christoph Cremer

Meiner Grossmutter, die nicht mehr wissen kann, was sie mir bedeutet.

Bildanalyse von MFISH - Zusammenfassung

Multiplex-FISH ist eine vor wenigen Jahren entwickelte Methode, mit der sich alle menschlichen Chromosomen mit mehreren Farbstoffen kombinatorisch färben lassen [Spe96a, Spe96b, Sch96]. Bei Verwendung von mehr als vier Farben lassen sich bei 31 Kombinationsmöglichkeiten alle Chromosomen in einer eindeutigen kombinatorischen Weise färben und werden mit geeigneten spektroskopischen Methoden anhand ihrer spektralen Information unterscheidbar.

Im Rahmen dieser Arbeit habe ich ein automatisches Verfahren für die Analyse von Bildern M-FISH markierter Chromosomen und subchromosomaler Proben entwickelt, das signifikante Verbesserungen und robuste Analysen, sowie erweiterte Möglichkeiten der Analyse mit Hilfe der M-FISH Technologie bietet. Das Verfahren basiert auf Clusteranalyse im Farbraum und kombiniert Farbinformation mit Ortsinformation, um chromosomale Bereiche im Bildvolumen zu identifizieren. Das Verfahren ist für die Analyse von Bildern unterschiedlicher M-FISH Experimente -sowohl chromosomal und subchromosomal, als auch inter- und intrachromosomal-geeignet. Identifizierung einzelner Chromosomen, sowie Volumen und Positionsbestimmung bei dreidimensionalen Aufnahmen im Interphase-Zellkern sind im Prinzip ebenso möglich.

Image Analysis of MFISH - Summary

Multiplex-FISH is a combinatorial staining technique that allows the simultaneous detection and discrimination of all human chromosomes. Using at least five fluorochromes all chromosomes can be uniquely labeled in a combinatorial way and identified by their specific spectral signature.

Within this thesis I developed a novel approach for the automated analysis of M-FISH images, yielding robust classification results and allowing the analysis of M-FISH images of different experiments. The method combines spectral information with spatial information to tessellate the image into regions of similar color. Subsequently a cluster analysis in color space and a final classification step are performed to identify the biological targets. This approach is applicable to images of different M-FISH experiments, allowing the analysis of interchromosomal as well as intrachromosomal abnormalities in the genome. It also allows the 3D analysis of M-FISH labeled chromosomes in interphase nuclei.

Inhaltsverzeichnis

1	Biologische Grundlagen	14
1.1	Die DNA	14
1.2	Gene	16
1.3	Genexpression	17
1.4	Replikation von DNA-Molekülen	18
1.5	Der Zellzyklus	18
1.6	Genetische Veränderungen	19
1.7	In Situ Hybridisierung	20
1.8	Multiplex-Fluoreszenz in situ Hybridisierung	21
1.9	Comparative Genomic Hybridization (CGH)	22
2	Physikalische Grundlagen	24
2.1	Mikroskopie	24
2.1.1	Linsen	24
2.1.2	Auflösungsvermögen	25
2.1.3	Axiale Auflösung	26
2.1.4	Konfokale Mikroskopie (CLSM)	27
2.2	Laserinduzierte Fluoreszenz (LIF)	28
2.3	Grenzen der konfokalen Mikroskopie	30
2.3.1	Photonenstatistik	30
2.3.2	Sättigungs- und Bleicheffekte	31
2.3.3	Nyquist Theorem	32
2.4	Aufnahme von M-FISH-Bildern	32
2.4.1	Das MFISH System	32
2.4.2	Das SKY System	34
3	Grundlagen der Bildverarbeitung	37
3.1	Klassifizierung	37
3.2	Clusteranalyse	37
3.2.1	K-Means-Algorithmus	39
3.2.2	Fuzzy-Clusteranalyse	39

3.3	Regionenwachstum	41
3.3.1	Einfache Verkettung	42
3.3.2	Hybride Verkettung	42
3.3.3	Zentroide Verkettung	42
3.4	Rauschen	43
3.5	Faltung und Filterung	44
3.5.1	Der Gaussfilter	45
3.5.2	Der Medianfilter	46
3.5.3	Anisotroper Diffusionsfilter	47
3.6	Dekonvolution	49
4	Methodik	51
4.1	Maximum-Projektion	52
4.2	Hintergrundkorrektur	53
4.3	Farbinformation und Datenraum	55
4.4	Bedeutung der Farbe bei M-FISH Bildern	56
4.4.1	Winkel als Distanzmass	57
4.4.2	Euklidischer Abstand als Distanzmass	58
4.5	Modellbasierte Segmentierung	59
4.5.1	Farbmodell	60
4.6	Tesselierung	61
4.6.1	Wahl der Startpunkte beim Regionenwachstum	62
4.7	Clustersuche	63
4.8	Klassifizierung der Regionen	64
4.8.1	Merging klassifizierter Regionen	66
4.9	Zusammenfassung der Methodik	66
5	Anwendungen	69
5.1	Anwendungen in 2D	69
5.1.1	Normale Metaphase	69
5.1.2	Aberrante Metaphase	70
5.1.3	Chromosomale Barcodes	72
5.1.4	Telomere	73
5.2	Anwendungen in 3D	74
5.2.1	3-Farben Experiment mit sieben Chromosomenpaaren in der Interphase	75
6	Ergebnisse und Diskussion	78
6.1	Auflösung und Hybridisierungsschema	79
6.2	Klassifikationsgüte von M-FISH im Vergleich mit SKY für Metapha- sen in 2D	81

6.3	Methodik	82
6.3.1	Binäre Klassifikation	82
6.3.2	Klassifikation mit Euklidischem Abstand	84
6.3.3	Regionenwachstum mit Euklidischem Abstand als Ähnlich- keitsmass	85
6.3.4	Clustersuche über Euklidischen Abstand	87
6.4	Chromosomale Barcodes	89
6.5	Telomere	90
6.6	Anwendungen in 3D	91
6.7	Zusammenfassende Diskussion	95
7	Zusammenfassung	98
8	Bedienung der Software	102
8.1	Die Datei classes	102
8.2	Die Datei menu	104
9	Publikationen im Rahmen dieser Dissertation	109

We dance round in a ring and suppose, but the secret sits in the middle and knows.
Robert Frost

Einleitung

”Der Mensch, die Krone der Schöpfung, das klügste und stolzeste Geschöpf, das die Evolution hervorgebracht hat, steht, was seinen genetischen Bauplan angeht, fast auf einer Stufe mit Würmern und Fliegen.” Diesen bemerkenswerten Aussage konnte man am 12.02.2001 in einer grossen deutschen Tageszeitung lesen. Die Entschlüsselung des menschlichen Genoms brachte zu Tage, dass das menschliche Erbgut offenbar nur zwischen 26000 und 40000 Gene besitzt, etwas mehr als der Fadenwurm, der einen Millimeter lang ist und praktisch über kein Gehirn verfügt. Noch vor einem Jahr waren Forscher mit einer Zahl von bis zu 130000 Genen ausgegangen. Diese Zahl beruhte auf der Vorstellung, dass ein Gen für die Synthese eines Proteins zuständig ist, die Zahl der Gene in etwa derjenigen der Proteine entspricht. Diese Vorstellung ist damit hinfällig.

Den Genen stehen also drei- bis zehnmal so viele Proteine gegenüber. Offenbar kann ein Gen den Bauplan für verschiedene Proteine liefern. Mit dieser Erkenntnis gewinnen weitere Zweige der Molekularbiologie an Bedeutung: Post- Genomics, das sich mit der Aktivierung und Deaktivierung von Genen und dem Ableseprozess durch Botenmoleküle beschäftigt, und die Proteomik, die Untersuchung der Proteine, ihrer Produktion und ihres komplexen Wechselspiels in den menschlichen Zellen. Die Erforschung von Proteinen ist jedoch sehr viel komplexer, da ihre Wirkung nicht nur auf ihrer molekularen Struktur, sondern auch auf ihrer Dynamik, ihrem Zusammenspiel beruht. Eiweisse werden in der Zelle synthetisiert, modifiziert, angehäuft und transportiert, sie beeinflussen andere Proteine ebenso wie Gene. Die Annahme, dass Gene und Proteine sich in geradlinigen kausalen Wirkungsketten verhalten, dass zum Beispiel Gen A für die Synthese von Protein A sorgt, das dann Protein B aktiviert, welches wiederum Eiweiss C anschaltet, ist nach dem heutigen Stand der Erkenntnis überholt. Ihr Verhalten ist viel zufälliger und unstrukturierter. Der Biologie fehlt jedoch der theoretische Rahmen, um solche Systeme zu beschreiben, während in der Mathematik und Physik schon zahlreiche Methoden existieren, um solches Verhalten zu erfassen.

Doch nicht nur die geringe Zahl der Gene ist erstaunlich, sondern auch ihre Verteilung innerhalb des Genoms. Da gibt es ausgedehnte Bereiche, in denen sich gar keine Gene befinden, während sie in anderen Abschnitten dicht an dicht liegen. Auch die Verteilung der Gene auf die Chromosomen ist nicht gleichmässig. Chromosom 18 beispielsweise ist besonders genarm und besteht über die Hälfte aus solchen Leerstellen.

Der Mithilfe von Bioinformatikern ist es zu verdanken, dass man heute ca. 10000 menschlichen Genen eine Funktion zuordnen kann. Etwa 1100 menschliche Gene sind dafür bekannt, eine Krankheit auslösen zu können, wenn sie mutieren, also ihr genetischer Code verändert wird. Glücklicherweise gibt es zwar mehr als einhundert Gene, die Reparaturwerkzeuge bereitstellen, deren Aufgabe es ist, solche Defekte

zu detektieren und auszubessern. Ein viertel aller Todesfälle in Deutschland ist auf bösartige Tumore zurückzuführen. Mehr über die Ursachen der Krebsentstehung zu erfahren und entsprechende Therapien zu entwickeln, ist daher eine der wichtigsten Aufgaben der medizinischen Forschung.

Die bei Krankheiten auftretenden genetischen Anomalien sind dabei vielfältig. Gene können überrepräsentiert sein oder ganz fehlen, aber auch ganze Chromosomen können in kranken Zellen fehlen oder in mehr als zwei Kopien auftreten. Sehr häufig kommt es auch zu inter- und intrachromosomalen Umbauten: Ein DNA Abschnitt von Chromosom A findet sich „plötzlich“ auf Chromosom B wieder, oder auf einem versetzten Bereich desselben Chromosoms.

Dabei gibt es verschiedene experimentelle Methoden, um die genomische Struktur und Aktivität auf unterschiedlicher genomischer Auflösung sichtbar zu machen. Allen im folgenden aufgeführten Methoden liegt die Technik der *Hybridisierung* zugrunde. Man macht sich hier die Eigenschaft der DNA-Bausteine (Nukleinsäuren) bzw. Sequenzen zunutze, mit in ihrer Basenabfolge komplementären Sequenzen Bindungen einzugehen: Interessierende Bereiche auf den DNA Einzelsträngen können sichtbar gemacht werden, in dem der diesem Bereich komplementäre Abschnitt mit radioaktiven oder fluoreszierenden Molekülen markiert wird.

Dazu zählen unter anderem die *Comparative Genomische Hybridisierung* (CGH) zur Analyse der Kopienanzahl chromosomaler Regionen mit einer Auflösung von etwa 10Mbp und die *Serielle Analyse von Genexpression* zur Untersuchung der Aktivität einzelner Gene, mit hoher cytogenetischer Auflösung im Bereich von einigen tausend Basenpaaren. Sie finden ihre Anwendung im Bereich der Bio-Chip-Technologie unter anderem bei den sogenannten *DNA-Microarrays*. Ein DNA-Microarray oder DNA-Chip besteht aus einem Träger mit einer meist gläsernen Oberfläche, auf den rasterförmig unterschiedliche Nukleinsäuren als Hybridisierungsziele aufgetragen und fixiert werden. Anschliessend gibt man in Lösung befindliche komplementäre Nukleinsäure-Moleküle hinzu. Dabei handelt es sich in der Regel um die Bausteine zweier unterschiedlicher DNAs, die zuvor mit unterschiedlichen Farbstoffen markiert wurden, wobei eine als Referenz dient. Zwei wichtige Anwendungen dieser Technologie sind die Untersuchung *genomischer Imbalancen in der Tumorgenese* mit genomischer DNA, und die *Analyse der Genexpression* mit mRNA (**m**essenger **R**ibo **N**ucleic **A**cid), das Aufschluss gibt über Stoffwechselwege, Reaktionen auf Arzneimittel und Umweltbedingungen. Parallele Abfrage einer grossen Anzahl von Zielstellen in der DNA, Automatisierung der Chipproduktion, Automatisierung der Auswertung und die Verifizierbarkeit einzelner Daten mit herkömmlichen molekularbiologischen Methoden sind dabei wichtige ökonomische Vorzüge der Chip-Technologie.

Eine experimenteller Ansatz mit niedrigerer genomischer Auflösung im Bereich von einigen Millionen Basenpaaren ist die sogenannte *Fluoreszenz in situ Hybridisierung*

(FISH). Sie beruht auf der Hybridisierung fluoreszenzmarkierter, chromosomenspezifischer DNA-Proben. Eine Weiterentwicklung hiervon ist die sogenannte *Multiplex-Fluoreszenz in situ Hybridisierung* (M-FISH), eine Technik die es erlaubt, mit Hilfe mehrerer Farbstoffe das gesamte Genom -also alle 24 Chromosomen-, in *einem* Experiment mit unterschiedlichen Farben und Farbkombinationen zu markieren. Mit geeigneten spektralen Techniken wird die Identifizierung aller Chromosomen einer Zelle eines Patienten und damit ihre Sortierung in ein sogenanntes Karyogramm möglich (*Karyotypisierung*), das Aufschluss über die Kopienanzahl der Chromosomen und die interchromosomalen Aberrationen gibt. Aber auch intrachromosomale Aberrationen und Anomalien in spezifischen Bereichen innerhalb der Chromosomen können mit dieser Technologie spezifisch gefärbt und untersucht werden.

Mit solchen Analysen können Zusammenhänge zwischen Krankheitsbild (*Phänotyp*) und genomischer Struktur (*Genotyp*) erstellt werden. Frühdiagnosen und Prognosen zum Krankheitsverlauf werden auf der Basis dieser Informationen möglich. Rechtzeitige therapeutische Massnahmen können zum Teil schon vor Ausbruch der Krankheit eingeleitet werden, was die Chancen einer erfolgreichen Behandlung drastisch erhöht. Anomalien in der DNA können auf unterschiedlichen genomischen Auflösungen auftreten. Es kann daher sein, dass mehrere Experimente notwendig sind, um alle bei einer Krankheit spezifisch auftretenden genetischen Aberrationen zu erfassen. Gewonnene experimentelle Daten müssen daher in Datenbanken gespeichert, aktualisiert und mit neuen experimentellen Resultaten erweitert und synchronisiert werden. Mit Verfahren aus dem Bereich der künstlichen Intelligenz lassen sich aus solchen Datenbanken Strukturen detektieren, die Rückschlüsse auf funktionale Zusammenhänge liefern.

Mit fortschreitender Entwicklung in der Biotechnologie, der zunehmenden Datenflut und Komplexität der zu bewältigenden Aufgaben ist in den letzten Jahren auch der Bedarf an automatischen Analyseverfahren zur Früherkennung genetisch bedingter Krankheiten rapide angestiegen. Damit erhebt sich die Notwendigkeit interdisziplinärer Forschung und Entwicklung - diese Arbeit ist ein Schritt in diese Richtung.

Kern dieser Dissertation war die Entwicklung einer robusten, flexiblen, automatischen Analyseverfahren von M-FISH Bildern. Es sollte nicht nur die zuverlässige Karyotypisierung von normalen und aberranten Zellen, sondern auch die Analyse von alternativen, auf M-FISH basierenden Experimenten mit kleinen DNA-spezifischen Proben und Experimenten zur Untersuchung von intrachromosomalen Anomalien erlauben. Eine Analyse von dreidimensionalen Bildaufnahmen von M-FISH markierten Chromosomen im Interphase-Zellkern ist ebenso möglich. Morphologische Grössen der einzelnen Chromosomen und ihre relativen Positionen zueinander können bestimmt werden. Während im zweidimensionalen das Resultat direkt verifiziert werden kann, ist dies im dreidimensionalen Fall schwieriger, da das Signal-Rausch-Verhältnis hier schlechter und die Farbinformation damit diffuser ist.

Weiterhin führen Überlappungen zwischen Chromosomen innerhalb des Zellkerns zu irreführenden Farbsignalen, die zu erkennen nur schwer wenn nicht gar unmöglich ist. Und nicht zuletzt die dreidimensionale Struktur macht eine Verifizierung über alle Bereiche aufwendig.

Zentrale Bedeutung bei dieser Arbeit hatte die Analyse im mehrdimensionalen Farbraum und die Untersuchung der Farbraumcharakteristik bei unterschiedlichen Experimenten. Dabei ging es nicht nur um die Frage einer robusten Identifikation aus bildverarbeitungstechnischer Sicht, sondern auch um Fragen der experimentellen Möglichkeiten, die farbliche Unterscheidung der Chromosomen und damit die „experimentelle“ Auflösung zu optimieren, um eine optimale Identifikation zu erreichen. Wichtig aber war es auch, Grenzen dieser Technik und Fehlerquellen der akkuraten Identifikation zu erkennen und aufzuzeigen. Zu Beginn dieser Arbeit handelte es sich bei MFISH um eine sehr neue Technologie. Auf folgende Arbeiten konnte diese Dissertation aufsetzen: [Spe96a, Sch96, Gar96, Eil98].

Bei Hybridisierungsexperimenten treten schwer quantifizierbare Rauschquellen auf. Das ideale Hybridisierungsexperiment sollte eine homogene Färbung der biologischen Präparate gewährleisten. Das ist jedoch nur selten der Fall, meistens sind die Intensitätsschwankungen innerhalb kleiner Bereiche sehr stark. Die Gründe für diese Inhomogenitäten sind im Detail noch nicht genau verstanden. Bei Mehrfachhybridisierungen, wie es bei M-FISH Experimenten der Fall ist, kommen noch weitere Schwierigkeiten hinzu. Bestimmte Fluoreszenzkombinationen führen bei einigen Chromosomen zu sehr schwachen Hybridisierungen. Die Hybridisierungsqualität der unterschiedlichen Farbstoffe und die Bildaufnahme der unterschiedlichen Farbkanäle müssen aufeinander abgestimmt werden. Standardisierte Färbeprotokolle und Probenkits haben sich erst kürzlich etabliert, weshalb jedes Labor, das auf diesem Gebiet arbeitet, zunächst sein eigenes Hybridisierungsprotokoll und -Schema entwickelte. Das wiederum stellte hohe Ansprüche an die Bildverarbeitung und erforderte flexible Analysemethoden, um der unterschiedlichen Qualität der anfallenden Daten gerecht zu werden.

Da es sich bei der Analyse genomischer Imbalancen mit Hilfe der CGH gleichfalls um ein Mehrfluoreszenzexperiment handelt, konnten wesentliche Teile der für die Analyse von M-FISH Bildern entwickelten Methoden auch auf die Analyse von DNA-Microarrays angewendet werden. Dazu wurde eigens eine Software im Rahmen einer Diplomarbeit entwickelt [Rud00]. Obwohl der Schwerpunkt dieser Arbeit nicht auf dieser Entwicklung beruht, werden auch für diese Anwendung die für das Verständnis notwendigen biologischen und experimentellen Grundlagen dargelegt.

Aufbau

Zunächst lege ich die für diese Arbeit relevanten biologischen und für die Bildaufnahme wichtigen physikalischen Grundlagen in Kapitel 1 und 2 dar. Dazu zählen vor allem die Technik der Multiplex - Fluoreszenz in situ Hybridisierung und die spektrale Bildaufnahme am Mikroskop. Anschliessend gebe ich im dritten Kapitel eine Einführung in relevante grundlegende Bildverarbeitungsoperationen. In Kapitel 4 beschreibe ich ausgehend von der Analyse des Farbraums die von mir entwickelte Methodik, die sich zusammengefasst am Ende des Kapitels wiederfindet. Anwendungen, die ich in Kapitel 5 präsentiere, sind im vierten Kapitel zum Teil schon vorweggenommen, um den Weg der Entwicklung anschaulich und nachvollziehbar zu dokumentieren. Ergebnisse meiner Methodik diskutiere ich in Kapitel 6. Eine etwas ausführlichere Zusammenfassung findet sich am Ende dieser Arbeit und ist bilingual gehalten.

Kapitel 1

Biologische Grundlagen

Die Zelle ist die strukturelle Basis für jeden Organismus und weist Stoffwechsel sowie die Fähigkeit, sich zu reproduzieren auf. Man unterscheidet zwischen Prokaryonten und Eukaryonten, letztere weisen einen Zellkern auf und bestehen im wesentlichen aus dem *Nukleus* (Zellkern), dem *Cytoplasma* und der Zellmembran. Im Zellkern befindet sich das Chromatin, die Gesamtheit der Chromosomen, und als kompakte Struktur das *Nukleolus* (Kernkörperchen), in dem die *ribosomale RNA* (rRNA) synthetisiert wird. Das Zellkernplasma (Karyoplasma) wird durch die Kernmembran vom Cytoplasma getrennt. In den Zellen aller Lebewesen ist die „Erbgutinformation“ auf dem selben Makromolekül -der DNA- gespeichert. Durch Untersuchungen an Bakterien und ihren Viren¹ ist in den vierziger Jahren erstmals der eindeutige Nachweis für die Funktion der DNA als Träger dieser Informationen, der Gene, gelungen.

1.1 Die DNA

Die Desoxyribonukleinsäure ist ein langes Polymermolekül, das aus kettenartig miteinander verbundenen Purinbasen Adenin, Thymin, Guanin und Cytosin, sowie weiteren Proteinen, die als Gerüststoffe dienen, aufgebaut ist. Der Grundbaustein der DNA ist das *Nukleotid*, das sich aus drei Einheiten zusammensetzt:

- einem Zucker, der Desoxyribose
- einer stickstoffhaltigen Purinbase
- und einer Phosphorsäure.

Im Falle der DNA ist der Zucker eine *Desoxyribose* und die möglichen Basen sind Adenin (A), Guanin (G), Cytosin(C) und Thymin (T). Bei der RNA dagegen ist

¹Das entscheidende Experiment dazu stammt von O.Avery und Mitarbeitern.

der Zucker eine *Ribose* und die Base Thymin ist gegen Uracil ausgetauscht. In den fünfziger Jahren hat E.Chargaff festgestellt, dass das Verhältnis von A zu T(U) und das von G zu C immer gleich ist, weshalb die prozentuale Basenzusammensetzung jeder DNA vorausgesagt werden kann.

In der Folgezeit zeigten Röntgenstruktur-Analysen, dass die DNA eine regelmäßige Periodik aufweist und aus zwei Einzelsträngen aufgebaut ist. 1953 schlossen J.D.Watson und F.Crick auf die Doppelhelixstruktur.

In Abb. 1.2 b) ist die Doppelhelix als Doppelspirale gezeigt, deren Einzelstränge über Stufen (Wasserstoffbrückenbindungen) miteinander verbunden sind. Die Bänder enthalten die Zucker-Phosphat-Teile, während die Stufen durch die Basen beschrieben werden. Allerdings sind nur ganz spezielle Basenpaarungen, nämlich A-T bzw. G-C, möglich.

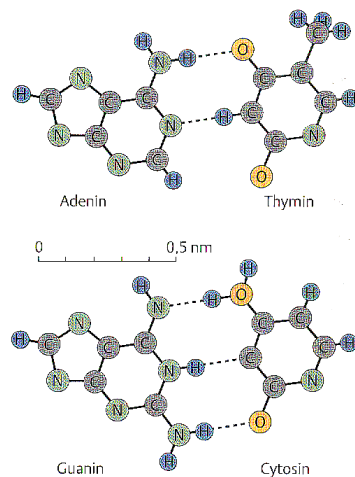


Abb. 1.1: Die Basen der DNA und ihre Paarung (Quelle:Kni95)

Folglich sind die zwei Basenstränge komplementär zueinander. Aufgrunddessen kann von der Nukleotid-Sequenz des einen Strangs die des anderen abgeleitet werden.

Beim Menschen ist die gesamte Erbgut auf 23 Chromosomenpaaren gespeichert. Man unterscheidet dabei zwischen den 44 *Autosomen* und den zwei *Gonosomen*, den Geschlechtschromosomen. Der Unterschied zwischen Mann und Frau besteht in genau einem Chromsom: die Frau hat zwei X-Chromsomen, während der männliche Karyotyp ein X- und ein Y-Chromosom aufweist. Menschliche Chromsomen enthalten im Mittel $3 \cdot 10^9$ Basenpaare. Das entspricht einem DNA-Strang mit einer Länge von 5-10 cm.

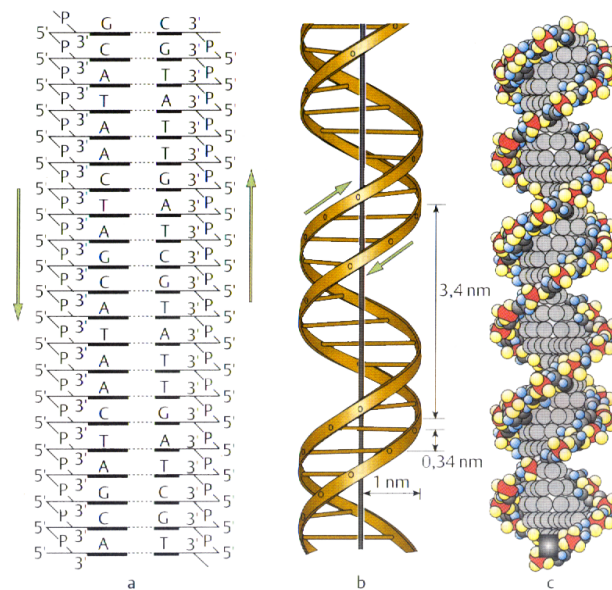


Abb. 1.2: **a** Die beiden Stränge der DNA laufen antiparallel. **b** Dimension der Doppelhelix. **c** sog. Kalottenmodell-Darstellung (blau: H, gelb: O, grau: C in der Phosphodiesterkette, hellgrau: C bzw. N in den Basen, rot: P). (Quelle:Kni95)

1.2 Gene

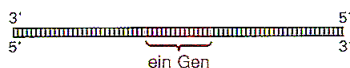
Gene sind Abschnitte auf den DNA-Strängen. Die biologische Funktion des Gens ist dabei in der Sequenz seiner Basenpaare kodiert. Diese Information umfasst eine Reihe von Anweisungen zur Synthese eines RNA-Moleküls, das anschliessend entweder die Synthese von Proteinmolekülen (z.B. Enzyme) veranlasst oder selbst in der Zelle eine Funktion übernimmt. Dieser Vorgang wird als *Genexpression* bezeichnet (s. Abs. 1.3) [Bro93]. Die Entschlüsselung des menschlichen Genoms brachte überraschend zu Tage, dass der Mensch „nur“ zwischen 25000 und 40000 Gene besitzt. Man war bisher von einer weit höheren Zahl ausgegangen. Der Vergleich mit der Zahl von bis zu 130000 Proteinen, lässt den Schluss zu, dass ein Gen für die Synthese mehrerer Proteine zuständig sein kann, entgegen der bisherigen Vorstellung.

Die Verteilung der Gene innerhalb der DNA ist dabei alles andere als homogen. Da gibt es ausgedehnte Wüsten, in denen sich gar keine Gene befinden (*intergene DNA*), während sie in anderen sogenannten Hotspots dicht an dicht liegen (siehe Abb. 1.3). In höheren Organismen können diese intergenen Regionen so gross sein, dass der Anteil der Gene an der gesamten zellulären DNA nur einige Prozent beträgt.

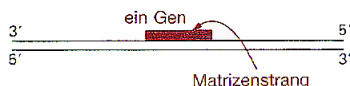
Abb. 1.3 c) veranschaulicht, dass die entsprechende Nukleotidsequenz in lediglich einem der beiden Stränge der Doppelhelix enthalten ist. Entsprechend seiner Funktion bei Replikationsprozessen nennt man diesen *Matrizenstrang*. Im Durchschnitt be-

Abb. 1.3: Gene sind Abschnitte der DNA, in c) geben die Pfeile die Richtungen an, in der die biologische Information während der Genexpression gelesen wird. (Quelle:Bro93)

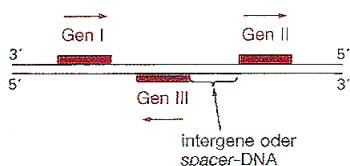
a) ein Gen ist ein Abschnitt eines DNA-Moleküls



b) die biologische Information ist nur in einem Polynucleotid enthalten



c) Gene können auf verschiedenen Polynucleotiden liegen



steht ein Gen aus einigen hundert (n) Basenpaaren, was insgesamt ca. 4^n Möglichkeiten ergibt, von denen der grösste Teil tatsächlich biologische Informationen enthält. Die DNA im Zellkern des Menschen etwa ist circa 3000 Mbp lang [Kni95].² Die funktionellen Einheiten der DNA sind Basentriplets. Diese kodieren die 20 unterschiedlichen Aminosäuren, welche die Bausteine der Proteine darstellen [Lin89]. Genbereiche auf der DNA sind durch wohldefinierte Start- und Stopsequenzen markiert. Das *Start-Codon* besteht für alle Gene aus den drei Basen (A T G). Ein der Startsequenz vorgelagerter *Promotor* besteht aus einer bestimmten Basensequenz, an die Ablesenzyme in stabiler Weise binden. Zwischen Start- und Stop-Codon liegen die die Aminosäuren kodierenden Tripletts. Interessant ist, dass mehrere Tripletts in ein und dieselbe Aminosäure kodieren, da es $4^3 = 64$ mögliche Dreier-Kombinationen der 4 unterschiedlichen Basen gibt. Drei der 64 möglichen Kombinationen sorgen für das Ende der Protein-Transkription. Sie lauten UAA, UAG und UGA.

1.3 Genexpression

Die Genexpression ist der Vorgang, der die in den Genen enthaltene Information der Zelle zugänglich macht. Dabei wird die DNA des Gens zuerst in RNA und anschliessend in ein Protein übersetzt. Dies wird auch als das *zentrale Dogma* der Molekulargenetik bezeichnet [Cri70].

Die erste Stufe der Genexpression, die *Transkription*, wird von allen Genen durchlaufen. Während dieses Prozesses wird ein RNA-Molekül synthetisiert³, wobei eine

²Megabasenpaare

³In einer Zelle sind immer Bausteine beider genetischer Säuren vorhanden.

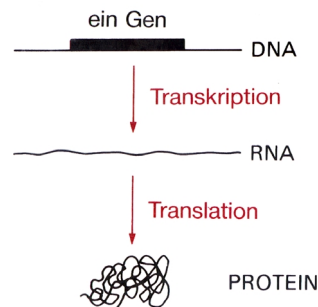


Abb. 1.4: Die beiden Stufen der Genexpression (Quelle:Bro93)

komplementäre Abschrift des Matrizenstrangs des jeweiligen Gens erfolgt. Die zweite Stufe ist die *Translation*. Diese wird allerdings nur von solchen Genen durchlaufen, für die das Endprodukt der Expression nicht das *RNA-Transkript* selbst darstellt. Das RNA-Molekül, nun als mRNA bezeichnet, veranlasst bei der Translation die Synthese eines Makromoleküls, dessen Aminosäuresequenz durch die Nukleotidsequenz der mRNA bestimmt ist. Die Endprodukte der Genexpression sind die *Proteine*, man spricht von Bioproteinsynthese. Die Aktivität der Gene, die durch komplexe Prozesse reguliert wird, ist demnach auf die Expression zurückzuführen. Quantitative Veränderungen der Genexpression wirken sich auf das Mengenverhältnis der resultierenden Proteine aus, was die Entstehung eines Tumors zur Folge haben kann.

1.4 Replikation von DNA-Molekülen

Der grundlegende Prozess bei der Entwicklung von Organismen ist die Zellteilung; bei jeder Teilung muss eine vollständige Kopie aller Gene hergestellt werden. Dabei ist ein hohes Maß an Genauigkeit notwendig, da schon sehr geringe Fehlerquoten (z.B. 1 Fehler pro 10000 Nukleotiden) eine beträchtliche Anhäufung an Veränderungen innerhalb der Gene verursachen können. Dadurch besteht die Gefahr, dass lebenswichtige Informationen verloren gehen.

Aufgrund der speziellen Basenpaarungen dient ein DNA-Strang dem anderen als Matrize, die Doppelhelix kann sich folglich aus dem in der Zelle vorhandenen biochemischen Material selbst replizieren.

1.5 Der Zellzyklus

Der Zellzyklus lässt sich in vier Phasen G1, S, G2 und M unterteilen. Die ersten drei werden zur *Interphase* zusammengefasst. Die M-Phase heisst auch *Mitose*, die Phase der Zellteilung. Sie besteht aus *Prophase*, *Metaphase*, *Anaphase* und *Telophase*. Die Gesamtheit dieser Phase beschreibt den zyklischen Vorgang der

DNA-Duplikation und anschliessender Zellteilung, also die Zellvermehrung durch Replikation der DNA und Vererbung der genetischen Information auf die Tochterzellen.

In der Interphase liegen die Chromosomen im Zellkern in der weniger stark kondensierten Form vor. Die eigentliche DNA-Synthese findet in der S-Phase statt. G1 und G2 sind Übergangsphasen zwischen der Synthesephase und der Mitose. In der G1-Phase steigt die biologische Aktivität der Zelle mit einem Anstieg der RNA Produktion. Nach der DNA-Synthese in der S-Phase wird die Zelle in der G2 Phase durch Bildung von Teilungsproteinen auf die Mitose vorbereitet. In der mitotischen Prophase werden Kernmembran und Nucleolen aufgelöst und der Spindelapparat um die Zellpole ausgebildet. Anschliessend bilden sich zwischen den Polen Spindelfasern, die mit einer Anordnung der Chromosomen durch den Spindelapparat einhergeht. Die Metaphaseplatte entsteht (*Metaphase*). Als Anaphase bezeichnet man den folgenden Vorgang der Trennung der Schwesterchromosomen hin zu den Zellpolen. In der Telophase findet der eigentliche Teilungsprozess statt: Kernmembran und Nucleolen werden neu gebildet, die Chromosomen dekondensieren, die Zellplatte wird gebildet, die Zelle teilt sich. Der Zellzyklus ist durchlaufen und die DNA-Synthese in der Interphase kann neu beginnen.

Durch Aufschraubungs- und Faltungsmechanismen kondensieren die Chromosomen in der Mitose und erreichen eine mittlere Länge von 5-10 μm [Czi81]. Sie sind nach Färbung im Mikroskop als Chromosomen in ihrer bekannten Form sichtbar. Durch den Ansatzpunkt der Spindelfasern, dem *Centromer* wird das Chromosom in einen kurzen **p**- und einen langen **q**-Arm unterteilt.

1.6 Genetische Veränderungen

Der Entstehung karzinogener Zellen gehen verschiedene genetische Veränderungen der DNA voraus, die man in zwei Klassen unterscheidet. *Mutationen* sind *Punktveränderungen*, d.h. Austausch einer einzigen Base oder Veränderungen, die sich auf einen sehr begrenzten Bereich der DNA beschränken. Sie entstehen häufig während des Replikationsprozesses oder durch chemische und physikalische *Mutagene*. Um die Zahl dieser oft auf natürliche Weise entstehenden Mutationen gering zu halten, existieren *Reparatur-Mechanismen*. Darunter sind Enzyme zu verstehen, die beim „Abfahren“ der Einzelstränge die DNA erneuern, aber auch gegebenenfalls den „programmierten Zelltod“ (*Apoptose*) einleiten können. Im Gegensatz zu Mutationen bezeichnet man mit *Rekombinationen* die Umordnungen von DNA-Abschnitten. Diese sind zum Teil erwünscht, da auf diese Weise genetische Vielfalt entsteht und erhalten wird.

Solche genetische Veränderungen können die Regulationsmechanismen derart in ihrer Funktion stören, dass sich ein ungebremstes Wachstum entwickelt und letztlich

Tumoren entstehen. Abhängig vom Krankheitstyp sind bei karzinogenen Zellen *Deletion* (Verluste) und *Amplifikation* (Vervielfältigungen) gewisser Gensequenzen typische Merkmale.

1.7 In Situ Hybridisierung

Ende der sechziger Jahre wurde von Gall und Pardue eine Technik entwickelt, die es ermöglicht, einzelne Regionen der DNA zu markieren. Dazu wird die DNA zunächst *denaturiert*, d.h. die Wasserstoffbrückenbindungen der Basenpaare werden aufgebrochen und die DNA aus ihrer doppelhelikalen Struktur in die beiden Einzelstränge getrennt. Auftrennen der DNA in ihre Einzelstränge (*Denaturierung*) geschieht z.B. durch Wärme [Str96]. Danach wird die DNA komplementären DNA-Proben wieder *renaturiert*. Den Vorgang der Bindung einer markierten DNA-Einzelstrangprobe an denaturierte chromosomale DNA bezeichnet man als *Hybridisierung*. Eine Bindung findet überall dort statt, wo sich komplementäre Basensequenzen befinden.

Im Falle der chromosomalen *in-situ*-Hybridisierung befindet sich die denaturierte DNA i.A. auf einem Objektträger, während der markierte Strang von einer zugesetzten Sonde stammt. Zunächst wurden zur Markierung radioaktive Stoffe verwendet, bspw. 3H und der Nachweis der hybridisierten und markierten DNA erfolgte über Autoradiographie. Allerdings ist dieses Signal nicht immer deutlich sichtbar und überdeckt zudem einen relativ grossen Teil des Chromosoms. Bei der *Fluoreszenz-in-situ-Hybridisierung* sind hingegen die Nukleotide mit fluoreszierenden Seitengruppen markiert. Die Fluoreszenzsignale, die mittels einem Fluoreszenzmikroskop nachgewiesen werden, sind aufgrund ihrer Strahlungseigenschaften besser geeignet; die dadurch erreichte höhere Auflösung ist gegenüber dem erstgenannten Verfahren von grossem Vorteil. Überdies lassen sich mit unterschiedlich markierten DNA-Sonden (klonierte DNA-Fragmente) verschiedene Genorte in einem Experiment anfärben.

Diese Methode macht eine *ortsspezifische* Bindung einer DNA-Probe allerdings nur dann möglich, wenn die zum markierten DNA-Stück komplementäre Nukleotidsequenz nur an einem Ort eines individuellen Chromosoms vorkommt. Es handelt sich in diesem Fall um eine *Einzelkopie-Sequenz*. Sollte die DNA-Probe jedoch neben ortsspezifischen Sequenzen auch hoch- oder mittelrepetitive Sequenzen enthalten, die auf allen Chromosomen vorkommen, hybridisiert die DNA-Probe an alle Chromosomen, so dass das ortsspezifische Signal nicht länger erkennbar sein wird. Man ist aus diesem Grund in den letzten Jahren zur *CISS-Hybridisierung* (Chromosomal In Situ Suppression) übergegangen [Cre88]. Zu einem geeigneten Zeitpunkt werden in die Hybridisierungslösung nicht markierte hochrepetitive DNA-Sequenzen hinzugegeben, die an die repetitiven Sequenzen der Chromosomen binden. Eine unspezifische Bindung der markierten Proben wird damit unterdrückt. Die Technik ermöglicht nicht nur die spezifische Markierung von Metaphase-Chromosomen,

sondern auch die komplette Färbung einzelner Chromosomen mit Hilfe von Proben-DNA, die alle Sequenzen eines bestimmten Chromosoms enthält (*Library-DNA*). Solche chromosomenspezifischen DNA-Bibliotheken konnten entwickelt werden, nachdem es gelang, individuelle Chromosomen mit dem Verfahren der fluoreszenzaktivierten Chromosomen-Sortierung anzureichern [Cre84].

1.8 Multiplex-Fluoreszenz in situ Hybridisierung

Da die Wahrscheinlichkeit zur Detektion genomischer Anomalien mit der Zahl gleichzeitig verwendeter DNA-Proben steigt, wurden in den letzten Jahren verschiedene Techniken entwickelt, um möglichst viele gleichzeitig hybridisierte Proben zu unterscheiden. FISH ist dabei die ideale Basis für diese Methoden, da eine Vielzahl spektral unterschiedlicher Farbstoffe für FISH zur Verfügung stehen.

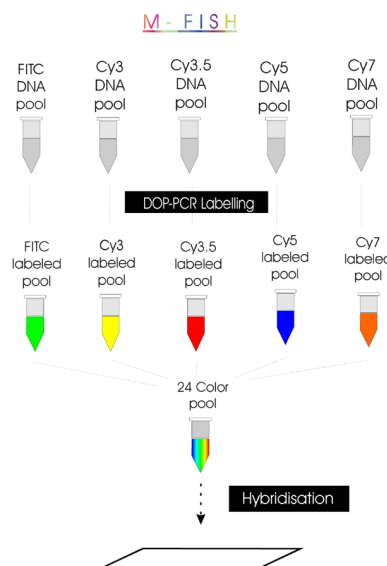


Abb. 1.5: Prinzip des M-FISH Verfahrens.(Quelle: [Eil98])

Bei M-FISH werden mehrere spektral unterschiedliche Farbstoffe zur Markierung der DNA verwendet. Dabei wird die Zahl unterscheidbarer Proben durch kombinatorische Färbung signifikant erhöht. Die ersten kombinatorischen Ansätze gab es schon Ende der achtziger, Anfang der neunziger Jahre [Ned89, Ned90, Rie92a, Rie92b, Len93, Pop93, Wie93]. Es wird nicht mehr nur eine einzige, sondern mehrere Fluorochrome gleichzeitig an die DNA-Proben hybridisiert, die dann an ihre komplementären Sequenzen binden. Da es bei n Farbstoffen $2^n - 1$ mögliche Kombination gibt, reichen 5 Farbstoffe aus ($2^5 - 1 = 31$), um das gesamte Genom, also alle 24 Chromosomen unterschiedlich zu färben und damit spektral unterscheidbar

zu machen. Abb. 1.6 zeigt ein typisches Hybridisierungsschema (*Labeling-Schema*) für ein 5-Farben Experiment.

Colour		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Grey																									
FITC		x			x		x		x	x		x		x			x		x			x			x
Cy3				x		x			x	x		x				x					x	x		x	x
Cy3.5		x		x	x			x			x	x						x			x				x
Cy5		x				x	x	x	x					x		x	x	x						x	
Cy7			x	x		x	x			x	x		x										x		x

Abb. 1.6: Typisches Hybridisierungsschema eines 5-Farben m-FISH Experiments. Jedes Chromosom erhält eine eindeutige Kombination von Farbstoffen und damit eine eindeutige spektrale Signatur. Die Darstellung der Chromosomen erfolgt über Falschfarben, die in der oberen Leiste zu sehen sind.

1.9 Comparative Genomic Hybridization (CGH)

CGH wurde zur Detektion ungewöhnlicher Kopienzahlveränderungen von Sequenzen der Nukleinsäure in einem oder mehreren Genen entwickelt [Kal92]. Grundlage dieser Methode ist das FISH-Verfahren. Mit CGH kann die relative Kopienzahlveränderungen von tumorspezifischen Gensequenzen als Funktion des Ortes der entsprechenden Sequenzen in einem Referenzgenom ausgedrückt werden [CGH95].

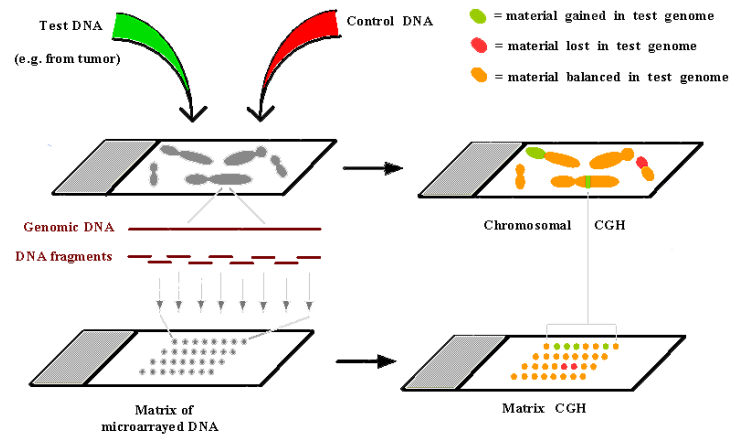
Als Probe wird hier einerseits die gesamte DNA des Tumorgewebes und andererseits die *normale* Referenz-DNA verwendet, wobei die Hybridisierung bzw. Ko-Hybridisierung auf aufgespreiteten *Metaphase-Chromosomen*⁴ gesunder Zellen stattfindet. Die Tumor-DNA wird dabei gewöhnlicherweise mit grünem Fluorochrom gelabelt, während die Referenz-DNA mit einem roten Fluoreszenzfarbstoff markiert wird. Gleiche Anteile von Tumor- und Referenz-DNA „konkurrieren“ nun um die gleiche Ziel-DNA (im folgenden mit *Target* bezeichnet). Ähnlich zu FISH werden die Fluoreszenzsignale für jedes einzelne Target mit einem Fluoreszenzmikroskop gemessen, im Anschluss berechnet man das Verhältnis von grünem Signal (Tumorgewebe) zu rotem (gesundes Gewebe). Rot-dominierte Verhältnisse deuten auf Deletionen und grün-dominierte auf Amplifikationen in der Tumor-DNA hin [CGH95]. Abb. 1.7 zeigt ein typisches Beispiel.

Der Vorteil dieser Methode besteht darin, dass in einem einzigen Experiment anhand der Ungleichverteilung der Signale (Imbalancen) entlang der Chromosomen-Achsen genomweit Veränderungen in den Kopienzahlen einzelner Bereiche identifiziert werden können. Zudem sind hier keine Vorinformationen über Veränderungen nötig. Durch CGH konnten insbesondere hämatologische Tumoren erfolgreich analysiert werden (Übersicht, [Lic00]).

⁴Bezeichnung für die organisierteste Form der Chromosomenstruktur

Comparative genomic hybridisation (CGH)

Abb. 1.7:
Das Prinzip der
CGH. (Quelle: [Rud00])



Diese Methode hat aber auch Nachteile. So ist für eine effiziente Analyse ein Mindestanteil von ca. 35% an Tumorzellen in der Gewebeprobe notwendig [DuM95], was sich zum Teil jedoch als problematisch erweisen kann. Auch sind diese Experimente nur bedingt als Analysemethoden mit hohem Durchsatz geeignet, weil sowohl die Herstellung als auch die anschließende Auswertung manuell erfolgen müssen, was einen hohen Arbeits- und Zeitaufwand bedeutet. Desweiteren schränkt die Gestalt des Targets die Auflösung ein. Kleinere Variationen, die durchaus für eine Klassifizierung und Prognose von entscheidender Bedeutung sein können, bleiben womöglich unentdeckt.

Kapitel 2

Physikalische Grundlagen

2.1 Mikroskopie

2.1.1 Linsen

Eine Linse ist im Prinzip eine fokussierende Blende. Licht, das durch einen Spalt tritt, wird gebeugt. Es entsteht ein Beugungsbild mit Intensitätsmaxima und -minima höherer Ordnungen, deren Lage von der Wellenlänge und der Geometrie des Spalts abhängen. Durch die Form der Linse werden die optischen Weglängen für die Lichtstrahlen so verändert, dass die Maxima höherer Ordnung unterdrückt werden und nur für das Hauptmaximum konstruktive Interferenz auftritt. Das Beugungsbild eines Spaltes ist dabei die Fouriertransformierte der Spaltfunktion, die die Form des Spalts beschreibt. Das leuchtet unmittelbar ein, wenn man sich das Intensitätsprofil in der Betrachtungsebene als Summe der Schwingungszentren im Spalt nach dem Huygensschen Prinzip vorstellt. Die Amplitudenfunktion $U(a, b)$, die das Beugungsbild in der Bildebene beschreibt, verdeutlicht diesen Zusammenhang:

$$U(a, b) = \int \int_{x,y} s(x, y) \exp^{i(ax+by)} dx dy \quad (2.1)$$

Die Koordinaten in der Blendenebene werden durch (x, y) , die Blendenfunktion durch $s(x, y)$ beschrieben, die für einen Spalt den Wert 1 hat; a und b sind die räumlichen Frequenzen in der Bildebene. Sind (X, Y) die Ortskoordinaten in der Bildebene und L die Brennweite der Linse, sowie k die Wellenzahl, lassen sich a und b schreiben als $a = kX/L$ und $b = kY/L$. Die vollständige theoretische Herleitung dieser Formel ergibt sich aus dem *Kirchhoffschen Integral Theorem*.

Beim Durchgang durch das Objektiv werden aufgrund der begrenzten räumlichen Ausdehnung höhere räumliche Frequenzen am Rand abgeschnitten. Dies führt dazu, dass das Bild eines Objekts nach Durchgang durch eine Linse einen Informationsverlust erleidet und das Bild nicht vollständig reproduziert wird. Die gerade für die

Ränder wichtigen hohen räumlichen Frequenzen fehlen und die tiefen dominieren mit der Folge, dass das Beugungsbild an den Rändern unscharf wird und verschmiert. Die Form der bewirkt, dass Licht, das von einem bestimmten Punkt vor der Linse ausgeht, in einen definierten Punkt hinter der Linse gebeugt wird, und zwar genau in den Punkt, für den die optischen Weglängen der von der Punktquelle ausgehenden Lichtstrahlen gleich, was zu konstruktiver Interferenz führt. Die normalisierte laterale Intensitätsverteilung, die ein weit entfernter Punkt in der Fokalebene der Linse erzeugt lautet:

$$I(\rho) = \left[\frac{2J_1(\rho)}{\rho} \right]^2 \quad (2.2)$$

$J_1(\rho)$ ist die Besselfunktion erster Ordnung, ρ der normierte Radius mit

$$\rho = kR \sin \alpha, \quad (2.3)$$

wobei R der Radius der Blende, α der Raumwinkel und $k = 2\pi/\lambda$ die Wellenzahl ist. Für $\rho \rightarrow 0$ konvergiert der Quotient $2J_1(\rho)/\rho$ gegen $1/2$, der gesamte Term also gegen 1 . Das Hauptmaximum heisst *Airy Disk*. Diese wird durch die ersten Nullstellen der Besselfunktion definiert, ihr Radius lässt sich mit $\sin \alpha = R/f$ schreiben als

$$r_{Airy} = 0.61 \frac{\lambda_0}{NA}. \quad (2.4)$$

Linsen weisen Abbildungsfehler auf, die die theoretische Auflösung in der Mikroskopie verschlechtern. Die Abbildungseigenschaften von Linsen verschlechtern sich dabei im allgemeinen mit zunehmendem Abstand von der optischen Achse. Unter *chromatischer Aberration* versteht man die Dispersion verursachte unterschiedliche Brechung und damit Abbildung von Licht unterschiedlicher Wellenlänge. Licht höherer Frequenz wird stärker gebrochen, damit liegt der Fokus solchen Lichts vor längerwelligem. Der Unterschied in der Lage des Brennpunkts kann im sichtbaren Spektrum bis zu einer Wellenlänge betragen. *Sphärische Aberrationen* sind Abweichungen und Fehler in der Oberfläche der Linse. Ein einfaches Beispiel ist der Astigmatismus, der durch einen anisotropen Krümmungsradius verursacht wird. Bei moderne Mikroskopen sind solche Abbildungsfehler weitgehend korrigiert. Für eine ausführliche Behandlung zu diesem Thema muss an dieser Stelle auf weiterführende Spezialliteratur der Optik verwiesen werden.

2.1.2 Auflösungsvermögen

Zu Beginn der Mikroskopie war das Auflösungsvermögen vor allem durch technische Probleme bei der Fertigung von Linsensystemen bestimmt. Vor allem chromatische und sphärische Abberationen waren die Hauptursachen für die Begrenzung des Auflösungsvermögens. Heutzutage sind diese Probleme grösstenteils behoben, so

dass die Auflösung allein durch die Beugung des Lichts begrenzt ist. Aufgrund seiner Welleneigenschaft ist es nicht möglich, das Licht durch eine endlich grosse Linse in einem beliebig kleinen Punkt zu bündeln. Die Verteilung der Lichtintensität im Fokus hat deshalb eine endliche Grösse.

Definiert man die Auflösungsgrenze als den Abstand der Bilder zweier Punktquellen, bei dem das Maximum des einen im ersten Minimum des anderen Beugungsbildes ist, so ergibt sich mit $\rho = 3.832$, für die erste Nullstelle der Besselfunktion

$$\sin \alpha = \frac{3.832}{kR} = 1.22 \frac{\lambda}{2R} = 0.61 \frac{\lambda_0}{nR} \quad (2.5)$$

Hier ist λ die Wellenlänge n der Brechungsindex des Mediums, in dem sich das Objekt befindet und α der halbe Öffnungswinkel des Objektivs. Das Produkt $n \sin \alpha$ wird als die **Numerische Apertur** (NA) des Objektivs bezeichnet und lässt sich auch schreiben als $NA = nR/f$. R und f sind respektive der Radius und die Brennweite der Linse. Für eine hohe Auflösung sind also eine kleine Wellenlänge und eine hohe numerische Apertur erforderlich. Man verwendet daher Immersionsöle mit einem hohen Brechungsindex ($n \approx 1.5$). Bei kurzen Wellenlängen (kleiner als 450 nm) ändert sich der Brechungsindex von Glas stark, so dass die Objektive zwar auf ein monochromatisches Beleuchtungslicht korrigiert werden können, jedoch das breitbandige Fluoreszenzlicht nicht hochauflösend zurück abgebildet werden kann. Die maximale theoretische Apertur beträgt 4π , dem sphärischen Raumwinkel, deren Realisierung an praktische Grenzen stösst. Schritte in die Richtung, die Numerische Apertur optischer Systeme zu erhöhen, wurden mit der Entwicklung der 4π und $4\pi - \text{Theta}$ Mikroskopie vollzogen, die insbesondere Auflösungsverbesserungen in axialer Richtung erzielen [Hel94a, Hel94b]. Die theoretische Auflösung in axialer Richtung konnte experimentell verifiziert werden. Die vorausgesagte laterale Auflösung wurde hingegen nicht erreicht.

2.1.3 Axiale Auflösung

Eine ähnliche Definition für die axiale Auflösung ergibt sich mit dem ersten Minimum des Intensitätsprofils der PSF (Point Spread Function) in axialer Richtung [BW80].

$$z_{min} = \frac{2n\lambda_0}{NA^2} \quad (2.6)$$

Die quadratische Abhängigkeit der axialen Auflösung im Nenner führt zu drastischer Diskriminierung in axialer Richtung bei Erhöhung der Numerischen Apertur, die experimentell durch die 4π -Mikroskopie bestätigt werden konnte. Das Verhältnis von axialer zu lateraler Auflösung beträgt

$$\frac{z_{min}}{r_{Airy}} = \frac{3.28n}{NA} \quad (2.7)$$

2.1.4 Konfokale Mikroskopie (CLSM)

Das erste *Confocal-Stage-Scanning-Optical-System* für transmittiertes und reflektiertes inkohärentes Licht wurde bereits 1957 von M.Minsky zum Patent angemeldet und in den folgenden Jahren weiterentwickelt. Die biologische Anwendung konnte durch die Entwicklung der konfokalen Laser Scanning Fluoreszenz Mikroskopie erheblich erweitert werden ([Cre78], [Cox82], [Res85]).

Bei einer konfokalen Anordnung ist der Beleuchtungs- und Detektionslichtweg symmetrisch. Der prinzipielle Aufbau eines CLSM besteht (Abb. 2.1) aus kohärenter Punktlichtquelle, Beleuchtungsobjektiv, Kollimatorlinse, Lochblende und Detektor. Der kohärente Lichtstrahl der Laserquelle wird mit Hilfe eines dichroischen Spiegels und des Mikroskopobjektivs in die farbstoffmarkierte Probe fokussiert; ggf. wird hier ein erstes Pinhole zur räumlichen Filterung verwendet. Ein Teil des von der angeregten Probe ausgehenden Fluoreszenzlichtes wird per Objektivlinse durch ein Pinhole in den Detektor fokussiert.

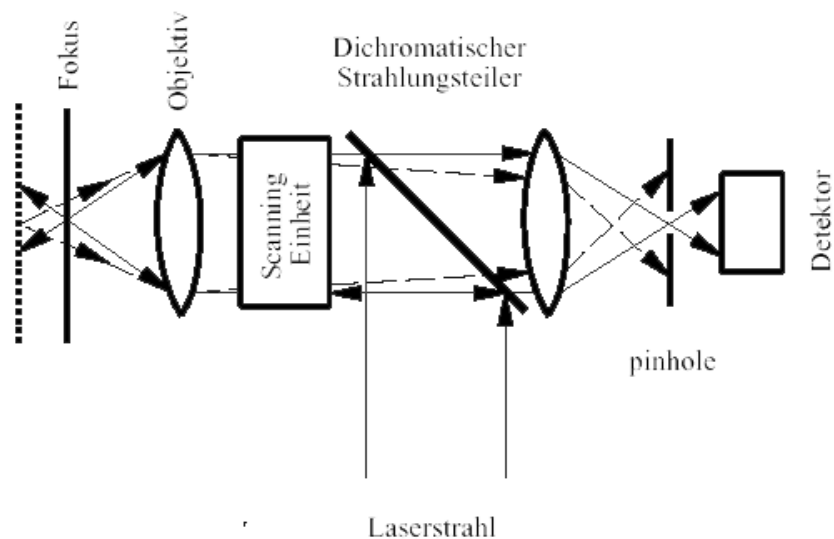


Abb. 2.1: Schematischer Aufbau des konfokalen Laser Scanning Mikroskops

Als Folge der Anordnung wird eine erhöhte laterale Auflösung von bis zu 200 nm und eine verbesserte Tiefenauflösung von etwa 500 nm erzielt, die es ermöglicht, dünne Schnittbildaufnahmen aus Proben zu erhalten. Durch wiederholtes laterales Scannen bei verschiedenen Tiefen entlang der optischen Achse erhält man so ein dreidimensionales Datenvolumen, das mit Hilfe der digitalen Bildverarbeitung ausgewertet werden kann.

Im Vergleich zum konventionellen Fluoreszenzmikroskop gilt für die Intensitätsver-

teilung bei der konfokalen Anordnung in erster Näherung

$$I(z) = \left[\frac{2J_1(z)}{z} \right]^4 \quad (2.8)$$

also im Prinzip das Quadrat von (3.4). Das liegt daran, dass sich die Punkt-Abbildungs-Funktion von Punktlichtquelle und Punktdetektor im Fokus der Linse überlagern. Das laterale Auflösungsvermögen ist aufgrund der vergrößerten Numerischen Apertur mit der Intensitätsverteilung in obiger Gleichung gegenüber konventionellen Aufnahmen um etwa 27% verbessert. Die verbesserte Auflösung entlang der optischen z-Achse ist im Wesentlichen der Lochblende zu verdanken, die direkt vor dem Detektorsystem positioniert wird. Durch sie wird Fluoreszenzlicht aus oberen und unteren angeregten Schichten weitgehend diskriminiert.

Der Einsatz von Lasern als Lichtquelle in konfokalen Mikroskopie in nicht unbedingt notwendig, hat sich aber bewährt, da Laser intensives quasi-monochromatisches Licht liefern, das sich gut fokussieren lässt. Damit werden hohen Photonenraten erzielt, um bei hohen räumlichen Frequenzen nahe der Auflösungsgrenze einen hohen Kontrast, Signal-Rausch Verhältnis und damit Bilder hoher Qualität zu erhalten. Ein Problem von Laserlicht ist seine hohe Kohärenzlänge, die zu Interferenzen ausserhalb des Fokus und in der Bildebene, sogenannten *speckles*, führt, die die Bildqualität beeinträchtigen. Ein Pinhole vor dem Detektor schafft Abhilfe, da nur Licht aus dem Fokussiervolumen in den Detektor gelangen kann.

Ein MFISH taugliches konfokales Laser Scanning Mikroskop steht noch nicht zur Verfügung. Mit herkömmlichen CLSMs lassen sich 3-Farben Experimente jedoch analysieren, wenn man sich auf sehr wenige Farbstoffe beschränkt. Das LEICA TCS 4D Mikroskop besitzt zum Beispiel einen Filterschieber für DAPI, FITC, Cy3 und Cy5. Diese Farbstoffe lassen sich mit den vorhandenen Linien der HeNe und Ar-Ionen Laser des Mikroskops anregen (488 und 568 nm).

Die Analyse dreidimensionaler Interphase Experimente mit mehr als drei Farbstoffen (s. Abs. 6.6, Abb. 6.21) ist zur Zeit nur mit Epifluoreszenzmikroskopie mit anschliessender Dekonvolution für jeden einzelnen Kanal möglich. Ein solches nicht konfokales Mikroskop für die Analyse von zweidimensionalen Aufnahmen wird in Abschnitt 2.4.1 beschrieben.

2.2 Laserinduzierte Fluoreszenz (LIF)

Durch Absorption eines Photons mit der Energie $h\nu$ kann ein Atom oder Molekül in einen höheren Energiezustand angeregt werden.

Die Anregungsenergie kann durch Emission eines Photons $h\nu'$ wieder abgegeben werden. Als *Fluoreszenz* wird die *spontane Emission* von Strahlung bezeichnet. Es besteht auch die Möglichkeit, Moleküle im optisch angeregten Niveau durch Stösse

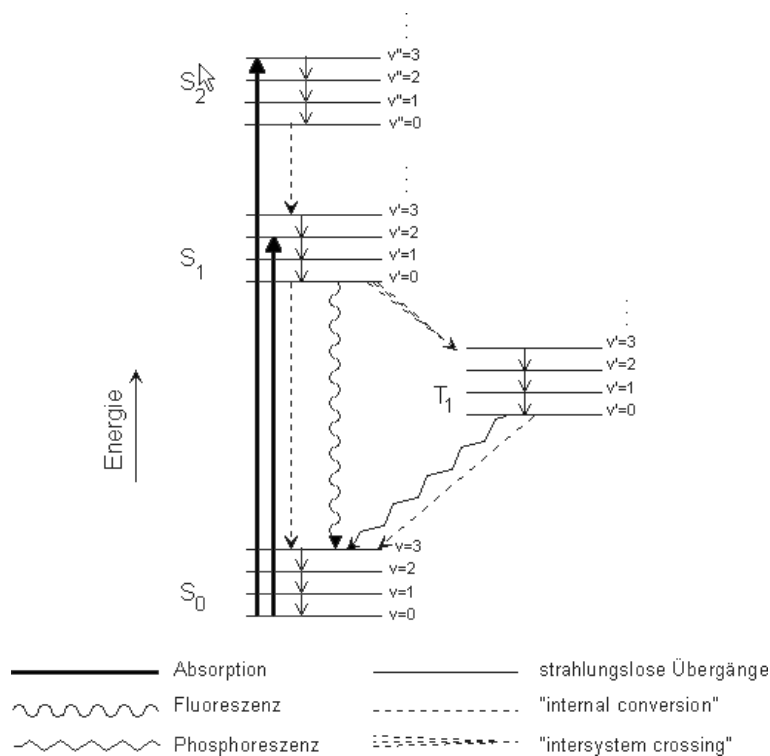


Abb. 2.2: Term-schemata für Laser-induzierte Fluoreszenz.

in andere angeregte Niveaus zu bringen. In diesem Fall spricht man von Phosphoreszenz. Bei der Fluoreszenz entsteht ein quantenmechanisch erlaubter Übergang, während bei der Phosphoreszenz ein quantenmechanisch verbotener Übergang, meistens von einem Triplett- zu einem Singulettzustand, entsteht. Dies äussert sich in der durchschnittlichen Lebenszeit des angeregten Zustands, die bei der Fluoreszenz im Bereich von 10^{-8} s liegt. Typische Lebenszeiten von Phosphoreszenz liegen bei Millisekunden bis Sekunden.

Wenn jedoch die Stossaktivierung vernachlässigt werden kann, wird für jedes absorbierte Photon $h\nu_a$ ein Fluoreszenzphoton $h\nu_{Fl}$ (mit $\nu_{Fl} \leq \nu_a$) ausgesandt. Der Bruchteil aller angeregten Moleküle, die nicht strahlungslos deaktiviert werden, sondern ein Fluoreszenzphoton aussenden, wird *Quantenausbeute* ($\varepsilon \leq 1$) genannt [Dem96].

Benutzt man als Detektionsgerät einen Photomultiplier benutzt, so erhält man bei N_a absorbierten Photonen innerhalb eines Raumwinkels ΔQ

$$N_e = N_a \eta \varepsilon (\Delta Q / 4\pi) \quad (2.9)$$

Photoelektronen, die zu N_e Signalpulsen führen.

- N_e : Anzahl der Photoelektronen
- η : Sekundäremissionskoeffizient
- N_a : Anzahl der absorbierten Elektronen
- ε : Quantenausbeute
- δQ : Raumwinkel

Bei der LIF-Spektroskopie wird die Laserwellenlänge λ_L kontinuierlich durchgestimmt und die vom Detektor erfasste Fluoreszenzleistung $P_{Fl}(\lambda_L)$ als Funktion von λ_L gemessen. Das so erhaltene Spektrum nennt sich *Anregungsspektrum*. Die grösste Empfindlichkeit wird für $\varepsilon = 1$, also stossfreie Bedingungen, erhalten.

Typisches Merkmal aller Farbstoffe ist ein delokalisiertes π -Elektronensystem, das durch Kohlenstoffringstrukturen entsteht. Die Elektronen dieses Systems können leicht angeregt werden und in Abhängigkeit der Energie der emittierten Elektronen entstehen dann Farben verschiedener Wellenlängen.

2.3 Grenzen der konfokalen Mikroskopie

In der (vor allem dreidimensionalen) Mikroskopie gilt es, einen möglichst kleinen Bereich eines Präparats mit einer möglichst hohen Photonendichte anzuregen, ohne es dabei zu beschädigen, und nur aus diesem optisch angeregten Volumen eine möglichst hohe Photonenzahl zu detektieren. Die Anzahl detektierter Photonen pro Pixel beträgt dabei nur einen Bruchteil ($< 1\%$) der erzeugten Anregungsprozesse [Paw95]. Ein fundamentales Verständnis dieser Prozesse ist für einen effizienten Umgang mit einem konfokalen Mikroskop unerlässlich, um optimale Auflösung und Bildqualität zu erzielen. Wie sich zeigen wird, herrscht eine Art Unschärfepinzipp bei der Bildaufnahme, das den Anwender zwingt, einen bestmöglichen Kompromiss zwischen Auflösung, Geschwindigkeit, Kontrast, Beschädigung der Probe, Beleuchtungsintensität, Quantenausbeute, sowie Sättigungs- und Bleicheffekten der Farbstoffe zu finden.

2.3.1 Photonenstatistik

Um bei einer Messung eine hohe Genauigkeit zu erzielen, ist es notwendig, so viele Photonen wie möglich zu detektieren. Photonen unterliegen der Poissonstatistik. Die Poisson-Verteilung beschreibt die Verteilung von unendlich vielen Messungen seltener Ereignisse, deren Nichteintreten wesentlich wahrscheinlicher ist, als deren Eintreten.

$$P(k) = \frac{n^k}{k!} \exp^{-n} \quad (2.10)$$

Die Wahrscheinlichkeit, dass eine gemessene Zählrate bei einer mittleren Zählrate von n Photonen zwischen $n - \sqrt{n}$ und $n + \sqrt{n}$ liegt, beträgt 63%, die Standardab-

weichung wird allein durch den Mittelwert n bestimmt ($\sigma = \sqrt{n}$). Hat man beispielsweise eine mittlere Photonenzahl von 100 gemessen, liegen 63% der Messungen zwischen 90 und 110. Die Messung in diesem Beispiel weist eine 10%ige (\sqrt{n}/n) Statistik auf. Weitaus bessere Statistiken erhält man mit hohen Photonenzahlen. Mit $n = 90000$ ist die Statistik auf 0.3% verbessert. Die Genauigkeit der Messung hängt von der Anzahl der gemessenen Photonen ab. Die Photonostatistik stellt also eine Prozessinterne Rauschquelle dar, die hohe Zählraten für zuverlässige Messungen notwendig macht [Paw95]. Bei hohen Photonenzahlen kann die Poissonverteilung durch eine Gaussverteilung angenähert werden.

2.3.2 Sättigungs- und Bleicheffekte

Bei nicht zu starken Beleuchtungsintensitäten ($< 1mW$), ist eine Möglichkeit, höhere Emissionsraten und damit höhere Detektionsraten zu erhalten, die Beleuchtungsintensität zu erhöhen. Die Absorptionsrate hängt linear von der Intensität ab, mit der Folge, dass mehr Farbstoffmoleküle pro Volumeneinheit angeregt werden, und mit wachsender Besetzung angeregter Zustände steigt die Anzahl spontaner Photonenemissionen. Damit stimulierte Emissionsprozesse die Anregungseffizienz nicht beeinträchtigen, müssen die angeregten Elektronen schneller als die spontane Übergangswahrscheinlichkeit ($\approx 10^{-8}s$) in tiefere Zustände des Anregungsbandes relaxieren ($\approx 10^{-12}s$), um von dort spontan unter Emission eines Photons in den Grundzustand überzugehen. Andernfalls ständen möglicherweise nicht alle Atome zur spontanen Emission und damit zur Detektion zur Verfügung.

Bei Anregungsintensitäten $> 1mW$ ändert sich die lineare Abhängigkeit der Absorptionsraten im Fokusbereich, da mit zunehmender Intensität immer mehr Moleküle pro Volumeneinheit angeregt und schliesslich ein Zustand erreicht wird, in dem im Fokalbereich nicht mehr genügend Farbstoffmoleküle zur Anregung zur Verfügung stehen. Da aber die Anregungsrate von Farbstoffmolekülen ausserhalb des Fokus und damit das Anregungsvolumen ansteigt, bedeutet dies eine zunehmende Verschlechterung der Auflösung. Wichtig ist, dass das Fluoreszenzsignal nicht mehr nur von der Farbstoffkonzentration abhängt, aber genau auf dieser Annahme baut die Digitale Bildanalyse auf! Ein weiteres Problem ergibt sich durch Bleicheffekte. Farbstoffmoleküle verändern sich photochemisch und verlieren ihre Eigenschaft zu fluoreszieren. Wenn mit zunehmender Intensität die Bleichrate ansteigt, wird dies vor allem bei Volumenaufnahmen zum Problem. Da beim optical sectioning ein Teil des Anregungslichtes für jeden abgetasteten Punkt immer durch das gesamte Präparat läuft, verschlechtert sich die Photonenausbeute und damit die Bildqualität mit zunehmender abgetasteter Schicht.

2.3.3 Nyquist Theorem

Eine bei Aufnahmen am Mikroskop oft nicht beachtete Tatsache ist, dass es für ein optisches System mit einer Auflösung, die (unter der Voraussetzung, dass das System aberrationsfrei ist) durch die Numerische Apertur der Linsen, den Brechungsindex des Mediums zwischen Objektiv und Präparat und die Anregungswellenlänge definiert wird, *nur eine* optimale Pixel-/Voxelgrösse existiert. Die Pixelgrösse ist durch lateralen Abstand zweier benachbarter Abtastungen gegeben. Das Nyquist Abtasttheorem besagt nun, *dass, um eine Struktur bestimmter Grösse auflösen zu können, diese Struktur 2.3 Mal abgetastet werden muss*. Für die Pixelgrösse bei Aufnahmen bedeutet das nun folgendes:

Bei gewähltem Objektiv, und damit definierter Auflösung, muss das Präparat in einem Abstand vom Auflösung/2.3 abgetastet werden. Eine etwas kleinere Pixelgrösse führt in der Regel zu einer verbesserten Bildqualität, da Übergänge weicher werden, allerdings ohne Informationsgewinn. Starkes Überabtasten kostet viel Zeit und begünstigt Sättigungs- und Bleicheffekte im Präparat. Unterabtastung hingegen bedeutet einen Informationsverlust und sollte vermieden werden. Durch Unterabtastung können Strukturen abgebildet werden, die nicht vorhanden sind. Diesen Effekt nennt man *Aliasing*. Die grösste räumliche Frequenz der Aufnahme ergibt sich aus $1/(2.3 \cdot \text{Auflösung})$.

2.4 Aufnahme von M-FISH-Bildern

Zwei Mikroskop-Systeme mit unterschiedlicher Detektionstechnik haben sich am Markt durchgesetzt. Das MFISH System von Leica Microsystems, und das SKY System von Applied Spectral Imaging (ASI). Ersteres verwendet spezifische Filtersätze, um die Spektralinformation in jedem Pixel zu detektieren. Die Anzahl der gescannten Bildern entspricht dabei der Anzahl verwendeter Fluorochrome. Die im Rahmen dieser Arbeit analysierten Zellen waren ausschliesslich von diesem Typ. Das SKY System beruht auf der Fourier-Spektroskopie. Für jedes Pixel wird die Autokorrelationsfunktion des Signals gemessen. Die Fouriertransformierte dieser Funktion ergibt das gesamte Spektrum in jedem Pixel. Damit ist die spektrale Auflösung dieses Systems wesentlich höher (bis zu $\approx 5nm$). Beide Systeme haben ihre vor und Nachteile, die im Kapitel 6 (*Diskussion*) besprochen werden.

2.4.1 Das MFISH System

Dieses System (z.B. Leica DMRXA-RF8 Epifluoreszenz Mikroskop) verwendet einen Satz spezifischer Anregungs- und Emissionsfilter sowie dichroischem Strahlteiler, die zur automatischen Bildaufnahme auf einer motorisierten Drehscheibe aufgebracht sind (bis zu 8 Filter). Abbildung 2.3 zeigt das Prinzip der Aufnahme schematisch.

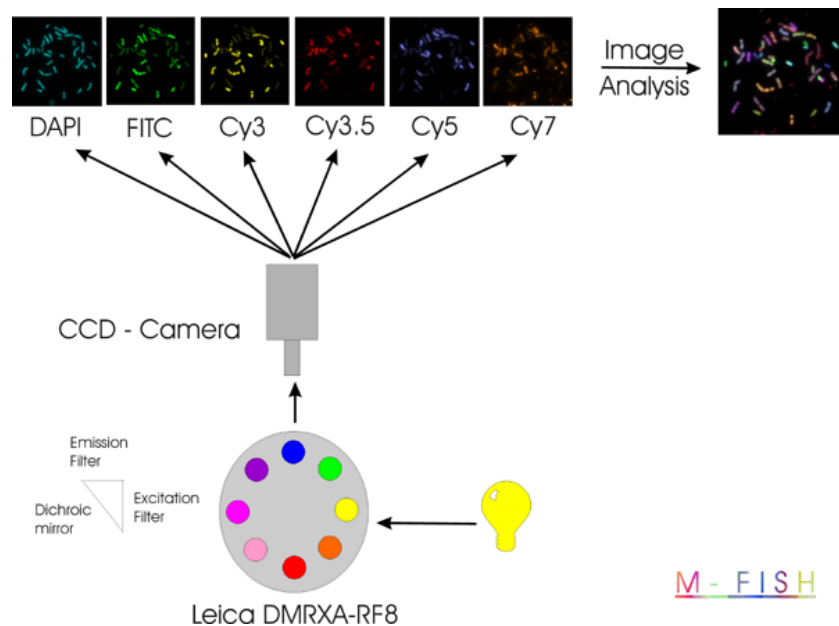


Abb. 2.3: Prinzip der Bildaufnahme am Leica DMXR-8.(Quelle: [Eil98])

Die Bildaufnahme einer kompletten Metaphase wird dadurch in wenigen Sekunden möglich. Diese Filter sind jeweils spezifisch für die im M-FISH Experiment verwendeten Fluorochrome, um optimale Anregungs- und Emissionsraten, bei minimalem Crosstalk zu erzielen, da sich die Anregungs- und Emissionsspektren der Fluorochrome in der Regel überlagern. Diese Bandpassfilter dürfen jedoch nicht zu schmalbandig sein, da die Belichtungszeiten mit abnehmender Bandbreite zunehmen ($\approx 2min$) [Eil98]. Folgende Fluorochrome haben sich zum Einsatz in M-FISH Experimenten bewährt:

- DAPI, ein intensiver Farbstoff, der zur Gegenfärbung der gesamten DNA verwendet wird, mit Absorptions- und Emissionsmaxima bei 350 und 460nm
- DEAC (426 und 480nm)
- Fluorescein (FITC) (490 und 520nm)
- sowie die Cyanine Cy3 (554 und 568nm)
- Cy3.5 (581 und 588nm)
- Cy5 (652 und 672nm)
- Cy5.5 (682 und 703nm)

- Cy7 (755 und 778nm)

Die Bildaufnahme erfolgt über eine Sensys CCD-Kamera von Photometrics (Kodak KAF 1400 Chip). Standardmässig wird ein Objektiv mit 100x Vergrößerung (Plan Apo, NA 1.4) verwendet. Bei ausgedehnten Metaphasen sind 63x (Plan Apo, NA 1.32) Objective üblich. Als Lichtquellen besitzt das Mikroskop eine 75W Xenon Bogenlampe und eine 100W Quecksilberdampf-Lampe. Tabelle (2.1) zeigt die Spezifikationen der spezifischen Filtersätze für ausgewählte Fluorochrome.

Komponente	DAPI	FITC	Cy3	Cy3.5	Cy5	Cy5.5	Cy7
Anregungsfilter	360±20	475±15	546±5.5	580±5	602±19	682±6	740±17.5
Strahlteiler	400	497	557	593	647	697	765
Emissionsfilter	460±20	522±20	567±7.5	612±15	667±15	720±20	790±20

Tab. 2.1: Epicube Filterkonfiguration für das Leica DMRXA-RF8 Mikroskop in nm

Die Belichtungszeiten für die einzelnen Fluorochrome hängen im wesentlichen von der Wahl der Filtersätze und ihrer Bandbreite, der Empfindlichkeit der Kamera, dem Spektrum der Lichtquelle sowie dem Hybridisierungsprotokoll, also der Färbefizienz der Probe ab. Typische Belichtungszeiten für die Fluorochrome sind 0.5s für DAPI, 3s für FITC, 3s für Cy3, 0.5s für Cy3.5, 4s für Cy5 und 7s für Cy7. Die gesamte Zeit für die Aufnahme einer Metaphase beträgt somit etwa 20s. Mit diesen Belichtungszeiten wird etwa die Hälfte der 12Bit Dynamik der CCD Kamera (4096 max. Intensität) erreicht. Die Belichtungszeiten mit der Quecksilberlampe sind in der Regel kürzer als die mit der Xenon Bogenlampe. Sie wird daher vorzugsweise bei Aufnahmen verwendet.

Pixelshifts können bei diesem System durch den Wechsel der Filtersätze entstehen und betragen im Mittel 1-2 Pixel in x- und y-Richtung. Sie werden automatisch durch geeignete Korrektur-Algorithmen beseitigt [Spe96a, Eil98].

2.4.2 Das SKY System

Das SKY System (SpectraCube) beruht auf einem anderen Ansatz [Sch96, Gar96]. Statt spezifische Filtersätze zu verwenden, wird mit Hilfe der Fourier-Spektroskopie das gesamte Spektrum in jedem Pixel gemessen. Dazu wird das Fluoreszenzlicht mit Hilfe eines Kollimators in ein Sagnac Interferometer geführt, und anschliessend in eine CCD Camera fokussiert. Interferometer spielen eine zentrale Rolle in der Fourier Spektroskopie. Das Prinzip wird im folgenden erläutert.

Fourier Spektroskopie

Der einfallende Lichtstrahl wird im Interferometer in zwei kohärente Strahlen aufgeteilt, für die eine variable optische Weglänge generiert wird. Dadurch wird der eine Strahl gegenüber der Referenzwelle in der Phase verschoben. Die kohärenten Strahlen werden wieder vereint, die Intensität für unterschiedliche optische Wegunterschiede detektiert und so das Autokorrelationsignal in jedem Pixel abgetastet. Eine anschließende Fouriertransformation liefert das gesuchte Spektrum in jedem Pixel.

Eine (nicht monochromatische) Lichtwelle mit der Feldstärke $E(t)$ lässt sich beschreiben durch:

$$E(t) = \int_k A(k) \cos(2\pi kx - \omega t) \delta k \quad (2.11)$$

Hier ist $k = 1/\lambda$ die Wellenzahl und $A(k)$ die Amplitude der Wellenkomponente mit der Wellenzahl k . Die mittlere Intensität dieser Lichtwelle ist gegeben durch das Integral

$$\bar{I} = \int_k A(k)^2 \delta k \quad (2.12)$$

Die Intensität für zwei interferierende Strahlen (derselben Lichtwelle) mit optischem Wegunterschied (phasenverschobene Wellen) L beträgt

$$I(L) = 0.5 \cdot \int_k A(k)^2 \delta k + 0.5 \cdot \int_k I(k) \cos(2\pi kL) \delta k \quad (2.13)$$

Der erste Summand ist nichts anderes als die Konstante \bar{I} . Der zweite Term ist der Interferenzterm, der die Spektralinformation in Form des optischen Wegunterschieds L enthält. Dieser ist aber nichts anderes als der Realteil der Fouriertransformierten von $I(k)$. Die Gleichung für die gemessene Intensität beim Wegunterschied L lässt sich also schreiben als:

$$I(L) = C + \Re(FT[I(k)]) \quad (2.14)$$

Misst man die Intensität in einem Pixel sukzessive für unterschiedliche L , erhält man ein Interferogramm (Autokorrelationsfunktion), das die Fouriertransformierte des Spektrums ist. In der Praxis wird die Intensität für unterschiedliche (diskrete) Wegunterschiede gemessen, so dass das Spektrum in jedem Pixel die Fouriertransformierte einer diskreten Funktion $I(L)$ ist:

$$I(k) = \sum_L I(L) \exp(-i2\pi kL) \Delta l \quad (2.15)$$

Der diskreten Fouriertransformation vorgeschaltet werden eine Phasenkorrektur, Apodisierung und das Auffüllen mit Nullen. Diese Schritte sind notwendig, um

numerische Probleme zu umgehen und das Hineinfalten nicht vorhandener Frequenzen zu unterdrücken.

Das Sagnac Interferometer basiert auf dem Prinzip rotierender Spiegel, um einen optischen Wegunterschied zu erzeugen. Das Prinzip rotierender (paralleler) Spiegel in Interferometern wurde 1981 von [Yas81] eingeführt. Für kleine Winkel ($\alpha \ll 1 \text{ rad}$) ist der optische Wegunterschied dem Winkel proportional:

$$L \cong c \cdot \alpha \quad \text{für } \alpha \ll 1 \text{ rad} \quad (2.16)$$

Die spektrale Auflösung beträgt für das Sagnac Interferometer etwa 0.01λ , also 4nm bei 400nm. In der Praxis wird aber mit einer wesentlich geringeren spektralen Auflösung gearbeitet, d.h die Intensität $I(L)$ wird seltener (bei weniger optischen Weglängen) gemessen, um die Zeit für die Bildaufnahme zu verkürzen. Üblich sind 10-15nm bei 400nm, man erhält damit 5-10 mal soviel Farbinformation pro Pixel wie beim Filterbasierten MFISH System. Die Aufnahmezeit für ein Bild bei einer Auflösung von 10nm beträgt ca. 60s und steigt linear mit der Auflösung auf 120s bei 5nm an [Gar96].

Kapitel 3

Grundlagen der Bildverarbeitung

3.1 Klassifizierung

Als *Klassifizierung* bezeichnet man die Zuweisung von Objekten anhand bestimmter Merkmale zu bekannten Klassen. Sie ist nicht spezifisch für die Bildanalyse und findet eine breite Anwendung auf verschiedenen Gebieten. In der Bildanalyse lassen sich zwei Typen von Klassifizierungen unterscheiden: die Pixelorientierte und Objektorientierte. In komplexen Fällen gelingt die Objektsegmentierung nicht mit einem einzigen Merkmal. Dann müssen mehrere Eigenschaften verwendet werden, sowie ein Klassifizierungsprozess, der die Bildpunkte den Objekten zuweist [Jäh97]. Falls sich die verschiedenen Objekte gut vom Hintergrund unterscheiden und sich nicht berühren und überlappen, ist die sehr viel einfachere objektbasierte Klassifikation die bessere Wahl. Der Daten- und Rechenaufwand ist sehr viel geringer, da die auf Pixeln beruhenden Merkmale der Objekte über das gesamte Objekt gemittelt und als Objekteigenschaft verwendet werden können. Parameter wie Grösse und Form, sowie die Orientierung sind weitere Eigenschaften, die als Merkmale dienen können. Manchmal ist auch die Kombination beider Verfahren sinnvoll: Die pixelbasierte zur Trennung der Objekte und anschliessend die objektbasierte zur Klassifizierung. Die Analyse von M-FISH Bildern lässt sich als Klassifizierungsproblem verstehen. Interpretiert man das Labeling-Schema (s. Abb. 1.6) als Klassen, die die Chromosomen repräsentieren, besteht die Hauptaufgabe darin, jedes Pixel anhand seiner Farbinformation (Merkmale) einer dieser Farbklassen zuzuordnen.

3.2 Clusteranalyse

Die n unterschiedlichen Merkmale eines Datensatzes spannen einen n -dimensionalen Raum, den *Merkmalsraum* auf. Jedes Pixel oder Objekt wird in diesem Raum durch einen *Merkmalsvektor* (oder auch *Datum*) repräsentiert. Wird ein Objekt oder eine

Objektklasse durch die Merkmale gut beschrieben, sollten alle Merkmalsvektoren dieser Klasse im Merkmalsraum eng beieinander liegen. Besitzt eine Objektklasse eine enge Verteilung im Merkmalsraum, spricht man von einem *Cluster*. Sind die Cluster gut getrennt, lassen sich die Objekte gut in bestimmte Klassen separieren (Abb. 4.6). Sind die Merkmale hingegen schlecht, kann dies zu fehlenden oder überlappenden Clustern führen, was eine fehlerfreie Klassifizierung erschwert.

Verfahren zur Clusteranalyse teilen einen nicht klassifizierten Datensatz in homogene Cluster ein. Ähnliche Daten werden also demselben Cluster zugeordnet. Die Daten werden dabei den Clustern aufgrund von Zugehörigkeitsgraden zugeordnet. Während diese bei deterministischen Verfahren binär sind, beschreiben sie bei probabilistischen und fuzzybasierten Methoden die Wahrscheinlichkeit der Zugehörigkeit eines Datums (oder Merkmalsvektors) zu einem Cluster.

Zwei Arten von Clusterverfahren gilt es dabei zu unterscheiden: die *überwachte* und die *unüberwachte* Clusteranalyse. Bei ersteren wird apriori Wissen in das System eingearbeitet. Unüberwachte Verfahren hingegen clustern den Datenraum blind, und kommen immer dann zum Einsatz, wenn kein Vorwissen über die Daten und das zu erwartende Ergebnis vorhanden ist. Ist hingegen solches Wissen vorweg bekannt, so sollte dieses unbedingt bei einer Clusteranalyse verwendet werden. Unüberwachte Verfahren sind in der Regel sehr viel rechenaufwendiger und liefern in solchen Fällen selten gleichwertige Ergebnisse.

Die Analyse der Daten mit Clusterverfahren basiert auf den Abständen der Daten zu den Clusterzentren, den sogenannten *Centroiden*. Die Cluster werden dabei so bestimmt, dass für alle Cluster die Summe der Abstände der Punkte zum Centroiden innerhalb eines Clusters minimal, und gleichzeitig der Abstand der Cluster untereinander maximal wird. Dabei liegt die Annahme zugrunde, dass alle Daten das gleiche Gewicht haben und einem Cluster zugeordnet werden, und dass alle Cluster Daten enthalten.

Die Clusteranalyse eines Datensatzes $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ in k Cluster löst also folgendes Optimierungsproblem:

$$C^* = \underset{c_i}{\operatorname{argmin}} \sum_{i=1}^k \sum_{j=1}^n d^2(\mathbf{c}_i, \mathbf{x}_j) \quad (3.1)$$

Dabei ist $C = \{\mathbf{c}_1, \dots, \mathbf{c}_k\}$ die Menge der Cluster bzw. ihrer Centroide, und $d(\mathbf{c}_i, \mathbf{x}_j)$ der Abstand zwischen Cluster \mathbf{c}_i und Datum \mathbf{x}_j .

Dazu muss ein geeignetes *Distanzmass* eingeführt werden, das das Clusterergebnis je nach Datenstruktur erheblich beeinflusst. Sehr häufig wird die Euklidische Metrik verwendet, oder bei objektbasierten Klassifizierungsproblemen auch ein binäres Abstandsmass. Die Minimierung des Euklidischen Abstands bedeutet im übrigen nichts anderes als die Berechnung des Mittelwertes (Schwerpunktes).

3.2.1 K-Means-Algorithmus

Das bekannteste Clusterverfahren ist der k-means Algorithmus. Er unterteilt den Datensatz in Cluster gleicher Form, die durch das Abstandsmass bestimmt wird. Häufig wird hier der Euklidische Abstand verwendet. Das führt dazu, dass der Datenraum in hyperkugelförmige Cluster unterteilt wird. Die Cluster werden nur durch ihre Centroiden \mathbf{c}_i beschrieben. Sie berechnen sich aus:

$$\mathbf{c}_i = \frac{\sum_{j=1}^n \mathbf{x}_j}{n} \quad (3.2)$$

Der k-means Algorithmus weist gute Klassifikationsresultate aus, wenn die Cluster eine ähnliche Form und Grösse besitzen, und ihre Form vor der Analyse bekannt ist.

3.2.2 Fuzzy-Clusteranalyse

Fuzzy-Cluster-Algorithmen zeichnen sich dadurch aus, dass die Daten den Clustern aufgrund von Zugehörigkeitsgraden zugeordnet werden, die nicht binär sind, sondern Werte von 0...1 annehmen können. Handelt es sich um *probabilistische* Clusterverfahren, können sie als Wahrscheinlichkeiten der Zugehörigkeit eines Datums zu einem Cluster interpretiert werden. Das Optimierungsproblem in Gl.(3.1) wird zu:

$$C^* = \operatorname{argmin}_{\mathbf{c}_i} \sum_{i=1}^k \sum_{j=1}^n w_{ij}^m d^2(\mathbf{c}_i, \mathbf{x}_j) \quad (3.3)$$

Dabei müssen die $\{w_{ij}\}$ folgende Nebenbedingungen erfüllen:

$$\sum_{j=1}^n w_{ij} > 0 \quad \forall i, i = 1 \dots k \quad (3.4)$$

$$\sum_{i=1}^k w_{ij} = 1 \quad \forall j, j = 1 \dots n \quad (3.5)$$

Die $\{w_{ij}\}$ sind die Menge der Zugehörigkeitsgrade der Daten zu den Clustern, $C = \{\mathbf{c}_1, \dots, \mathbf{c}_k\}$ die Menge der Cluster bzw. ihrer Centroiden, und $d(\mathbf{c}_i, \mathbf{x}_j)$ der Abstand zwischen Cluster \mathbf{c}_i und Datum \mathbf{x}_j . Der Exponent m ist ein Fuzzifizier, der den Einfluss von Daten auf Cluster mit geringem Zugehörigkeitsgrad bestimmt.

Eine umfassende Behandlung zum Thema Fuzzy-Clusteranalyse findet sich in [Hpp99]. Einige bekannte Vertreter dieser Clusterverfahren stelle ich im folgenden vor.

Fuzzy-C-Means-Algorithmus

Das bekannteste Fuzzy-Clusterverfahren ist der Fuzzy-C-Means-Algorithmus. Auch er unterteilt, wie der k-means, den Datensatz in Cluster gleicher Form, die durch das Abstandsmass bestimmt wird. Meistens wird auch hier der Euklidische Abstand verwendet. Der Datenraum wird dementsprechend in hyperkugelförmige Cluster unterteilt. Die Cluster c_i werden nur durch ihre Centroide \mathbf{c}_i beschrieben, die sich berechnen aus:

$$\mathbf{c}_i = \frac{\sum_{j=1}^n w_{ij}^m \mathbf{x}_j}{\sum_{j=1}^n w_{ij}} \quad (3.6)$$

Der Fuzzy-C-Means-Algorithmus weist gute Klassifikationsresultate aus, wenn die Cluster eine ähnliche Form und Grösse besitzen, und ihre Form vor der Analyse bekannt ist. Trifft jedoch eines dieser Aspekte auf den Datensatz nicht zu, erreicht man mit Clusterverfahren, die neben dem Centroiden auch noch die Form und Grösse der Cluster bestimmen, wesentlich bessere Resultate.

Gustafson-Kessel-Algorithmus

Einen wesentlich flexibleren Ansatz als das Fuzzy-C-Means-Verfahren bietet der Gustafson-Kessel-Algorithmus. Er ist in der Lage, ellipsoide Cluster gleicher Grösse zu erkennen, die sich aber in ihrer ellipsoiden Form unterscheiden. Für jeden Cluster c_i wird neben dem Centroiden \mathbf{c}_i auch eine $n \times n$ Normmatrix \mathbf{A}_i bestimmt, die bei der Berechnung der Abstände verwendet wird (Gl.(3.9)). Die Clusterzentren \mathbf{c}_i werden wie beim Fuzzy-C-Means nach Gl.(3.6) berechnet. Die Normmatrix wird mit Hilfe der Kovarianzmatrix \mathbf{C}_i der Centroide bestimmt:

$$\mathbf{A}_i = \frac{(\det \mathbf{C}_i)^{1/n}}{\mathbf{C}_i^{-1}} \quad (3.7)$$

$$\mathbf{C}_i = \frac{\sum_{j=1}^n w_{ij}^m (\mathbf{x}_j - \mathbf{c}_i)(\mathbf{x}_j - \mathbf{c}_i)^T}{\sum_{j=1}^n w_{ij}} \quad (3.8)$$

Jedes Cluster wird nun durch das Tupel $(\mathbf{c}_i, \mathbf{C}_i)$ beschrieben. Der Abstand von Datum \mathbf{x}_j zu Cluster c_i wird durch

$$d^2(\mathbf{x}_j, \mathbf{c}_i) = (\mathbf{x}_j - \mathbf{c}_i)^T \mathbf{A}_i (\mathbf{x}_j - \mathbf{c}_i) \quad (3.9)$$

berechnet, um die Zielfunktion in Gl.(3.3) zu minimieren.

Gath-Geva-Algorithmus

Der Gath-Geva-Algorithmus stellt eine weitere Verfeinerung des Fuzzy-C-Means-Algorithmus dar. Bei diesem Verfahren wird der Datenraum in ellipsoide Cluster unterschiedlicher Grösse unterteilt. Form und Grösse jedes Clusters werden dabei individuell ermittelt.

Dieser Ansatz interpretiert den Datenraum als p -dimensionale Normalverteilung: Die Annahme ist, dass die *a posteriori* Wahrscheinlichkeit, dass ein Datum \mathbf{x}_j eine Realisierung der i -ten Normalverteilung ist, sich umgekehrt proportional zum Abstand zu diesem Cluster c_i verhält. Die *a priori* Wahrscheinlichkeit P_i der Zugehörigkeit eines beliebigen Datums zu einem Cluster c_i ist hier proportional zu der relativen Grösse des Clusters und wird aus dem Quotienten der Anzahl der Daten im Cluster c_i zu der Gesamtanzahl an Daten geschätzt:

$$P_i = \frac{\sum_{j=1}^n w_{ij}^m}{\sum_{i=1}^k \sum_{j=1}^n w_{ij}^m} \quad (3.10)$$

Die Berechnung der Abstände erfolgt durch:

$$d^2(\mathbf{x}_j, \mathbf{c}_i) = \frac{(\det \mathbf{C}_i)^{1/2}}{P_i} \exp \frac{(\mathbf{x}_j - \mathbf{c}_i)^T \mathbf{C}_i^{-1} (\mathbf{x}_j - \mathbf{c}_i)}{2} \quad (3.11)$$

Dabei werden die Kovarianzmatrix \mathbf{C}_i und die Centroiden \mathbf{c}_i wie beim Gustafson-Kessel-Algorithmus nach Gleichungen (3.8) und (3.6) berechnet.

3.3 Regionenwachstum

Die Aufgabe der Objekterkennung und Segmentierung besteht darin, auf der Grundlage eines Schwellwertes ein Bild in mehrere disjunkte Bildregionen zu zerlegen. Dies kann beispielsweise durch kantenbasierte Verfahren, bei denen die Zerlegung durch das Auffinden von Kanten im Bild erfolgt, oder durch regionenbasierte Verfahren, bei denen die Zerlegung in Flächen erfolgt, geschehen [Kle92], [Pinz94]. Objekte sind ihrer Natur nach zusammenhängend, daher lassen sich solche strukturellen Informationen in der Regel durch Betrachtung der Grauwertrelationen benachbarter Pixel erkennen.

Das grundlegende Prinzip des Regionenwachstums ist, benachbarte Pixel, für die eine Ähnlichkeitsrelation zutrifft, zu bestimmten Bereichen zusammenzufassen. Dazu muss mindestens ein Startpunkt pro zu segmentierender Region bestimmt werden. Anschliessend findet eine wiederholte Zuordnung derjenigen Nachbarpixel statt, für die das festgelegte Ähnlichkeitskriterium zutrifft, bis alle Bildpunkte verkettet sind [Dic98]. Die unterschiedlichen Arten des Region Growing werden im folgenden vorgestellt.

3.3.1 Einfache Verkettung

Die einfache Verkettung ist ein sehr einfaches und schnelles Verfahren. Jedes Pixel ist dabei ein Knoten in einem Graphen. Benachbarte Pixel, die sich in einem Ähnlichkeitskriterium weniger als eine bestimmte Schwelle unterscheiden, werden im Graphen durch eine Kante verbunden. Dieses Verfahren führt jedoch sehr häufig zu Verkettungsfehlern. Farblich unterschiedliche Regionen können leicht durch eine Folge langsam ineinander übergehender Farben verbunden werden. Es genügt eine einzige Kante im Graphen, um zwei benachbarte Regionen miteinander zu verschmelzen.

Aufgrund dieser unangenehmen Eigenschaft ist dieses Verfahren für exakte Analysen bei komplexen Bildszenen untauglich.

3.3.2 Hybride Verkettung

Hier wird die Entscheidung, ob zwei benachbarte Pixel miteinander verknüpft werden, anhand der Eigenschaft der $N \times N$ Nachbarschaft der betrachteten Pixel getroffen. Beiden Pixeln wird jeweils der Eigenschaftsvektor ihrer Nachbarschaft zugewiesen. Diese Art der Verkettung hat einen glättenden Effekt: Die Ähnlichkeit der Pixel wird durch die Ähnlichkeit der Nachbarschaft bestimmt.

Ein Beispiel für eine solche Verkettung ist die der gemeinsamen nächsten Nachbarn, die von Levine und Leemet vorgeschlagen wurde [Lev76]. Aus den Listen der $N \times N$ Nachbarschaft beider Pixel werden zunächst die jeweils ähnlichsten Pixel in einer Nachbarschaftsliste bestimmt. Die beiden Pixel werden nun verknüpft, wenn alle Pixel einer Nachbarschaftsliste auch in der Nachbarschaftsliste des anderen Pixels stehen und es genug Pixel gibt, also die Anzahl gemeinsamer Nachbarn genügend gross ist. Die Anzahl der Verkettungsfehler ist bei diesem Verfahren deutlich geringer als bei der einfachen Verkettung.

3.3.3 Zentroide Verkettung

Bei dieser Art des Region Growing werden nicht mehr, wie bei den beiden zuvor beschriebenen Methoden, paarweise Pixel miteinander verglichen. Statt dessen wird ein Pixel mit dem Eigenschaftsvektor einer schon bestehenden benachbarten Region verglichen, die nicht notwendigerweise fertig zu sein braucht. Ist der Unterschied im Ähnlichkeitskriterium unterhalb einer bestimmten Schwelle, wird das Pixel mit der Region verschmolzen, mit ihr durch eine Kante im Graphen verkettet. Der Eigenschaftsvektor der Region wird anschliessend neu berechnet.

Levine und Shaheen haben ein Ähnlichkeitsmass für Farbbilder definiert [Lev81]. Bei ihnen wird das Pixel der Region hinzugefügt, wenn der Unterschied zwischen Pixel-Farbvektor und Regionen-Farbvektor in jedem Farbkanal unterhalb einer gewissen

Schwelle liegt. Diese Schwelle wird dabei dynamisch an die Standardabweichung der aktuellen Farbverteilung der Region angepasst. Sie wird mit wachsender Streuung der Verteilung kleiner.

In seinem Überblick über Region Growing Algorithmen weist Zucker auf einige Probleme solcher zentroid verketteter Algorithmen hin [Zuc76]. Der Vorteil dieser sequentiellen Algorithmen liegt darin, dass schon gewonnene Information nachfolgende Schritte beeinflussen kann. Die Wahl des Startpunktes für eine Region beeinflusst die Wahl des Startpunktes für die nächste. Die Reihenfolge der verarbeiteten Bildpunkte beeinflusst das Ergebnis der Segmentierung. Für den Erfolg des Verfahrens ist die Wahl des Ähnlichkeitskriteriums und die Wahl der Schwellen von entscheidender Bedeutung.

3.4 Rauschen

Rauschen ist ein fundamentales Problem der Bildanalyse. Die tatsächliche Bildinformation wird durch Rauschprozesse überlagert und damit verfälscht. Dabei kann es verschiedene Formen und Quellen des Rauschens geben, oft sind sie nur schwer zu erkennen und nicht quantifizierbar. Die bei der Bildaufnahme detektierten Photonen unterliegen der Poissonstatistik. Man muss bei der Aufnahme daher dafür sorgen, möglichst viele Photonen für jeden Bildpunkt aufzufangen, um eine zuverlässige Aussage über den Informationsgehalt am betrachteten Punkt des Objekts treffen zu können. Durch mehrfaches Aufnehmen und Mitteln wird ein besseres Signal-Rausch-Verhältnis erzielt. Das Ausbleichen der Farbstoffe setzt dieser Form der Rauschminderung jedoch praktische Grenzen. In den letzten Jahren haben sich CCD Chips als Detektionsmedien immer mehr durchgesetzt. Sie haben den Vorteil, dass nur eine Aufnahme vom gesamten Bildvolumen gemacht wird. Ihre Geometrie bestimmt dabei die Anzahl der aufgenommenen Bildpunkte. Die Auflösung pro Pixel wird dabei durch die Wahl der Vergrößerung des Objektivs und der Geometrie des Chips definiert. Diese Chips sind in der Regel gekühlt, um thermisches Rauschen in den elektronischen Bauteilen zu eliminieren. Dieses sind nur einige - wenn auch wichtige - Beispiele für Rauschquellen und deren Behandlung auf der Aufnahmeseite. Eine ausführliche Behandlung zum Thema Rauschen und die dazu von der Bildanalyse entwickelten Modelle und Methoden finden sich in [Jäh97] oder [Jai89]. Bei Hybridisierungsexperimenten kommen weitere schwer quantifizierbare Rauschquellen hinzu. Das optimale Hybridisierungsexperiment sollte eine homogene Färbung der biologische Präparate gewährleisten. Das ist jedoch nur selten der Fall, meistens sind die Intensitätsschwankungen innerhalb kleiner Bereiche sehr stark. Man kann dies in Abb. 4.2 erkennen: Obwohl hier die Maximalwerte der Farben pro Pixel dargestellt sind, lassen sich starke Intensitätsgradienten innerhalb der Chromosomen beobachten. Die Gründe für diese Inhomogenitäten sind im Detail noch nicht genau

verstanden. Bei Mehrfachhybridisierungen, wie es bei M-FISH Experimenten der Fall ist, kommen noch weitere Schwierigkeiten hinzu. Bestimmte Fluoreszenzkombinationen führen bei einigen Chromosomen zu sehr schwachen Hybridisierungen. Man fasst diese Phänomene oft unter dem Begriff *biologisches Rauschen* zusammen. Die Hybridisierungsqualität der unterschiedlichen Farbstoffe und die Bildaufnahme der unterschiedlichen Farbkanäle müssen aufeinander abgestimmt werden. Ist die Hybridisierung in einem der Kanäle vergleichsweise schwach, führt dies zu einem Kippen der Cluster einer Klasse im Farbraum (s. Abb. 4.6), was Überlappungen benachbarter Cluster nach sich ziehen kann. Standardisierte Färbeprotokolle und Probenkits haben sich erst vor kurzem etabliert, weshalb jedes Labor, das auf diesem Gebiet arbeitet, sein eigenes Hybridisierungsprotokoll und -Schema entwickelt hat. Das wiederum stellt hohe Ansprüche an die Bildverarbeitung und erfordert flexible Analysemethoden, um der unterschiedlichen Qualität der anfallenden Daten gerecht zu werden.

3.5 Faltung und Filterung

Vor der eigentlichen Bildanalyse kommen in der Regel rauschmindernde Verfahren, sogenannte Rauschfilter zum Einsatz, um diese Fehlerquellen zu minimieren. Filter können dabei verschiedenste Eigenschaften haben. Lineare Filteroperationen zeichnen sich vor allem dadurch aus, dass sie unabhängig von der Position einer Bildstruktur stets dasselbe Resultat liefern (Verschiebungsinvarianz). Mathematisch sind diese Filter als Faltung einer Bildfunktion $B(x,y)$ mit einer diskreten Filterfunktion $F(a,b)$ im Bildortsbereich (x,y) definiert. Die Filtermaske ist durch die Koeffizienten der Filterfunktion bestimmt und entspricht daher den Abtastwerten der Impulsantwort des Operators mit Koordinatenursprung in (x,y) [Zim96]. Ihre Realisierung beruht auf dem Konzept der *Faltung*. Die Faltung ist eine Nachbarschaftsoperation, oder Verknüpfung benachbarter Pixel, und wird (im linearen Fall) beschrieben durch

$$G = F * B \quad (3.12)$$

G ist das gefilterte Bild, F der Faltungsfilter und B das Originalbild. Der Faltungsfilter ist durch eine Maske bestimmter Grösse realisiert. Seine Einträge entsprechen den Wichtungsfaktoren, mit dem die Bildpunkte des Originalbildes multipliziert werden. Die Werte werden aufsummiert, um den neuen Wert des Bildpunktes im gefilterten Bild zu berechnen, der sich im Maskenzentrum befindet. Diese Berechnung wird für jeden Bildpunkt durchgeführt, dabei wandert das Zentrum der Maske über alle Pixel des Originalbildes. Die Berechnung an einem Ort (x,y,z) geschieht also durch:

$$G_{x,y,z} = \sum_{a=-s}^s \sum_{b=-s}^s \sum_{c=-s}^s F_{a,b,c} B_{x-a,y-b,z-c} \quad (3.13)$$

wobei die Maskengrösse durch $(2s - 1) * (2s - 1) * (2s - 1)$ für den hier dargestellten dreidimensionalen Fall gegeben ist. Mit Hilfe des *Faltungstheorems* lässt sich zeigen, dass die Fouriertransformierte von G sich aus dem Produkt der Fouriertransformierten der Filtermaske und des Originalbildes berechnen lässt:

$$F(G) = F(F)F(B) \tag{3.14}$$

Diese Form ist analytisch zugänglicher als Gl. (3.1). Aus ihr lässt sich bei bekanntem Faltungsoperator aus einem Bild das Originalbild theoretisch zurückrechnen:

$$F(G) = F(F)F(B) \Leftrightarrow F(B) = F(G)F(F)^{-1} \tag{3.15}$$

Dieser Zusammenhang spielt bei Dekonvolutionsalgorithmen eine wichtige Rolle und wird im Kapitel über Dekonvolution noch einmal aufgegriffen.

Das Faltungstheorem stellt den Zusammenhang zwischen Orts- und Frequenzraum her. Beide Repräsentationen sind gleichwertig und stellen den gleichen Sachverhalt nur unterschiedlich dar. Man spricht von der *Dualität* von Orts- und Frequenzraum. Kleinen Strukturen im Ortsraum, also im Bild, entsprechen hohe Frequenzen im Fourierraum (bzw. kleine Wellenlängen). Umgekehrt korrelieren grosse Strukturen im Bild mit tiefen Frequenzen im Fourierraum.

Indem Rauschfilter im Ortsraum die Bildinformation verändern, beeinflussen sie auch das Fourierspektrum. Entfernt ein Glättungsfilter beispielsweise kleine Strukturen im Bild, während er grosse Objekte erhält, wirkt er als Tiefpassfilter.

Drei wichtige Rauschfilter werden im folgenden vorgestellt.

3.5.1 Der Gaussfilter

Dieser Filter ist einer gebräuchlichsten Rauschfilter überhaupt. Er gehört zur Klasse der *linearen Filter* [Jäh97]. Lineare Filter zeichnen sich durch Verschiebungsinvarianz aus, d.h. sie liefern unabhängig vom Ort das gleiche Resultat. Seine Wichtungsfaktoren entsprechen der einer Gaussverteilung.

Interessant ist, dass die Fouriertransformierte einer Gaussfunktion wiederum eine Gaussfunktion mit reziproker Halbwertsbreite ist (dies gilt auch für den Gaussfilter). Man erkennt daher aus Gl. (3.14) die Wirkung des Gaussfilters: Während grobe Strukturen (tiefe Frequenzen) erhalten werden, entfernt er die hohen Frequenzen im Fourierraum. Oft wird der Gaussfilter dazu verwendet, das *Abtasttheorem* nachträglich zu erzwingen, sofern es auf der Aufnahmeseite noch nicht erfüllt wurde. Allerdings muss darauf geachtet werden, im Ortsraum zu Filtern, da sonst nicht ausgeschlossen werden kann, dass hohe Frequenzen in den niederfrequenten Bereich des Spektrums hineingefaltet werden.

3.5.2 Der Medianfilter

Dieser Filter ist ein Rangordnungsfiler, der zur Klasse der *nichtlinearen Filter* [Jäh97] gehört. Betrachtungen im Fourierraum sind beim Medianfilter daher nicht möglich. Rangordnungsoperatoren jedoch ermöglichen eine elegante Lösung, vor allem weil sie sich als robuste und wenig rechenaufwendige nichtlineare Transformationen im Ortsraum erwiesen haben [Zim96].

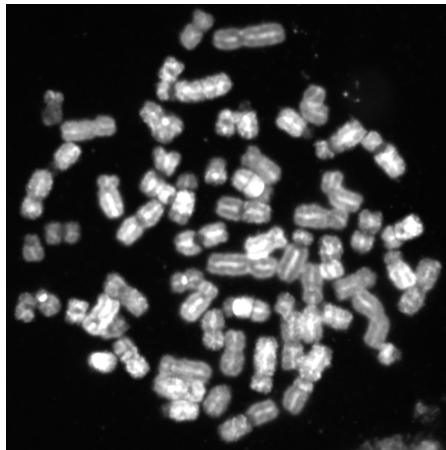


Abb. 3.1: Median gefiltertes Maximum-Projektionsbild aus Abb. 4.2.

Die Werte innerhalb des Maskenbereichs werden bei dieser Methode ihrer Grösse nach sortiert. Der Rückgabewert bestimmt sich dann als eine Funktion der auf diese Weise entstehenden „Rangordnung“

$$p_0 = K \sum_{k=0}^m w_k Q(k), \quad (3.16)$$

wobei K einen einfachen Normierungsfaktor und $Q(k)$ die nach der Grösse geordneten Grauwerte darstellen.

Je nach Wahl der Koeffizienten w_k können unterschiedliche Rangordnungsfiler mit verschiedenen Eigenschaften konstruiert werden. Beim Medianfilter wird der mediale Wert, also der Wert zurückgegeben, der in der Rangfolge der Grauwerte in der Mitte steht (also $w_M = 1$, mit $M = (N + 1)/2$, alle anderen $w_i = 0$) [Zim96]. Der Medianfilter besitzt die Eigenschaft, Änderungen innerhalb des Maskenbereichs herauszufiltern, besonders Peakrauschen. Grauwertkanten und -plateaus bleiben hingegen erhalten. Kanten werden nicht wie beim Gaussfilter verschmiert. Mehrfaches Anwenden des Medianfilters führt zu einem stationären Ergebnis, das sich bei weiteren medialen Filteroperationen nicht mehr ändert.

Der mediale Wert ist statistisch robuster als der Mittelwert eines Ensembles, das

heisst, er ist unempfindlicher gegenüber Störungen. Ein einfaches Beispiel soll dies illustrieren: Angenommen man hat 9 gemessene Werte (oder Pixel), von denen 7 den Wert 50 und zwei den Wert 200 besitzen. Dann beträgt der mediale Wert 50, während der mittlere 83.3 beträgt. Bei statistischen Analysen rechnet man in der Regel jedoch mit Mittelwerten, da die Nichtlinearität des medialen Werts ihn für analytische Betrachtungen untauglich macht.

3.5.3 Anisotroper Diffusionsfilter

Dem *Diffusionsfilter* liegt das physikalische Prinzip des Konzentrationsausgleichs bei vorhandenem Gefälle zugrunde. Diffusion findet immer dann statt, wenn sich die Teilchenzahldichte u benachbarter Orte unterscheidet. Bei diesen Prozessen werden Teilchen weder erzeugt noch vernichtet. Dieser Konzentrationsausgleich wird durch das *1.Ficksche Gesetz* ausgedrückt:

$$\mathbf{j} = -\mathbf{D}\nabla\mathbf{u} \quad . \quad (3.17)$$

Das Konzentrationsgefälle $\nabla\mathbf{u}$ bewirkt einen Teilchenstrom, der dem Gefälle entgegenwirkt. Dieser wird durch die Teilchenstromdichte \mathbf{j}_u beschrieben, der $\nabla\mathbf{u}$ entgegengerichtet ist. Die Diffusionskonstante \mathbf{D} beschreibt den Zusammenhang zwischen Fluss und Gefälle. Im mehrdimensionalen Fall wird sie zu einem Tensor, einer positiv definiten, symmetrischen Matrix. Der isotope Fall, wenn \mathbf{j} und $\nabla\mathbf{u}$ parallel sind, ist ein Spezialfall der allgemeineren *anisotropen* Diffusion, wenn \mathbf{j} und $\nabla\mathbf{u}$ nicht parallel sind.

Die Erhaltung der Teilchenzahl wird mit der *Kontinuitätsgleichung* beschrieben

$$\partial_t\mathbf{u} = -\text{div}\mathbf{j} \quad . \quad (3.18)$$

Aus Gl. (3.17) und (3.18) folgt dann die *allgemeine Diffusionsgleichung*

$$\partial_t\mathbf{u} = \text{div}(\mathbf{D}\nabla\mathbf{u}) \quad . \quad (3.19)$$

Um Diffusion auf Bilder anwenden zu können, wird die Teilchenkonzentration mit dem Grauwert an einem bestimmten Punkt identifiziert. Der Gradient beschreibt dann die relative Änderung der Grauwerte benachbarter Punkte. Im linearen Fall (3.19) ist die Diffusion unabhängig von dem sich ändernden Bild. Oft ist es jedoch vorteilhafter, den Diffusionstensor als eine Funktion der sich entwickelnden lokalen Bildstruktur anzusehen, was zu nichtlinearen Diffusionsfiltern führt [Wei96].

Dem nichtlinearen Diffusionsfilter liegt das Modell von Perona und Malik [Per87] zugrunde, die anstelle von Homogenität im linearen Skalenraum scharfe Regionengrenzen fordern. Die regioneninterne Glättung soll dabei der Glättung über Regionengrenzen vorgezogen werden. Der Prozess diffundiert ein Bild in Regionen mit

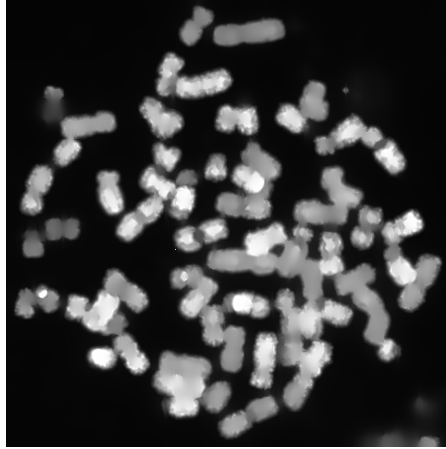


Abb. 3.2: Nichtlinearer anisotroper Diffusionsfilter angewendet auf das Maximum-Projektionsbild. Geeignete Parameter wurden zuvor interaktiv bestimmt.

konstantem mittleren Grauwert und in solche mit sich stärker ändernder Grauwertintensität, um das Rauschen zu eliminieren. Gesteuert wird der Diffusionsprozess durch die Form der sogenannten „Kanten-Stop-Funktion“ g , die der Diffusivität entspricht, sowie durch ihre Abhängigkeit vom Betrag des Gradienten. In diesem Fall hat die Diffusionsgleichung die folgende Form:

$$\partial_t \mathbf{u} = \operatorname{div}(g(|\nabla \mathbf{u}|^2) \nabla \mathbf{u}) \quad (3.20)$$

Die Koeffizienten von \mathbf{D} werden nun durch Funktionen beschrieben.

In verschiedenen Arbeiten zu diesem Thema wurde gezeigt, dass mit Einführung der Regularisierung in (3.20) eine höhere Stabilität des Perona-Malik-Prozesses erreicht werden kann (vgl. hierzu [WB96]). Nach einem Vorschlag von Catt, Lions, Morel und Coll [Cat92] wird die Diffusion nun durch

$$\partial_t \mathbf{u} = \operatorname{div}(g(|\nabla \mathbf{u}_\sigma|^2) \nabla \mathbf{u}) \quad (3.21)$$

beschrieben.

Dabei wurde $g(|\nabla \mathbf{u}|^2)$ unter Einbeziehung der Äquivalenz von linearem Diffusionsfilter und Gaussfilter durch eine mit einer Gausschen Glockenfunktion geglätteten Version $g(|\nabla \mathbf{u}_\sigma|^2)$ ersetzt, mit σ als Varianz. Der nichtlineare **isotrope** Diffusionstensor ist dann von der Gestalt

$$\mathbf{D}^\sigma = \begin{bmatrix} g(|\nabla \mathbf{u}_\sigma|^2) & 0 \\ 0 & g(|\nabla \mathbf{u}_\sigma|^2) \end{bmatrix} \cdot \quad (3.22)$$

In der nichtlinearen anisotropen Diffusion wird die Kanten-Stop-Funktion durch einen **anisotropen** Tensor ersetzt, womit für den nichtlinearen **anisotropen** Dif-

fusionsfilter das folgende gilt:

$$\mathbf{D}^\sigma = \begin{bmatrix} g(|\nabla \mathbf{u}_\sigma|^2) & 0 \\ 0 & 1 \end{bmatrix} \quad (3.23)$$

Für detaillierte Ausführungen zum Thema anisotroper Diffusion in der Bildverarbeitung wird auf [Wei96] verwiesen.

3.6 Dekonvolution

Die Dekonvolution ist das inverse Problem zur Faltung und spielt eine wichtige Rolle bei der Aufnahme von dreidimensionalen Bildstapeln am Mikroskop. Standard für solche Aufnahmen ist das konfokale Laser-Scanning-Mikroskop (*CLSM*). Man nimmt dabei an, dass die Punktabbildung am Mikroskop B sich aus der Faltung einer Punktantwort P mit dem Punktsignal des Originalbildes O ergibt:

$$B = P * O \quad (3.24)$$

Dieses Modell gilt jedoch nur für lichtdurchlässige Objekte (bei der entsprechenden Wellenlänge) und ist daher eine starke Vereinfachung der tatsächlichen experimentellen Bedingungen, wird jedoch als gute Näherung an das tatsächliche Abbildungsverhalten akzeptiert. Darüber hinaus können durch vergleichsweise einfache mathematische Operationen die Abbildungseigenschaften von Mikroskopen untersucht werden.

Die Punktantwort ist dabei die Antwort eines Filters auf ein Punktbild und kann für ein gegebenes Mikroskop-Setup theoretisch berechnet, oder experimentell bestimmt werden, indem ein Bild von einem fluoreszierenden Objekt (in der Regel winzige Kügelchen, sog. Beads), dessen Grösse weit unterhalb der Auflösung liegt, aufgenommen wird. Die zentrale Bedeutung der Punktantwort beruht auf der Tatsache, dass die Faltungsoperation linear ist. Kennen wir die Antwort auf ein Punktbild, können wir die Antwort auf jedes beliebige Bild berechnen, da Bilder aus Punktbildern zusammengesetzt werden können [Jäh97].

Durch die Faltung wird die ursprüngliche Information in O verwischt. Um diese in jedem Punkt wieder zu rekonstruieren, gibt es verschiedene Entfaltungsverfahren, die Gl. (3.24) hinreichend genau erfüllen. Alle diese Methoden haben unterschiedliche Vor- und Nachteile in Bezug auf numerische Stabilität, Speicherverbrauch und Geschwindigkeit. Das wesentliche Problem zur Lösung dieses linearen Gleichungssystems besteht darin, dass es zwar durch den vollen Rang von \mathbf{P} invertierbar ist, jedoch $\det(\mathbf{P}) \approx 0$ ist, was zu numerischen Schwierigkeiten führen kann. Auf eine ausführliche Behandlung dieses Themas muss auf Spezialliteratur verwiesen werden. Hier sollen nur zwei einfache Möglichkeiten vorgestellt werden, um das Prinzip zu

illustrieren.

Im Fourierraum wird (3.24) zu

$$F(B) = F(P)F(O) \Leftrightarrow F(O) = F(B)F(P)^{-1} \quad (3.25)$$

Durch Berechnung der Fouriertransformierten lässt sich $F(O)$ also im Prinzip leicht berechnen und durch Rücktransformation in den Ortsraum das Originalbild O rekonstruieren. Das Problem ist allerdings, dass die Werte in $F(P)$ oft nahe bei Null sind, was im diskreten Fall zu Divisionen durch Null führt. Um dieses Problem zu umgehen gibt es zwei Möglichkeiten.

Entweder man addiert eine kleine Konstante zum Divisor (*Wiener Filter*):

$$F(O) = \frac{F(B)}{F(P) + c} \quad (3.26)$$

Oder man bricht die Division unterhalb einer kleinen Schwelle ab (*direkte Entfaltung*):

$$F(O) = \begin{cases} \frac{F(B)}{F(P)} & \text{falls } F(P) > \varepsilon \\ 0 & \text{sonst} \end{cases} \quad (3.27)$$

Beide Verfahren sind sehr schnell und weisen gutartiges Verhalten bei grossen Datenmengen und Dimensionen auf. Der Nachteil ist, dass bei der Rücktransformation in periodische Funktionen Artefakte in Form von Wellen an Objektkanten entstehen, die im Ursprungsbild nicht vorhanden sind.

Kapitel 4

Methodik

Untersuchungen im Rahmen dieser Arbeit haben gezeigt, dass die Zuordnung der Pixel zu ihren Klassen (Chromosomen) allein auf der Basis der Farbinformation zu fehlerhaften Resultaten führt. Dafür werden in diesem und in folgenden Kapiteln zahlreiche Beispiele angeführt. Dies hat mehrere Gründe. Zum einen weicht die tatsächliche Farbinformation je nach Experiment in einigen Bereichen und bei bestimmten Kombinationen zum Teil erheblich von der idealen ab. Zum anderen wird die Farbinformation mit abnehmender Grösse der markierten Bereiche und Proben allein aus statistischen Gründen schlechter. Je weniger Pixel die Bereiche umfassen, um so stärker machen sich Unregelmässigkeiten in der Hybridisierung und mögliche *Fokalshifts* in lateraler Richtung, die beim Filterwechsel entstehen können, sowie unterschiedlich grosse Signale einer Probe in den einzelnen Kanälen bemerkbar.

Um diesen Problemen zu begegnen, habe ich folgenden Ansatz gewählt (Abb. 4.1): Zunächst ermittle ich auf einer Segmentierungsmaske die wichtigen Bildbereiche. Dies bringt einen signifikanten Geschwindigkeitsvorteil für den folgenden Zerlegungsalgorithmus, da in der Regel mehr als 80% der Bildinformation aus Hintergrund bestehen. Die Hintergrundkorrektur kann zudem mit relativ einfachen Bildverarbeitungsoperationen bei gutem Segmentierungsergebnis erzielt werden. Anschliessend zerlege ich diese Bildbereiche in kleine Bereiche annähernd konstanter Farbe (*Tessellierung*). Dieser Schritt hat einen rauschmindernden Effekt. Kritische Übergangsbereiche zwischen Bildregionen unterschiedlicher Farbe werden zuverlässiger identifiziert. Gleichzeitig hat dieser Schritt für die folgende Clusteranalyse einen wichtigen positiven Effekt: Pixel kritischer Übergangsbereiche werden im Farbraum zusammengefasst. Damit werden Überlappungseffekte farblich ähnlicher Cluster reduziert, und damit eine zuverlässigere Clusteranalyse möglich. Bereiche ähnlicher Farbe fasse ich anschliessend im Farbraum zu Gruppen zusammen (*Clustersuche*). Dabei suche ich die Cluster (Centroide), indem ich innerhalb eines Clusters die Summe der Winkel bzw. der orthogonalen Abstände bezüglich einer Richtung minimiere, um eine optimale Richtung zu finden. Dadurch erhalte ich direktionale Cluster, die

den Datensätzen gerecht werden. Das originäre Hybridisierungsschema wird dabei dynamisch an die Bildinformation angepasst. Im letzten Schritt werden diese wieder durch eine Zuordnung zu den Klassen identifiziert (*Klassifizierung*). Die von mir entwickelten Algorithmen und Modelle stelle ich für den zweidimensionalen Fall vor. Für den dreidimensionalen Fall gelten sie analog.

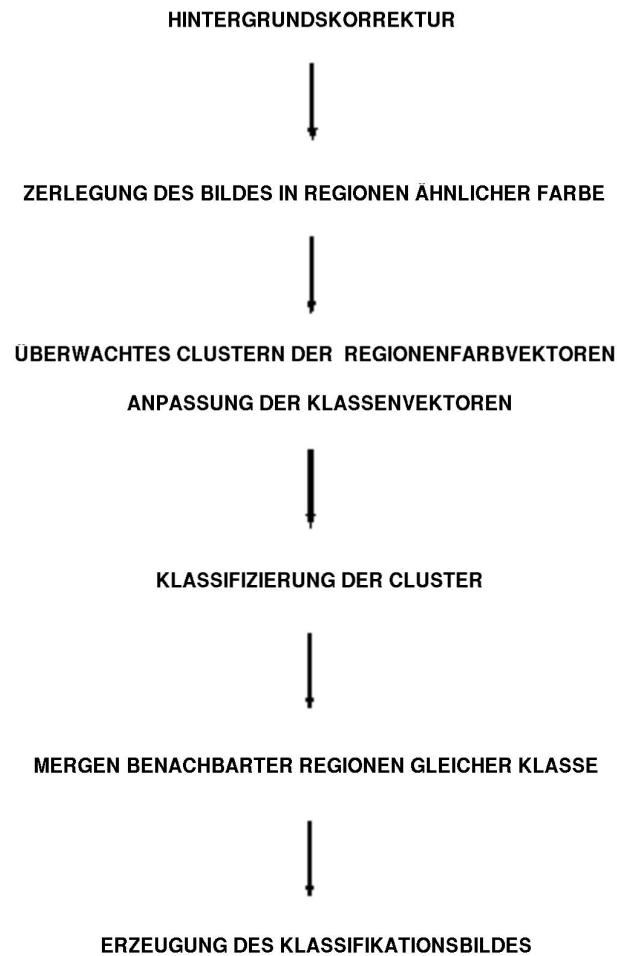


Abb. 4.1: Schematischer Ablauf der Methodik. Die Schritte werden im Verlauf dieses Kapitels im einzelnen beschrieben.

4.1 Maximum-Projektion

Die Maximum-Projektion ist eine einfache Verknüpfungsoperation zweier oder mehrerer Bilder. Jedem Pixel P_i an der Stelle i wird der grösste Wert aus allen Bildern zuordnet (vgl. Abb. 4.2).



Abb. 4.2: Maximum-Projektionsbild der Aufnahmen eines typischen 5-Farben M-FISH Experiments. Jedem Pixel wird der grösste Wert aus den 5 Farbkanälen zugeordnet. Zu sehen sind Chromosomen in der Metaphase.

Das Maximumbild bietet einige wichtige Vorzüge für Hintergrundkorrekturalgorithmen bei der Analyse von M-FISH Bildern. Zum einen beschränkt sich die Parameterwahl auf ein Bild, anstatt für jeden Farbkanal unabhängig einen eigenen Parametersatz zu finden. Das erleichtert die Automatisierung und Segmentierung. Zum anderen können schwache Hybridisierungssignale in einem Kanal bei mehrfachem Labeling durch einen anderen Farbkanal, der ein besseres Hybridisierungssignal liefert, kompensiert werden. Das erleichtert das Auffinden von Objektinformation, mit anderen Worten wird die Detektionswahrscheinlichkeit erhöht.

4.2 Hintergrundkorrektur

Sämtliche Filteroperationen und Hintergrundkorrekturen führe ich auf dem Maximum-Projektionsbild durch, wenn keine DAPI-Gegenfärbung aller relevanten biologischen Zielobjekte durchgeführt wurde (vgl. Abb. 4.2, 4.3). Bei dreidimensionalen Aufnahmen von Interphase-Zellen oder Barcode-Experimenten oder Experimenten bei denen nur subchromosomale Bereiche angefärbt werden ist dies der Fall. Bei zweidimensionalen Aufnahmen von Metaphase-Chromosomen hingegen existiert normalerweise eine DAPI-Gegenfärbung, so dass die Segmentierung auf dem DAPI Bild erfolgen kann.

Ein optimiertes Segmentierungsverfahren kann möglicherweise zu etwas besseren Resultaten führen. Wesentlich bleibt aber immer noch die Farbinformation im Bild. Ich habe daher darauf Wert gelegt, keine Filteroperationen im Farbraum durchzuführen, da nicht klar ist, wie die mehrdimensionale Farbinformation im einzelnen manipuliert und damit möglicherweise zerstört wird. Glättungseffekte lassen sich

jedoch indirekt erzielen, in dem für ein Pixel ein gemittelter Farbwert aus seiner unmittelbaren Nachbarschaft berechnet wird. Für sämtliche zweidimensionalen Bilder habe ich auf dem Hintergrundbild (DAPI oder Maximum) einen Medianfilter mit einer 3x3 Maske verwendet, um Peakrauschen im Bild zu entfernen. Optional kam der anisotrope Diffusionsfilter [Tva00] zum Einsatz. Für die anschließende Segmentierung der Farbbereiche in 2D Bildern reichte ein einfacher Schwellwert. Die beiden Bilder besitzen in der Regel ausreichend Kontrast, um ein sehr gutes Segmentierungsergebnis zu erzielen. Auf den hier definierten Bereichen, setzt die Farbsegmentierung ein, die ich im Folgenden beschreibe.

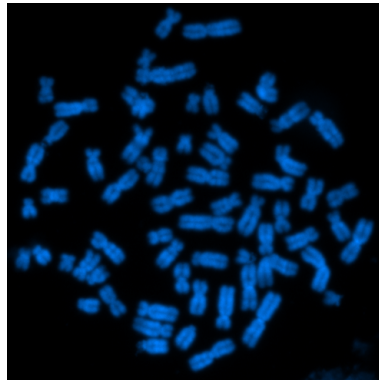


Abb. 4.3: DAPI Bild eines 5-Farben Experiments einer aberranten Zelle. Die Gegenfärbung mit DAPI ist in der Regel vorhanden, so dass dieses Bild als Grundlage für eine Vorabsegmentierung verwendet werden kann. (Vgl. Abb. 4.2)

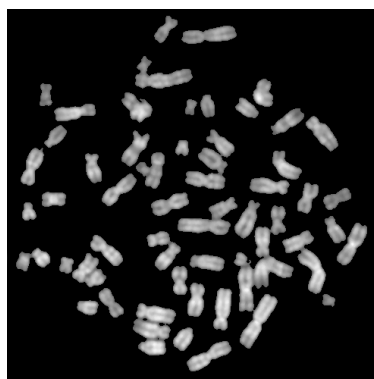


Abb. 4.4: DAPI Bild nach einer Schwellwert-Segmentierung. Der Schwellwert beträgt das 1.5 fache der Standardabweichung des Gesamtbildes.

4.3 Farbinformation und Datenraum

Durch kombinatorische Färbung erhalten die biologischen Präparate eine jeweils spezifische spektrale Signatur. Diese Farbinformation kann durch geeignete spektrale Aufnahmetechniken aufgelöst werden (s. Kapitel 2). Je nach Aufnahmemethode erhält jedes Pixel nicht mehr nur einen einzigen Intensitätswert, oder eine eindimensionale Farbinformation, wie es bei Grauwertbildern der Fall ist, sondern eine mehrdimensionale Farbinformation. Bei der M-FISH-Technologie, die spezifische Bandpassfilter für die Aufnahme verwendet, entspricht die Farbtiefe pro Pixel der Anzahl der verwendeten Fluorochrome, da für jedes Fluorochrom ein Bild aufgenommen wird.

Bei der interferometrischen Methode, wie sie beim SKY-System zum Einsatz kommt, ist die spektrale Auflösung sehr viel höher, und man erhält eine in der Größenordnung fünffach höhere Farbtiefe. Nachteile dieser Methode sind jedoch, dass zum einen Ergebnisse nicht visuell verifiziert werden können, zum anderen die Aufnahmezeiten sehr viel länger sind als bei M-FISH. Weiterhin ist ein Abgleich der einzelnen Farbkanäle durch unterschiedliche Belichtungszeiten nicht möglich, was zu wesentlichen Kontrastverbesserungen führen kann, da durch unterschiedliche Bindungs- und Fluoreszenzcharakteristika der unterschiedlichen Fluorochrome die kombinatorische Information oft stark verfälscht wird. Abb. 4.5 zeigt drei Kanäle des Farbraums eines typischen 5-Farben Experiments, in der die Metaphasen-Chromosomen einer aberranten Zelle zu analysieren waren.

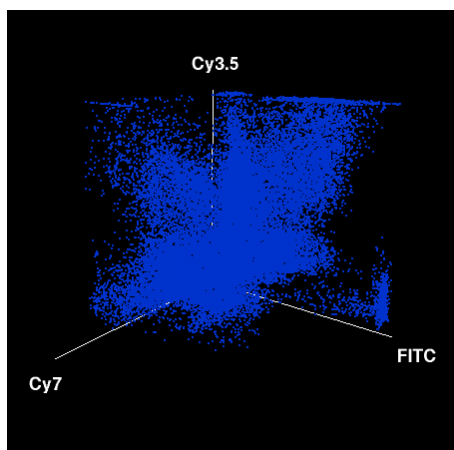


Abb. 4.5: Farbraum eines 5-Farben Experiments. Dargestellt sind die ersten drei Farbkanäle.

Ich habe im Rahmen dieser Arbeit nur mit M-FISH Daten gearbeitet, bei denen spezifische Filter zur Detektion verwendet wurden. Die Anzahl gleichzeitig verwendeter Farbstoffe pro Experiment betrug nicht mehr als sieben. In keinem

Hybridisierungsschema, unabhängig von der Form des Experiments, traten mehr als drei Simultan-Hybridisierungen für ein biologisches Target auf. Für solche Experimente hat sich die folgende Darstellung der Farbinformation oder des Farbraums bewährt: Drei Farbkanäle werden nach der Klassifikation ausgewählt, die die Achsen repräsentieren, und nur diejenigen Datenpunkte dargestellt, die von solchen Objekten stammen, die *nur* in diesen drei Kanälen gefärbt sind, also keine Hybridisierung in einem der nicht dargestellten Farbkanal besitzen. Abb. 4.6 zeigt einen typischen Fall eines 5-Farben Experiments, in dem die Metaphase einer aberranten Zelle analysiert wurde.

Wie sich im Verlauf dieser Arbeit herausstellen wird, gibt es für den Experimentator gute Gründe, auf mehr als drei simultane Hybridisierungen in seinem Labeling-Schema zu verzichten. Diese von mir gewählte Form der Darstellung des Farbraums von M-FISH Daten eignet sich jedoch nicht für Bilddaten aus SKY Experimenten, da einem Label, bzw. einer Hybridisierung mehr als ein Farbkanal zugeordnet wird. Der Farbraum stellt im übrigen einen Unterraum des *Merkmalsraums* dar. Der von mir gewählte Satz von Merkmalen schliesst neben der Farbinformation auch die Ortsinformation ein.

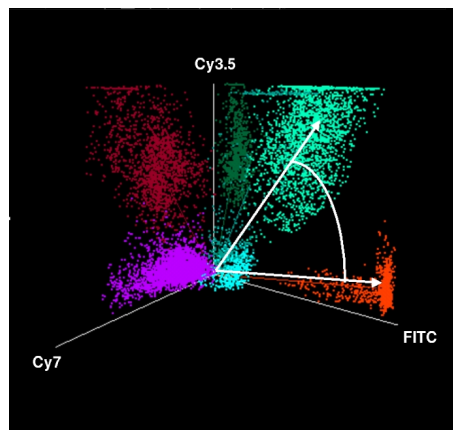


Abb. 4.6: Farbraum eines analysierten 5-Farben Experiments. Die unterschiedlichen Farben sind frei gewählte Falschfarben für unterschiedliche Chromosomen. Analysiert wurde die Metaphase einer hochaberranten Zelle. Dargestellt sind die ersten drei Farbkanäle. Die Kodierung der Falschfarben entspricht der in Abb. 1.6.

4.4 Bedeutung der Farbe bei M-FISH Bildern

Schaut man sich das Hybridisierungs-Schemema eines M-FISH Experiments an, so leuchtet unmittelbar ein, dass Pixel, die beispielsweise nur in FITC und Cy3 gefärbt sind, *unabhängig von ihrer Intensität* Chromosom 18 zugeordnet werden sollten (vgl.

Abb. 1.6). Pixel eines Chromosoms sollten daher näherungsweise durch ein *konstantes Verhältnis der Werte in den verschiedenen Achsen* charakterisiert sein. Solch ein konstantes Achsenverhältnis hat im Vektorraum die Bedeutung einer *Richtung*. Aus diesem Grund wird folgende Darstellung der Pixel eingeführt:

$$\vec{p}_{x,y} = \vec{e} I_{x,y} \quad (4.1)$$

Das Pixel an der Stelle (x,y) $\vec{p}_{x,y}$ setzt sich aus einem auf Eins normierten Vektor, der *Farbrichtung* \vec{e} , und seiner Intensität zusammen.

Im Experiment wird der Idealfall aufgrund von Intensitätsschwankungen der unterschiedlichen Farbkanäle untereinander und unspezifische Hybridisierungen praktisch nicht erreicht. Trotzdem sollte dieses Merkmal als wesentliches Charakteristikum erhalten bleiben, so dass in der Praxis die Pixel, die zu Chromosom 18 gehören, durch einen mittleren Farbvektor, zum Beispiel $(180, 212, 23, 45, 40)$, beschrieben werden können. Man erkennt diesen Zusammenhang bereits in der Struktur der Daten im Farbraum in Abb. 4.6. Richtungen und Winkelunterschied für zwei Klassen ist in der Abbildung angedeutet. Aus den Rohdaten in Abb. 4.5 ist dieser wiederum nur schwer zu erahnen.

4.4.1 Winkel als Distanzmass

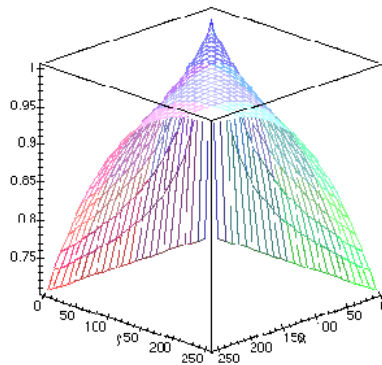


Abb. 4.7: Das normierte Skalarprodukt aller Vektoren im zweidimensionalen Raum mit $(255,255)$. Das Skalarprodukt ist in z-Richtung aufgetragen. Man erkennt die Intensitätsunabhängigkeit der Richtung $(255,255)$.

Mit der Einführung der Farbrichtung wird der Winkel zwischen zwei Farbvektoren das entscheidende Kriterium zu ihrer Unterscheidung. Mit anderen Worten, der Winkel zwischen den Farbvektoren zweier Pixel bestimmt ihre Ähnlichkeit und ihre Zugehörigkeit zu einer Klasse, unabhängig von ihrer Intensität. Es dient als geeignetes *Distanzmass* und *Klassifikationskriterium* bei der Analyse von M-FISH

Bildern. Der Winkel α zwischen zwei Vektoren oder Pixeln \vec{p}_1 und \vec{p}_2 im Farbraum ist leicht zu berechnen:

$$\cos \alpha = \frac{\vec{p}_1 \vec{p}_2}{\|\vec{p}_1\| \|\vec{p}_2\|} = \vec{e}_1 \vec{e}_2 \quad (4.2)$$

Dies ist nichts anderes als das normierte Skalarprodukt zwischen diesen Vektoren (Abb. 4.7 und 4.8). So vorteilhaft die Intensitätsunabhängigkeit ist, der Winkel besitzt die unangenehme Eigenschaft, dass er bei kleinen Intensitäten sehr anfällig für Rauschen ist. Kleine Änderungen in der Intensität bewirken bei kleinen Intensitäten grosse Änderungen in der Richtung. Dies berücksichtige ich jedoch durch die Einführung eines geeigneten *Farbmodells*, das Pixel mit höheren Werten stärker gewichtet. Zudem müssen Pixel, die in keinem der Farbkanäle Information besitzen, ausgeschlossen werden, da der Winkel für Vektoren mit einer Norm gleich Null nicht definiert ist (Gl. 4.2).

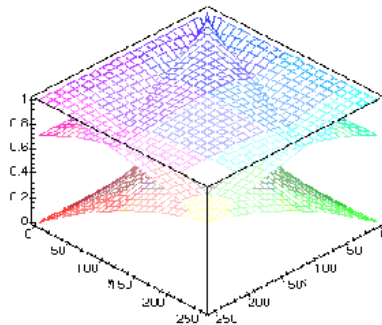


Abb. 4.8: Das normierte Skalarprodukt aller Vektoren im zweidimensionalen Raum mit $(255,0)$, $(255,255)$ und $(0,255)$. Das Skalarprodukt ist in z -Richtung aufgetragen. Jedes Pixel erhält daher drei Werte in z -Richtung. Die Schnittgeraden der obersten Ebenen definieren die Klassengrenzen zwischen diesen Vektoren.

4.4.2 Euklidischer Abstand als Distanzmass

Viele Verfahren beruhen auf dem euklidischen Abstand als Distanzmass. Für die Analyse von M-FISH Bildern hat es sich jedoch als nicht tauglich herausgestellt. Das liegt an der Struktur der Daten.

In der Literatur gibt es zahlreiche Hinweise darauf, dass sich kantenbasierte Verfahren für die Segmentierung von Dreikanal-(Echt-)Farbbildern nicht zu eignen scheinen [Rer97]. Das liegt möglicherweise daran, wie die Farbkanten berechnet wurden. Eine gängige Methode bei mehrdimensionaler Kantenberechnung ist die Summe der Quadrate der Kantenwerte der einzelnen Kanäle zu berechnen ([Jäh97],

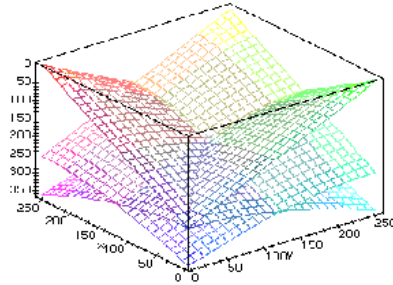


Abb. 4.9: Euklidischer Abstand als Klassifikator. In z-Richtung ist der Euklidische Abstand aller Vektoren im zweidimensionalen Raum zu den Vektoren $(255,0)$, $(255,255)$ und $(0,255)$. Jedes Pixel erhält daher drei Werte in z-Richtung. Die Schnittgeraden der obersten Ebenen definieren die Klassengrenzen zwischen diesen Vektoren (vgl. 4.8).

s. 362 ff.). Letztlich berechnet man auf diese Art *Euklidische Distanzen im Merkmalsraum*, da der (diskrete) Gradient wiederum aus einer Differenz zweier Grauwerte berechnet wird.

Dies könnte ein Hinweis darauf sein, dass der von mir gewählte Ansatz auch bei klassischen Farbbildern gute Segmentierungsergebnisse bringt.

4.5 Modellbasierte Segmentierung

Nach der Hintergrundkorrektur sind die interessierenden Bereiche im Bild definiert. Um ein möglichst zuverlässiges Analyseergebnis zu erhalten, sollten soviel nicht redundante (orthogonale) Merkmale wie möglich verwendet werden. Hier bietet sich zusätzlich zur Farbinformation die Orts- bzw. Nachbarschaftsinformation der Pixel an. Benachbarte Pixel in unmittelbarer Umgebung gehören mit hoher Wahrscheinlichkeit der gleichen Klasse an.

Um die Ortsinformation mit zu berücksichtigen, habe ich die pixelierte Bildinformation durch ein regionenorientiertes Modell ersetzt: Das Bild wird nicht mehr durch Pixel, sondern durch Regionen konstanter Farbe zusammengesetzt. Daher erfolgt zunächst die Zerlegung des Bildes in Bereiche ähnlicher Farbe. Für diese Bereiche werden Farbvektoren berechnet, die die einzelnen Regionen repräsentieren.

Diese Methode hat mehrere Vorteile: Das Bild wird durch diesen Schritt stark komprimiert, da es nicht mehr durch Pixel, sondern in der Regel einigen hundert regionalen Farbvektoren repräsentiert, was bei der Weiterverarbeitung einen enormen Geschwindigkeitsgewinn bedeutet. Weiterhin konnte nach der Zerlegung in Regionen eine robuste Konvergenz beim Clustern (s.u.) beobachtet werden, der bei Anwendung des weiter unten beschriebenen Cluster-Verfahrens auf die reinen Pixeldaten

nicht auftrat. Ausserdem werden durch die Hinzunahme von Ortsinformation die durch den Klassifikator (Abb. 4.8) definierten Klassengrenzen aufgelöst, was im kritischen Fall sich leicht überlappender Klassen im Farbraum unter Umständen zu besseren Klassifikationsergebnissen führt. Sehr wichtig ist zudem, dass Übergangsregionen zwischen benachbarten Regionen unterschiedlicher Farbe sehr viel besser identifiziert werden. In solchen Bereichen kommt es durch sog. *Flaring* zu einer Überlagerung der benachbarten Farben, die bei einer pixelbasierten Klassifikation zu falschen Ergebnissen führt. Dieser Effekt wird durch den regionenbasierten Ansatz stark reduziert (Abb. 4.10 und 4.11). Die Farbsensitivität kann durch geeignete Wahl des Ähnlichkeitskriteriums (Winkel) variiert und dem Bildinhalt angepasst werden (s.u).

4.5.1 Farbmodell

Das zugrundeliegende Farbmodell, das das Bild durch Bereiche konstanter Farbe beschreibt, beruht auf der schon vorgestellten Farbrichtung. In diesem Modell beschreibe ich einen Bereich durch eine konstante Farbrichtung \vec{e} , einem normierten Vektor und einer variierenden Intensität. Ein gemessenes Pixel an der Stelle $\vec{p}_{x,y}$ innerhalb dieses Bereichs stellt dann eine Realisierung von \vec{e} mit einer Intensität $I_{x,y}$ und einem Fehlervektor $\vec{f}_{x,y}$ dar.

$$\vec{p}_{x,y} = \vec{e}I_{x,y} + \vec{f}_{x,y} \quad (4.3)$$

Oder in Komponenten für ein Experiment mit drei Farben:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{x,y} = \begin{pmatrix} e_r \\ e_g \\ e_b \end{pmatrix} I_{x,y} + \begin{pmatrix} f_r \\ f_g \\ f_b \end{pmatrix}_{x,y} \quad (4.4)$$

Die Berechnung des Farbvektors für eine Region erfolgt nun in der Art, dass die Summe der Fehlerquadrate minimal wird (*least squares, Methode der kleinsten Quadrate*). Dazu wird für eine Region R_n folgender Energieterm bezüglich (e_r, e_g, e_b) und $I_{x,y}$ minimiert, der die gesuchte Farbrichtung \hat{e} und die Intensitäten $I_{x,y}$ liefert:

$$(e_r^*, e_g^*, e_b^*, I_{x,y}^*) = \underset{e_r, e_g, e_b, I_{x,y}}{\operatorname{argmin}} \sum_{(x,y) \in R_n} \left[\begin{pmatrix} R \\ G \\ B \end{pmatrix}_{x,y} - \begin{pmatrix} e_r \\ e_g \\ e_b \end{pmatrix} I_{x,y} \right]^2 \quad (4.5)$$

Die Lösungen $(e_r^*, e_g^*, e_b^*, I_{x,y}^*)$, die diese Gleichung im Minimum erfüllen lauten:

$$(e_r^*, e_g^*, e_b^*, I_{x,y}^*) = \frac{\sum \sum_{R_n} I_{x,y}^* (R, G, B)_{x,y} \Delta x \Delta y}{\sum \sum_{R_n} I_{x,y}^{*2} \Delta x \Delta y} \quad (4.6)$$

$$I_{x,y}^* = (e_r^*, e_g^*, e_b^*)(R, G, B)_{x,y}^T \quad (4.7)$$

wobei $e_r^{*2} + e_g^{*2} + e_b^{*2} = 1$ gilt.

Die Lösung für die Intensitäten zeigt, dass sie nichts anderes sind als die Projektionen der gemessenen Pixel auf die Richtung \vec{e} . Die Projektion ist das Skalarprodukt von \vec{e} und $(R, G, B)_{x,y}^T$. Die Fehlervektoren $(f_r, f_g, f_b)_{x,y}$ liegen alle in einer Ebene und sind zur Richtung \vec{e} orthogonal. Diese Methode hat gegenüber Methoden, die mit normalisierten Datenvektoren arbeiten, also $\vec{P}_{x,y} \mapsto \frac{\vec{P}_{x,y}}{\sqrt{(R^2+G^2+B^2)_{x,y}}}$, den Vorteil, dass Orte mit kleiner Intensität bei der Bestimmung der Richtung \vec{e} geringer gewichtet werden, da bei gleicher Farbrichtung die Fehlervektoren niedriger Intensitäten kleiner sind und bei der Berechnung daher geringeres Gewicht haben. Beide Methoden haben den Nachteil, dass die Intensitäten herausdividiert werden und daher Intensitätskanten nicht detektiert werden können. Dieser Nachteil kann jedoch in Kauf genommen werden, da die Vorzüge der intensitätsunabhängigen Segmentierung bei der Analyse solcher Daten bei weitem überwiegen.

4.6 Tessellierung

Die Zerlegung in Regionen konstanter Farbe habe ich als Region Growing Algorithmus mit *zentroider Verkettung* implementiert (Abs. 3.3).

Zu Beginn wird für die erste Region ein Startpixel gewählt. Alle diesem Startpixel (oder Voxel) benachbarten Pixel werden darauf geprüft, ob sie gemäss Gl. 4.2 dem vorgegebenen Ähnlichkeitskriterium genügen. Diejenigen Pixel, deren Winkelunterschied zum Startpixel kleiner als die gewählte Schwelle ist, werden mit dem Startpixel zu einer Region verschmolzen. Für die Region wird nun nach Gl. 4.6 ein Regionen-Farbvektor berechnet. Wieder werden alle dieser Regionen benachbarten Pixel, deren Winkelunterschied zum Regionen-Farbvektor kleiner als die gewählte Schwelle ist, mit der Region verschmolzen und danach der Farbvektor nach Gl. 4.6 aktualisiert. Dieser Prozess wird solange iteriert, bis es keine benachbarten Pixel mehr gibt, die dem Ähnlichkeitskriterium genügen. Wieder wird ein neues Startpixel für die nächste Region gesucht und seine Nachbarschaft wie oben beschrieben durchsucht. Iterativ wird das Bild auf diese Weise in Regionen (Bereiche zusammenhängender Pixel) zerlegt, bis alle Pixel Regionen zugeordnet worden sind.

Abb. 4.10 und 4.11 illustrieren die Abhängigkeit der Zerlegung von der Wahl des Winkelkriteriums. Je kleiner der Winkel, um so kleinere und mehr Regionen entstehen. Im Grenzfall $\alpha \rightarrow 0$ erhält man wieder die Pixeldarstellung, in der jedes Pixel (Voxel) für sich genommen eine Region darstellt. Abb. 4.11 zeigt, dass ein Winkelunterschied von 10-15 Grad reicht, um die Inhomogenitäten innerhalb einer

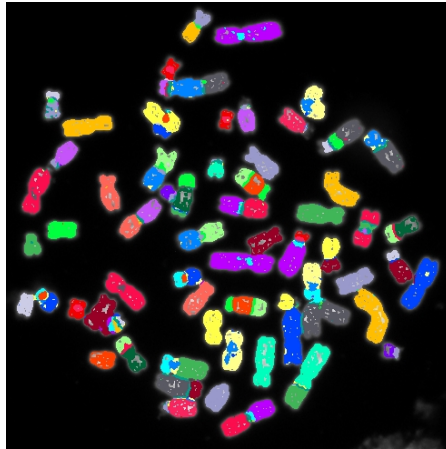


Abb. 4.10: Regionwachstum mit einem Ähnlichkeitskriterium von 3 Grad. Dargestellt sind nur die Regionengrenzen. Bereiche, die ausgefüllt erscheinen, sind Bereiche, in denen viele sehr kleine Regionen beieinander liegen. Bei diesem Kriterium erhält man fast nur Regionen, die aus einem Pixel bestehen.

Farbregion aufzufangen. In der Praxis war eine Schwelle von 5-15 Grad für eine gute Tessellierung bei der Analyse von Chromosomen ausreichend. Für sehr kleine Proben hingegen muss in der Regel ein sehr viel grösseres Winkelkriterium gewählt werden, da die Farbinformation weniger ausgeprägt ist (siehe auch Kapitel 5 und 6).

Durch die Zerlegung in Regionen werden die Pixel im Farbraum nunmehr durch die Regionen-Farbvektoren ersetzt. Abb. 4.12 zeigt die Regionvektoren als Linien dargestellt, statt der Pixel, wie in Abb. 4.6.

4.6.1 Wahl der Startpunkte beim Regionenwachstum

Die Wahl der Startpunkte beeinflusst die Zerlegung des Bildes. Die Frage nach dem für diese Art von Bildern optimalen Startpunkt ist nicht klar. Ich habe mich aus Performancegründen für eine sequentielle Suche nach Startpunkten entschieden. Der Abhängigkeit der Zerlegung von der Wahl der Startpunkte kann zum Teil durch einen anschliessenden Verschmelzungsschritt (*merge*) entgegengewirkt werden. Benachbarte Regionen, deren Farbvektoren sich um einen Winkel unterscheiden, der kleiner ist als das Ähnlichkeitskriterium, werden zu einer Region verschmolzen. In der Praxis reduziert dieser Schritt die Anzahl der Regionen je nach Experiment und gewähltem Kriterium um einige Prozent. Da weder geklärt ist, welcher Startpunkt für eine Region, noch welche Reihenfolge der Startpunkte grundsätzlich der optimale ist, und sich das Klassifikationsergebnis in der Praxis bei unterschiedlicher Wahl der Startpunkte nicht signifikant unterschied, habe ich aus Gründen der Geschwindigkeit

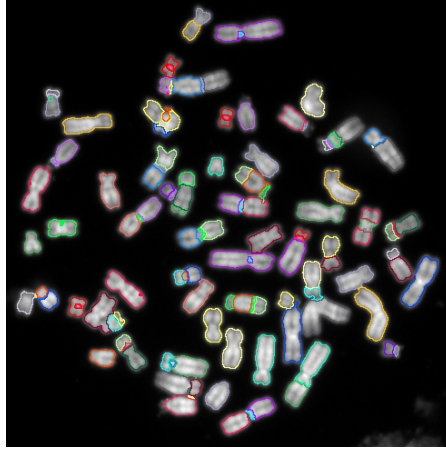


Abb. 4.11: Regionwachstum mit einem Ähnlichkeitskriterium von 15 Grad. Dargestellt sind nur die Regionengrenzen. Man erhält praktisch nur noch eine Region für ein zusammenhängendes Gebiet gleichfarbiger Pixel. Regionengrenzen benachbarter Bereiche sind besser klassifiziert.

das Bild wie zuvor beschrieben sequenziell durchgearbeitet.

4.7 Clustersuche

Da die Anzahl der unterschiedlichen Chromosomen oder chromosomalen Bereiche und damit der Klassen durch das Experiment (Hybridisierungsschema) vorgegeben ist, habe ich einen *überwachten* Clusteralgorithmus entwickelt, der auf der Basis der Farbrichtung und der Winkelunterschiede den Farbraum *der Regionenvektoren* in die vorgegebene Anzahl von Clustern unterteilt.

Initialisierung der Cluster: Zunächst werden die n Cluster durch die Wahl von Startpunkten für die n Centroide initialisiert. Dazu wähle ich von jeder der n Klassen ausgehend diejenige Region mit dem kleinsten Winkelabstand zu dieser Klasse als Centroid \vec{C}_n^0 .

Zuordnung der Regionen: Alle Regionen \vec{P}_i werden anhand ihrer Farbinformation jeweils den Centroiden \vec{C}_i^0 mit kleinstem Winkelabstand nach Gl. 4.2 zugeordnet.

Neuberechnung der Centroide: Auf der Basis der Farbe der zugeordneten Regionen und ihrer Pixel werden alle Centroide \vec{C}_n^1 nach Gl. 4.6 neu berechnet.

Konvergenz: Die Zuordnung und die Neuberechnung der Centroide wird solange iteriert, bis alle Centroide sich nicht mehr als um ein ε ändern (Abb. 4.12 und 4.13).

$$\| \vec{C}_i^k - \vec{C}_i^{k-1} \| < \varepsilon \quad \forall i, i = 1 \dots n \quad (4.8)$$

Mit der Clustersuche auf den Farbvektoren der Regionen konnte im Gegensatz zur

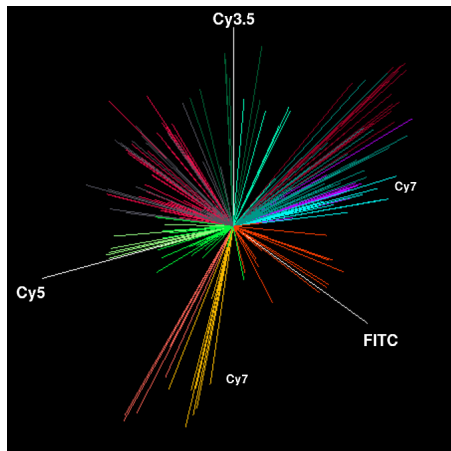


Abb. 4.12: Die Regionen-Farbvektoren im Farbraum. Die Regionen-Cluster sind durch gleiche Farbgebung angedeutet. Die negativen Achsen wurden für die Darstellung eines vierten Farbkanals verwendet. Die Kodierung der Falschfarben entspricht der in Abb. 1.6.

Clustersuche auf den Farbvektoren der Pixel ein robustes Konvergenzverhalten beobachtet werden. Pixel kritischer Übergangsbereiche werden schon beim Region Growing im Farbraum zusammengefasst. Überlappungseffekte farblich ähnlicher Cluster werden dadurch reduziert, und damit eine zuverlässigere Clusteranalyse möglich. Bei der Berechnung der Cluster-Centroide nach Gl. 4.6 wird die Summe der Winkel bzw. der orthogonalen Abstände für jeden Cluster bezüglich einer Richtung minimiert, um optimale Richtungen für die Centroiden zu finden. Dadurch erhalte ich direktionale Cluster (Abb. 4.13).

4.8 Klassifizierung der Regionen

Im letzten Schritt müssen die Cluster, repräsentiert durch ihre Centroiden, anhand ihrer Farbe den Klassen zugeordnet werden. Als sehr zuverlässig hat sich erwiesen, die Zuordnung der Centroiden bei der *Initialisierung* der Cluster zu übernehmen. Das Zuordnungsproblem wird damit einfach und effizient umgangen. Ist die Zuordnung der Cluster erfolgt, werden allen den Clustern zugeordneten Regionen und die dazugehörigen Pixel derselben Klasse zugeordnet. Anhand der Ortskoordinaten der Pixel kann das klassifizierte Bild aufgebaut werden, wobei den Pixeln die Falschfarben der Chromosomen zugewiesen werden (Abb. 4.14). Die Zuweisung von Falschfarben ist in der Regel nötig, um einen guten Farbkontrast im Klassifikationsbild zu gewährleisten, da Abbildung vom fünf und mehrdimensionalen Farbraum in den dreidimensionalen RGB-Raum nicht eineindeutig ist.

Die im Hybridisierungsschema definierten Klassen-Vektoren stellen den binären Idealfall dar. Im Experiment weicht die mittlere Farbinformation pro Klasse zum

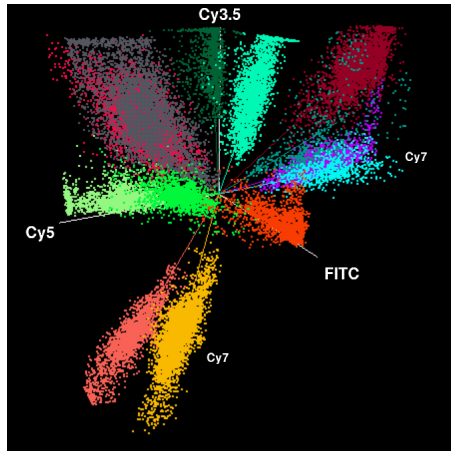


Abb. 4.13: Die Cluster im Farbraum. Dargestellt sind die Centroide der Cluster als Vektoren und die zu den Clustern gehörenden Datenpunkte der Pixel. Die negativen Achsen wurden für die Darstellung eines vierten Farbkanals verwendet. Die Kodierung der Falschfarben entspricht der in Abb. 1.6.

Teil erheblich davon ab. Diese mittlere Farbinformation pro Klasse wird bei diesem Ansatz durch die Centroide \vec{C}_n beschrieben. Die Abweichung manifestiert sich im Farbraum durch ein Wegkippen der gerichteten Cluster. Dieser Abweichung kann durch eine Anpassung der Klassen an die Bildinformation in den einzelnen Farbkanälen entgegengewirkt werden. Beispielsweise kann die Standardabweichung eines Farbkanals als Mass für den Hintergrund einer Klasse dienen, die in diesem Kanal nicht gefärbt ist. Der mittlere oder maximale Intensitätswert eines Kanals könnte als Wert für einen gefärbten Kanal gewählt werden. Allerdings sollte man hier darauf achten, keine Hintergrundpixel bei der Bestimmung dieser Werte zu berücksichtigen, da die Werte ansonsten sehr klein, und damit für Rauschen anfällig und somit möglicherweise unbrauchbar werden. Die Klassenvektoren für Chromosomen 1 und 4 aus dem Hybridisierungsschema in Abb. 1.6 könnten dann zum Beispiel so aussehen:

$$(1, 0, 1, 1, 0) \mapsto (210, 34, 160, 180, 60) \text{ für Chromosom 1}$$

$$(1, 0, 1, 0, 0) \mapsto (210, 34, 160, 24, 60) \text{ für Chromosom 4}$$

Extensive Untersuchungen an unterschiedlichen Datensätzen haben gezeigt, dass die Verkippungen so stark sein können, dass die Zuordnung der Cluster zu den Klassen nicht mehr eindeutig ist. Das heisst, dass weder die Zuordnung der Cluster zu ihren nächstenliegenden Klassen, noch die Zuordnung der Klassen zu ihren nächsten Clustern in jedem Fall eine korrekte Zuordnung liefert. Es muss daher über die Minimierung der Summe aller Winkel die richtige Zuordnung der Cluster \vec{C}_j zu den Klassen \vec{K}_i gefunden werden:

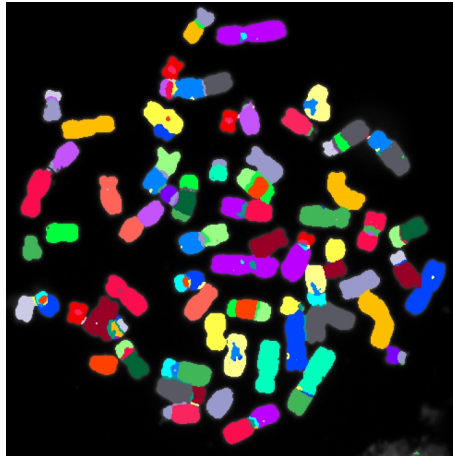


Abb. 4.14: Die Regionen-Farbvektoren werden den Klassen des Hybridisierungsschemas zugeordnet. Damit erhalten die zu den Regionen gehörenden Pixeln eine eindeutige Klasse, sind einem Chromosom zugeordnet. Anhand der Ortskoordinaten der Pixel kann das klassifizierte Bild aufgebaut werden.

$$\sum_i \mathcal{L}(\vec{K}_i, \vec{C}_j) \rightarrow \min \quad \forall ij, i, j = 1 \dots n \quad (4.9)$$

4.8.1 Merging klassifizierter Regionen

In einem abschliessenden Schritt werden benachbarte Regionen, die der gleichen Klasse zugeordnet wurden, zu einer Bildregion verschmolzen. Dies ist für dreidimensionale Untersuchungen notwendig, wenn man Volumina und (relative) Positionen von Chromosomen oder chromosomaler Bereiche bestimmen möchte, also eine Objektsegmentierung durchgeführt werden muss. Auf das Klassifikationsergebnis hingegen hat dieser Schritt keinen Einfluss.

4.9 Zusammenfassung der Methodik

Zum Abschluss dieses Kapitels möchte ich das Prinzip des Verfahrens noch einmal kurz zusammenfassen, und einige wichtige Anmerkungen anbringen. Es sollte klar sein, dass jeder Schritt, bis zu einem gewissen Grad vom Vorhergehenden beeinflusst wird: Die Art der Hintergrundkorrektur bestimmt die Punkte, die zu analysieren sind und damit die Verteilung der Datenpunkte im Farbraum. Eine sehr schlechte Vorabsegmentierung wird möglicherweise viele Pixel einbringen, die anhand ihrer Farbinformation nur schwer zu klassifizieren sind, und für Überlappungen nahe beieinander liegender Cluster verantwortlich sind. Die Wahl des Ähnlichkeitskriteriums

bei der Regionensuche bestimmt, in wieviele Regionen das Bild zerlegt wird und durch welche Farbe diese Regionen im Farbraum beschrieben werden. Der Einfluss auf die Anzahl und Verteilung der Regionenvektoren im Farbraum hat wiederum grossen Einfluss auf das Ergebnis der Clusterung der Regionen-Farbvektoren und der Separierbarkeit nahe liegender Cluster. Das wiederum ist entscheidend für ein korrektes Klassifikationsergebnis.

Das Verfahren bietet eine flexible Plattform für M-FISH Experimente jeglicher Art, für die die passenden Parameter und Einstellungen einmal gefunden werden müssen. Sind diese Parameter für ein Experiment ermittelt, liefert das Verfahren robuste Klassifikationsergebnisse, bei gleichbleibenden Einstellungen. Grob fehlerhafte Klassifikationsergebnisse waren fast immer auf ungünstige Parameterwahl zurückzuführen. Einzelne fehlerhafte Klassifikationen bei ansonsten überwiegend richtiger Klassifikation liefen praktisch immer auf ein Hybridisierungsproblem bei der Probe in mindestens einer Farbe hinaus, das ein so schwaches Signal lieferte, dass eine korrekte Identifizierung der Probe anhand der Farbsignatur nicht möglich war.

Da es bei aberranten Zellen häufig zu Umbauten kommt, ist die Gefahr gradueller Übergänge in der Farbe und damit die Verschmelzung benachbarter Regionen unterschiedlicher Farbe beim Region Growing gross. Deshalb war es wichtig, einen Region Growing Algorithmus *centroid verketteter* Art zu realisieren, um die Zerlegung in solchen Übergangsbereichen besser zu kontrollieren. Auf die Bedeutung der richtigen Wahl des Ähnlichkeitskriterium und der Schwelle bei solchen Algorithmen wurde hingewiesen (vgl. Abs. 3.3). Dies werde ich in den Kapiteln 5 und 6 dokumentieren und diskutieren. Werden diese richtig gewählt, ist die Abhängigkeit der Resultate von der Wahl der Startpunkte und ihrer Reihenfolge nach meinen Untersuchungen nicht signifikant.

Für alle hier vorgestellten Algorithmen, sowohl beim Regionenwachstum als auch bei der Klassifizierung, lassen sich homogenere Klassifikationsresultate erzielen, wenn für jedes Pixel ein gemittelter Farbvektor aus seiner Nachbarschaft berechnet wird (*hybride* Farbinformation). Damit lassen sich gewisse Rauschartefakte im Farbraum entfernen, ohne die originäre Farbinformation zu modifizieren.

Zusammenfassung der Methodik:

1. *Hintergrundkorrektur* auf DAPI Bild, sonst auf dem Maximum-Projektionsbild
2. *Tesslierung/Zerlegung* der relevanten Bildbereiche in Bereiche konstanter Farbe (Abb. 4.15B)
3. Überwachte *Clustersuche* im Farbraum auf den Regionen-Farbvektoren und Anpassung der Klassenvektoren (Abb. 4.15C)
4. *Klassifizierung* der Cluster
5. *Verschmelzen* (Merging) benachbarter Regionen gleicher Klasse
6. Erzeugung eines Klassifikationsbildes (Abb. 4.15D)

7. ggf. Sortierung in ein *Karyogramm*

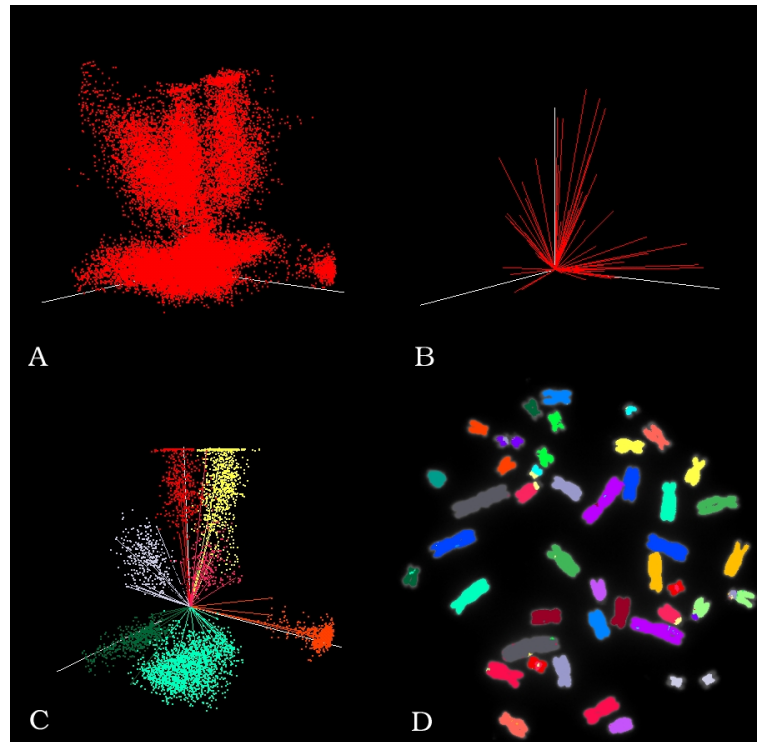


Abb. 4.15: Das Prinzip der Methodik in Bildern. Das Originalbild wird in Regionen konstanter Farbe zerlegt. Die Pixeldaten im Farbraum A) werden nun durch die Farbvektoren der Regionen repräsentiert B). Clustersuche auf den Regionenvektoren C). Dargestellt sind die Regionen als Vektoren mit ihren assoziierten Pixeln in der gleichen Farbe. Die Zuordnung der Cluster zu den Klassen liefert das Klassifikationbild D).

Kapitel 5

Anwendungen

In diesem Kapitel stelle ich kurz die unterschiedlichen Multicolor-FISH Experimente vor, die als Anwendung für das von mir entwickelte und in Kapitel 4 Verfahren in Frage kommen.

Die hier dargestellten typischen Beispiele sollen die unterschiedlichen Anwendungsbereiche der M-FISH Technologie, damit aber auch die Anforderungen an die Methodik und ihr Potenzial illustrieren. Für unterschiedliche M-FISH Experimente müssen gewisse Parameter angepasst werden. Eine ausführlichere Analyse nehme ich in Kapitel 6 vor. Dadurch wird schon einiges an Problemen, aber auch Ergebnisse vorweggenommen, was zu Überschneidungen mit dem nächsten Kapitel führen wird.

Eine Beschreibung der Präparation und Aufnahme mittels M-FISH findet sich zum Beispiel in [Eil98].

5.1 Anwendungen in 2D

Zunächst stelle ich Anwendungsbeispiele für M-FISH vor, die zweidimensionale Bilder liefern, die es zu analysieren gilt. Die Darstellung erfolgt in Falschfarben. Als Hintergrund blende ich das DAPI-Bild ein, was bei kleineren Proben die Anschaulichkeit erhöht. Die Analyse erfolgt bei Zellen, die sich in der Metaphase befinden. Die Chromosomen sind hier in kondensiertem Zustand, was zu intensiven Farbsignalen führt.

5.1.1 Normale Metaphase

Das Standard-Experiment für die M-FISH Technologie ist die Identifizierung der Chromosomen anhand ihrer spezifischen Farbsignatur. Im Metaphasenbild werden die Chromosomen identifiziert (Abb. 5.1), und in ein *Karyogramm* sortiert, wie es

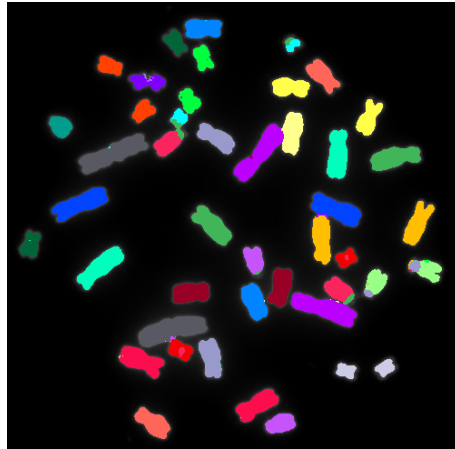


Abb. 5.1: Die klassifizierten Metaphase-Chromosomen eines gesunden Patienten.

in Abb. 5.2 zu sehen ist. Für Klassifikationen normaler Metaphasen (gesunder Zellen) liefert die Methodik mit einem Winkelkriterium von $5 - 15^\circ$ (vgl. Kapitel 4.5) sehr zuverlässige und homogene Ergebnisse, wenn die Hybridisierungen ausreichend gute Clusterstrukturen im Farbraum liefern, was bei den zahlreichen im Laufe dieser Arbeit durchgeführten Untersuchungen der Fall gewesen ist. Untersucht habe ich dabei in der Mehrzahl 5-Farben Experimente von peripheren Blutlymphocyten. Falls alle Chromosomen ausreichend getrennt sind, können die Regionen durch die DAPI Segmentierung ersetzt werden.

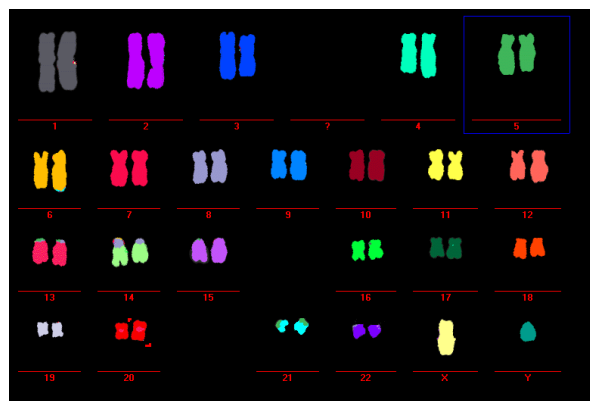


Abb. 5.2: Die Chromosomen werden in ein Karyogramm sortiert.

5.1.2 Aberrante Metaphase

Ein für die Mediziner interessanteres Beispiel einer aberranten Metaphase eines kranken Patienten zeigt Abb. 5.3. Man sieht wie komplex die *interchromosomalen* Umbauten sein können.

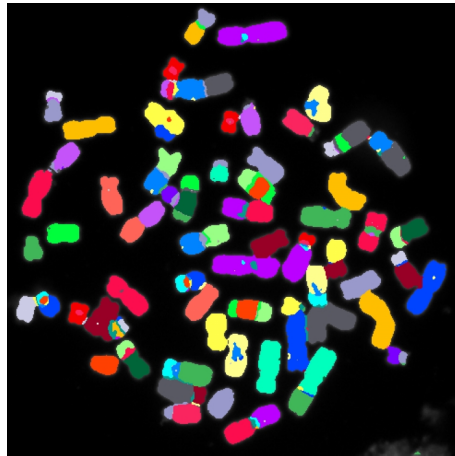


Abb. 5.3: Die Metaphase-Chromosomen eines kranken Patienten. Man beachte die komplexen interchromosomalen Umbauten.

Bei der Analyse aberranter Metaphasen mit vielen interchromosomalen Umbauten (Abb. 5.3) sind kleine Umbauten, die wenig Pixel umfassen und die Übergangsregionen zwischen unterschiedlichem DNA Material die kritischen Bereiche. Hier kann es zu Überlagerungen der Farbe kommen, was eine richtige Zuordnung erschwert. Für das Winkelkriterium gilt ein Bereich von $5 - 15^\circ$. Man sollte bei solchen Zellen in der Wahl des Winkels (zur Erinnerung: Winkelkriterium ist entscheidend für die Zerlegung in Regionen) vorsichtig sein, da ein zu grobes Kriterium Umbauten verdecken kann. Im gezeigten Beispiel weisen die X-Chromosomen (hellgelb) in ihrem mittleren Abschnitt Kreuzhybridisierungen von Chromosom 9 (hellblau) auf. Wird das Kriterium $\geq 15^\circ$ gewählt, verschwinden diese im Klassifikationsbild. Andererseits werden die Übergänge benachbarter Bereiche unterschiedlichen Chromosomen-Materials bei grösseren Winkeln besser klassifiziert. Man muss daher einen guten Kompromiss zwischen Homgenität und Farbsensitivität vorab interaktiv finden. Abb. 5.4 zeigt die Clusterstruktur und ist typisch für M-FISH Metaphasen Experimente in 2D.

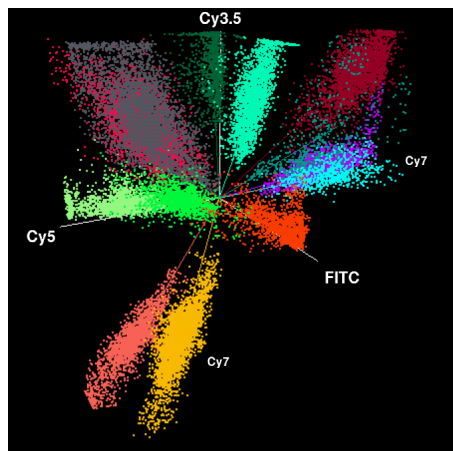


Abb. 5.4: Der Farbraum eines 5-Farben M-FISH Experiments einer aberranten Metaphase. Gezeigt werden der erste, dritte und vierte Farbkanal. Die negativen Achsen wurden für eine weitere Farbe verwendet. Die Kodierung der Falschfarben entspricht der in Abb. 1.6.

5.1.3 Chromosomale Barcodes

Für Untersuchungen *intrachromosomaler* Umbauten eignen sich oben genannte Experimente nicht, da es sich bei den verwendeten Proben, obwohl chromosomen-spezifisch, um sog. *whole chromosome paints* handelt, die überall an ein spezifisches Chromosom binden. Es gibt jedoch Chromosomen-spezifische DNA-Proben, sog. YAC Klone (*Yeast Artificial Chromosomes*, cytogenetische Auflösung etwa 500 kBp), mit denen sich gewisse Bereiche innerhalb eines Chromosoms anfärben lassen. Markiert man ein Chromosom mit Hilfe solcher Proben von Anfang bis Ende in unterschiedlichen Farben oder Farbkombinationen, erhält man ein farbiges Streifenmuster wie *Barcodes*, das die Untersuchung intrachromosomaler Aberrationen ermöglicht. Abb. 5.5 zeigt ein Beispiel für ein Chromosom mit Originalbild, Klassifikationsbild und die Reihenfolge der Markierungen in Falschfarben.

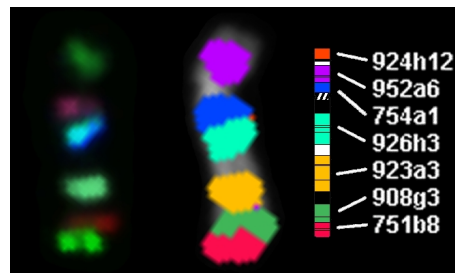


Abb. 5.5: Chromosom im Original (links), seine Klassifikation (mitte) und den Reihenfolge der unterschiedlich markierten Hybridisierungen (rechts). (Quelle: [Sar01])

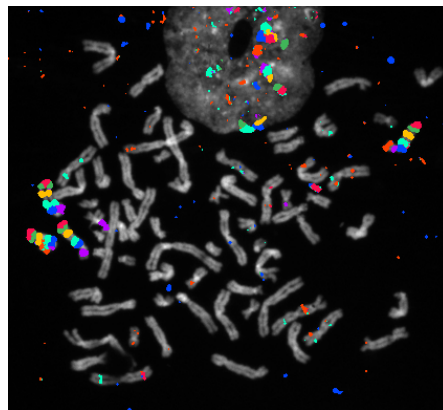


Abb. 5.6: Die mit der Barcode-Technik gefärbten Metaphase-Chromosomen eines kranken Patienten. Gefärbt sind nur die Chromosomen 12

Abb. 5.6 zeigt das Klassifikationsbild für eine Metaphase. Da für ein Chromosom wesentlich mehr Farbstoffe benötigt werden als für interchromosomale Untersuchungen, werden nicht mehr alle Chromosomen angefärbt. In diesem Experiment wurden die Chromosomen 12 markiert. Die Metaphase besitzt vier davon, von denen zwei normal sind und zwei eine Deletion eines Abschnitts aufweisen.

Wichtig für ein aussagefähiges und zuverlässiges Ergebnis ist hier die Wahl der Proben. Die Bindungsgrenzen der Proben sollten so wenig wie möglich überlappen, um

akkurate Analysen durchführen zu können. In der Praxis können je nach Hybridisierungsschema etwas grössere Winkelkriteria von bis zu 30° gewählt werden, um bessere Ergebnisse in den Grenzbereichen zu erhalten. Da im Experiment in der Regel weniger unterschiedliche Bereiche markiert werden, reduziert sich die Anzahl der Hybridisierungen pro Probe. Damit erhöht sich der Winkelabstand zwischen den Klassen, was grössere Toleranzen beim Winkelkriterium zulässt.

5.1.4 Telomere

M-FISH Experimente mit Chromosomen-spezifischen Proben sind in der cytogenetischen Auflösung beschränkt, unabhängig von der Aufnahmetechnik und vom Design der Proben. Diese kann jedoch durch die Verwendung ortsspezifischer Proben erhöht werden. Die Auflösung ist dann nur noch durch die Grösse der Proben limitiert.

Telomere nennt man die Endregionen von Chromosomen. Sie können durch spezielle telomerische und subtelomerische Proben kombinatorisch angefärbt werden¹. Durch ihre geringe Grösse umfassen sie in den zu analysieren Bildern nur wenige Pixel, und haben daher schlechte statistische Eigenschaften. Jedes Pixel besitzt ein grosses Gewicht. Dies macht sie für Rauschen und Störungen jeglicher Art sehr anfällig und die Analyse damit sehr schwierig, da die Farbinformation sehr gestört sein kann. Insbesondere führen unterschiedlich grosse Signale in den einzelnen Farbkanälen zu sehr starken Farbgradienten innerhalb kleiner Bereiche. Das hat für meine Methodik die Konsequenz, dass das Winkelkriterium sehr viel grösser ausfallen sollte, üblicherweise $30 - 50^\circ$. Sind alle Signale gut getrennt, kann auch hier 90° beim Regionenwachstum gewählt werden, um möglichst grosse und homogene Signale im Klassifikationsbild zu erhalten. Ist dies nicht der Fall, muss ein Kompromiss interaktiv gefunden werden.

Wird das Kriterium zu klein gewählt, wird das richtig klassifizierte Signal von falsch klassifizierten Bereichen umgeben sein, so dass eine Identifizierung erschwert wird. Abb. 5.7 zeigt ein Klassifikationsbild für ein solches Experiment.

Es gibt noch weitere Konsequenzen. Interphase-Kerne, die zusammen mit der Metaphase aufgenommen wurden, sollten möglichst entfernt werden, da sie durch ihre Grösse die Bildinformation dominieren, und damit die Analyse im Farbraum wesentlich erschweren können. Die Farbinformation im Kern muss nicht zwangsläufig mit derjenigen der kleinen Telomere identisch sein. Das ist aber noch nicht alles. Die Grösse der Signal und die auftretenden Störungen führen dazu, dass es im Farbraum praktisch zu keiner Clusterausbildung kommt (Abb. 5.8). Dieser Schritt kann daher in der Analyse ausgelassen werden: die Regionen werden direkt den durch das Hybridisierungsschema definierten Klassen zugeordnet.

¹Die cytogenetische Auflösung wird durch die Verwendung sog. BAC, PAC und P1 Klone auf etwa 500 kBp erhöht.

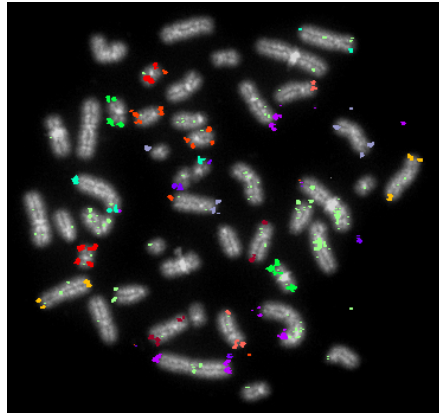


Abb. 5.7: Klassifikationsbild von vielmarkierten Telomeren. Gefärbt wurde nur die Hälfte des Chromosomensatzes.

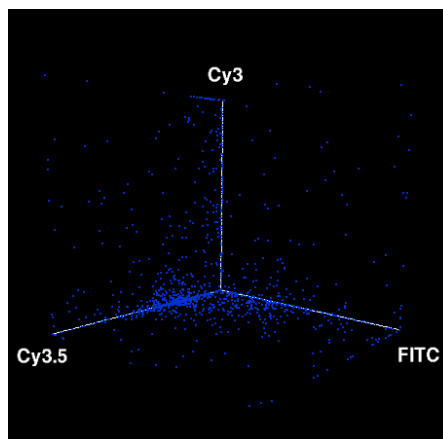


Abb. 5.8: Farbraum von Abb. 5.7. Man erkennt die mangelnde Clusterausbildung.

Mit Hilfe eines speziell entwickelten subtelomerischen Probensatzes (**M-TEL**) konnte mit der von mir entwickelten Methodik eine chromosomale Aberration von Chromosom 2 auf 7 bei einem Patienten mit geistiger Zurückgebliebenheit detektiert werden, die mit konventionellen M-FISH Methoden nicht aufzulösen war [Bwn01, Sar01]. Das Karyogramm der Metaphase in Abb. 5.7 zeigt Abb. 5.9. Das verwendete Hybridisierungsschema ist in Abb. 5.10 zu sehen.

5.2 Anwendungen in 3D

Die 3D Anwendung von M-FISH in Interphase Kernen steckt noch in den Kinderschuhen. Sie hat mehr biologische als medizinische Relevanz. Hier gilt das Hauptinteresse der Verteilung der Chromosomen im Zellkern. Man möchte durch gleichzeitige relative Positionbestimmung möglichst vieler Chromosomen (am besten aller 24 auf einmal) Rückschlüsse ziehen, ob die Verteilung zufällig ist oder nicht.

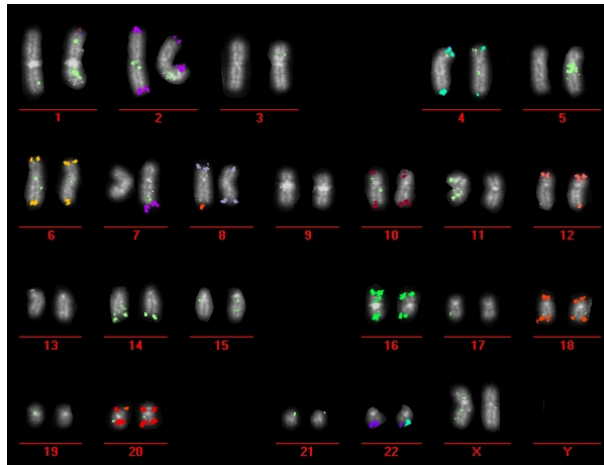


Abb. 5.9: Karyogramm eines Patienten mit geistiger Zurückgebliebenheit. Man erkennt die detektierte Translokation von Chromosom 2 auf 7. (Quelle: [Sar01])

Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
Colour	grey	purple	blue	cyan	green	yellow	red	grey	blue	red	yellow	red	red	green	purple	green	green	orange	grey	red	cyan	purple	yellow	green
FITC		x		x				x				x											x	
Cy3						x				x	x					x								
Cy3.5				x						x				x							x	x		
Cy5.5							x									x		x		x	x			
Description																								

Abb. 5.10: Hybridisierungsschema eines 4-Farben M-FISH Experiments mit 12 Telomer-Proben. (Quelle: [Sar01])

5.2.1 3-Farben Experiment mit sieben Chromosomenpaaren in der Interphase

Abb. 5.11 zeigt ein 3-Farben Experiment mit sieben nach dem Schema in Tab. 5.1 gefärbten Chromosomenpaaren. Das Klassifikationsbild ist in Abb. 5.12 zu sehen.

Chromosom	1	3	7	9	12	15	20
FITC	x	-	x	x	-	-	x
TRITC	x	x	-	x	x	-	-
Cy5	x	x	-	-	-	x	x

Tab. 5.1: Hybridisierungsschema für das gezeigte 3-Farben Experiment in der Interphase

Die von mir untersuchten Zellen waren fast ausschliesslich Fibroblasten. Ein konfokales Laser Scanning Mikroskop für M-FISH Experimente steht noch nicht zur Verfügung. Mit herkömmlichen CLSMs lassen sich 3-Farben Experimente jedoch analysieren, wenn man FITC, Cy3 und Cy5 als Markierungsfarbstoffe wählt. Diese lassen sich mit den vorhandenen Linien der HeNe und Ar-Ionen Laser anregen (488

und 568 nm). Die Analyse dreidimensionaler Interphase Experimente mit mehr als drei Farbstoffen (s. Abs. 6.6, Abb.6.21) ist zur Zeit nur mit Epifluoreszenzmikroskopie mit anschliessender Dekonvolution für jeden einzelnen Kanal möglich.

Die Dekondensation der Chromosomen in der Interphase führt zu grösseren Farbschwankungen und damit zu ausgedehnten Clustern im Farbraum. Damit werden Überlappungen wahrscheinlicher. Gleichzeitig führen Überlappung und Durchdringung der Chromosomen zu ausgedehnten Bereichen im Bild, in den sich die Signale überlagern und die Analyse erschweren (s. Abb. 6.22).

Auch hier muss das Winkelkriterium ausbalanciert sein ($10 - 30^\circ$), da die Objekte geographisch zusammenhängen und das Verfahren nur Farbunterschiede, aber keine Intensitätskanten detektieren kann. Zu kleine Winkel führen leicht zu sehr vielen fragmentierten Regionen mit schlechter Farbcharakteristik, während zu grosse Winkel Objektverschmelzungen zur Folge haben können. Der Toleranzbereich des Winkelkriteriums hängt hier sehr stark vom Hybridisierungsschema ab. Hier sollte aufgrund der Grösse und Variabilität der Farbcluster unbedingt darauf geachtet werden, einen möglichst grossen Abstand zu erzielen, also möglichst viele Farbstoffe bei gleichzeitig möglichst wenigen Simultan-Hybridisierungen pro Target (vgl. Abs. 6.1).

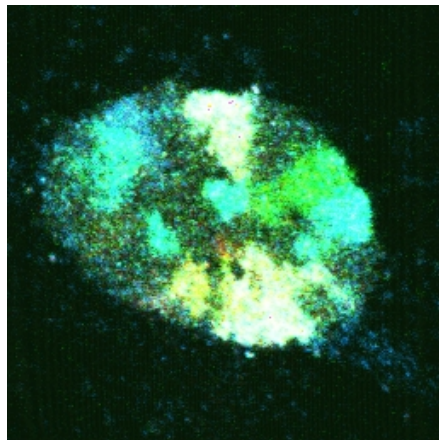


Abb. 5.11: 3-Farben M-FISH Experiment mit 7 unterschiedlich kombinatorisch gefärbten Chromosomenpaaren.



Abb. 5.12: Klassifikationsbild von Abb. 5.11. Die Falschfarben der Chromosomen sind entsprechend Abb. 1.6

Kapitel 6

Ergebnisse und Diskussion

Die Beispiele im vorhergehenden Kapitel haben einiges an Ergebnissen schon vorweggenommen. Sie liessen schon deutlich werden, dass die Klassifikationsgüte stark vom Grad der Kondensation der Proben, sowie der Güte der Hybridisierung, die wiederum einen direkten Einfluss auf die Farbinformation und damit die Farbraumcharakteristik hat, abhängt. Weiterhin wurde der Zusammenhang zwischen abnehmendem Farbinformationsgehalt bei abnehmender Probengrösse an Beispielen illustriert, was eine korrekte Klassifikation ohne Ortsinformation praktisch unmöglich macht, wenn eine gewisse Spotgrösse erreicht werden soll. Die Hinzunahme der Ortsinformation ermöglicht robusteres Klassifikationsverhalten. Überlappungen von Clustern im Farbraum können unter Umständen aufgelöst werden, da die durch den Klassifikator vorgegebenen rigiden Klassifikationsgrenzen diffuser werden. Das ist bei pixelbasierter Klassifikation nicht möglich. Die Clustersuche hat einen weiteren Vorteil. Sie hat für die Klassifikation die Bedeutung einer indirekten Anpassung der Klassenvektoren an die Bildinformation. Klassifikatoren jeglicher Art, die nur auf den idealen Klassenvektoren arbeiten, werden bei der gegenwärtigen Güte der Bilder keine befriedigenden Resultate liefern. Auf diese Zusammenhänge wird in diesem Kapitel noch einmal eingegangen.

Alle in diesem Abschnitt dargelegten Beispiele sind repräsentativ und verallgemeinerbar. Sie dienen zur Illustration meiner Ergebnisse.

Die Methodik mitsamt ihrer biologischen Anwendung und Ergebnissen wurde in mehreren Publikationen veröffentlicht, u.a. in [Bwn01, Bwn00, Sar01]. Darüber hinaus wurde das Klassifikationsmodul als Bestandteil in die *Leica QFISH/MCK* Karyotypisierungs-Software integriert. Die Klassifikationsgüte des Verfahrens beträgt nach Studien unterschiedlicher Gruppen bei (2D) aberranten Metaphasen nahe 100% [Ren00, Soi00]. Die Güte bei Telomeren ist nach internen unveröffentlichten Studien ebenfalls nahe 100%, und hängt praktisch nur vom Erfolg der Hybridisierung ab. Dazu mehr im Abs. 6.5 in diesem Kapitel. Auch die Klassifikation chromosomaler Barcodes, vor allem in den Übergangsbereichen ist zuverlässig.

6.1 Auflösung und Hybridisierungsschema

Um ein korrektes Klassifikationsresultat zu erzielen, muss auch das Experiment bestimmte Anforderungen erfüllen. Bei Hybridisierungsexperimenten treten schwer quantifizierbare Rauschquellen auf. Das ideale Hybridisierungsexperiment sollte eine homogene Färbung der biologischen Präparate gewährleisten. Das ist jedoch nur selten der Fall, meistens sind die Intensitätsschwankungen innerhalb kleiner Bereiche sehr stark. Die Gründe für diese Inhomogenitäten sind im Detail noch nicht genau verstanden. Bei Mehrfachhybridisierungen, wie es bei M-FISH Experimenten der Fall ist, kommen noch weitere Schwierigkeiten hinzu. Bestimmte Fluoreszenzkombinationen führen bei einigen Chromosomen zu sehr schwachen Hybridisierungen. Die Hybridisierungsqualität der unterschiedlichen Farbstoffe und die Bildaufnahme der unterschiedlichen Farbkanäle müssen aufeinander abgestimmt werden. Durch optimierte Hybridisierungsprotokolle kann der Experimentator daher die Grundlage für zuverlässige Analysen schaffen. Es gibt jedoch ein weiteres sehr wichtiges Kriterium, das im Experiment Beachtung finden sollte, das jedoch oft genug übersehen wird.

Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
Colour																								
FITC	x			x		x		x	x		x		x		x		x		x		x		x	
Cy3			x		x			x	x		x		x		x				x	x			x	x
Cy3.5	x		x	x			x			x	x				x		x		x					x
Cy5	x				x	x	x	x				x		x	x	x							x	
Cy5.5		x	x		x	x			x	x		x										x		x
Description																								

Abb. 6.1: Typisches Hybridisierungsschema eines 5-Farben m-FISH Experiments mit 24 Proben. Der minimale Winkelunterschied zwischen benachbarten Klassen 35.3°

Die maximale theoretische Auflösung¹ bei M-FISH Experimenten wird erzielt, wenn die Anzahl der verwendeten Fluorochrome der Anzahl der Proben entspricht. Möchte man 24 unterschiedliche Chromosomen oder Proben unterscheiden, wäre es optimal, 24 spektral unterscheidbare Fluorochrome, für jede Probe eines, zu verwenden. Leider gibt es nicht so viele Fluorochrome, die die notwendigen unterschiedlichen spektralen Eigenschaften besitzen, um solche Experimente zu ermöglichen. Gleichzeitig müssten auch geeignete, an die Fluoreszenzcharakteristika der Farbstoffe angepasste Bandpassfilter entwickelt werden, deren Realisierbarkeit an dieser Stelle noch in Frage gestellt werden muss. Kurzum: der Umsetzung stehen noch viele technische Unwägbarkeiten im Wege. Trotzdem ist die Entwicklung weiterer Fluorochrome neben der Verbesserung gegenwärtiger Hybridisierungsprotokolle eine der vordringlichsten Aufgaben der Forschung auf dem Gebiet der M-FISH Technologie.

Selbst bei Erhöhung der Anzahl verfügbarer Fluorochrome wird man aber auch in

¹Mit Auflösung ist hier nicht die Genomische gemeint, die durch die verwendeten DNA Proben gesetzt ist (sie beträgt bei M-FISH einige Mbp), sondern die Auflösung im Sinne der Klassifikation.

der Zukunft auf kombinatorische Färbestrategien angewiesen sein, da damit auch die Anzahl gleichzeitig unterscheidbarer Proben steigt, um komplexere Untersuchungen durchzuführen. Hat man also eine bestimmte Anzahl von Fluorochromen n vorgegeben, lassen sich durch Wahl des Schemas der Hybridisierung die Winkel der Klassen untereinander maximieren, um so einen grösstmöglichen Unterschied zwischen den Klassen zu erhalten. Dies ist für eine nachfolgende Klassifizierung in der Regel von grossem Vorteil. Betrachten wir ein binäres Hybridisierungsschema, dann ist der Cosinus des Winkels zwischen zwei solchen Vektoren \vec{a} und \vec{b} definiert als

$$\cos \alpha = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2} \sqrt{\sum_i b_i^2}} = \frac{s_{a,b}}{\sqrt{n_a} \sqrt{n_b}} \quad (6.1)$$

wobei $s_{a,b}$ die Anzahl gleicher Labels von \vec{a} und \vec{b} ist; n_a und n_b sind die Anzahl der Labels von \vec{a} und \vec{b} .

Der Winkel zwischen zwei Vektoren nimmt also mit der Anzahl gleicher Labels ab und mit der Anzahl der verwendeten simultanen Labels zu. Den Idealfall stellt also $s = 0$ für alle Paare dar, was gleichbedeutend damit ist, dass alle Klassenvektoren orthogonal zueinander sind. Der *worst case* für ein Experiment mit n Fluorochromen ist der Winkel zwischen dem Vektor einer Klasse mit n Hybridisierungen und einem mit $n - 1$, bzw. bei einem Experiment mit maximal m simultanen Hybridisierungen der Winkel zwischen einem solchen Vektor und einem mit $m - 1$ gleichen Labels, also hat man

$$\cos \alpha_{worst} = \sqrt{\frac{(m-1)}{m}} \quad (6.2)$$

als grössten auftretenden Winkelcosinus (und damit kleinsten Winkel) im Schema. Betrachten wir als Beispiel das binäre Hybridisierungsschema Abb. 6.1. Die maximale Anzahl simultaner Hybridisierungen beträgt hier drei. Die Klassen 11 und 13 bilden den kleinstmöglichen Winkel in diesem Schema (unter anderen). Mit $s_{11,13} = 2$ und $n_{11} = 3, n_{13} = 2$ ergibt sich $\cos \alpha = 2/(\sqrt{2}\sqrt{3}) = \sqrt{2/3}$. Dies entspricht einem Winkel von 35.3° . Bei einem 7 Farben Experiment mit 24 unterschiedlichen Proben hingegen kommt man ohne dreifache Hybridisierungen aus (Abb. 6.2). Der minimale auftretende Winkel zwischen den Klassen beträgt 45° , also einen im Mittel 27% verbesserten Winkelabstand gegenüber dem 5 Farben Experiment mit 24 Proben. Die minimalen auftretenden Winkel für verschiedene Anzahlen von Fluorochromen bei 24 Proben ist in folgender Tabelle nochmals zusammengefasst.

Bei gleicher Anzahl unterschiedlicher Proben kann durch Erhöhung der Anzahl der Fluochrome die Anzahl simultaner Hybridisierungen reduziert, und damit die Diskriminierung im Farbraum erhöht werden. Abb. 6.2 zeigt ein Hybridisierungsschema für ein 7 Farben Experiment mit 24 unterschiedlichen Proben. Jede Probe hat nur noch zwei simultane Färbungen. Das bedeutet einen minimalen Winkelunterschied

Anzahl Fluorochrome	Max. Anzahl Labels	Min. Winkel im Farbraum
5	3	35.3°
5	4	30°
5	5	26.6°
7	2	45°
8	alle 2	60°

Tab. 6.1: Abhängigkeit des minimalen Winkelabstands der Klassenvektoren von der Anzahl der verwendeten Fluorochrome und der maximalen Anzahl der simultanen Hybridisierungen

zwischen den Klassen von 45°.

Ein 7 Farben M-FISH Experiment samt Protokoll und Ergebnissen wird beschrieben in [Azo00]. Ausführlichere Diskussionen zum Thema Auflösung und Genauigkeit bei Klassifikationen von M-FISH Bildern findet sich in [Azo00, Gar99, Cas00, Sar01].

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	ps	
Color																										
DEAC						x				x		x	x									x			x	x
SG			x		x	x			x						x						x					
Cy3					x		x	x						x		x	x								x	x
Cy3.5									x		x		x				x		x					x		x
Cy5										x				x					x				x			
Cy5.5	x			x				x				x			x					x						
Cy7		x		x			x				x							x			x					

Abb. 6.2: Typisches Hybridisierungsschema eines 7-Farben m-FISH Experiments mit 24 Proben. Jede Probe hat nur noch zwei simultane Färbungen. Das bedeutet einen minimalen Winkelunterschied zwischen benachbarten Klassen von 45°.

6.2 Klassifikationsgüte von M-FISH im Vergleich mit SKY für Metaphasen in 2D

Zwei unterschiedliche Systeme zur Aufnahme und automatischen Analyse und Karyotypisierung von vielfarben markierten Metaphasen haben sich auf dem Markt durchgesetzt. Zum einen das **SKY** System von Applied Spectral Imaging (ASI) [Sch96, Gar96, Gar99], sowie das **MFISH** System von Leica [Spe96a, Spe96b, Eil98], welches als Klassifizierungsmodul das von mir entwickelte Verfahren einsetzt. Die Methode der Bildanalyse, die im SKY System zum Einsatz kommt, ist mir nur soweit bekannt, wie es in [Sch96, Gar96] beschrieben wird. Mögliche Weiterentwicklungen kann ich dabei nicht ausschliessen.

Das SKY System verwendet ein Interferometer und Fourierspektroskopie, um das

Spektrum in jedem Pixel der CCD Kamera zu detektieren. Nur eine Aufnahme wird benötigt, die jedoch eine längere Belichtung erfordert. Abgesehen von einem dreifachen Bandpass-Filter, der das Anregungslicht blockt, sind bei diesem System keine weiteren spezifischen Filter nötig. Das MFISH System verwendet Paare spezifischer Anregungs- und Detektionsfilter für jedes eingesetzte Fluorochrom, um die Spektren zu detektieren. Dabei werden mehrere Bilder, mindestens fünf, aufgenommen. Der Wechsel der Filter erfolgt automatisch.

Beide Systeme haben ihre Vor- und Nachteile. Während das SKY System mit einer Aufnahme auskommt, benötigt das M-FISH System mindestens fünf unterschiedliche Aufnahmen. Garini et al. [Gar99] argumentieren, dass das SKY System aufgrund seiner höheren spektralen Auflösung auch eine sehr viel zuverlässigere Klassifikation als die filterbasierte Methode zulässt, da mehr Datenpunkte aus dem Spektrum für jedes Pixel gemessen werden.

Das SKY System benötigt hochwertige optische Bauteile, die chromatische Aberrationen kompensieren, um sicherzustellen, dass alle Chromosomen bei allen Wellenlängen im Fokus sind. Durch die wesentlich längeren Belichtungszeiten können sensitive Farbstoffe leicht ausgebleicht werden, die eine eventuelle Wiederaufnahme unmöglich machen. Das M-FISH System ist flexibler. Falls ein Fluorochrom ein schwaches Signal liefert, kann die Belichtungszeit für diesen Kanal individuell justiert werden. Die aufgenommenen Einzelbilder ermöglichen eine Überprüfung der Klassifizierung, was beim SKY System nicht möglich ist. Beim Wechsel der spezifischen Filtersätze können jedoch laterale Verschiebungen des Fokus auftreten, die korrigiert werden müssen. Problematisch wird dieser Effekt bei sehr kleinen Proben. Direkte Vergleiche dieser beiden Systeme wurden von verschiedenen Labors durchgeführt und auf der Konferenz *QMC 2000* vorgestellt [Ren00, Soi00]. Die Resultate zeigen, dass beide Systeme ähnlich gute Ergebnisse -fast 100%- vorweisen können, wobei das Leica System in einigen Fällen sogar etwas bessere Klassifizierungen liefert, obwohl das System auf einer Aufnahmetechnik mit geringerer spektraler Auflösung beruht. Diese Ergebnisse sind in grobem Widerspruch zu einer vorausgesagten zehnfach höheren Auflösung des von SKY zu MFISH [Gar99].

6.3 Methodik

6.3.1 Binäre Klassifikation

Ein direkte und vielleicht naheliegenste Möglichkeit M-FISH Bilder zu klassifizieren wäre, jeden Kanal mit geeigneten Rauschfiltern zu glätten, einen optimalen Schwellwert zu finden und zu setzen (Abb. 6.3), dies für jeden Kanal durchzuführen und anschliessend die Pixel anhand ihrer binären Farbinformation zu klassifizieren (Abb. 6.4).

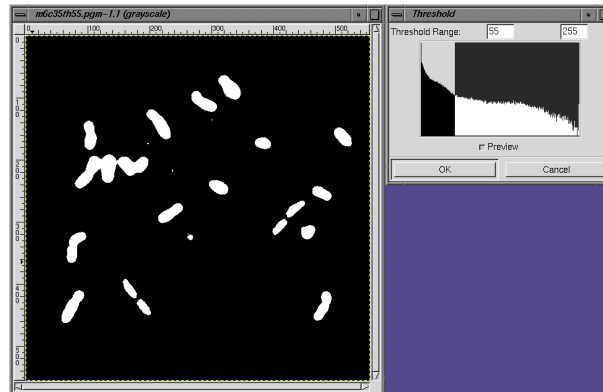


Abb. 6.3: Für jeden Farbkanal wird ein Schwellwert im Histogramm gefunden. Er beträgt hier 55 für den ersten Farbkanal. Man erhält ein binäres Bild.



Abb. 6.4: Klassifikation einer normalen Metaphase mit dem binären Ansatz.

Das Klassifikationsbild (Abb. 6.4) wurde durch bestmögliche, interaktive Schwellwertfindung für jeden Kanal berechnet. Trotzdem zeigt die Klassifikation einen Ringeffekt bei den Chromosomen, der von der unterschiedlichen Segmentierungsgrösse in den Einzelkanälen herrührt. Weiterhin ist eine zum Teil starke Fragmentierung, speziell bei Chromosom 2 zu erkennen, das auf ein schwaches Hybridisierungssignal in den inneren Bereichen zurückzuführen ist. Selbst wenn für diesen Fall ein besserer Satz von Schwellwerten gefunden werden sollte, ist eine Automatisierung dieses Verfahrens ausgesprochen schwierig. Abb. 6.6 und Abb. 6.5 zeigen, wie anfällig das Verfahren ist, wenn der Schwellwert nur für einen Kanal schlecht gesetzt wird.

Auch wenn es einen solchen optimalen Algorithmus gibt, der Ringeffekt wird in jedem Fall bleiben. Selbst bei kantenbasierten Verfahren, die unabhängig vom Schwellwert die Chromosomen in jedem Kanal segmentieren, wird der Ringeffekt nicht vermieden, da die Kanäle unabhängig betrachtet werden. Dies ist bei mei-

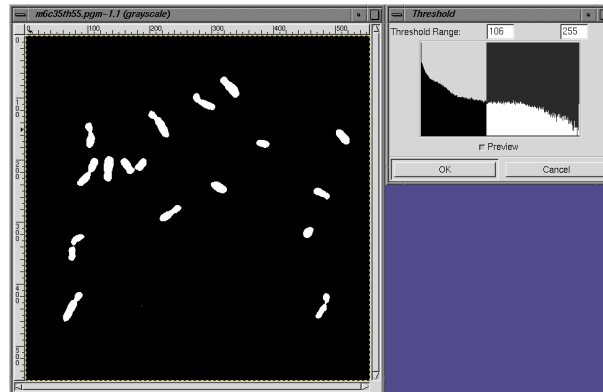


Abb. 6.5: Der Schwellwert für den ersten Farbbkanal ist etwas zu hoch. Die Chromosomen 1 fehlen.



Abb. 6.6: Der fehlerhafte Schwellwert im ersten Farbbkanal (Abb. 6.5) führt zu einer Fehlklassifikation der Chromosomen 1 (hier Grün). Sie werden als Chromosomen 16 identifiziert.

nem Ansatz nicht der Fall. Abb. 6.7 zeigt das Ergebnis mit der hier entwickelten Methodik.

6.3.2 Klassifikation mit Euklidischem Abstand

Dieser Abschnitt wird belegen, dass die Zuordnung der Pixel und Regionen zu den Klassen über den Winkel die geeignete Klassifikationsmethode ist. Um dies zu demonstrieren, führe ich die Klassifikation auf Pixeln durch, um den Einfluss der Zerlegung in Regionen und die Art und Weise der Berechnung der Regionenfarbvektoren auszuschliessen.

Zunächst liefere ich zwei repräsentative Beispiele direkter Klassifikation, über den Winkel und über den euklidischen Abstand (Abb. 6.8 und 6.9). Im anschließenden Fall clustere ich auf den Pixeln, und ordne die Cluster anschliessend über den Winkel und über den Euklidischen Abstand den Klassen zu, um allgemein bessere



Abb. 6.7: Zum Vergleich das Klassifikationsresultat mit der hier entwickelten Methodik.

Klassifikationsergebnisse zu erhalten (Abb. 6.10 und 6.11). Die besseren Ergebnisse bei der Zuordnung über Winkelabstände sind augenfällig. Als Referenz dient Abb. 6.12, das mit dem in dieser Arbeit entwickelten Verfahren erzeugt wurde (vgl. auch Abb. 4.8, 4.9).

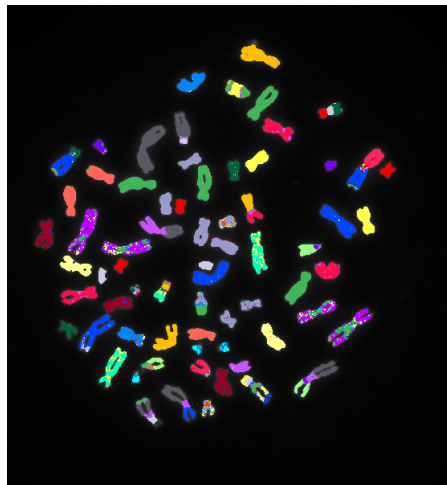


Abb. 6.8: Klassifikation der Pixel über den Winkel zu den Klassen des Hybridisierungsschemas. Vergleiche auch Abb. 6.12.

6.3.3 Regionenwachstum mit Euklidischem Abstand als Ähnlichkeitsmass

Dieser Vergleich wird zeigen, dass der Winkelansatz homogenere Regionen bei vergleichbaren Kriterien liefert. Zugleich dokumentiert dieser Abschnitt die Abhängigkeit der Zerlegung einer zentroiden Verkettung beim Region Growing vom Ähnlichkeitskriterium (Abs. 3.3). Erfolgt die Klassifikation der Regionen über den Winkel, sind die Ergebnisse trotzdem vergleichbar gut. Die Beispiele sind repräsentativ für

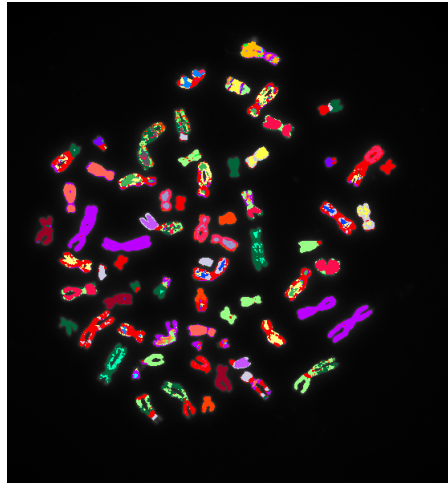


Abb. 6.9: Klassifikation der Pixel über den Euklidischen Abstand zu den Klassen des Hybridisierungsschemas. Die Zuordnung ist wesentlich schlechter als in Abb. 6.8. Vergleiche auch Abb. 6.12.

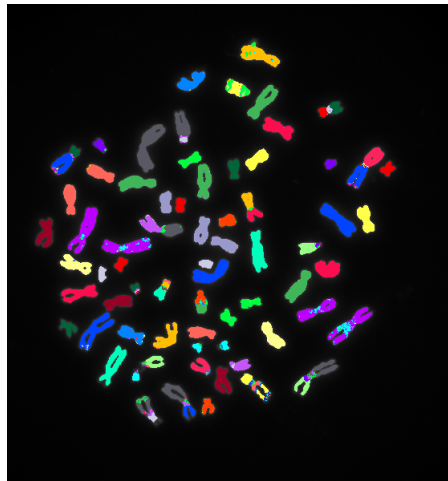


Abb. 6.10: Klassifikation der geclusterten Pixel über den Winkel zu den Klassen des Hybridisierungsschemas. Vergleiche auch Abb. 6.12.

die im Rahmen dieser Arbeit untersuchten Daten (wie auch die anderen in dieser Arbeit gezeigten).

Dafür setze ich zunächst beim Regionenwachstum als Ähnlichkeit nicht den Winkel zwischen benachbarten Pixeln, sondern den Euklidischen Abstand ihrer Farbvektoren. Die Berechnung der Regionenfarbvektoren erfolgt durch Mittelwertberechnung, also Minimierung des Euklidischen Abstands der Farbvektoren zum Regionenfarbvektor innerhalb einer Region. Als Ähnlichkeit habe ich hier 80 Grauwerteinheiten zwischen den Farbvektoren gewählt. Als Vergleich dient das Regionenwachstum mit einem Winkelunterschied von 15° . In beiden Fällen clustere und klassifiziere ich über den Winkel. Abb. 6.13 und 6.14 zeigen die vergleichbar guten Klassifikationsergebnisse. Dargestellt sind nur die Regionengrenzen in Klassifikationsfarben. Die Regionen für das Regionenwachstum über den Winkel sind in der Regel homogener.

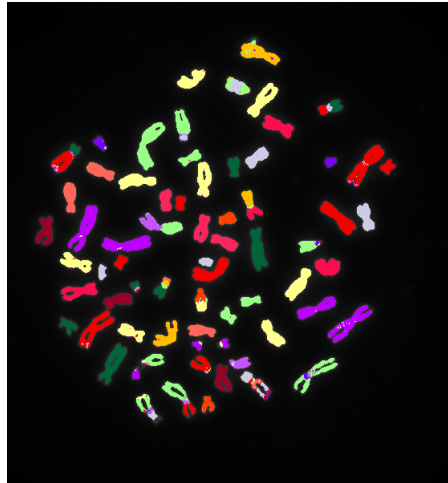


Abb. 6.11: Klassifikation der geclusterten Pixel über den Euklidischen Abstand zu den Klassen des Hybridisierungsschemas. Die Zuordnung ist wesentlich schlechter als in Abb. 6.10. Vergleiche auch Abb. 6.12.

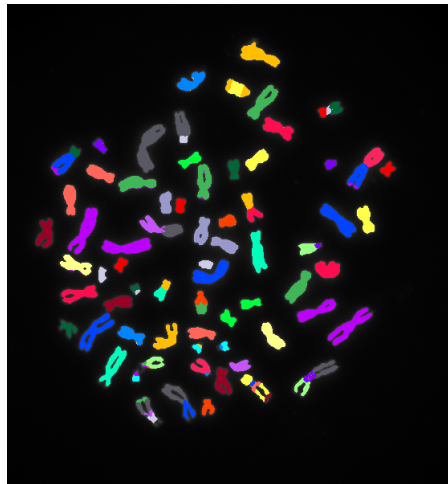


Abb. 6.12: Zum Vergleich das Klassifikationsresultat mit der hier entwickelten Methodik.

6.3.4 Clustersuche über Euklidischen Abstand

Hier werde ich belegen, dass beim Clustern, wie bei der Klassifikation, der entscheidende Faktor der Winkel als Distanzmass ist. Eine Einschränkung allerdings gibt es hier zu erwähnen. Durch die Zerlegung in Regionen wird das Bild komprimiert. Von typischerweise 30000 zu untersuchenden Pixel bleiben nach der Zerlegung je nach Kriterium nur noch einige wenige Hundert Regionen, die es zu clustern und klassifizieren gilt. Wählt man einen sehr grossen Ähnlichkeitswinkel, kann es dazu führen, dass die Kompression so stark ist, dass nur noch sehr wenige Regionen im Farbraum verbleiben, damit der Clustereffekt möglicherweise verschwindet. Damit reduziert sich aber auch die erwähnte Abhängigkeit der Clusterung vom Abstandsmass, und verschwindet sogar völlig. In der Praxis wird dies aber nicht vorkommen, da man vor allem an der Untersuchung von aberranten Zellen mit chromosomalen Umbauten interessiert ist. In solchen Fällen aber darf der Winkel beim Regionenwachstum nicht

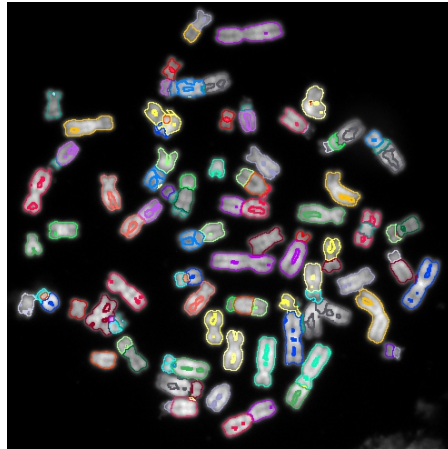


Abb. 6.13: Regionenwachstum mit (Euklidischem) Abstandskriterium von 80 (s. Text). Dargestellt sind nur die Regionengrenzen in Klassifikationsfarben. Vergleiche auch Abb. 6.14.

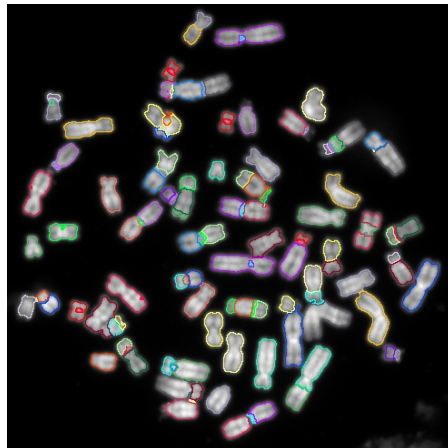


Abb. 6.14: Regionenwachstum bei einem Winkelkriterium von 15° . Dargestellt sind nur die Regionengrenzen in Klassifikationsfarben. Vergleiche auch Abb. 6.13.

zu gross gewählt werden, da man mit grösseren Winkeln an Farbempfindlichkeit einbüsst, was bei kleinen Insertionen durchaus Probleme schaffen kann.

Für diese Beispiele wählte ich daher ein Winkelkriterium von 5° für das Regionenwachstum, um eine nicht zu starke Kompression des Bildinhaltes zu gewährleisten, und den Effekt beobachten zu können. Die Regionen werden wie üblich nach dem Farbmodell in Abs. 4.4.1 berechnet. Nun wird im einen Fall standardmässig über den Winkel, im anderen über den Euklidischen Abstand im Farbraum geclustert, und in beiden Fällen über den Winkel klassifiziert (Abb. 6.15 und 6.16). Die Clustersuche über den Winkel ist wesentlich besser als über den Euklidischen Abstand. Es handelt sich hierbei auch nicht um einen methodischen Artefakt. Eine Klassifikation über Euklidischen Abstand, oder alternative Berechnungen der Regionen-Farbvektoren oder Regionenwachstum mit Euklidischem Abstand als Ähnlichkeitsmass, bringen keine Verbesserung des Resultats als in Abb. 6.16, sondern liefern eher schlechtere Klassifikationsergebnisse.

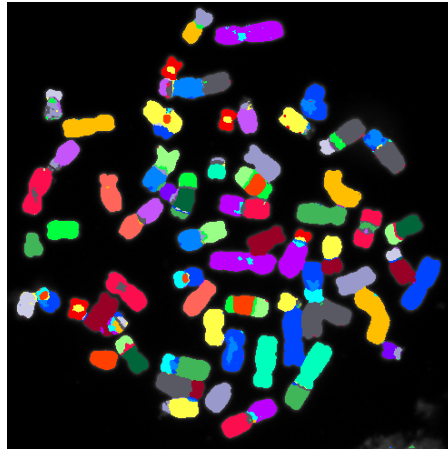


Abb. 6.15: Clustersuche über den Winkel. Das Bild wurde über eine Winkelorientiertes Regionewachstum zerlegt. Die Cluster werden über den Winkel klassifiziert.

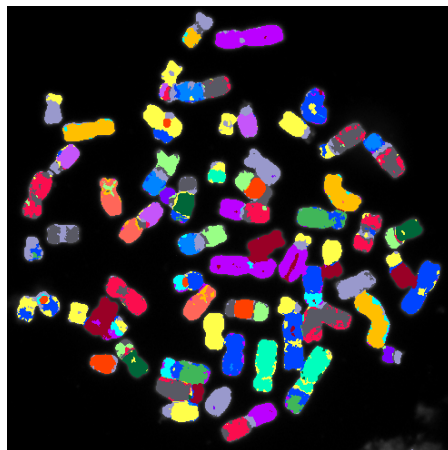


Abb. 6.16: Clustersuche über den Euklidischen Abstand. Die Clustersuche wurde auf den gleichen Regionen wie in Abb. 6.15 durchgeführt. Die gefundenen Cluster werden über den Winkel klassifiziert.

6.4 Chromosomale Barcodes

Das Verfahren zeigt für Barcode Experimente sehr gute Resultate. Dies wurde in mehreren Versuchsreihen bestätigt. Durch den regionenorientierten Ansatz lassen sich die Bruchkanten (Übergangsbereiche) zwischen zwei benachbarten Regionen gut klassifizieren. Die Wahl der Proben ist hier sehr wichtig. Die Bindungsbereiche benachbarter Proben sollten so wenig wie möglich überlappen. Das gutartige Verhalten in den Übergangsbereichen ist in Abb. 6.17 zu erkennen. In dem gezeigten Bild kann man ein normales Chromosom 12 erkennen. Die anderen beiden weisen eine Deletion auf.

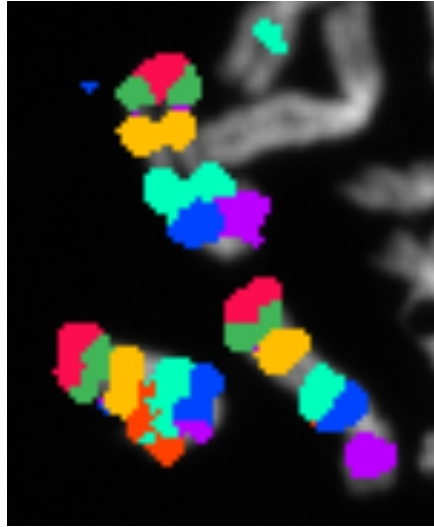


Abb. 6.17: Die mit der Barcode-Technik gefärbten Metaphase-Chromosomen eines kranken Patienten. Gefärbt sind nur vier Chromosomen 12. Gezeigt ist hier ein Ausschnitt aus 5.6.

6.5 Telomere

Wie schon mehrfach angeführt, macht die geringe Grösse der Telomere (und anderer kleiner chromosomaler Bereiche) sie für Rauschen und Störungen jeglicher Art sehr anfällig und die Analyse damit sehr schwierig. Unterschiedlich grosse Signale in den einzelnen Farbkanälen führen bei mehrfachen Hybridisierungen zu sehr starken Schwankungen in der Farbrichtung (Abb. 6.18). Das Winkelkriterium muss bei diesen Anwendungen daher in der Regel sehr viel grösser ausfallen, um homogene Klassifikationen der Signale zu erhalten (Abb. 6.19).

Sind alle Signale gut getrennt, kann auch hier 90° beim Regionenwachstum gewählt werden, um möglichst grosse und homogene Signale im Klassifikationsbild zu erhalten.

Wird das Kriterium zu klein gewählt, ist das korrekt klassifizierte Signal oft von grossen, falsch klassifizierten Bereichen umgeben, was die Analyse erschwert. Selbst wenn durch geeignete Suchalgorithmen das richtige identifiziert, und die falschen entfernt werden, führt dies oft zu sehr kleinen Regionen im Bild, die kaum sichtbar sind. Auch ist es nicht einfach die richtige Region innerhalb eines solchen Bereichs finden, da selbst die höchste lokale Anzahl der Labels oder die Intensität, die diesen Bereich auszeichnen sollten, oft als geeignetes Suchkriterium scheitern.

Festzuhalten bleibt, dass eine korrekte Klassifikation auf der Farbe allein ohne die Berücksichtigung von Ortsinformation nicht möglich ist, wenn eine bestimmte Mindestgrösse des Signals im Klassifikationsbild erwünscht ist.

Durch die - durch die geringe Grösse der Signale bedingte - mangelnde Clusterbildung wird die Clustersuche überflüssig (Abb. 6.20). Die Regionen ordne ich daher direkt den durch das Hybridisierungsschema definierten Klassen zu, ohne die Clustersuche durchzuführen, da keine sinnvollen Ergebnisse zu erwarten sind, wenn

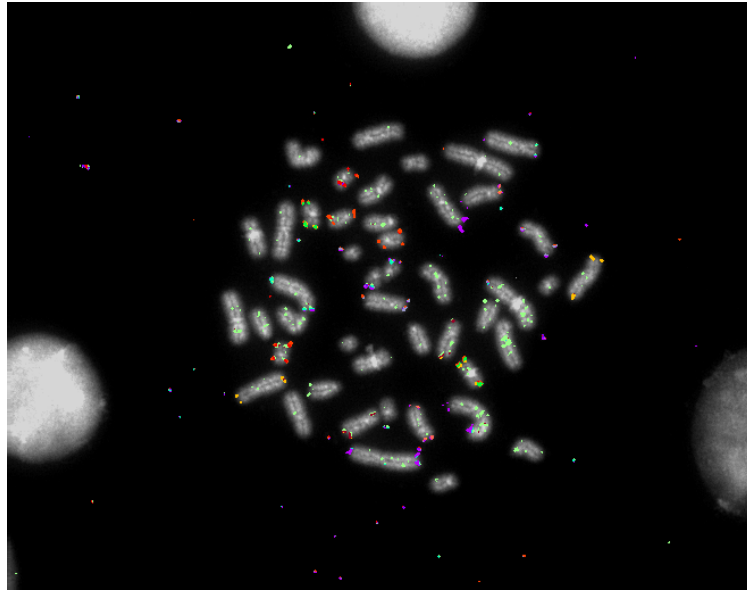


Abb. 6.18: Kleine Winkelkriterien (hier 15°) liefern bei kleinen Signalen schlechte Resultate. Unterschiedlich grosse Signale in den Einzelkanälen führen bei mehrfachen Hybridisierungen zu sehr starken Farbgradienten innerhalb kleiner Bereiche.

man den Farbraum betrachtet.

Abb. 5.10 zeigt das Hybridisierungsschema eines M-FISH Experiments mit spezifischen subtelomerischen Proben. Es fällt auf, dass hier nur die Hälfte des Chromosomensatzes angefärbt wurde. Das Problem rührt daher, dass Proben im allgemeinen nicht mit hundertprozentiger Wahrscheinlichkeit an ihre Komplementärsequenzen am Genom binden. Sie beträgt etwa 95%. Je kleiner die Proben sind, um so stärker macht sich dieser Effekt fehlerhafter Hybridisierung bemerkbar. Durch die Aufteilung in zwei Hybridisierungspools mit jeweils 12 und 11 Proben wird die Wahrscheinlichkeit einer erfolgreichen Hybridisierung signifikant erhöht. Damit reduziert sich die Anzahl der Metaphasen, die aufgenommen werden müssen, um eine mit sämtlichen Signalen zu erhalten, drastisch. Folgende Tabelle (6.2) illustriert diesen Sachverhalt für verschieden Fälle. Für den speziell entwickelten subtelomerischen Probensatz **M-TEL** in Kapitel 5 benötigt man im Mittel 2 Metaphasen, um für alle Telomere Signale zu erhalten [Sar01].

6.6 Anwendungen in 3D

Zum jetzigen Stand der Entwicklung ist nicht geklärt, ob und wie die Methodik erweitert werden muss, um stabilere Klassifikationsergebnisse für Untersuchungen an Interphase-Zellkernen zu erzielen. Im Interphase-Kern kommen einige Phänomene

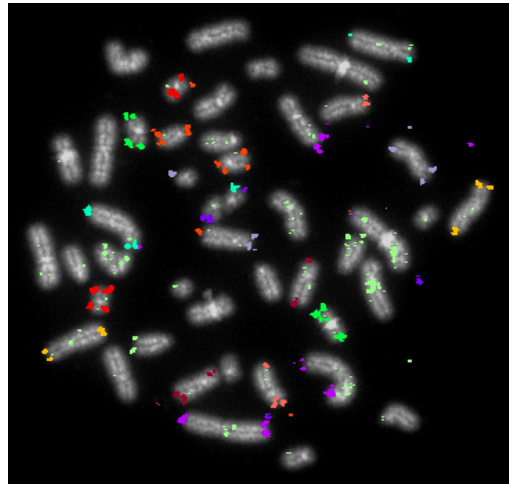


Abb. 6.19: Grosse Winkelkriterien liefern bei kleinen Signalen oft bessere Resultate

subtelomer. Proben	Hybr. Effizienz	Farben	Labels	Wahrscheinlichkeit
41	95%	6	96	0.7%
24	95%	5	49	8.1%
12(M-TEL12)	95%	4	22	32.4%
11(M-TEL11)	95%	4	19	37.7%
41(M-TEL 12+11)	95%	4	41	12%

Tab. 6.2: Abhängigkeit einer erfolgreichen Hybridisierung von der Anzahl der Proben. Die Wahrscheinlichkeit gilt für den Fall, das alle Hybridisierungen korrekt sind.

hinzu, die die Analyse erschweren. Die Dekondensation der Chromosomen führt zu grösseren Farbschwankungen und damit zu ausgedehnten Clustern im Farbraum. Damit werden Überlappungen wahrscheinlicher. Gleichzeitig führen Überlappung und Durchdringung der Chromosomen zu ausgedehnten Bereichen im Bild, in den sich die Signale überlagern und die Analyse erschweren (Abb. 6.22).

Das Winkelkriterium muss ausbalanciert werden, da die Objekte geographisch zusammenhängen und das Verfahren nur Farbunterschiede, aber keine Intensitätskanten detektieren kann (abgesehen von den Intensitätsschwelle, die auf dem Maximum-Projektionsbild angewendet wird). Zu kleine Winkel führen leicht zu sehr vielen fragmentierten Regionen mit schlechter Farbcharakteristik, während zu grosse Winkel Objektverschmelzungen zur Folge haben. Der Toleranzbereich des Winkelkriteriums hängt hier sehr stark vom Hybridisierungsschema ab. Hier sollte aufgrund der Grösse und Variabilität der Farbcluster unbedingt darauf geachtet werden, einen möglichst grossen Abstand zu erzielen, also möglichst viele Farbstoffe bei gleichzeitig

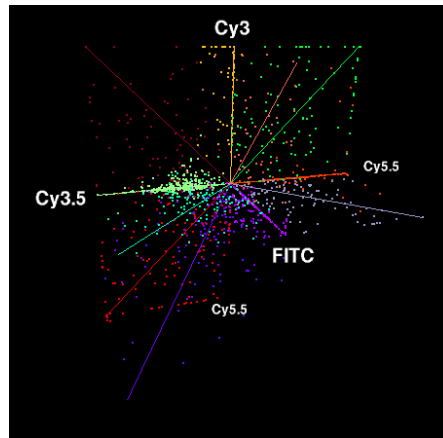


Abb. 6.20: Typischer Farbraum eines klassifizierten 4-Farben M-FISH Experiments mit 12 Telomer-Proben. Die Kodierung der Falschfarben entspricht der in Abb. 5.10.

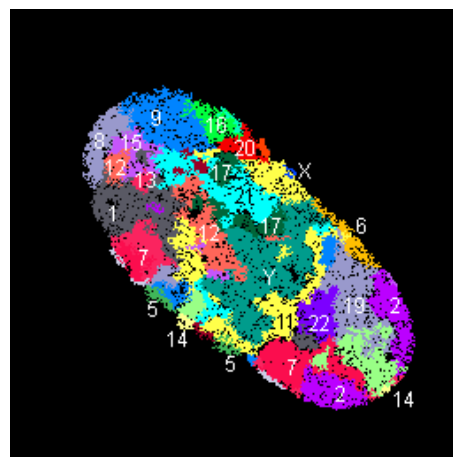


Abb. 6.21: Klassifikationsbild eines 7 Farben M-FISH Experiments aller 24 Chromosomen in der Interphase. Die Falschfarben der Chromosomen sind entsprechend Abb. 1.6

möglichst wenigen Simultan-Hybridisierungen pro Target (vgl. Abs. 6.1). Das Verfahren ist für 3D Anwendungen instabiler als im zweidimensionalen Fall. Es reagiert auf Veränderungen des Winkelkriteriums kritischer als bei der Analyse von Metaphasen. Der Grund ist, dass die durch Überlagerungen erzeugte Falschinformation sehr grosse Bereiche einnehmen kann, was bei kleineren Chromosomen zu Problemen in der korrekten Clusterfindung im Farbraum führen kann. Das Hineinwachsen in benachbarte Chromosomen ist ebenfalls kritisch zu beobachten, da die Farbinformation diffuser ist, Farbkanten daher weniger ausgeprägt sein werden. Es gibt einen weiteren Faktor, der für das instabile Verhalten verantwortlich ist. Die überwachte Clustersuche und anschliessende Klassifikation ist darauf angewiesen, dass alle Klassen, also Chromosomen im Bildvolumen vorhanden sind. Da bei Metaphasen die Segmentierung auf der DAPI Maske bzw. durch die intensiven Farbsignale auf der Maximumprojektion unkritisch ist, sind alle Klassen bzw. Chro-

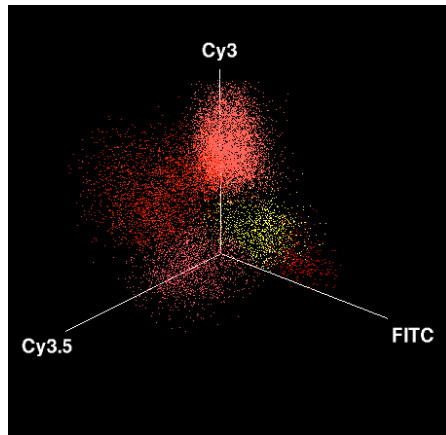


Abb. 6.22: Der Farbraum eines 3D M-FISH Experiments. Zu erkennen sind diffusere Cluster, die sich überlappen und eine eindeutige Identifizierung erschweren.

mosomen im Bildvolumen vorhanden. Bei Interphasebildern muss die Maximumprojektion für jede Ebene erzeugt werden. Die schwächere und diffusere Struktur der Chromosomen macht es jedoch sehr viel schwieriger, eine geeignete Schwelle zu finden, die alle Chromosomen, insbesondere auch die Kleinen, berücksichtigt. Fehlt auch nur eine einzige Klasse, hat das negative Auswirkungen für alle anderen. Mit unüberwachten Ansätzen, die blind den Farbraum in Cluster unterteilen, wird man nur schwer zu besseren Resultaten kommen, da Überlappungen vorkommen.

Abb. 6.23 zeigt den Farbraum nach der Zerlegung in Regionen. Dargestellt sind die Regionenfarbvektoren. Die Abbildung verdeutlicht, dass Überlappungen der Cluster durch Hinzunahme der Ortsinformation bei der Zerlegung in Regionen ähnlicher Farbe zum Teil wieder aufgelöst werden können. Dies gilt ebenso für Anwendungen in zwei Dimensionen.

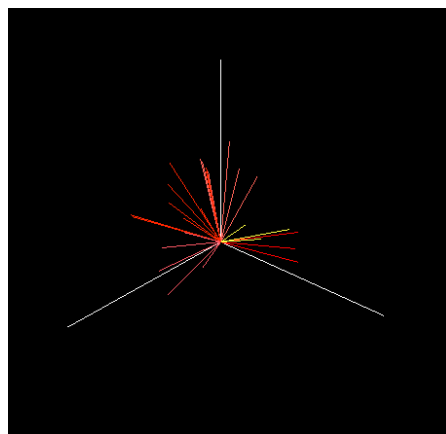


Abb. 6.23: Der Farbraum der Regionenfarbvektoren. Überlappungen von Clustern können durch Hinzunahme der Ortsinformation bei der Zerlegung in Regionen ähnlicher Farbe zum Teil wieder aufgelöst werden.

6.7 Zusammenfassende Diskussion

Nachdem ich in Kapitel 4 die von mir entwickelte Methodik ausführlich geschildert, in Kapitel 5 und 6 Anwendungsbeispiele geliefert und die Methodik diskutiert habe, möchte ich zum Abschluss dieses Kapitels eine übergreifende, zusammenfassende Diskussion anbringen:

Bei Klassifikationsproblemen jeglicher Art haben sich *Neuronale Netze* als leistungsfähigste Methode erwiesen. Das Problem bei allen neuronalen Klassifikationen ist jedoch, dass diese Netze *trainiert* werden müssen. Dazu benötigt man aber zahlreiche unterschiedliche Datensätze, bei denen das korrekte Resultat schon bekannt ist. Solche Daten liegen aber noch nicht vor, müssen also von Methoden wie dieser hier geliefert werden. Dann wäre das Problem aber nur für einen bestimmten Experimenttyp mit einem definierten Hybridisierungsschema und -protokoll gelöst. Für jedes andere Experiment bestünde das Problem erneut. Daher ist man vorerst auf flexible zuverlässige Methoden angewiesen, die sich an Experiment, Schema und Informationsstruktur anpassen können.

Binäre Klassifikationsmethoden sind im Prinzip denkbar und auch realisiert [Spe96b]. Dazu müssen die Objekte in jedem Farbkanal getrennt segmentiert werden. Eine halbwegs automatische Segmentierung über Schwellwerte ist jedoch nicht immer einfach und wird bei kleinen Segmentierungsfehlern im Einzelkanal sehr grosse Fehler im Klassifikationsbild nach sich ziehen. Vorstellbar sind auch auf *fuzzy-logic* basierende Klassifikationsmethoden, die über Wahrscheinlichkeitsverteilungen den Pixeln eine Wahrscheinlichkeit der Zugehörigkeit zuordnen. Letzlich ist der Biologe oder Anwender jedoch auf eine Entscheidung angewiesen. Zudem ist nicht klar, wie robust die Wahl der Parameter solcher Systeme auf das Ergebnis auswirken. Die Frage, wie zuverlässig und brauchbar solche Systeme arbeiten, kann daher an dieser Stelle nicht beantwortet werden, und bietet aufgrund ihres Umfangs Raum für weitere Untersuchungen in der Zukunft.

Die hier entwickelte Methodik stellt eine Plattform für die Analyse von M-FISH Bildern dar, die sich flexibel den experimentellen Gegebenheiten durch die Clustersuche im Farbraum anpasst. Sie beschränkt sich auf die Analyse der Ort-Farbinformation und liefert bei einmaliger Anpassung bestimmter Parameter robuste Klassifikationsergebnisse. Der überwachte Ansatz der Clustersuche bietet gegenüber unüberwachten Methoden den klaren Vorteil der Geschwindigkeit. Auch die Ergebnisse werden in der Regel bei der überwachten Variante besser sein, speziell, wenn es zu Überlappungen grosser Cluster kommt, wird das Verhalten unüberwachter Algorithmen nur schwer vorhersagbar sein. Der Nachteil der überwachten Methode ist jedoch, dass die Klassen a priori bekannt sein müssen. Fehlt auch nur eine einzige Klasse, hat das negative Auswirkungen für alle anderen, da mehr Cluster gesucht werden, als möglicherweise vorhanden sind. Und schliesslich werden diese einer Überzahl an Klassen zugeordnet, was zu schwerwiegenden Fehlern führen wird. Trotzdem überwiegen die

Vorteile des überwachten Ansatzes bei weitem.

Die Farbinformation bei vielfarbmarkierten Chromosomen in der Interphase ist durch die weniger kompakte Struktur der Chromosomen diffuser. Speziell hier muss auf das Design des Hybridisierungsschemas grosser Wert gelegt werden, um maximalen Farbkontrast zu erreichen, um einen Überlapp der Cluster im Farbraum möglichst gering zu halten. Ein weiteres Problem sind die zum Teil sehr grossen räumlichen Überlappungen oder Durchdringungen der Chromosomen im Nukleus, die die tatsächliche Farbinformation in diesen Bereichen zerstören und eine korrekte Identifizierung anhand der Farbe unmöglich machen. Die Bereiche haben jedoch einen erheblichen Einfluss auf die Verteilung im Farbraum, und beeinflussen damit das Ergebnis der Clustersuche in negativer Weise. Durch die Berücksichtigung der Ortsinformation durch den regionenorientierten Ansatz können die durch diese Bereiche entstehenden Artefakte zum Teil aufgefangen werden. Hinzu kommen noch nicht untersuchte und quantifizierte Artefakte, die durch das dreidimensionale Aufnahmeverfahren mit der begrenzten axialen Auflösung entstehen. Fragen zu geeigneten Dekonvolutionsalgorithmen bei Mehrfarbbildern sind noch offen. Die vorgestellten Anwendungen für 3D-Experimente zeigen die Probleme mit diesen Daten auf, die Resultate sind vielversprechend, aber noch nicht hundertprozentig zufriedenstellend. Da diese Experimente erst gegen Ende meiner Arbeit durchgeführt wurden, handelt es sich um den gleichen Ansatz wie im zweidimensionalen Fall, den ich auf drei Dimensionen erweitert habe. Weiteres Entwicklungspotenzial sehe ich hier vor allem durch die Verbindung der hier entwickelten Methoden mit bewährten 3D-Segmentierungsverfahren. Ein möglicher Ansatz wäre zum Beispiel, eine 3D-Voronoi Tesselierung zu entwickeln, die auf der (regionalen) Farbrichtung den Split durchführt. In [Eil98] wurde ein ähnlicher Ansatz schon für die zweidimensionale Anwendung vorgenommen.

Die Datenraumstruktur, also die Farbinformation von Multicolor-FISH Bilder bei unterschiedlichen M-FISH Experimenten wurde dargestellt und damit die auftretenden Probleme zur korrekten Klassifikation illustriert. Es konnte gezeigt werden, dass die Richtung bei kombinatorischen Bildern ein wesentlich besseres Distanzmass zur Klassifikation darstellt als der euklidische Abstand. Dieser Befund ist im Einklang mit anderen Ergebnissen, die kantenbasierten Verfahren schlechte Resultate bei der Analyse von Farbbildern bescheinigen. Ein möglicher Grund ist, dass die Berechnung von Kanten häufig auf dem euklidischen Distanzmass beruhen. Der Zusammenhang zwischen Kondensationsgrad und Farbraumstruktur wurde deutlich und damit die Schwierigkeit der Analyse von dekondensierten Chromosomen in Interphase Zellkernen rein auf Basis der Farbinformation für eine korrekte Identifizierung aufgezeigt. Ähnlich ist das Problem bei sehr kleinen Proben, bei denen aufgrund ihrer Grösse und damit verbundenen schlechten statistischen Eigenschaften vor allem unterschiedliche Segmentierungsmasken in den Farbkäneln zu einem starken Verfälschen der Farbinformation führen können und eine Identifizierung nur

anhand der Farbe sehr schwer machen, wenn eine gewisse Grösse erwünscht ist. Die mangelnde Clusterbildung im Farbraum belegt dies.

Die Gefahr, dass durch graduelle Übergänge in der Farbe benachbarter unterschiedlicher Regionen fehlerhafte Verschmelzungen auftreten, ist beim Region Growing gross. Deshalb war es wichtig, einen Region Growing Algorithmus *zentroid verketteter* Art zu realisieren, um die Zerlegung in solchen Übergangsbereichen besser zu kontrollieren. Das von mir entwickelte Farbmodell mit seinem (intensitätsunabhängigen) Regionenansatz führt zu robusten Tesselierungen, speziell in kritischen Übergangsbereichen, sowie stabilem Verhalten bei der Clusteranalyse. Die Farbrichtung und der Farbwinkel als Ähnlichkeitsmass haben sich als ausgezeichnete Kriterien erwiesen, und belegen den schon mehrfach angeführten funktionalen Zusammenhang dieser Grössen (Abs. 3.3, 4.6). Intensitätskanten müssen durch zusätzliche Verfahren detektiert werden. Dies gilt im besonderen für 3D-MFISH Experimente von Chromosomen in der Interphase.

Die im Verlauf dieser Arbeit aufgezeigten Zusammenhänge im Farbraum und Probleme, die bei der Analyse unterschiedlicher M-FISH Experimente auftreten können, haben gezeigt, dass Analysemethoden, die rein auf der Farbinformation beruhen, nicht in jedem Fall zuverlässige Ergebnisse erwarten lassen. Die Ergebnisse und Anwendungen belegen, dass der von mir gewählte Ansatz der Kombination von (direktionaler) Farbinformation mit lokaler Information hingegen ein geeignetes Konzept für die Analyse von MFISH-Bildern unterschiedlicher Experimente ist, das robuste Klassifikationen liefert. Das habe ich durch die Modellierung des Bildes in Bereiche konstanter Farbe realisiert. Insbesondere kritische Übergangsbereiche zwischen Bereichen unterschiedlich gefärbter chromosomaler Abschnitte werden zuverlässig identifiziert.

Meine Ergebnisse weisen ferner die (intensitätsunabhängige) Farbrichtung und den Winkelunterschied von Farbvektoren als entscheidende Merkmale bei Analyse aus. Die aufgrund der unterschiedlichen Datentypen erforderliche Flexibilität der Methodik ohne gravierenden Performanceverlust ist durch den adaptiven Ansatz mit der überwachten Clusteranalyse bei meiner Methodik gewährleistet. Die Analyse einer Metaphase dauert im Durchschnitt nur wenige Sekunden.

Die im Rahmen dieser Dissertation entwickelte Methodik liefert durch seinen Ansatz kombinierter Orts- und Farbinformation und modularen Aufbau eine geeignete Basis für Weiterentwicklungen, speziell im Rahmen von Analysen M-FISH markierter Chromosomen in Interphase Kernen, wo die Nachbarschaft der Voxel im Ortraum an Bedeutung gewinnt. Sie eröffnet zudem Anwendungsmöglichkeiten in anderen biologischen Experimenten, wie es zum Beispiel in der Entwicklung einer Software für die Analyse von DNA-Microarrays realisiert wurde [Rud00].

Kapitel 7

Zusammenfassung

Multiplex-FISH ist eine vor wenigen Jahren entwickelte Methode, mit der sich alle menschlichen Chromosomen mit mehreren Farbstoffen kombinatorisch färben lassen [Spe96a, Spe96b, Sch96]. Bei Verwendung von mehr als vier Farben lassen sich bei 31 Kombinationsmöglichkeiten alle Chromosomen in einer eindeutig kombinatorischen Weise färben und werden mit geeigneten spektroskopischen Methoden anhand ihrer spektralen Information unterscheidbar.

Im Rahmen dieser Arbeit habe ich ein automatisches Verfahren für die Analyse von Bildern M-FISH markierter Chromosomen und subchromosomaler Proben entwickelt, das signifikante Verbesserungen und robuste Analysen, sowie erweiterte Möglichkeiten der Analyse mit Hilfe einer M-FISH Technologie bietet. Das Verfahren basiert auf Clusteranalyse im Farbraum und kombiniert Farbinformation mit Ortsinformation, um chromosomale Bereiche im Bildvolumen zu indentifizieren. Das Verfahren ist für die Analyse von Bildern unterschiedlicher M-FISH Experimente - sowohl chromosomal und subchromosomal, als auch inter- und intrachromosomal geeignet. Identifizierung einzelner Chromosomen, sowie Volumen und Positionsbestimmung bei dreidimensionalen Aufnahmen im Interphase-Zellkern sind im Prinzip ebenso möglich.

Der Zusammenhang zwischen Kondensationsgrad und Farbraumstruktur wurde aufgezeigt und damit die Schwierigkeit der Analyse von dekontensierten Chromosomen in Interphase-Zellkernen rein auf Basis der Farbinformation für eine korrekte Identifizierung nachgewiesen. Bei sehr kleinen Proben, bei denen aufgrund ihrer geringen Grösse und damit verbundenen schlechten statistischen Eigenschaften unterschiedliche Signalgrössen in den Farbkanälen zu starken Farbgradienten führen, ist eine Identifizierung nur anhand der Farbe kaum möglich, wenn eine gewisse Grösse erwünscht ist.

Das vorgestellte Farbmodell mit seinem Regionenansatz liefert zuverlässige Zerlegungen des Bildvolumens. Durch die Gewichtung von Pixeln höherer Intensität wird die richtungsabhängige Farbinformation der Regionen vor allem in Bereichen

niedriger Intensitäten zuverlässiger bestimmt. Die Hinzunahme der Ortsinformation ermöglicht robustes Klassifikationsverhalten. Überlappungen von Clustern im Farbraum können unter Umständen aufgelöst werden, da die durch den Klassifikator vorgegebenen rigiden Klassifikationsgrenzen diffuser werden. Das ist bei pixelbasierter Klassifikation nicht möglich. Die Clustersuche an sich hat die Bedeutung einer Anpassung der Klassen, die im Hybridisierungsschema definiert sind, an die Bildinformation. Damit werden die Ergebnisse signifikant verbessert.

Die intensitätsunabhängige Farbrichtung und damit der Winkel zwischen den Farbvektoren wurden als entscheidende Merkmale der Pixel identifiziert. Sie haben in allen Schritten der in dieser Arbeit entwickelten Methodik die besten Ergebnisse geliefert.

Die Bedeutung des Labelingschemas, sowie der Zusammenhang zwischen Anzahl simultaner Hybridisierungen und der Anzahl der Farben für die experimentelle Auflösung im Sinne der Klassifikation wurde aufgezeigt. Dem Experimentator werden damit wichtige Hinweise aufgezeigt, sein Experiment schon im Vorfeld auf die Klassifikation hin zu optimieren.

Es wurde aufgezeigt, dass bei kombinatorischen Bildern eine Hintergrundkorrektur auf dem Maximum-Projektionsbild der Farbkanäle erfolgreich erfolgen kann, wenn eine DAPI Färbung nicht zur Verfügung steht, oder als Segmentierungsmaske nicht verwendet werden kann, wie es für Telomere oder chromosomale Barcodes der Fall ist. Auch für dreidimensionale Anwendungen liefert sie eine gute Basis als Segmentierungsmaske. Im Gegenzug sollte bei nur einfach gelabelten Bildern jeder Kanal für sich getrennt verarbeitet werden.

Die Variabilität der verschiedenen M-FISH Experimente erfordert eine flexible und vielseitige Analysemethode um der unterschiedlichen Qualität und Charakteristik der anfallenden Daten gerecht zu werden. Die hier entwickelte Methodik stellt eine Plattform für die Analyse von M-FISH Bildern dar, die sich flexibel den experimentellen Gegebenheiten anpassen lässt. Sie beschränkt sich auf die Analyse der Orts-Farbinformation und liefert bei einmaliger Anpassung bestimmter Parameter robuste Klassifikationsergebnisse. Der überwachte Ansatz der Clustersuche bietet gegenüber unüberwachten Methoden den klaren Vorteil der Geschwindigkeit. Auch die Ergebnisse werden in der Regel bei der überwachten Variante besser sein, speziell, wenn es zu Überlappungen grosser Cluster kommt, wird das Verhalten unüberwachter Algorithmen nur schwer vorhersagbar sein.

Summary

Multiplex-FISH is a combinatorial staining technique that allows the simultaneous detection and discrimination of all human chromosomes. Using at least five fluorochromes all chromosomes can be uniquely labeled in a combinatorial way and identified by their specific spectral signature.

With this work I have developed a novel approach for the automated analysis of M-FISH images, yielding robust classification results and allowing the analysis of M-FISH images of different experiments. The method combines spectral information with spatial information to tessellate the image into regions of similar color. Subsequently a cluster analysis in color space and a final classification step are performed to identify the biological targets. This approach is applicable to images of different M-FISH experiments, allowing the analysis of interchromosomal as well as intrachromosomal abnormalities in the genome. It also allows the 3D analysis of M-FISH labeled chromosomes in interphase nuclei, however, the accuracy of the results is still left for further investigation.

The severe influence of chromosomal condensation and size on the topology of clusters in color space has been shown. Classification of decondensated chromosomes becomes therefore a difficult task due to less expressed spectral information. This problem is closely related to the analysis of small probes. Including only a few pixels. Such targets show weak statistical properties. Therefore lateral focal shifts and different spot sizes in the color channels lead to significant color gradients within small scales, thus very likely clusters in color space cannot be observed in such applications, and classification on color information alone fails, if an appreciable size in the classification image is desired.

The color model with the region tessellation approach presented here, yields reliable spatial tessellations of the image space. Higher weight of pixels with higher intensities allows a more reliable determination of spectral information especially within areas of lower intensities. The combination of spatial with color information leads to robust behaviour in the clustering step. Overlaps of clusters may possibly be resolved in this way, as the rigid borders defined by the classifier get more diffusive by this approach, allowing correct classification of pixels that not classifiable when classification is performed pixelwise. The cluster step can be regarded as a way of

matching the classes, which are defined by the labeling scheme, to image content. This is an important step that significantly improves classification results.

Direction of color vectors and angles between them so far have been identified as the main features of pixels in M-FISH experiments. They have proved to yield best results in each of the steps of the method developed here.

The importance of deliberate labeling design has been pointed out. It allows an optimization of the hybridization scheme in such a way, that color distribution in the overall image achieves maximum discrimination in the terms of classification. In cases where a DAPI image is not present or cannot be used for intensity based pre-segmentation, the maximum projection image of all color images can be used. This is the case for barcode experiments as well as experiments, where only telomeric regions of chromosomes have been painted. It is also true for 3d images of painted chromosomes in the interphase nucleus.

The variety of M-FISH experiments require a flexible and versatile image analysis approach to account for the versatility of the imaged data. The method developed within this work is a flexible methodological platform for the analysis of M-FISH images, that is adjustable to different types of M-FISH experiments. It combines spatial with color information yielding robust classification results after experimental specific parameters have been determined. Using a supervised clustering algorithm seems to be the right choice, as the number of clusters is a priori known. With color direction and angle between color vectors readily being identified as crucial features, the use of unsupervised clustering algorithms do not promise to deliver better results at the cost of much longer computation time. What is more, behaviour may be unpredictable in cases where overlaps occur.

Kapitel 8

Bedienung der Software

Die entwickelte Methodik ist in einem Programm **mfish3d** integriert und steht für zwei Plattformen, Linux und IRIX zur Verfügung. Das Programm ist Kommandozeilen orientiert und benötigt zwei zusätzliche Dateien zur Laufzeit, mit denen die Software gesteuert werden kann. Diese beiden ASCII Dateien heissen **menu** und **classes** und sollten sich im Verzeichnis befinden, aus dem das Programm aufgerufen wird, da für jedes Experiment und jeden Datentyp diese Parameter individuell eingestellt werden müssen. Die *menu*-Datei ist für die Steuerung des Programms zuständig, während in der *classes*-Datei das Hybridisierungsschema enthält.

8.1 Die Datei classes

Eine typische *classes*-Datei für ein 5-Farben Experiment mit 24 Chromosomen zeigt Tab. 8.1.

Jede Zeile beschreibt ein hybridisiertes biologisches Objekt. Die erste Zahl ist ein Identifier, oder wie hier das gefärbte Chromosom, gefolgt von den durch Komma getrennten binären Labeln. In diesem Beispiel sind die Chromosomen 1 in den Kanälen 1, 3 und 4 angefärbt worden. Die Reihenfolge muss dabei der Reihenfolge der Argumente beim Programmaufruf entsprechen. Würden im Experiment die Farbstoffe FITC, Cy3, Cy3.5, Cy5 und Cy7 verwendet, wären die Chromosomen 1 also in FITC, Cy3.5 und Cy5 markiert, müsste der Aufruf lauten:

```
mfish3d FITC-Bild.tif Cy3-Bild.tif Cy3.5-Bild.tif Cy5-Bild.tif Cy7-Bild.tif output-  
name DAPI-Bild.tif
```

Das DAPI Bild wird nur in der Kommandozeile angegeben, falls es vorhanden ist. Der Output der Software ist das Klassifikationsbild im TIFF Format. Die Falschfarben im Klassifikationsbild sind fest kodiert und entsprechen denen in Abb. 6.1.

1 1, 0, 1, 1, 0
2 0, 0, 0, 0, 1
3 0, 1, 1, 0, 1
4 1, 0, 1, 0, 0
5 0, 1, 0, 1, 1
6 1, 0, 0, 1, 1
7 0, 0, 1, 1, 0
8 1, 1, 0, 1, 0
9 1, 1, 0, 0, 1
10 0, 0, 1, 0, 1
11 1, 1, 1, 0, 0
12 0, 0, 0, 1, 1
13 1, 1, 0, 0, 0
14 0, 0, 0, 1, 0
15 0, 1, 1, 1, 0
16 1, 0, 0, 1, 0
17 0, 0, 1, 0, 0
18 1, 0, 0, 0, 0
19 0, 1, 1, 0, 0
20 0, 1, 0, 0, 0
21 1, 0, 0, 0, 1
22 0, 1, 0, 1, 0
23 0, 1, 0, 0, 1
24 1, 0, 1, 0, 1

Tab. 8.1: Die Datei **classes**.

Der Aufruf erfolgt in der genannten Art allerdings nur für 2D Aufnahmen, bei denen für jeden Kanal nur 1 Bild aufgenommen wurde. Bei dreidimensionalen Aufnahmen von M-FISH markierten Zellkernen in der Interphase ändert sich an der Reihenfolge nichts. Man sollte jedoch die Bildstapel der Übersicht halber in eigene Unterverzeichnisse ablegen. Der Aufruf könnte dann wie folgt aussehen:

*mfish3d FITC/namefitc Cy3/namecy3 Cy3.5/namecy35 Cy5/namecy5 Cy7/namecy7
outputname*

Wichtig ist, dass die Endung *.tif* nicht mit angegeben wird, da dies ansonsten als 2D Experiment interpretiert wird.

8.2 Die Datei menu

Eine typische *classes*-Datei für ein 5-Farben Experiment mit 24 Chromosomen zeigt Tab. 8.2.

<i>COLORS</i>	5
<i>MERGE?</i>	0
<i>SINGLELABELS?</i>	0
<i>MEDIANMAXIMAGE?</i>	1
<i>ZSCALE</i>	2.5
<i>ANGLE</i>	10
<i>TAU</i>	0.0001
<i>MAXSTEP</i>	0
<i>LAMBDA</i>	1
<i>BACKGROUND</i>	1
<i>BACKGROUNDDAPI</i>	1
<i>MAXITERATION</i>	30
<i>BORDERS?</i>	0
<i>SMALLSIGNALS?</i>	0
<i>REGIONCORRECTION?</i>	0
<i>PERFORMCLUSTERING?</i>	1
<i>IVCHANNEL1</i>	0
<i>IVCHANNEL2</i>	2
<i>IVCHANNEL3</i>	3
<i>POINTSIZ</i>	1
<i>INVENTORCOLSPACE?</i>	1
<i>INVENTORDAPI?</i>	0
<i>MINSIZE</i>	2
<i>MAXSIZE</i>	0
<i>MINNORM</i>	0

Tab. 8.2: Die Datei **menu**.

Alle mit einem Fragezeichen behafteten Parameter sind binärer Art, d.h. nur 0 und 1 -für nein und ja- sind in diesen Fllen mögliche Eingaben. Dieses Beispiel (Tab. 8.2) ist typisch für die Analyse von Chromosomen in der Metaphase in 2D.

Die Bedeutung der Parameter im einzelnen:

COLORS Die Anzahl der verwendeten Fluorochrome.

MERGE? Ist dieser Wert auf 1 gesetzt, werden benachbarte Regionen, die der glei-

chen Klasse zugeordnet wurden, miteinander verschmolzen. Für das Klassifikationsresultat hat dieser Schritt keinen Einfluss. Bedeutung hat dieser Schritt bei dreidimensionalen Daten, die eine Farbsegmentierung der Objekte erfordern, um Positionen oder Volumina zu bestimmen.

SINGLELABELS? Sind keine kombinatorischen Färbungen vorhanden, ist es vorteilhaft, die einzelnen Kanäle bei der Hintergrundkorrektur einzeln zu behandeln. Dies wird durch den Wert 1 bei diesem Parameter erreicht.

MEDIANMAXIMAGE? Mit 1 wird das Hintergrundbild (DAPI oder Maximum-Projektion) mit einem 3x3 Medianfilter behandelt.

ZSCALE Dieser Wert im double-Format gibt das Verhältnis von axialer zu lateraler Voxelgrösse an. Er ist nur für dreidimensionale Anwendungen von Interesse, und wird zur verhältnistreuen Darstellung im Klassifikationsbild im OpenInventor-Format benötigt.

ANGLE Das Winkelkriterium in Grad. Sinnvolle Werte für ein 5 Farben Metaphasen Experiment mit 24 gefärbten Chromosomen sind im Bereich 5-15, für 3D Anwendungen 10-30 und kleine Signale 30-90 Grad, je nach Experiment und Hybridisierungsschema. Die angegebenen Werte sind empirische Richtwerte aus den zur Zeit der Entwicklung vorliegenden Daten.

TAU Dieser und die nächsten beiden folgenden Werte steuern den anisotropen nicht-linearen Diffusionsfilter. Der aktuelle ist ein künstlich eingeführter Zeitschritt beim Diffusionsprozess und kann in der Regel unverändert beim aktuellen Wert 0.0001 bleiben.

MAXSTEP Ist dieser Wert auf 0 gesetzt, wird der Filter nicht angewendet. Im üblichen Wertebereich von 1-5 gibt er die Anzahl der iterativen Diffusionsschritte an.

LAMBDA Hat Bedeutung für die Edge-Stopping-Function und sollte im Bereich der Rauschens liegen. Je grösser die Objekte, die es zu glätten gilt, um so grösser dieser Wert. Üblich sind Werte zwischen 1 und 10.

BACKGROUND Mit diesem Parameter gibt steuert man die Schwelle beim Hintergrundbild. Der Schwellwert berechnet sich aus der Standardabweichung des Hintergrundbildes multipliziert mit diesem double-Wert. Bei kleinen Signalen kann dieser Wert durchaus 6.0-8.0 erreichen.

MAXITERATION Gibt die Maximalzahl der Schritte bei der Clustersuche an, und sollte in der Regel grösser als 30 sein, da das Verfahren üblicherweise nach spätestens 25 Iterationen konvergiert. Für Sonderfälle kann dieser Parameter als weiterer Steuermechanismus dienen.

BORDERS? Bewirkt, das im Klassifikationsbild nur die Randpixel/-Voxel der Regionen herausgeschrieben werden. Interessant, um die Wirkung des Verfahrens und der Parameter zu untersuchen.

SMALLSIGNALS? Ich gestehe: die Nomenklatur ist irreführend. Ist dieser Wert auf 1 gesetzt wird in jedem Fall die Hintergrundkorrektur auf der Maximum-

Projektion durchgeführt. Ist das DAPI Bild vorhanden und mit angegeben (was zu empfehlen ist), wird es beim Klassifikationsbild als Hintergrund verwendet. Dieser Wert sollte also bei Barcodes, Telomeren und kleineren Signalen, bei denen nicht auf der DAPI Maske segmentiert werden soll auf jeden Fall auf 1 gesetzt werden. Bei 3D Anwendungen oder in Fällen, wo kein DAPI vorhanden ist bzw. nicht als Parameter beim Programmaufruf mit angegeben wurde, wird die Hintergrundkorrektur automatisch auf dem Maximum-Projektionsbild durchgeführt.

REGIONCORRECTION? Für Analysen von M-FISH Bildern in der Regel nicht interessant und sollte daher 0 sein. Bewirkt, dass nur lokal grösste Regionen erhalten bleiben. Sehr wichtig bei der Analyse von DNA-Chips, wie es in [Rud00] beschrieben wird.

PERFORMCLUSTERING? Gibt an, ob die Clustersuche durchgeführt werden soll. In der Regel sollte dieser Wert auf 1 bleiben, allein, bei kleinen Signalen, die keine Clusterbildung im Farbraum aufweisen, kann durch das Umgehen der Clustersuche mit 0 unter Umständen ein besseres Ergebnis erzielt werden. Die Regionen werden in diesem Fall direkt über den Winkel den Klassen zugeordnet.

IVCHANNEL1 Mit diesem und den nächsten beiden folgenden Werten wählt man die darzustellenden Farbkanäle aus, die als Farbraum im OpenInventor-Format herausgeschrieben werden. Dabei werden nur diejenigen Pixel und Regionen herausgeschrieben, die zu Klassen gehören die nur in diesen Farben kombinatorisch gefärbt wurden. Die Darstellung von Klassen mit mehr als drei Simultanlabels ist hier nicht möglich. Die Zählung beginnt hier bei 0 für den ersten Farbkanal.

IVCHANNEL2 siehe oben.

IVCHANNEL3 siehe oben.

POINTSIZe Gibt die Punktgrösse in den Inventordateien an. Sollte nicht grösser als drei gewählt werden, da die Visualisierung der Punkte klobig wirkt.

INVENTORCOLSPACE? Mit 0 werden keinerlei Inventor-Dateien herausgeschrieben. Bei zweidimensionalen Anwendungen also keine Farbraumdateien, im dreidimensionalen Fall sowohl keine Farbraumdateien, als auch kein Objektbild.

MINSIZe Gibt die untere Schwelle für Regionen an, die zu behandeln sind. Richtwerte sind schwer zu nennen, da es von Experiment zu Experiment stark variieren kann. Ist das Winkelkriterium klein gewählt, oder hat man trotz grossen Kriteriums viele kleine Regionen, sollte man mit diesem Wert vorsichtig sein, da schon bei kleinen Werten viel wichtige Information aus dem Bildvolumen entfernt werden kann, was unter Umständen zu sehr unterschiedlichen Klassifikationsergebnissen führen kann. Der Wert 0 oder 1 deaktiviert diesen Grössenfilter.

MAXSIZe Ähnliches gilt für diesen Wert. Er setzt die obere Schwelle für die Regionengrösse und eignet sich besonders gut dazu, bei Experimenten mit kleinen Signalen, bei dem ein relativ grosser Winkel gesetzt wurde, Interphase-Kerne aus dem Bild zu entfernen. Der Wert 0 deaktiviert diesen Grössenfilter.

MINNORM Dieser Wert setzt eine untere Schwelle für die Intensität von Regio-

nen, und kann eine fehlerhafte Hintergrundkorrektur möglicherweise korrigieren, in dem Regionen mit geringer Intensität, die zum Hintergrund gehören, entfernt werden.

BACKGROUNDDAPI Nur für 3D interessant. Ist eine Segmentierung des Zellkerns notwendig, der in DAPI gefärbt wurde, um beispielsweise den Mittelpunkt zu bestimmen, entspricht dieser Parameter dem BACKGROUND Parameter, der weiter oben beschrieben ist. Dieser Wert multipliziert mit der Standardabweichung des DAPI Bildes definiert die Intensitätsschwelle für den Hintergrund.

INVENTORDAPI? Gibt an, ob im 3D Bild der DAPI Kern mit herausgeschrieben werden soll. Diese letzten beiden Schritte sind in der Regel zeitaufwendiger, da der Kern als Region sehr gross ist. Möchte man nur die Randvoxel der Regionen herauschreiben, was im Dreidimensionalen optisch oft sinnvoll ist, kann es aufgrund der Grösse durchaus einige Minuten dauern. Daher Vorsicht beim Umgang.

Zum Schluss noch eine typische menu-Datei für ein Telomer-Experiment.

<i>COLORS</i>	4
<i>MERGE?</i>	0
<i>SINGLELABELS?</i>	0
<i>MEDIANMAXIMAGE?</i>	1
<i>ZSCALE</i>	2.5
<i>ANGLE</i>	50
<i>TAU</i>	0.0001
<i>MAXSTEP</i>	3
<i>LAMBDA</i>	2
<i>BACKGROUND</i>	4
<i>BACKGROUNDDAPI</i>	1
<i>MAXITERATION</i>	30
<i>BORDERS?</i>	0
<i>SMALLSIGNALS?</i>	1
<i>REGIONCORRECTION?</i>	0
<i>PERFORMCLUSTERING?</i>	0
<i>IVCHANNEL1</i>	0
<i>IVCHANNEL2</i>	2
<i>IVCHANNEL3</i>	3
<i>POINTSIZ</i>	1
<i>INVENTORCOLSPACE?</i>	1
<i>INVENTORDAPI?</i>	0
<i>MINSIZE</i>	3
<i>MAXSIZE</i>	0
<i>MINNORM</i>	0

Tab. 8.3: Die Datei **menu**.

Kapitel 9

Publikationen im Rahmen dieser Dissertation

1: **Saracoglu K**, Brown J, Kearney L, Uhrig S, Azofeifa J, Fauth C, Speicher MR, Eils R. New concepts to improve resolution and sensitivity of molecular cytogenetic diagnostics by multicolor fluorescence in situ hybridization. *Cytometry*. 2001 May 1;44(1):7-15.

2: Brown J, **Saracoglu K**, Uhrig S, Speicher MR, Eils R, Kearney L. Subtelomeric chromosome rearrangements are detected using an innovative 12-color FISH assay (M-TEL). *Nat Med*. 2001 Apr;7(4):497-501.

3: Brown J, Horsley SW, Jung C, **Saracoglu K**, Janssen B, Brough M, Daschner M, Beedgen B, Kerkhoffs G, Eils R, Harris PC, Jauch A, Kearney L. Identification of a subtle t(16;19)(p13.3;p13.3) in an infant with multiple congenital abnormalities using a 12-colour multiplex FISH telomere assay, M-TEL. *Eur J Hum Genet*. 2000 Dec;8(12):903-10.

4: Castleman KR, Eils R, Morrison L, Piper J, **Saracoglu K**, Schulze MA, Speicher MR. Classification accuracy in multiple color fluorescence imaging microscopy. *Cytometry*. 2000 Oct 1;41(2):139-47.

5: Eils R, Uhrig S, **Saracoglu K**, Sätzler K, Bolzer A, Petersen I, Chassery J, Ganser M, Speicher MR. An optimized, fully automated system for fast and accurate identification of chromosomal rearrangements by multiplex-FISH (M-FISH). *Cytogenet Cell Genet*. 1998;82(3-4):160-71.

Literaturverzeichnis

- [Abe00] **D. Abels**, *GoTo C++Builder*, Addison-Wesley, 2000
- [Alb94] **B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter**, *Molecular Biology of the Cell*, Garland Publishing. 3. Edition 1994.
- [Alb98] **B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter**, *Essential Cell Biology*, Garland Publishing. 1998.
- [Azo00] **J. Azofeifa et al.**, *An optimized probe set for the detection of small inter-chromosomal aberrations by 24-color FISH.*, Am. J. hum. Genet., 66, 1684-1688, 2000
- [Bal98] **Pierre F. Baldi, S. Brunak**, *Bioinformatics: the machine learning approach*, Bradford Book, MIT Press, 1998
- [Bed98] **Jörg Bewersdorf, Rainer Pick and Stefan W. Hell**, *Multifocal multi-photon microscopy*, Opt. Lett., 23, pp 655-657. 1998.
- [Bof97] **Harald Bornfleth, K. Sätzler, R. Eils and C. Cremer**, *High precision distance measurements and volume conserving segmentation of objects near and below the resolution limit in three dimensional confocal fluorescence microscopy*, J. Microsc., 189, pp 118-136. 1997.
- [BW80] **M. Born and E. Wolf**, *Principles of Optics*, John Wiley, 1980
- [Bro93] **T. Brown**, *Moderne Genetik*, Spektrum, 1993
- [Bwn01] **J. Brown, K. Saracoglu, S. Uhrig, M. R. Speicher, R. Eils, L. Kearney**, *Subtelomeric chromosome rearrangements are detected using an innovative 12-colour FISH assay (M-TEL)*, Nat Med., 7(4), 497-501, 2001
- [Bwn00] **Brown J, Horsley SW, Jung C, Saracoglu K, Janssen B, Brough M, Daschner M, Beedgen B, Kerkhoffs G, Eils R, Harris PC, Jauch A, Kearney L.**, *Identification of a subtle t(16;19)(p13.3;p13.3) in an infant with*

- multiple congenital abnormalities using a 12-colour multiplex FISH telomere assay, M-TEL., Eur J Hum Genet., 8(12), 903-10, 2000*
- [Cas00] **K.R. Castleman, R. Eils, L. Morrison, J. Piper, K. Saracoglu, M. Schulze, M.R. Speicher**, *Classification accuracy in multiple color fluorescence imaging microscopy*, Cytometry, 41, 139-147, 2000
- [Cat92] **F.Catt, P.-L.Lions, J.-M.Morel, T.Coll**, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., Vol.29, 182-193, 1992
- [Cel90] **Mehmet Celenk**, *A Color Clustering Technique for Image Segmentation*, Computer Vision, Graphics and Image Processing, 52, pp 145-170. 1990.
- [CGH95] **Inventors: D.Pinkel, J.Gray, A.&O.Kallioniemi, F.Waldman**, *Comparative genomic hybridization*, Patent No.:5,665,549, 1997
- [Chh81] **Robert F. Churchhouse (Ed.)**, *Handbook of Applicable Mathematics*, Volume III: Numerical Methods. John Wiley, Chichester, New York, Brisbane, Toronto. 1981.
- [Cox82] **I.J.Cox et al.**, *Super-resolution by confocal fluorescent microscopy*, Optik, 60, 391-396, 1982
- [Cre78] **C.Cremer and T. Cremer**, *Considerations on a laser-scanning microscope with high resolution and depth*, Microsc. Acta, 81, 31-44
- [Cre84] **C.Cremer, J. W. Gray and H. H. Ropers**, *Flow cytometry characterization of a chinese hamster X man hybrid cell line retaining the human Y chromosome.*, Hum. Genet, 60, 262-266
- [Cre88] **T.Cremer, P. Lichter, J. Borden, D. C. Ward and L. Manuelidis**, *Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome specific library probes.*, Hum. Genet, 80, 235-246
- [Cri70] **F.Crick**, *Central Dogma of Molecular Biology*, Nature, 227, 561-563, 1970
- [Czi81] **G. Czihak, F. Langer, H. Ziegler**, *Biologie*, Springer, 1996
- [Dem96] **W.Demtröder**, *Atome, Moleküle und Festkörper*, Springer, 1996
- [Dic98] **Dickhaus**, *Skript zur Vorlesung Bildverarbeitung*, Universität Heidelberg, nicht veröffentlicht

- [Dnk90] **Winfried Denk, James Strickler and Watt W. Webb**, *Two-Photon Laser Scanning Fluorescence Microscopy*, Science, 248, pp 73-76. 1990.
- [DuM95] **S.DuManoir et al.**, *Quantitative analysis of comparative genomic in-situ-hybridization*, Cytometry, 19, 27-41, 1995
- [Eil98] **R. Eils, S. Uhrig, K. Saracoglu, K. Sätzler, A. Bolzer, I. Petersen, J. Chassery, M. Ganser, M. R. Speicher**, *An optimized, fully automated system for fast and accurate identification of chromosomal rearrangements by multiplex-FISH (M-FISH)*, Cytogenet. Cell Genet., 82, 160-171, 1998
- [Eis99] **M.Eisen**, *ScanAlyze User Manual*, Stanford University, 1999
- [Elk94] **H. Erlenkötter, V. Reher**, *Objektorientierte Programmierung in C++*, Rororo, 1994.
- [Erm99] **E.Ermann**, *Herstellung von biomolekularen Arrays - eine technologische Herausforderung*, medgen, 6-11, 1999
- [Fim96] **Eugene Fiume**, *Scientific Computing*, Dpunkt - Verlag für digitale Technologie GmbH, Heidelberg. 1996.
- [Fow75] **Grant R. Fowles**, *Introduction to Modern Optics*, Holt, Rinehart and Wilson, New York. 2. Edition 1975.
- [Gar96] **Y. Garini et. al.**, *Spectral karyotyping*, Bioimaging, 4, 65-72, 1996
- [Gar99] **Y. Garini, A. Gil, I. Bar-Am, D. Cabib, N. Katzir**, *Signal to noise analysis of multiple color fluorescence imaging microscopy*, Cytometry, 35, 214-226, 1999
- [Ger99] **D.Gerhold et.al.**, *DNA chips: promising toys to have become powerful tools*, Techniques, 4, 168-173, 1999
- [Gra00] **J.Gray, C.Collins**, *Genome changes and gene expression in human solid tumors*, Carcinogenesis, 21,3, 443-452, 2000
- [Hän94] **Pekka Hänninen and Stefan W. Hell**, *Femtosecond pulse broadening in the focal region of a two-photon fluorescence microscope*, Bioimaging, 2, pp 117-121. 1994.
- [Har98] **John W. Harris, Horst Stocker**, *Handbook of Mathematical and Computational Science*, Springer-Verlag, New York. 1998.

- [Hel94a] **S.W. Hell, S. Lindek, C. Cremer and E.H.K. Stelzer**, *Measurement of the 4Pi confocal point spread function proves 75 nm axial resolution*, Appl. Phys. Lett., 64 (11), pp 1335-1337. 1994.
- [Hel94b] **Stefan W. Hell, Steffen Lindek, Christoph Cremer and Ernst H.K. Stelzer**, *Confocal microscopy with an increased detection aperture: type-B 4Pi confocal microscopy*, Opt. Lett., 19, pp 222-224. 1994.
- [Hel95] **Stefan W. Hell, Jori Soukka and Pekka Hänninen**, *Two- and multiphoton detection as an imaging mode and means of increasing the resolution in far-field light microscopy: a study based on photon-optics*, Bioimaging, 3, pp 64-69. 1995.
- [Her97] **D.Hermann**, *C++ für Naturwissenschaftler*, Addison-Weseley, 1997
- [Hpp99] **F. Höppner, F. Klawonn, R. Kruse, T. Runkler**, *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*, John Wiley, 1999
- [Iwa95] **H.Iwabuchi et al.**, *Genetic analysis of benigne, low-grade and high-grade ovarian tumors*, Cancer Research, 55, 6172-6180, 1995
- [Jäh97] **B.Jähne**, *Digitale Bildverarbeitung*, Springer, 4.Auflage, 1997
- [Jai89] **Anil K. Jain**, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989
- [Kal92] **A.Kallioniemi et al.**, *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors*, Science, 258, 818-821, 1992
- [Kle92] **R.Klette, P.Zamperoni**, *Handbuch der Operatoren für die Bildbearbeitung*, Vieweg, 1992
- [Kni95] **R.Knippers**, *Molekulare Genetik*, Thieme, 6.Auflage, 1995
- [Lam00] **S.Lampel et al.**, *Disease-specific microarrays for the automated analysis of genetic imbalances by matrixCGH*, unveröffentlicht
- [Len93] **C. Lengauer et al.**, *Chromosomal bar codes constructed by fluorescence in situ hybridization with Alu-PCR products of multiple YAC clones*, Hum. Mol. Genet., 2, 505-512, 1993
- [Lev76] **M. D. Levine and J. Leemet**, *A method for non-purposive picture segmentation*, Proceedings of the Third International Joint Conference on Pattern Recognition, 494-497, 1976

- [Lev81] **M. D. Levine and S. I. Shaheen**, *A modular computer vision system for picture segmentation and interpretation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 3, 540-556, 1981
- [Lew97] **B.Lewin**, *Genes*, Oxford University Press, 1997
- [Lic00] **P.Lichter et al.**, *Comparative genomic hybridization. Uses and limitations Seminars in Hematology*, in press, 2000
- [Lin89] **H.Linder**, *Biologie*, Metzler, 20.Auflage, 1989
- [Mue99] **C.Mühlhardt**, *Der Experimentator: Molekularbiologie*, Gustav-Fischer-Verlag Stuttgart, 1999
- [Net97] **Hans Netten**, *Automated Image Analysis of FISH-Stained Nuclei*, Doktor Thesis, Tech. University of Delft, 1997
- [Ned89] **P. M. Nederlof et al.**, *Three color fluorescence in situ hybridization for the simultaneous detection of multiple neucleic acid seyquences*, Cytometry, 10, 20-27, 1989
- [Ned90] **P. M. Nederlof, S. van der Flier, J. Wiegant, A. K. Raap, H. J. Tanke**, *Multiple fluorescence in situ hybridization*, Cytometry, 11, 126-131, 1990
- [Paw95] **James B. Pawley (Ed.)**, *Handbook of Confocal Microscopy*, Plenum Press, New York. Second Edition 1995.
- [Pea94] **A.Pease et al.**, *Light-generated oligonucleotide arrays for rapid DNA sequence analysis*, Proc.Natl.Acad.Sci, 91, 5022-5026, 1994
- [Per87] **P.Perona, J.Malik**, *Scale space and edge detection using anisotropic diffusion*, Proc.IEEE Comp.Soc. Workshop on Computer Vision(Miami Beach, Nov.30-Dec.2), 16-22, IEEE Computer Society Press, 1987
- [Pez96] **J.-Ph. Prez**, *Optik*, Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford. 1996.
- [Pin98] **J.Pinkel**, *High resolution analysis of DNA copy-number variations using comparative genomic hybridization to microarrays*, Nature Genetics, 20, 207-211, 1998
- [Pinz94] **A.Pinz**, *Bildverstehen*, Springer, 1994
- [Pol99] **J.Pollack et al.**, *Genome-wide analysis of DNA copy-number changes using cDNAmicroarrays*, Nature Genetics, 23, 41-46, 1999

- [Pop93] **S. Popp et al.**, *A strategy for the characterization of minute chromosome rearrangements using multiple color fluorescence in situ hybridization with chromosome specific DNA libraries and YAC clones*, Hum. Genet., 92, 527-532, 1993
- [Prs92] **W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery**, *Numerical Recipes in C*, Cambridge University Press. Second Edition 1992.
- [Rao92] **S. S. Rao**, *Optimization*, Wiley Eastern Limited. Second Edition 1992.
- [Ras00] **H. Rashidi, L. Buehler**, *Bioinformatic Basics*, CRR Press LLC, 2000
- [Rei99] **K. Reisdorph et al.**, *Borland C++Builder Unleashed*, SAMS Borland Press, 1999
- [Ren00] **W. Rens, F. Yang, P.C.M. O'Brien, N. Solanky, M.A. Ferguson-Smith**, *Comparison of Spectral Karyotyping (SKY) and mFISH in the analysis of chromosomal re-arrangements*, QMC 2000, 170-173, 2000
- [Rer97] **Volker Rehrmann**, *Segmentierung von Farbbildern*, 1. Heidelberger Bildverarbeitungsforum: Segmentierung. 1997.
- [Res85] **W. van Resandt et al.**, *Optical fluorescence microscopy in three dimension: microtomoscopy*, J. Microsc., 138, 29-34, 1985
- [Rie92a] **T. Ried, A. Baldini, T. C. Rand and D. C. Ward**, *Simultaneous visualization of seven different DNA probes by in situ hybridization using combinatorial fluorescence and digital imaging microscopy*, Proc. Nat. Acad. Sci. USA, 89, 1388-1392, 1992
- [Rie92b] **T. Ried, G. Landes, W. Dackowski, K. Klinger and D. C. Ward**, *Multicolor fluorescence in situ hybridization for the simultaneous detection of probe sets for chromosomes 13, 18, 21, X and Y in cultured amniotic fluid cells*, Hum. Mol. Genet., 1, 307-313, 1992
- [Rud00] **E. Rudigier**, *Entwicklung eines automatisierten Bildverarbeitungssystems zur Auswertung unregelmässiger Bildpunkte auf DNA-Chips*, Diploma Thesis, University of Heidelberg, 2000
- [Rus95] **John C. Russ**, *The Image Processing Handbook*, CRC Press. Second Edition 1995.
- [Sac92] **L. Sachs**, *Angewandte Statistik*, Springer, 7. Auflage, 1992
- [Sac93] **L. Sachs**, *Statistische Methoden*, Springer, 7. Auflage, 1993

- [Sar01] **K. Saracoglu, J. Brown, L. Kearney, S. Uhrig, J. Azofeifa, C. Fauth, M. R. Speicher, R. Eils**, *New concepts to improve resolution and sensitivity of molecular cytogenetic diagnostics by multicolor fluorescence in situ hybridization*, *Cytometry*, 44, 7-15, 2001
- [Sch96] **E. Schröck et. al.**, *Multicolor spectral karyotyping of human chromosomes*, *Science*, 273, 494-497, 1996
- [Scw91] **Norbert Schwarz**, *Einführung in TeX*, Addison-Wesley. 3. Auflage 1991.
- [Sha96] **D. Shalon et al.**, *A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization*, *Genome Research*, 6, 639-645, 1996
- [Sp96a] **Michael R. Speicher, Stephen G. Ballard and David C. Ward**, *Karyotyping human chromosomes by combinatorial multi-fluor FISH*, *Nature Genet.*, 12, pp 368-75. 1996.
- [Sp96b] **Michael R. Speicher, Stephen G. Ballard and David C. Ward**, *Computer image analysis of combinatorial multi-fluor FISH*, *Bioimaging*, 4, pp 52-64. 1996.
- [Spe96a] **M.R. Speicher, S.G. Ballard, D.C. Ward**, *Karyotyping human chromosomes by combinatorial multi-fluor FISH.*, *Nature Genet.*, 12, 368-375, 1996
- [Spe96b] **M.R. Speicher, S.G. Ballard, D.C. Ward**, *Computer image analysis of combinatorial multi-fluor FISH.*, *Bioimaging*, 4, 52-64, 1996
- [Soi00] **M. Ahlstedt-Soini, S. Kytölä, C. Larsson, J. Isola, R. Karhu**, *Comparison of Spectral Karyotyping (SKY) and mFISH in the analysis of chromosomal re-arrangements*, *QMC 2000*, 170-173, 2000
- [Sol97] **S. Solinas-Toldo et al.**, *Matrix-based Comparative Genomic Hybridization: Biochips to Screen for Genomic Imbalances Genes*, *Genes Chromosomes & Cancer*, 20, 399-407, 1997
- [Sta00] <http://rana.lbl.gov>
- [Str96] **T. Strachan, A. Read**, *Human Molecular Genetics*, Bios Scientific Publishers, 1996
- [Sup92] **Bjarne Stroustrup**, *Die C++ Programmiersprache*, Addison-Wesley. 2. Auflage 1992.

- [Tva00] **W.Tvarusko**, *Zeitaufgelöste Analyse und Visualisierung in lebenden Zellen*, PhD-Thesis, Universität Heidelberg, 2000
- [War95] **Michael S. Waterman**, *Introduction to Computational Biology*, Chapman and Hall/CRC, 1995
- [WB96] **J.Weickert, B.Benhamouda**, *A semidiscrete nonlinear scale-space theory and its relation to the Perona-Malik paradox*, Theoretical foundations of computer vision(Dagstuhl, März 18-22),1996
- [Wei96] **J.Weickert**, *Anisotropic Diffusion in Image Processing*, Universität Kaiserslautern, 1996
- [Wie93] **j. Wiegant et al.**, *Multiple and sensitive fluorescence in situ hybridization with rhodamine-, fluorecein-, and coumarin-labeled DNAs*, Cytogenet. Cell Genet., 63, 73-76, 1993
- [Xu96] **Chris Xu and Watt W. Webb**, *Measurement of two-photon excitation cross sections of molecular fluorophores with data from 690 to 1050 nm*, J. Opt. Soc. Am., 13, pp 481-491. 1996.
- [Yas81] **Zafer A. Yasa and Nabil M. Amer**, *A rapid-scanning autocorrelation scheme for continuous monitoring of picosecond laser pulses*, Opt. Commun., 36, pp. 406-408, 1981
- [Zhu95] **Song Chun Zhu and A.L. Yuille**, *Region Competition and its Analysis*, Harvard Robotics Lab. Technical Report no. 95-07. 1995.
- [Zim96] **W.Zimmer, E.Bonz**, *Objektorientierte Bildverarbeitung*, Hanser, 1996
- [Zuc76] **S. W. Zucker**, *Region Growing: Childhood and adolescence*, Computer Graphics and Image Processing, 5, 382-399, 1976

Abbildungsverzeichnis

1.1	Mögliche Basenpaarungen	15
1.2	DNA in der Doppelhelix-Darstellung	16
1.3	Gen sind DNA-Abschnitte	17
1.4	Die beiden Stufen der Genexpression	18
1.5	M-FISH	21
1.6	Hybridisierungsschema	22
1.7	schematischer Ablauf der vergleichenden genomischen Hybridisierung	23
2.1	konfokales Laser Scanning Mikroskop	27
2.2	Termschema	29
2.3	Bildaufnahme von MFISH	33
3.1	Medianfilter	46
3.2	Diffusionsfilter	48
4.1	Flowchart der Methodik	52
4.2	Maximum-Projektionsbild	53
4.3	DAPI Bild	54
4.4	DAPI Schwellwertbild	54
4.5	Datenraum	55
4.6	Farbraum	56
4.7	Normiertes Skalarprodukt	57
4.8	Winkel als Klassifikator	58
4.9	Euklidischer Abstand als Klassifikator	59
4.10	Regionwachstum 3 Grad	62
4.11	Regionwachstum 15 Grad	63
4.12	Regionen im Farbraum	64
4.13	Cluster im Farbraum	65
4.14	Klassifiziertes Bild	66
4.15	Methodik	68
5.1	Normale Metaphase	70

5.2	Karyogramm	70
5.3	Aberrante Metaphase	71
5.4	Farbraum Metaphase	71
5.5	Barcode Färbung eines Chromosoms	72
5.6	Barcode Klassifikation Metaphase	72
5.7	Klassifikation von Telomeren	74
5.8	Farbraum von Telomerbildern	74
5.9	Karyogramm eines Patienten mit geistiger Zurückgebliebenheit.	75
5.10	Hybridisierungsschema mit 4 Farben	75
5.11	3-Farben M-FISH in der Interphase	76
5.12	3-Farben Klassifikation in der Interphase	77
6.1	Hybridisierungsschema mit 5 Farben	79
6.2	Hybridisierungsschema mit 7 Farben	81
6.3	Schwellwert pro Kanal	83
6.4	Binäre Klassifikation	83
6.5	Schwellwert pro Kanal schlecht gewählt	84
6.6	Fehlerhafte binäre Klassifikation	84
6.7	Korrekte Klassifikation	85
6.8	Winkel als Klassifikator	85
6.9	Euklidischer Abstand als Klassifikator	86
6.10	Winkel als Klassifikator	86
6.11	Euklidischer Abstand als Klassifikator	87
6.12	Aberrante Metaphase	87
6.13	Regionenwachstum mit Euklidischem Abstand	88
6.14	Regionenwachstum mit Winkel	88
6.15	Clustersuche über Winkel	89
6.16	Clustersuche über Euklidischen Abstand	89
6.17	Ausschnitt Barcode Metaphase	90
6.18	Focal shifts	91
6.19	Grosses Winkelkriterium bei Telomeren	92
6.20	Farbraum eines Telomereperiments	93
6.21	7-Farben Klassifikation in der Interphase	93
6.22	Farbraum 3D M-FISH	94
6.23	Farbraum der Regionen 3D	94

Danksagung

Ich möchte mich sehr herzlich bei Herrn Prof. Dr. Bernd Jähne für die Betreuung dieser Arbeit bedanken. Den Mitgliedern seiner interdisziplinären Forschergruppe danke ich für den vielfältigen Austausch und die stets anregenden Diskussionen. Ebenso herzlich bedanke ich mich bei Herrn Prof. Dr. Dr. Christoph Cremer für seine freundliche Bereitschaft, diese Arbeit als Koreferent zu begutachten.

Herrn Dr. Roland Eils, ohne den diese Arbeit nicht möglich gewesen wäre, und der gesamten Arbeitsgruppe „Intelligente Bioinformatiksysteme“ möchte ich für die grosse Unterstützung und die angenehme Arbeitsatmosphäre am DKFZ danken.

Besonders erwähne ich an dieser Stelle Dr. Kurt Sätzler, den ich über die Jahre weit mehr als nur einen hilfsbereiten und kompetenten Kollegen kennen und schätzen gelernt habe. Bei Frau Eveline Rudigier, die mich gut ein Jahr als ihren Betreuer ertragen musste, bedanke ich mich für ihre Geduld und die gute geleistete Arbeit. Meine Gedanken sind bei meinen Eltern und meinem Bruder, Quelle meiner Inspiration.

Erklärung

Ich versichere, da ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 20. September 2001

.....
(Kaan Saracoglu)