

# Inhaltsbasierte Bildsuche mittels visueller Merkmale



Vom Fachbereich Informatik  
der Technischen Universität Darmstadt  
genehmigte

## DISSERTATION

zur Erlangung des akademischen Grades eines  
Doktor-Ingenieurs (Dr.-Ing.)

von

Dipl.-Ing. Stephan Volmer  
geboren in Lich

Referenten der Arbeit: Prof. Dr. J.L. Encarnaç o  
Prof. Dr. W.-D. Fellner

Tag der Einreichung: 1. Februar 2006

Tag der m ndlichen Pr fung: 24. M rz 2006

Darmst dter Dissertation

D17



*Meiner lieben Kirstin*



# Danksagungen

Ein Doktorand wird im Hinblick auf seine Arbeit oft als Einzelkämpfer titulierte. Tatsächlich findet die fachliche Auseinandersetzung mit dem Thema überwiegend in der Klausur des „stillen Kämmerleins“ statt. Dennoch kann ein solches Vorhaben niemals ohne den fachlichen Rat und die moralische Unterstützung einer Vielzahl anderer Personen zu einem fruchtbaren Ende geführt werden. Daher möchte ich allen Beteiligten meinen persönlichen Dank aussprechen.

Meinem Doktorvater Herrn Professor Dr. José Luis Encarnação gilt ein besonderer Dank für die Betreuung meiner Promotion an seinem Lehrstuhl für Graphisch-Interaktive Systeme an der Technischen Universität Darmstadt. Herrn Professor Dr. Wolf-Dietrich Fellner vom Lehrstuhl für Computergraphik und Wissensvisualisierung der Technischen Universität Graz danke ich für die Übernahme des Koreferats.

Die vorliegende Dissertation entstand im Wesentlichen während meiner Zeit als wissenschaftlicher Mitarbeiter am Fraunhofer Institut für Graphische Datenverarbeitung. Ich danke meinem Abteilungsleiter Herrn Professor Dr. Georgios Sakas, der es mir ermöglicht hat, in einem erstklassigen wissenschaftlichen Umfeld zu arbeiten und zu forschen. Meinen Abteilungskollegen danke ich für die grandiose Zusammenarbeit in freundschaftlicher Atmosphäre. Bedanken möchte ich mich auch bei allen Studenten, die mich als wissenschaftliche Hilfskraft, Praktikant oder Diplomant unterstützt haben. Ohne ihr Engagement wäre es nicht möglich gewesen, alle meine Ideen umzusetzen.

Persönlich bedanke ich mich bei Dr. Sven Gürke, René Cavet, Stefan Wesarg, Dr. Peter Neugebauer, Konrad Klein, Dr. Volker Roth, Dr. Matthias Finke und Kai Richter für die vielen wissenschaftlichen Diskussionen und ihre Bereitschaft, mit hilfreichen Anregungen zur Seite zu stehen.

Ein Dank gilt auch meiner Korrekturleserin Katharina Krickow, die meine Arbeit bezüglich ihrer Konformität mit der neuen deutschen Rechtschreibung überprüft hat.

Zu guter Letzt möchte ich meiner Frau Kirstin, die einen nicht zu benennenden Teil zur Fertigstellung dieser Arbeit beigesteuert hat, von ganzem Herzen danken. Du hast mir die Kraft gegeben, diesen langen Weg gehen zu können!



# Abstrakt

Die ständig wachsende Menge an verfügbaren digitalen Daten erfordert neuartige Methoden, die einen gezielten Zugriff auf relevante Information ermöglichen. Eine zentrale Rolle spielt in diesem Zusammenhang die automatische Erschließung bildlicher Information in digitalen Bilddaten. Der klassische Ansatz – die manuelle Annotation des Bildinhaltes mittels alphanumerischer Texte – hat sich in der Vergangenheit als zu fehleranfällig und zu kostenintensiv erwiesen.

Im Rahmen dieser Arbeit wird ein alternativer Ansatz entwickelt, der es ermöglicht, große Mengen von digitalen Bilddaten mittels merkmalsbasierter Verfahren zu erschließen. Dies geschieht unter der Annahme, dass das zugrundeliegende Bildmaterial weder aufgrund seines Erscheinungsbildes, noch aufgrund seiner Bedeutung irgendeiner Einschränkung unterliegt.

Zunächst wird ein allgemeingültiges Modell für die merkmalsbasierte Suche nach visuellen Inhalten in digitalen Bildern definiert. Dieses Modell stellt den formalen Rahmen für die Entwicklung und Kombination neuartiger Algorithmen zur Merkmalsextraktion und -indexierung dar. Das Modell ermöglicht eine inhaltsbasierte Bildsuche auf der Basis eines Systems mit einheitlicher Architektur und standardisierten Schnittstellen. Ein solches System kann durch die Entwicklung einzelner anwendungsspezifischer Bausteine für eine bestimmte Problemstellung erweitert werden.

Startpunkt für die Entwicklung eines Algorithmus zur Merkmalsextraktion ist die sinnvolle Interpretation der Farben einzelner diskreter Bildpunkte. Der Mensch unterscheidet im Gegensatz zur technischen Darstellung nur zwischen einer handvoll verschiedener Farben. Daher wird im Rahmen dieser Arbeit die neuartige Farbrepräsentation vorgestellt, die die Farbinformation eines Bildes auf der Basis von – für den Menschen bedeutungsvollen – Farbnamen zugänglich macht. Das zugrundeliegende mathematische Gerüst ermöglicht einen einfachen und schnellen Vergleich von Farben. Eine solche Interpretation von Farbinformation kann nahezu für jede Aufgabe im Bereich der digitalen Bildverarbeitung nutzbringend eingesetzt werden. Auf dieser Farbrepräsentation aufbauend werden einige universelle Extraktionsalgorithmen vorgestellt, die gewisse visuelle Aspekte eines digitalen Bildes kompakt beschreiben.

Große Datenmengen bedingen entsprechend lange Verarbeitungszeiten bei der Suche nach Information. Daher wird im letzten Teil der Arbeit ein Indexierungsverfahren vorgestellt, das die proportionale Beziehung zwischen Datenmenge und Verarbeitungszeit aufbricht. Das Indexierungsverfahren basiert auf der Lokalisierung der Suche in der unmittelbaren Umgebung der Suchanfrage im Merkmalsraum. Durch das Eingrenzen des Suchraums kann eine signifikante Beschleunigung der Suche erreicht werden. Da ein solches Verfahren immanent mit einer gewissen Ungenauigkeit behaftet ist, werden experimentelle Ergebnisse präsentiert, die den Nutzen des Verfahrens im praktischen Einsatz dokumentieren.

# Abstract

The constantly growing quantity of available digital data requires novel methods in order to enable direct access to relevant information. The automatic extraction of pictorial information from digital images plays a central role in this context. The traditional approach – the manual annotation of the images' content by means of alphanumeric text – has proven in the past as too error-prone and too cost-intensive.

In the scope of this work an alternative approach that enables the handling of large quantities of digital image data by means of feature-based algorithms is pursued. It is assumed that the images are not constrained by their visual appearance or their semantic content.

For this purpose a generic model for feature-based image retrieval is defined. The model constitutes a formal framework for the development and combination of new algorithms for feature extraction and similarity indexing. It enables content-based retrieval systems to be built on basis of a uniform software architecture with standardized interfaces. Such a system can be easily extended by individually developed components with regard to customer-specific problems.

Human visual recognition and recall are highly dependent on color as a visual cue. A key step of the feature extraction process is the meaningful interpretation of individual pixel colors. Contrary to the potential of technical representations, humans can only differentiate between a few different colors. Therefore, a novel color model is introduced that makes color information accessible by means of – for humans meaningful – color names. Its underlying mathematical foundation enables a simple and fast comparison of colors. Such a representation of color can be leveraged profitably for almost any task in the domain of Digital Image Processing. Some feature extraction algorithms based on this color model that describe certain visual aspects of a digital image are presented exemplarily.

Large amounts of data inherently require longer processing times when performing search tasks. Consequently, a similarity indexing scheme is presented that breaks the proportional relationship between the amount of data searched and the time needed

to perform the search. The indexing scheme is based on the principle of localization of the search in feature space. A significant acceleration can be achieved by reducing the search space to the immediate neighborhood of the query. Since such an approach is immanently afflicted with a certain imprecision, experimental results are presented that document its usefulness in a realistic scenario.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Motivation . . . . .	2
1.2	Zielsetzung . . . . .	3
1.3	Organisation der Arbeit . . . . .	5
<b>2</b>	<b>Inhaltsbasierte Suche</b>	<b>7</b>
2.1	Daten – Wissen – Information . . . . .	7
2.2	Information Retrieval . . . . .	8
2.2.1	Eingabe von Information . . . . .	9
2.2.2	Suche nach Information . . . . .	11
2.2.2.1	Suchanfrage . . . . .	13
2.2.2.2	Suchergebnis . . . . .	13
2.2.3	Retrieval Modelle . . . . .	14
2.3	Visual Information Retrieval . . . . .	16
2.3.1	Textbasierte Verfahren . . . . .	16
2.3.1.1	Freier Volltext . . . . .	17
2.3.1.2	Verschlagwortung . . . . .	19
2.3.2	Merkmalsbasierte Verfahren . . . . .	21
2.3.2.1	Merkmale . . . . .	24
2.3.2.2	Ähnlichkeit . . . . .	26
2.3.2.3	Kombination von Merkmalen . . . . .	27
2.3.2.4	Semantische Lücke . . . . .	28

2.3.2.5	Suchparadigmen . . . . .	29
2.3.2.6	Aktuelle Forschungsthemen . . . . .	30
2.4	Evaluation . . . . .	32
2.4.1	Spezielle Probleme des VIR . . . . .	34
2.4.2	Benchmarking . . . . .	36
<b>3</b>	<b>Modellbildung</b>	<b>37</b>
3.1	Bilder . . . . .	38
3.2	Farbbilder . . . . .	39
3.3	Merkmalsextraktion . . . . .	41
3.4	Ähnlichkeit . . . . .	43
3.5	Merkmalsbasierte Suche . . . . .	44
<b>4</b>	<b>Farbe</b>	<b>49</b>
4.1	Stand der Technik . . . . .	49
4.1.1	RGB-Farbsystem . . . . .	53
4.1.2	Munsells Farbordnungssystem . . . . .	55
4.1.3	HSL-Farbsystem . . . . .	57
4.2	BCR-Farbrepräsentation . . . . .	61
4.2.1	Grundform . . . . .	65
4.2.2	Ähnlichkeit . . . . .	73
4.2.3	Approximation der Grundform . . . . .	74
4.2.4	Auswahl der Basisfarben . . . . .	77
4.2.4.1	RGB9-Basisfarben . . . . .	77
4.2.4.2	DIN5023-Basisfarben . . . . .	80
4.2.4.3	Evaluation . . . . .	86
<b>5</b>	<b>Merkmalsextraktion</b>	<b>89</b>
5.1	Stand der Technik . . . . .	89
5.1.1	Farbe . . . . .	89
5.1.2	Textur . . . . .	92

5.1.3	Form . . . . .	93
5.1.4	Räumliche Anordnung . . . . .	95
5.1.5	CBIR-Systeme . . . . .	96
5.2	Merkmale . . . . .	97
5.2.1	Histogramme . . . . .	97
5.2.1.1	Verteilungen . . . . .	100
5.2.1.2	Verteilungsmomente . . . . .	104
5.2.1.3	Co-Occurrence Matrizen . . . . .	106
5.2.2	Signaturen . . . . .	110
5.2.2.1	Wavelet-Transformation . . . . .	112
5.2.3	Evaluation . . . . .	117
<b>6</b>	<b>Merkmalsindexierung</b>	<b>119</b>
6.1	Stand der Technik . . . . .	120
6.1.1	Multidimensionale Zugriffsmethoden . . . . .	121
6.1.1.1	Vektorräume . . . . .	121
6.1.1.2	Metrische Räume . . . . .	125
6.1.1.3	„Fluch der Dimensionen“ . . . . .	126
6.1.2	Dimensionsreduktion . . . . .	127
6.1.3	Ungefähre Suche . . . . .	128
6.2	Bojen-Indexierung . . . . .	129
6.2.1	Indexerzeugung . . . . .	131
6.2.1.1	Äußere Zwänge . . . . .	133
6.2.1.2	Dynamische Aktualisierung . . . . .	134
6.2.2	Bojen-indexierte Nächste-Nachbar-Suche . . . . .	137
6.2.3	Evaluation . . . . .	139
6.2.3.1	Indexerzeugung . . . . .	140
6.2.3.2	Bojen-indexierte Nächste-Nachbar-Suche . . . . .	141
<b>7</b>	<b>Zusammenfassung</b>	<b>147</b>
7.1	Ausblick . . . . .	148

<b>A Delaunay-Zerlegungen</b>	<b>151</b>
<b>B Ergebnisse von Suchanfragen</b>	<b>153</b>
<b>Literaturverzeichnis</b>	<b>157</b>
<b>Veröffentlichungen</b>	<b>173</b>
<b>Lebenslauf</b>	<b>175</b>

# Abbildungsverzeichnis

2.1	Daten, Wissen und Information . . . . .	7
2.2	Schematische Darstellung eines IR-Systems . . . . .	9
2.3	Schematische Darstellung des Eingabeprozesses . . . . .	10
2.4	Schematische Darstellung des Retrievalprozesses . . . . .	12
2.5	Sandro Botticellis „Die Geburt der Venus“ . . . . .	18
2.6	Bild der Getty Images Bildkollektion . . . . .	20
2.7	Prinzipielle Funktionsweise merkmalsbasierter Verfahren des VIR . . . . .	23
2.8	Konkurrierende Faktoren eines IR-Systems . . . . .	34
3.1	Übersicht des Spektrums elektromagnetischer Strahlung . . . . .	39
3.2	Digitales RGB-Farbbild . . . . .	40
3.3	Schematische Darstellung der Nachbarschaftssuche . . . . .	47
4.1	Relative spektrale Empfindlichkeit der <i>S</i> -, <i>M</i> - und <i>L</i> -Zapfen . . . . .	50
4.2	RGB-Spektralwertkurven eines CIE-Standardbeobachters . . . . .	52
4.3	XYZ-Spektralwertkurven eines CIE-Standardbeobachters . . . . .	52
4.4	CIE <i>xy</i> -Chromazitätsdiagramm . . . . .	53
4.5	RGB-Farbraum . . . . .	54
4.6	RGB-Farbkanäle eines digitalen Bildes . . . . .	55
4.7	Munsell-Farbkreis . . . . .	56
4.8	Value- und Chroma-Skalen der Munsell-Haupt- und Zwischenfarbtöne . . . . .	58
4.9	HSL-Farbraum . . . . .	60
4.10	HSL-Farbkanäle eines digitalen Bildes . . . . .	61

4.11	Beispiel einer BCR-Farbrepräsentation . . . . .	64
4.12	Beispiel alternativer BCR-Farbrepräsentationen . . . . .	66
4.13	Vier Basisfarben als Eckpunkte eines tetraederförmigen Volumens umhüllen eine Farbe im dreidimensionalen Bezugsfarbraum . . . . .	67
4.14	Schematische Darstellung der Dualität des Voronoi-Diagramms und der Delaunay-Zerlegung . . . . .	69
4.15	Beispiel einer BCR-Farbrepräsentation in Grundform . . . . .	72
4.16	Beispiel einer Berechnung der Ähnlichkeit zweier BCR-Farbrepräsentationen . . . . .	74
4.17	Projektion einer Farbe in Richtung der neutralen Farbe auf die konvexe Hülle der Basisfarbmenge . . . . .	75
4.18	Farbmuster der RGB9-Basisfarbmenge . . . . .	78
4.19	RGB9-Farbkanäle eines Bildes in BCR-Grundform . . . . .	79
4.20	Unterschiedliche Auffassungen einzelner Individuen bei der Wahl prototypischer Farbmuster . . . . .	81
4.21	Farbmuster der DIN5023-Basisfarbmenge . . . . .	83
4.22	DIN5023-Farbkanäle eines Bildes in BCR-Grundform . . . . .	84
4.23	Konvexe Hülle der DIN5023-Basisfarbmenge im RGB-Farbraum . . . . .	85
4.24	Rekonstruktion eines Bildes aus DIN5023-Farbkanälen . . . . .	87
5.1	Schematische Darstellung der Abbildung des Merkmalsraums mittels Histogrammzuweisungsfunktionen . . . . .	100
5.2	Farbhistogramm eines digitalen Bildes . . . . .	102
5.3	Merkmalsbasierter Vergleich zweier digitaler Farbbilder . . . . .	103
5.4	Farbmomente eines digitalen Bildes . . . . .	105
5.5	Empirische Verteilung der Werte eines BCR-Farbkanals . . . . .	106
5.6	Bildliche Darstellung der Deutung der Elemente einer Co-Occurrence Matrix . . . . .	107
5.7	Unschärfe Abbildung der Distanzwerte auf diskrete Klassen . . . . .	108
5.8	Co-Occurrence Matrix eines digitalen Bildes . . . . .	109
5.9	Mehrstufige Quadraturspiegelfilterbank einer Wavelet-Transformation . . . . .	113
5.10	Zweidimensionale Wavelet-Transformation eines Grauwertbildes . . . . .	114

5.11	Extraktion einer Signatur aus Wavelet-Koeffizienten aus einem digitalen Farbbild . . . . .	116
6.1	Schematische Darstellung der Unterteilungsstrategien verschiedener multidimensionaler Zugriffsmethoden . . . . .	124
6.2	Veranschaulichung des Phänomens „Fluch der Dimensionen“ . . . . .	127
6.3	Schematische Darstellung der Partitionierung des Merkmalsraums mittels Bojen . . . . .	130
6.4	Iterative Optimierung mit äußerem Zwang . . . . .	134
6.5	Trajektorien der Bojen während der iterativen Optimierung . . . . .	135
6.6	Iterative Optimierung des $K$ -Medians-Clustering . . . . .	136
6.7	Schematische Darstellung des Ausschlusskriteriums . . . . .	138
6.8	Experimenteller Aufbau . . . . .	139
6.9	Histogramme der Verteilungen von Clustergrößen in Abhängigkeit des oberen Grenzwertes $W_{\max}$ . . . . .	144
6.10	Gegenüberstellung von Genauigkeit und Dauer von bojen-indexierten 20-Nächste-Nachbar-Suchen in Abhängigkeit der Parameter $K$ und $q$ . . . . .	145
B.1	Resultate verschiedener Suchanfragen für unterschiedliche Merkmale . . . . .	154



# Tabellenverzeichnis

4.1	Normalisierte RGB-Werte der RGB9-Basisfarbmenge . . . . .	78
4.2	Relative Anteile der RGB9-Basisfarben an der Grundgesamtheit aller Farben am Beispiel einer typischen Bilddatenbank . . . . .	80
4.3	Normalisierte RGB-Werte der DIN5023-Basisfarbmenge . . . . .	83
4.4	Mittlerer Fehler pro DIN5023-Farbkanal . . . . .	87
6.1	Einfluss der Parameter $N$ , $K$ und $W_{\max}$ auf die Indexerzeugung . . . .	140
A.1	Delaunay-Zerlegung der RGB9-Basisfarbmenge . . . . .	151
A.2	Delaunay-Zerlegung der DIN5023-Basisfarbmenge . . . . .	152



# Abkürzungen und Akronyme

AAT	<i>Art and Architecture Thesaurus</i>
BCR	<i>Base Color Ratio</i>
CBIR	<i>Content-Based Image Retrieval</i>
CIE	<i>Commision International de l'Eclairage</i>
DARPA	<i>Defense Advanced Research Projects Agency</i>
DIN	<i>Deutsches Institut für Normung</i>
GNAT	<i>Geometric Near-Neighbor Access Tree</i>
HSL	<i>Hue, Saturation &amp; Luminance</i>
HVC	<i>Hue, Value &amp; Chroma</i>
IR	<i>Information Retrieval</i>
ISCC	<i>Inter-Society Color Council</i>
ISO	<i>International Standardisation Organization</i>
JPEG	<i>Joint Photographic Experts Group</i>
LCTGM	<i>Library of Congress Thesaurus of Graphical Material</i>
MBB	<i>Minimum Bounding Box</i>
NBS	<i>National Bureau of Standards</i>
NIST	<i>National Institute of Standards and Technology</i>
NN	<i>Nearest Neighbor</i>
QBIC	<i>Query by Image Content</i>
QBF	<i>Query-by-Feature</i>
QBE	<i>Query-by-Example</i>
RDBMS	<i>Relational Database Management System</i>
RGB	<i>Red, Blue &amp; Green</i>
SQL	<i>Structured Query Language</i>
TCP/IP	<i>Transmission Control Protocol / Internet Protocol</i>
TREC	<i>Text Retrieval Conference</i>
VIR	<i>Visual Information Retrieval</i>
WWW	<i>World Wide Web</i>
XYZ	<i>Primary Colors X, Y &amp; Z</i>



## *Ours is a visual age.*

*We are bombarded with pictures from morning till night. Opening our newspaper at breakfast, we see photographs of men and women in the news, and raising our eyes from the paper, we encounter the picture on the cereal package. The mail arrives and one envelope after the other discloses glossy folders with pictures of alluring landscapes and sunbathing girls to entice us to take a holiday cruise, or of elegant menswear to tempt us to have a suit made to measure. Leaving our house, we pass billboards along the road that try to catch our eye and play on our desire to smoke, drink or eat. At work it is more than likely that we have to deal with some kind of pictorial information: photographs, sketches, catalogues, blueprints, maps or at least graphs. Relaxing in the evening, we sit in front of the television set, the new window on the world, and watch moving images of pleasures and horrors flit by. Even the images created in times gone by or in distant lands are more easily accessible to us than they ever were to the public for which they were created. Picture books, picture postcards and colour slides accumulate in our homes as souvenirs of travel, as do the private mementos of our family snapshots.*

*No wonder it has been asserted that we are entering a historical epoch in which the image will take over from the written word. In view of this claim it is all the more important to clarify the potentialities of the image in communication, to ask what it can and what it cannot do better than spoken or written language. In comparison with the importance of the question the amount of attention devoted to it is disappointingly small.*

*Ernst Hans Gombrich, (1909–2001), Kunsthistoriker*



# Kapitel 1

## Einleitung

Der Begriff der Information hat in den letzten Jahrzehnten eine ungeheure Aufwertung erfahren. Die Information ist heute – neben Arbeit und Kapital – der wichtigste Produktionsfaktor unserer Zivilisation. Die Information wird in allen Bereichen des Lebens gebraucht (z.B. Wissenschaft, Wirtschaft und Freizeit). Sie entsteht durch die Kreativität von Menschen, besteht aus der Originalität des Inhaltes und wird durch die Kommunikation zwischen Menschen verbreitet. Die Bereitstellung ihrer Verfügbarkeit und Nutzbarkeit für jedermann an jedem Ort zu jederzeit wird eine der Herausforderungen der Informationsgesellschaft der Zukunft werden.

Die Technologie, mit der in Zukunft Information erzeugt, gespeichert und kommuniziert wird, wird dabei maßgeblichen Einfluss auf die Ausgestaltung und Entwicklung der Gesellschaft haben. Die effiziente Nutzung von Information, insbesondere deren Verarbeitung zu Wissen, ist dabei der Schlüssel zur Vorbereitung der volkswirtschaftlichen und technologischen Rahmenbedingungen für den Weg unserer Gesellschaft in die Zukunft.

Der Trend der Informationsverarbeitung geht heute bereits eindeutig in Richtung digitaler Informations- und Kommunikationstechnologien. Die Ursache hierfür liegt nicht in der Präzisierung der inhaltlichen Qualität der Information, sondern vielmehr in der zunehmenden Möglichkeit, die Quantität des Informationsflusses mit Hilfe digitaler Systeme in nahezu unvorstellbaren Größenordnungen beherrschbar zu machen. Durch die Nutzung digitaler Technologien kann Information zunehmend in größeren Mengen bearbeitet und verbreitet werden. Dieser Trend wird sich auch im Hinblick auf den zunehmenden technischen Fortschritt bei digitalen Systemen und deren zunehmender Vernetzung in Zukunft noch weiter verstärken.

Die Handhabbarkeit größerer Informationsmengen eröffnet neue Möglichkeiten, jenseits der traditionellen textorientierten Informationsverarbeitung neue Wege der Informationsvermittlung zu beschreiten. Durch die Kombination unterschiedlicher diskreter Medien, wie Schrift, Bild, Audio, Animation und Video kann die Informationsaufnahme durch Ansprache verschiedener Auf- und Wahrnehmungsfähigkeiten des Menschen gesteigert werden. Diese Tatsache ist im Aphorismus

*„Ein Bild sagt mehr als tausend Worte“*

zusammengefasst. Die Art der Aufbereitung, Bearbeitung und Präsentation der Information unterschiedlicher Medien als Daten wird im allgemeinen unter dem Stichwort *Multimedia* zusammengefasst.

Multimediale Daten beinhalten in der Regel binärkodierte syntaktische (symbolische) Information, die durch technische Sensoren erzeugt werden. Im Unterschied dazu enthalten herkömmliche Textdaten, obwohl ebenfalls binärkodierte, semantische Information, die vom Menschen erzeugt wurden. Diese Tatsache bringt schwerwiegende Konsequenzen für die Art und Weise mit sich, wie Information in der Zukunft für die optimale Nutzung strukturiert und organisiert wird.

## 1.1 Motivation

Die Produktion von Bildern zum Zweck der Kommunikation ist bereits bekannt, seitdem unsere Vorfahren vor Jahrtausenden die Wände ihrer Höhlen mit Malerei versahen. Das erste Anfertigen von Dokumenten zur Verbreitung von bildlicher Information lässt sich auf die Zeit vor den Römern zurückdatieren. Das letzte Jahrzehnt hat jedoch eine nie zuvor dagewesene Zunahme in der Anzahl und Verfügbarkeit von Bildmaterial in allen Bereichen des menschlichen Lebens erfahren. Bildmaterial spielt heutzutage in verschiedenen Bereichen wie z.B. Medizin, Journalismus, Werbung, Design, Bildung und Unterhaltung eine bedeutende Rolle. Die Entwicklung und Kommerzialisierung neuer Technologien im Bereich digitaler graphischer Ein- und Ausgabegeräte haben dazu geführt, dass sich der Anteil von Bildern in Dokumenten in den letzten Jahren dramatisch erhöht hat.

Der technische Fortschritt im Bereich Netzwerktechnologie und ein größer werdendes öffentliches Interesse am Internet haben bewirkt, dass das *World Wide Web* (WWW) zunehmend als Medium zur Verbreitung solcher Dokumente genutzt wird. Das WWW besteht schon heute aus einer geschätzten Anzahl von mehreren Milliarden Webseiten. Diese Zahl wird sich in den nächsten Jahren voraussichtlich um das Zehnfache vergrößern. Durch den Einsatz von textbasierten Suchmaschinen wie Google, HotBot,

AltaVista, Excite, Infoseek und Lycos kann das WWW als eine Enzyklopädie mit mehreren Billionen Wörtern angesehen werden. Suchmaschinen wurden ursprünglich zur Indexierung einer statischen Ansammlung von Textdokumenten entwickelt. Das Vokabular einer Suchanfrage stimmt dabei in der Regel unmittelbar mit dem Vokabular der erfassten Inhalte dieser Dokumente überein.

Das WWW hat sich aber in den letzten Jahren zu einer dynamischen, schnell wachsenden Ansammlung multimedialer Dokumente entwickelt, dessen Information bereits zu 73% aus visuellen Inhalten besteht [92]. Textbasierte Suchmaschinen bieten jedoch typischerweise keine oder nur begrenzte Möglichkeiten an, nach visueller Information zu suchen. Visuelle Inhalte bestehen in der Regel aus digitalen Repräsentationen mit syntaktischer Natur, die von technischen Sensoren aufgezeichnet wurden. Eine automatische Identifikation von semantischen Bedeutungen, die für einen Vergleich mit einer herkömmlichen Suchanfrage in Textform benötigt wird, ist nach dem heutigen Stand der Technik nicht umfassend realisierbar. Daher ziehen textbasierte Suchmaschinen den im Umfeld eines Bildes befindlichen Text – soweit überhaupt vorhanden – zur Beschreibung des semantischen Inhaltes heran. Dieser Ansatz ist aber nur sehr unzureichend, da solche Texte in den seltensten Fällen speziell für diesen Zweck erzeugt wurden. Des Weiteren können Metadaten zur Suche verwendet werden (z.B. Dateinamen, Label, Unter-/Überschriften), die jedoch bei der Erstellung des Dokumentes einem Bild manuell hinzugefügt werden müssen.

Diese Tatsachen zwingen uns, die Art und Weise zu überdenken, wie Information in Form visueller Inhalte gefunden werden kann. Daher muss man nach neuen Methoden und Paradigmen suchen, die es ermöglichen, Dokumente mit relevanter visueller Information schnell und effizient zu finden.

## **1.2 Zielsetzung**

Das grundsätzliche Problem, das im Rahmen dieser Arbeit behandelt wird, ist die effiziente Suche nach digitalen Bildern in sehr großen – möglicherweise verteilten – Archiven. Es wird dabei angenommen, dass die Inhalte der Bilder keinerlei Einschränkungen unterliegen, d.h. sie können weder aufgrund ihres visuellen Erscheinungsbildes noch ihres semantischen Inhaltes einer bestimmten Domäne zugeordnet werden. Solche Bildarchive entstehen in der Regel durch eine Akkumulation von Bilddaten, wie sie täglich durch unterschiedliche Formen visueller Medien, wie z.B. Fotografien, Graphiken, Animationen und Videos, produziert werden.

Als Lösung des Problems wird ein System vorgeschlagen, das auf dem Einsatz von inhaltsbasierten Methoden basiert. Algorithmen und Methoden aus den Gebieten Bild-

verarbeitung, Mustererkennung und Computer Vision bilden dabei einen integralen Teil der Funktionsweise eines solchen Systems. Die Suche wird dabei nicht auf konzeptioneller Ebene wie bei den textbasierten Methoden durchgeführt, sondern auf wahrnehmungsorientierter Ebene unter Verwendung von objektiven Kriterien zur direkten Bewertung visueller Inhalte. Dabei werden Deskriptoren des visuellen Inhaltes eines Bildes durch eine vollautomatische Analyse der Verteilung der diskreten Bildpunkte und ihrer Eigenschaften erzeugt. Diese Deskriptoren sind in der Regel kompakte Darstellungen von Merkmalen, deren Elemente signifikante, auffällige Charakteristiken des Bildinhaltes repräsentieren.

Die Farbe eines Bildpunktes ist die wichtigste Quelle visueller Information in digitalen Bildern, die unmittelbar für die Verarbeitung mittels inhaltsbasierter Methoden zu Verfügung steht. Farbe ist jedoch ein Sinneseindruck, der für jeden Menschen als Individuum einzigartig ist. Um Farbe für die Beschreibung von Bildinhalten mittels Merkmalen nutzbar machen zu können, muss zunächst eine Darstellungsform gefunden werden, die der Art und Weise entspricht, wie ein menschlicher Betrachter die Farbinformation eines Bildes wahrnimmt und kommuniziert. Im Rahmen dieser Arbeit wird ein Ansatz präsentiert, der einen technischen Farbraum in eine Anzahl von Farbkategorien unterteilt, denen jeweils ein prototypisches Farbmuster und ein symbolischer Farbname zugeordnet wird. Durch diesen Ansatz wird die Nutzbarkeit des Mediums Farbe über die Grenze zwischen Mensch und Maschine ermöglicht.

Die große Menge an digitaler Information benötigt zur automatischen Verarbeitung eine Rechenleistung, die in den seltensten Fällen bereits zur Verfügung steht. Es ist zu erwarten, dass der technologische Fortschritt in der Zukunft noch leistungsfähigere Rechner hervorbringt, jedoch wird diese Entwicklung mit einer noch größeren Flut an digitaler Information, die es zu bewältigen gilt, einhergehen. Um die Komplexität von Verfahren im Bereich der inhaltsbasierten Suche reduzieren zu können, bedarf es der Erschließung der Topologie eines vorliegenden Datensatzes, um mittels geeigneter Indexstrukturen eine effizientere Suche zu ermöglichen. Dazu wird ein Verfahren vorgestellt, dass durch eine geeignete Auswahl von Elementen des Datensatzes die zu durchsuchende Datenmenge eingrenzt.

Um ein solche Vorgehensweise in einem technischen System realisieren zu können, wird zunächst der Prozess der inhaltsbasierten Informationsverarbeitung in Form eines Modells entworfen werden. Primäres Ziel der Modellbildung ist dabei, die wesentlichen Eigenschaften der visuellen Informationsverarbeitung des Menschen vereinfacht abzubilden. Sekundäres, aber nicht minder wichtiges Ziel ist die Einigung auf eine gemeinsame Sprache bzw. Symbolik in dieser Problemdomäne.

## **1.3 Organisation der Arbeit**

Diese Arbeit befasst sich mit der Entwicklung eines Ansatzes, wie visuelle Inhalte digitaler Bilder mittels merkmalsbasierter Methoden erschlossen und im Rahmen einer Suche nach Information genutzt werden können. In Kapitel 2 wird die Problemstellung zunächst präzisiert. Des Weiteren werden die Besonderheiten aufgezeigt, die merkmalsbasierte Verfahren von herkömmlichen textbasierten Verfahren unterscheiden. Darauf aufbauend wird in Kapitel 3 ein für die merkmalsbasierte Suche allgemeingültiges Modell definiert. Kapitel 4 beschäftigt sich mit der Fragestellung, wie die in einem digitalen Bild vorliegende Farbinformation in einer geeigneten, aus der Sicht des Menschen bedeutungsvollen, Weise erschlossen werden kann. In Kapitel 5 wird aufgezeigt, wie die so gewonnene Farbinformation für die Beschreibung visueller Inhalte mittels verschiedener Merkmale genutzt werden kann. In Kapitel 6 wird ein neuartiger pragmatischer Ansatz zur Indexierung von Merkmalen vorgestellt, der es ermöglicht, die Suche in sehr großen Datensätzen zu beschleunigen. Kapitel 7 schließt die Arbeit mit einer Zusammenfassung und einem kurzen Ausblick ab.



## Kapitel 2

# Inhaltsbasierte Suche

Ein Zugriff auf eine Sammlung von Daten wird im Allgemeinen zum Zweck der Suche nach Information durchgeführt. Die Frage, die es hier zu beantworten gilt, lautet: „Wie werden Daten zur Information?“ Um diese Frage beantworten zu können, müssen zunächst die Begriffe Daten (*engl. data*), Wissen (*engl. knowledge*) und Information (*engl. information*) erläutert werden.

### 2.1 Daten – Wissen – Information

In der Informationswissenschaft hat man sich auf eine einheitliche Terminologie geeinigt, die in Abbildung 2.1 graphisch dargestellt ist. Danach sind Daten als Werte auf syntaktischer Ebene zu verstehen. Nur in Verbindung mit einer anwendungsbezogenen Semantik lässt sich aus Daten Wissen ableiten. In diesem Sinne enthalten also Datenbanksysteme nicht nur Daten, sondern auch Wissen, falls zusätzlich zu den Daten zumindest ein Teil der Semantik des jeweiligen Anwendungsgebietes im System modelliert ist. Information ist schließlich auf pragmatischer Ebene definiert. Information ist die Teilmenge von Wissen, die von einer Person in einer konkreten Situation zur Lösung eines Problem es benötigt wird [86]. Da diese Information häufig nicht vorhanden ist, wird danach in externen Quellen gesucht. Hierzu dient ein Informationssystem, die benötigte Information aus dem gespeicherten Wissen zu extrahieren. Die Transformation von Wissen in Information ist ein Mehrwert erzeugender Prozess.



Abbildung 2.1: Daten, Wissen und Information

Ein wichtiger Unterschied zwischen Information und Wissen besteht darin, dass Information flüchtig ist, Wissen dagegen permanent.

**Beispiel 2.1 (Wissen – Information)** *Der Fahrplan der Deutschen Bundesbahn ist eine Form gespeicherten Wissens über Zugverbindungen. Dann handelt es sich bei der Aussage, dass um 13:15 Uhr, Gleis 9 ein Zug von Frankfurt nach Berlin abfährt um eine Information für denjenigen, der sich zu dieser Zeit in Frankfurt befindet und am Abend in Berlin sein möchte.*

Im Folgenden wird der Begriff „Information“ entsprechend der umgangssprachlichen Verwendung synonym sowohl für Information als auch für Wissen (im Sinne von gespeicherter Information) benutzt.

## 2.2 Information Retrieval

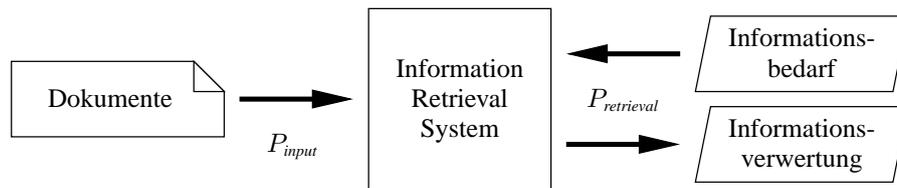
Die Disziplin Information Retrieval (IR) beschäftigt sich allgemein mit dem technisch-gestützten Prozess des Suchens und Findens von relevanter Information in großen Datensammlungen. In den meisten Fällen werden Datenbanksysteme verwendet, um informationstragende Objekte (Dokumente) zu verwalten. Die Datenbanksysteme selbst enthalten dabei die Dokumente nicht notwendigerweise selbst, sondern möglicherweise lediglich kompakte Beschreibungen (Deskriptoren) der Dokumente mit Verweisen auf die Originale. Diese Deskriptoren müssen die Eigenschaften und Inhalte der Dokumente so gut wiedergeben, dass die Dokumente voneinander unterscheidbar bleiben.

Eine Anfrage an ein solches IR-System muss in einer Art und Weise gestellt werden, dass auf der Grundlage dieser Deskriptoren gesucht werden kann. Dazu müssen folgende Prozesse im IR-System realisiert sein:

**Eingabe von Information** (*engl. information input*) – Beim Aufbau der Datenbank müssen für die Dokumente Deskriptoren erzeugt und in geeigneterweise in der Datenbasis gespeichert werden.

**Suche nach Information** (*engl. information retrieval*) – Bei einer Anfrage muss der Informationsbedarf des Nutzenden mit den in der Datenbasis gespeicherten Deskriptoren verglichen werden. Gefundene relevante Dokumente müssen dem Nutzenden zur weiteren Verwertung in geeigneter Form präsentiert werden.

Abbildung 2.2 stellt diese Situation schematisch dar. Während der Eingabe von Information werden Dokumente durch einen Prozess  $P_{input}$  in Deskriptoren überführt, die in dem IR-System gespeichert werden. Eine Anfrage zur Befriedigung eines Informationsbedarfs wird zunächst durch den Prozess  $P_{retrieval}$  in eine Repräsentation überführt,



**Abbildung 2.2:** Schematische Darstellung eines IR-Systems

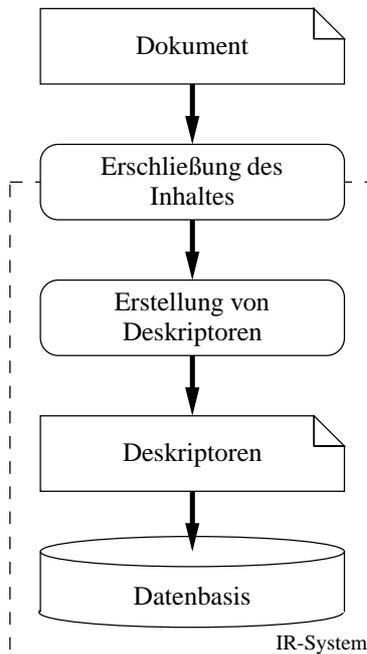
die mit den vorhandenen Deskriptoren der gespeicherten Dokumente verglichen wird. Das Ergebnis dieses Vergleichs wird dem Nutzenden zur weiteren Informationsverwertung dargestellt.

Es ist wichtig, festzustellen, dass in vielerlei Hinsicht Input und Retrieval inverse Prozesse sind und eng miteinander zusammenhängen. Sowohl die Eingabe als auch die Suche nach Information können mit unterschiedlichem Aufwand betrieben werden. Es besteht in vielen Fällen Wahlfreiheit, Probleme bei der Eingabe oder bei der Suche nach Information anzugehen – mit jeweils unterschiedlichen Vor- und Nachteilen. Generell gilt jedoch, dass ein geringerer Aufwand bei der Eingabe von Information keineswegs automatisch eine Einsparung bedeutet, da er in der Regel mit einem höherem Aufwand bei der Suche nach Information verbunden ist. Nicht mehr der Archivar überlegt sich, welche Beschreibungen für ein Dokument relevant sein sollten, sondern der Rechercheur muss sich überlegen, welche Beschreibungen vermutlich in relevanten Dokumenten auftauchen könnten.

### 2.2.1 Eingabe von Information

Der Prozess der Eingabe eines Dokumentes ist in Abbildung 2.3 schematisch dargestellt. Zunächst muss der Inhalt bzw. die in dem Dokument enthaltene Information erschlossen werden. Danach werden mehrere für das Datenbanksystem geeignete Deskriptoren erstellt, die in die Datenbasis des IR-Systems eingefügt werden.

Die besondere Schwierigkeit liegt dabei in der Umwandlung von Information in Daten. Die im Dokument enthaltene semantische Information muss unter Berücksichtigung des konzeptionellen Modells des IR-Systems in syntaktische Werte überführt werden. Dieser komplexe Prozess kann, je nach Grad der Automatisierung des IR-Systems, manuelle Unterstützung des Menschen erfordern. In der Regel ist die in der Datenbasis gespeicherte Informationsmenge wesentlich geringer als die in den ursprünglichen Dokumenten vorhandene, da die im Hinblick auf Kompaktheit erstellten Deskriptoren nur die hervorstechendsten inhaltlichen Aspekte der Dokumente verkörpern. Der Eingabeprozess ist daher bezüglich des Informationsgehaltes zwangsläufig verlustbehaftet.



**Abbildung 2.3:** Schematische Darstellung des Eingabeprozesses

Mit der Eingabe eines Dokumentes sind Kosten verbunden, die proportional zur der benötigten Zeit und den in diesem Zeitraum verwendeten Ressourcen sind. In besonderem Maße wird dabei eine manuelle Inhaltserschließung durch den Menschen kostenintensiv wirksam. Aus ökonomischer Sichtweise stellen die beim Aufbau einer Datenbasis entstehenden Kosten einen kritischen Faktor dar, da zum Zeitpunkt der Dokumenteneingabe noch nicht bekannt ist, ob die gerade eingespeicherte Information zukünftig wieder benötigt wird. Es handelt sich daher um eine Investition in die Datenbasis ohne Rentabilitätsgarantie bezüglich ihres zukünftigen Nutzens. Daher ist es aus dieser Sichtweise prinzipiell wünschenswert, die Kosten der Dokumentenerfassung auf ein Minimum zu reduzieren.

Der Aufwand zur Erstellung von Deskriptoren bei der Eingabe von Dokumenten in eine Datenbasis wird in der Regel pro Dokument genau einmal betrieben (die damit assoziierten Kosten entstehen entsprechend ebenfalls genau einmal). Die Datenbasis ist danach allerdings statisch: Aspekte, die erst zu einem späteren Zeitpunkt relevant werden, sind nicht berücksichtigt. Eine nachträgliche Änderung der Deskriptoren eines IR-Systems hat zur Folge, dass die in der Datenbasis bereits abgespeicherten Deskriptoren möglicherweise unbrauchbar werden (das benutzte Modell muss bei der Erstellung von Deskriptoren schließlich bekannt sein). Die im IR-System gespeicherten Dokumente müssen dann unter Berücksichtigung des geänderten Modells – verbunden mit den entsprechenden Kosten – komplett neu eingepflegt werden.

### 2.2.2 Suche nach Information

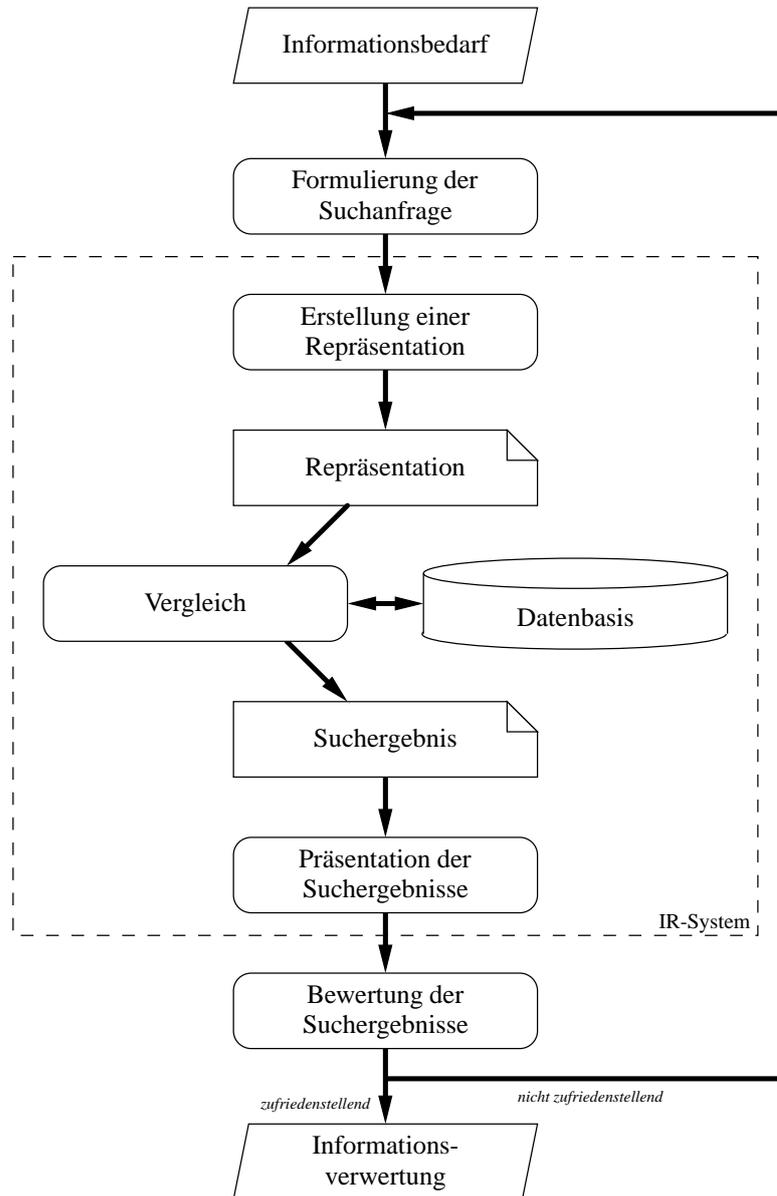
Der Prozess der Suche nach Information mittels eines IR-Systems ist in Abbildung 2.4 schematisch dargestellt. Das grundsätzliche Ziel dabei ist es, den Informationsbedarf eines Benutzers zu befriedigen. Im optimalen Fall bedeutet dies, dass das IR-System auf der Grundlage einer präzisen Suchanfrage ein Suchergebnis produziert, das eine übersichtliche Kollektion aller relevanten Dokumente, die für den Benutzer für die weitere Informationsverwertung und Problemlösung notwendig sind, enthält.

Die Realität unterscheidet sich jedoch bereits dadurch, dass auf Seiten des Benutzers in der Regel nur unsicheres Wissen bezüglich der gesuchten Problemlösung vorhanden und somit nur eine ungenaue Formulierung der Suchanfrage möglich ist („Wie soll man genau wissen, was man nicht weiß?“). Eine strenge Interpretation der vom Benutzer vorgegebenen Suchanfrage ist daher nicht adäquat, da diese eher vage Zielvorstellungen als harte Kriterien beinhaltet. Ein IR-System sollte dementsprechend die Suchanfrage vage interpretieren und im Suchergebnis auch solche Dokumente berücksichtigen, die den Kriterien der Suchanfrage nicht exakt entsprechen, ihnen aber nahe kommen. Dies gilt insbesondere dann, wenn kein einziges Dokument in der Datenbasis existiert, das der Suchanfrage hundertprozentig entspricht.

Oft sind die Probleme des Benutzers und der daraus resultierende Informationsbedarf von so komplexer Art, dass sie sich nur schrittweise präzisieren lassen. Daher stellt sich die Suchanfrage als iterativer Prozess dar, wobei die Bewertung eines Suchergebnisses und die daraus gewonnene Information als Feedback zur Verfeinerung oder Umformulierung der Suchanfrage benutzt wird. Der Benutzer hat somit die Möglichkeit, die Suche interaktiv durch Relevanzbeurteilung von Zwischenergebnissen zielgerichtet zu steuern, ohne a priori eine präzise Aussage über das gesuchte Ziel machen zu müssen.

Da es sich bei einer Suchanfrage um einen iterativen Prozess mit Interaktion des Benutzers handelt, ist die Zeit, die zwischen dem Stellen einer Suchanfrage und dem Bereitstellen des Suchergebnisses vergeht, ein kritischer Faktor. Zu lange Antwortzeiten des IR-Systems frustrieren den Benutzer und haben zur Folge, dass die Motivation, in sukzessiven Schritten nach der gesuchten Information zu suchen, sinkt. Daher ist nicht unbedingt nur die Qualität des Suchergebnisses, sondern auch eine „interaktive“ Antwortzeit für eine erfolgreiche Suche von großer Bedeutung. Zudem sind die bei der Suche eines Dokumentes entstehenden Kosten in erster Linie proportional zu dem für die Recherche benötigten Zeitraum.

Der Retrievalprozess ist im Gegensatz zum Eingabeprozess von seiner Natur her dynamisch und zielgerichtet. Trotzdem kann ein aus Kostengründen unterlassener Aufwand während des Eingabeprozesses oftmals nicht mit einem größeren Aufwand bei der Suche ausgeglichen werden. Die während des Eingabeprozesses geleistete Erschlie-



**Abbildung 2.4:** Schematische Darstellung des Retrievalprozesses

Bearbeitung kann zum Retrievalzeitpunkt auf praktikable Weise nicht gleichwertig rekonstruiert werden.

Die Qualität des Retrievalergebnisses hängt unmittelbar von mehreren Faktoren ab: der Vollständigkeit und Aktualität der Datenbasis, deren formaler und inhaltlicher Aufbereitung und dem Geschick und der Erfahrung des Rechercheurs.

### 2.2.2.1 Suchanfrage

Es ist offensichtlich, dass eine Suchanfrage (*engl. query*) immer nach inhaltlichen Kriterien ausgerichtet ist. Eine Anfrage kann jedoch weitergehend nach ihrer Art klassifiziert werden:

- **semantisch** – „Suche alle Information über Fußball!“
- **logisch** – „Suche alle Resultate des Vereins Eintracht Frankfurt im Jahr 2001!“
- **syntaktisch** – „Suche alle Bilder von Spielern mit schwarz-roten Trikots!“

Dieser Aspekt muss im konzeptionellen Modell berücksichtigt werden; Anfragen können sich nur auf vorhandene Deskriptoren einer entsprechenden Sicht beziehen (d.h. es muss z.B. eine passende syntaktische Sicht der Dokumente in der Datenbasis vorhanden sein, um eine entsprechende syntaktische Suchanfrage beantworten zu können).

Eine besondere Schwierigkeit bei der Formulierung der Suchanfrage liegt in der Tatsache, dass der Informationsbedarf des Benutzers dem System als formalisierte Anfrage übergeben werden muss, die eine ähnliche Struktur wie die in der Datenbasis gespeicherten Deskriptoren aufweist. Dabei muss die Syntax der unterstützten Anfragesprache berücksichtigt werden. Die Suchanfrage wird in der Folge mittels  $P_{retrieval}$  in eine Repräsentation überführt, die mit den in der Datenbasis abgespeicherten Deskriptoren verglichen werden kann. Ist die von  $P_{retrieval}$  generierte Repräsentation identisch mit dem von  $P_{input}$  generierten Deskriptoren, so spricht man von einem symmetrischen Vergleich, d.h. de facto wird die Suchanfrage wie ein zu archivierendes Dokument behandelt. Handelt es sich um einen asymmetrischen Vergleich, so unterscheidet sich die Repräsentation der Suchanfrage von den in der Datenbasis vorhandenen Deskriptoren der Dokumente.

### 2.2.2.2 Suchergebnis

Ein IR-System reagiert im Allgemeinen auf eine gestellte Suchanfrage mit einem Suchergebnis (*engl. result*) als Antwort. Es bestehen dabei grundsätzlich mehrere Möglichkeiten, eine gestellte Suchanfrage zu beantworten:

**Exakte Antwort** (*engl. exact match*) – Die Ergebnismenge beinhaltet alle Dokumente, deren Deskriptoren mit den in der Suchanfrage getroffenen Spezifikationen exakt übereinstimmen.

**Bereichsantwort** (*engl. range match*) – Die Ergebnismenge beinhaltet alle Dokumente, deren Deskriptoren mit der Suchanfrage gemäß einem spezifizierten Kriterium ähnlich sind.

**Nachbarschaftsantwort** (*engl. neighborhood match*) – Die Ergebnismenge beinhaltet eine festgelegte Anzahl von Dokumenten, deren Deskriptoren mit der Suchanfrage gemäß einem spezifizierten Kriterium am ähnlichsten sind.

**Ungefähre Antwort** (*engl. approximate match*) – Ähnlich der Bereichs- bzw. Nachbarschaftsantwort, jedoch ohne Garantie, dass alle Dokumente, die das spezifizierte Kriterium erfüllen, auch wirklich in der Ergebnismenge vorhanden sind. Die ungefähre Antwort wird häufig benutzt, um mit effizienter Ausnutzung von Ressourcen zu einem qualitativ akzeptablen Ergebnis zu kommen.

**Stöbern** (*engl. browsing*) – Oftmals hat ein Benutzer nur äußerst vage Vorstellungen von der von ihm gesuchten Information. In diesem Fall kann er nur sehr unspezifische Suchanfragen stellen. Hier ist es sinnvoll, den Benutzer einen Teil der Dokumentenbasis selbst erschließen zu lassen. Er kann dann solange stöbern, bis er etwas passendes gefunden hat. Das IR-System muss dazu eine Reihe sinnvoller Angebote liefern und insbesondere helfen, dass der Benutzer im Ergebnisraum navigieren kann.

Die Präsentation der Suchergebnisse ist ein wichtiger Teil des Retrievalprozesses, da der Benutzer aus dieser Darstellung unmittelbar seine Bewertung ableitet (siehe Abbildung 2.4). Ein Problem dabei ist die Auswahl einer vernünftigen Anzahl von Dokumenten, die dem Benutzer einen Überblick über interessant erscheinende Information geben soll. Eine zu kleine Zahl birgt die Gefahr, dass ausgerechnet das Dokument mit der maßgeblichen Information nicht mehr in der Ergebnismenge erscheint, eine zu große Zahl, dass der Benutzer den Überblick verliert.

### 2.2.3 Retrieval Modelle

IR-Systeme lassen sich generell danach unterscheiden, nach welchem Modell relevante Dokumente für die Ergebnismenge als Antwort auf eine Suchanfrage bestimmt werden. Folgende Ansätze sind hierbei hervorzuheben:

**Boole'sche Modelle** basieren auf dem Ansatz, dass Suchanfragen aus regulären Ausdrücken natursprachlicher Terme bestehen. Das IR-System liefert als Ergebnis

alle Dokumente, deren Deskriptoren mit dem regulären Ausdruck exakt übereinstimmen. Typischerweise steht bei der Konstruktion einer Suchanfrage die Verwendung Boole'scher Operatoren zur Verfügung. Der Vorteil dieser Methode liegt darin, dass sie einfach zu realisieren ist. Der Benutzer kann zudem unmittelbar beurteilen, ob und warum ein Dokument zur Ergebnismenge gehört oder nicht. Nachteilig ist, dass die Qualität der Suchergebnisse unmittelbar mit dem bei der Dokumenteingabe betriebenen Aufwand zusammenhängt. Außerdem erfolgt keine Reihung der Dokumente bezüglich ihrer Relevanz innerhalb der Ergebnismenge. Die Anzahl der Treffer auf eine Suchanfrage lässt sich vorher nur schwer abschätzen.

**Metrische Modelle** basieren auf dem Ansatz, dass sowohl Dokumente als auch Suchanfragen als Punkte in einem metrischen Raum aufgefasst werden. Als Grundlage für die Suche dient ein in diesem Raum definiertes Distanzmaß, mit dem die Ähnlichkeit zweier Punkte bestimmt werden kann. Der Vorteil dieses Ansatzes liegt in der Kombination und Gewichtung verschiedener Anfrageterme. Die Dokumente der Ergebnismenge lassen sich weiterhin gemäß ihrer Ähnlichkeit ordnen. Daher ist eine Rangfolgeerstellung der Dokumente bezüglich ihrer Relevanz möglich.

**Probabilistische Modelle** basieren auf statistischen Methoden. Die Wichtigkeit eines Dokumentes für eine Suchanfrage wird über die Wahrscheinlichkeit der Relevanz eines Dokumentes gemessen. Das Hauptproblem hierbei ist, dass eine große Anzahl von Daten gesammelt werden muss, um eine statistisch signifikante Aussage treffen zu können.

Die Wahl des Modells beeinflusst maßgeblich den Charakter der Dokumentdeskriptoren im IR-System. So müssen z.B. Deskriptoren des Boole'schen Modells auf Identität, Deskriptoren des Metrischen Modells dagegen auf Ähnlichkeit verglichen werden.

Verschiedene Modelle lassen sich grundsätzlich in einem IR-System kombinieren, soweit die entsprechenden Dokumentdeskriptoren in der Datenbasis vorhanden sind. Jedoch findet in der Regel keine echte, integrierte Kombination statt, sondern lediglich eine sukzessive Verkettung verschiedener Verfahren. Häufig werden Verfahren nach dem Boole'schen Retrievalmodell zur a priori Einschränkung des Suchraumes (im Sinne einer Filterung der Datenbasis) eingesetzt; ein Dokument der Datenbasis wird in folgenden Suchverfahren nur berücksichtigt, wenn es dem regulären Ausdruck natürlicherer Terme der vorgeschalteten Boole'schen Suchanfrage entsprochen hat.

## 2.3 Visual Information Retrieval

Die Disziplin Visual Information Retrieval (VIR) ist die Erweiterung des klassischen IR auf visuelle Medien.

Das Auge ist das wichtigste Sinnesorgan des Menschen. Physiologisch betrachtet liegen 70% aller Sinnesrezeptoren des menschlichen Körpers im Auge. In der Tat bewertet und versteht der Mensch seine Umwelt hauptsächlich dadurch, dass er sie sieht. Weit mehr als 80% aller Information, die im menschlichen Gedächtnis gespeichert sind, werden über das Auge aufgenommen [126].

Die Umwelt wird dabei als zweidimensionale Abbildung von Lichtintensitäten auf der Netzhaut des Auges wahrgenommen, die als Farben erkannt werden. Durch eine massive Parallelverarbeitung im nachgeschalteten Nervennetzwerk werden die optischen Reize in Zusammenhang mit der individuellen intellektuellen Erfahrung des Menschen durch assoziative Bindungen formal und räumlich zu einem Vorstellungsbild weiterverarbeitet. Daher besitzt der Mensch die Fähigkeit zum kognitiven Sehen.

Auf den ersten Blick erscheint die Suche nach visueller Information für den Menschen trügerisch einfach. Jedoch sind Bilder als Dokumente eines IR-Systems in ihrer ursprünglichen Form für die Suche nicht verwendbar, da die Bilder in Form einer abstrakten Datenrepräsentation eines von einem technischen Sensor produzierten Signals vorliegen. Bilddaten müssen zunächst dekodiert und in geeigneter Weise dargestellt werden, bevor der Bildinhalt vom Menschen wahrgenommen werden kann. Die Bilddaten sind somit nicht direkt für den Zugriff auf visuelle Information nutzbar. Für die inhaltliche Suche sind daher zusätzliche Zugriffswege auf die in den Bildern enthaltene Information anzulegen. In bisherigen Forschungsarbeiten kann man zwei unterschiedliche Arten der Erschließung von visuellen Inhalten aus Bildern für VIR finden: text- und merkmalsbasierte Verfahren. Beide Ansätze werden in den nächsten Abschnitten kurz vorgestellt und diskutiert.

### 2.3.1 Textbasierte Verfahren

Traditionell wird der Inhalt eines Bildes durch einen alphanumerischen Text beschrieben, der zu einem Bild hinzugefügt wird. Die Verwendung menschlicher Sprache bei der Verfassung dieser Texte ist dabei die unvermeidliche Grundlage [147] und erfolgt in der Regel durch die Interpretation des visuellen Bildinhaltes durch einen individuellen Betrachter. Die Beschreibung von Bildern in natürlicher Sprache kommt der Art der menschlichen Kommunikation sehr nahe, jedoch birgt sie einige Gefahren, die im Folgenden erläutert werden.

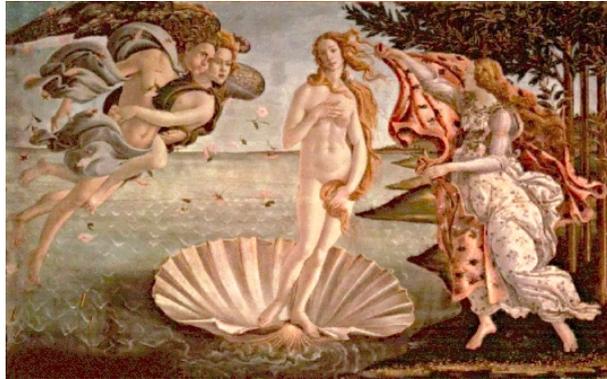
### 2.3.1.1 Freier Volltext

Die naheliegendste Form der Erschließung von Bildern ist die Verwendung von natur-sprachlichen Beschreibungstexten ohne Einschränkung des Vokabulars. Dieses Vorgehensweise entspricht der natürlichen Art und Weise, wie Information zwischen Menschen verbal ausgetauscht wird. Grundlegende Voraussetzung hierbei ist, dass alle, die an dem Informationsaustausch beteiligt sind, die verwendete Sprache beherrschen. Die Problematik von freiem Volltext entsteht in erster Linie dadurch, dass zwischen den Personen, die die Beschreibungen bei der Archivierung der Bilder erstellt haben, und denen, die die Suchanfrage formulieren, in der Regel kein offener Kommunikationskanal existiert. Unklarheiten im Beschreibungstext können daher nachträglich nicht beseitigt werden.

Textbasierte Verfahren basieren auf dem Boole'schen Retrievalmodell (siehe 2.2.3) und damit auf der Suche von exakten Übereinstimmungen von Zeichenfolgen in den Termen der Suchanfrage und den archivierten Bildbeschreibungen. Der Erfolg einer Suchanfrage hängt im Wesentlichen von der Kongruenz des Vokabulars von Archivar und Rechercheur ab. Diese Problematik wird verschärft, falls für einen Begriff anstelle eines geeigneten Wortes Umschreibungen verwendet werden. Häufige Ursache von Missverständnissen ist auch die Verwendung von Homonymen (eine Bedeutung bezeichnet mehrere Begriffe) und Synonymen (ein Begriff hat mehrere Bedeutungen). Zudem führen die Verwendung unterschiedlicher grammatikalischer Strukturen und Formulierungen in der formalisierten Suchanfrage und den Beschreibungstexten zu weiteren Problemen mit dem Boole'schen Retrievalmodell, da Begriffe nicht in eindeutigen Standard-Wortformen auftreten.

Freier Volltext beinhaltet einen hohen Grad an visueller Abstraktion und bezeichnet die im Bildinhalt manifestierten logischen und semantischen Zusammenhänge. Es liegt allein im Ermessen des Betrachters, welche inhaltlichen Aspekte wie erschlossen werden. Nicht jeder Aspekt eines Bildes wird vom Archivar während der Erstellung der Beschreibung berücksichtigt. Zudem sind einige Aspekte eines Bildes nur schwer in Worten zu beschreiben oder sie können durch unterschiedliche Beschreibungen gleich gut dargestellt werden. Folglich sind Szenen- und Objektbeschreibungen hochgradig abhängig von der subjektiven Interpretation des individuellen Betrachters und werden von vielen Faktoren, wie z.B. dem sozialen Umfeld, dem Bildungsstand und dem Hintergrundwissen beeinflusst [34]. Besitzt eine nach einem bestimmten Bild suchende Person nicht ähnliche Voraussetzungen wie die Person, die das Bild archiviert hat, so sind Übereinstimmungen zwischen Bildbeschreibung und der Suchanfrage eher zufällig, und der Erfolg der Suchanfrage ist gering. Syntaktische Bildinhalte finden sich nur selten in diesen Beschreibungen wieder. Diese Sachverhalte werden in Beispiel 2.2 verdeutlicht. Weitergehende Studien sind in [156] und [18] zu finden.

**Beispiel 2.2 (Freier Volltext)** *Das Gemälde von Sandro Botticelli (siehe Abbildung 2.5) wurde unter dem Namen „Die Geburt der Venus“ berühmt.*



**Abbildung 2.5:** Sandro Botticellis „Die Geburt der Venus“

*Generell ist es dem Betrachter möglich, den Inhalt des Bildes wiederzugeben. Eine Beschreibung könnte etwa wie folgt lauten:*

*„Im Bild sind vier Personen zu sehen. In der Mitte steht eine unbekleidete Frau, die ihre Scham bedeckt, aufrecht auf einer riesigen Muschel, die ihr offenbar als seetüchtiges Gefährt auf dem Meer dient. In der linken Bildhälfte schweben zwei leicht bekleidete Personen heran. Die zuvorderst platzierte Gestalt ist männlichen, die andere weiblichen Geschlechts. Mit aufgeblasenen Backen erzeugen sie den Wind, dessen Kraft das Gefährt der Frau auf die Küste zutreiben lässt. Auf der anderen, dem Land zugewandten Seite des Bildes, erwartet eine weiß gekleidete weibliche Gestalt die Ankunft der Frau.“*

*Ein Kunstkennner mit der nötigen Sachkenntnis über die dem Gemälde zugrundeliegenden literarischen Quellen würde jedoch das Bild etwa wie folgt beschreiben:*

*„Im Bild sind vier Figuren der Mythologie dargestellt. Die nackte Frau in der Mitte stellt Venus, die aus der blauen Tiefe der Meeres geborene und von den schäumenden Fluten herangezogene Göttin der Fruchtbarkeit, dar. In der linken Bildhälfte sind Zephyr und Aura, die Personifikationen eines wärmenden Frühlingswindes und einer sanften Brise, zu sehen. Rechts im Bild, am Ufer der Insel Zypern, fiebert die Nymphe Pomona, eine Nachkommin der Göttin der Fruchtbäume, sehnsüchtig die Ankunft der Venus herbei, die dort Blumen erblühen und die Vegetation zum Leben erwachen soll.“*

*Die beiden Beschreibungen sind von ihrer Art und Weise, sowie von der Bedeutung her völlig unterschiedlich. Dies zeigt, dass die Interpretation von visuellem Inhalt durch individuelle Betrachter von einer Anzahl externer Faktoren abhängig ist. Vielleicht*

*wurde das Bild nur mit der Maßgabe geschaffen, seine Auftraggeber durch seine Schönheit der Farben und Gestaltung zu erfreuen. Eine Eigenschaft, die sich in keiner der obigen Beschreibungen nur ansatzweise wiederfindet.*

### 2.3.1.2 Verschlagwortung

Eine Alternative zur Verwendung von freiem Text stellt die Beschreibung des Bildinhaltes durch Schlagwörter (*engl. keywords*) dar. Ein sachverständiger Bearbeiter wählt eine beliebige Anzahl von Schlagwörtern aus, die den Inhalt des Bildes aus seiner Sichtweise wiedergeben. Die Verschlagwortung ist im Vergleich zum freien Volltext generell weniger fehleranfällig, da sie nicht der Syntax der natürlichen Sprache unterliegt und keine individuelle Kreativität beim Formulieren von Sachverhalten erlaubt (Schlagwörter stehen nur in eindeutigen Standard-Wortformen zur Verfügung).

Problematisch ist allein das unbegrenzte Vokabular bei der Verschlagwortung. Die Freiheit wenig hilfreichen Einfallsreichtums bei der Auswahl von Schlagwörtern verursacht eine erhebliche Unsicherheit auf seiten der Recherche: Wie soll man bei dieser Bandbreite die im Einzelfall hilfreichen Suchbegriffe erraten können? Abhilfe schafft hier die Kontrolle der zu vergebenden Schlagwörter mittels Vorgabe eines verbindlichen Schlagwortkataloges. Technisch gesehen selektiert der Archivar nur noch eine beliebige Anzahl von Schlagwörtern aus einer vorgegebenen Menge.

Der für ein bestimmtes Bildarchiv verwendete Schlagwortkatalog wird dabei a priori so ausgewählt, dass er die gesamte Themenbreite aller Bilder für die entsprechende Domäne eines bestimmten Klientels abdeckt. Oftmals werden die Schlagwörter im Katalog zusätzlich in einer hierarchischen Baumstruktur angeordnet, so dass aus ihrer Anordnung zwischen Ober- bzw. Unterbegriffen unterschieden werden kann. Der Vorteil liegt eindeutig in dem von allen Beteiligten verwendeten identischen Vokabulars. Da zur Archivierung und Recherche dieselbe Schlagwortliste zur Verfügung steht, ist eine spätere Suche deutlich zuverlässiger.

Der Nachteil der kontrollierten Verschlagwortung liegt auf der Hand: die Auswahl an Schlagwörtern ist eingeschränkt, der Schlagwortkatalog unvollständig. Oftmals steht genau das Schlagwort, das den Bildinhalt optimal charakterisieren würde, nicht zur Verfügung. Eine undifferenzierte Erweiterung der Schlagwortliste führt paradoxerweise wieder zu einem unzuverlässigerem Retrievalergebnis. Denn werden zu viele Schlagwörter angeboten, leidet die Übersichtlichkeit, und für ein und denselben Sachverhalt gibt es zunehmend mehr in Frage kommende Schlagwörter. Die Auswahl aus zunehmend großen Schlagwortkatalogen hängt sehr von der Sorgfalt des Nutzers ab und ist daher oft auch vom Zufall geprägt.

Der Schlagwortkatalog ist aus verständlichen Gründen immer auf den thematischen Inhalt bzw. auf das spezifische Benutzerklientel der entsprechenden zugehörigen Bildkollektion zugeschnitten. Daher haben unterschiedliche Bildkollektionen in der Regel unterschiedliche Verschlagwortungen und lassen sich – wenn überhaupt – nur mit erheblichem Aufwand fusionieren. In den letzten Jahren hat es einige Anstrengungen gegeben, einen generischen Schlagwortkatalog zu entwerfen und zum Standard zu erheben [56]. Als Beispiele sind hier der *Getty's Art and Architecture Thesaurus* (AAT) und der *Library of Congress Thesaurus of Graphical Material* (LCTGM) zu nennen. Der AAT besteht beispielsweise aus insgesamt 120000 verschiedenen Begriffen, die zur Beschreibung von Bildern aus den Bereichen Kunst, Kultur, Geschichte und Architektur herangezogen werden können. Es hat sich aber gezeigt, dass sich ein standardisierter Schlagwortkatalog in einer vertretbaren Größenordnung nur im Hinblick auf eine bestimmte thematische Domäne realisieren lässt.

**Beispiel 2.3 (Verschlagwortung)** *Die Getty Images<sup>1</sup> Bildkollektion ermöglicht eine Bildsuche ausschließlich nach Schlagworten.*



**Abbildung 2.6:** Bild der Getty Images Bildkollektion

*Das in Abbildung 2.6 dargestellte Bild ist beispielsweise mit den Schlagworten*

*„nur eine Frau, Mountainbike, junger Erwachsener, Ganzansicht, Radfahren, Europäer, Menschen, eine Person, nur eine junge Frau, Sport, Fahrrad, Extremsport, Tag, Außenaufnahme, horizontal, Muße, Fotografie, Sport & Fitness, Farbe, Verkehr“*

*katalogisiert. Bei den Schlagwörtern „horizontal“, „Fotografie“ und „Farbe“ handelt es sich um formale Deskriptoren, die im eigentlichen Sinne keinen inhaltlichen Aspekt des Bildes wiedergeben und als Ordnungskriterien verwendet werden. Andere Schlagworte, wie z.B. „Extremsport“, „Muße“ und „Verkehr“, versuchen in dem*

<sup>1</sup>[www.gettyimages.de](http://www.gettyimages.de)

*Bild vorhandene semantische Konzepte zu vermitteln. Sie beschreiben den Bildinhalt nur vage und vermitteln in erster Linie einen subjektiven Eindruck des Archivars, der nicht immer konsistent mit der Erfassung eines anderen Betrachters ist. Schlagworte, wie z.B. „Ganzansicht“, „Tag“ und „Außenaufnahme“ versuchen den Bildinhalt auf syntaktischer Ebene zu charakterisieren – von einer detaillierten Beschreibung der im Bild enthaltenen Darstellung von Personen und Objekten der Bildszene kann allerdings nicht die Rede sein.*

*Auffällig ist, dass der zugrundeliegende Schlagwortkatalog in manchen Bereichen unerwartet detailliert aufgebaut ist (z.B. „Fahrrad“ und „Mountainbike“, „nur eine Frau“ und „nur eine junge Frau“). Daher liegt die Schlussfolgerung nahe, dass eine große Anzahl relevanter Bilder mit dieser speziellen Thematik in der Bildkollektion vorhanden ist, zwischen denen mit einer differenzierteren Auswahl an Schlagwörtern diskriminiert werden muss.*

Das System der Verschlagwortung ist in großem Maße unflexibel, da die Auswahl der Schlagworte a priori erfolgen muss und nachträglich nur mit großem Aufwand geändert werden kann. Trotzdem hat sich die Verschlagwortung in der Vergangenheit aufgrund mangelnder Alternativen und ihrer Vorteile gegenüber freiem Volltext durchgesetzt und ist heute Standard bei der kommerziellen Archivierung digitalen Bildmaterials.

Die Entwicklung von Methoden und Technologien zur automatischen Erkennung von semantischen Sachverhalten in Bildern durch natursprachliche Texte, steckt noch in den Anfängen. Die Forschung hat sich in der Vergangenheit darauf beschränkt, detaillierte Probleme zu lösen, die nur für einen eingeschränkten Bereich von Bildern mit bestimmten Eigenschaften gelten. Aus der Sicht eines Benutzers ist jedoch ein uneingeschränkter Zugriff auf vorhandenes Bildmaterial wünschenswert, so dass solche In-sellösungen nur wenig hilfreich sind. Trotz einer zu erwartenden rasanten Entwicklung im Bereich Mustererkennung und Computer Vision werden solche Methoden mittelfristig nicht zur Verfügung stehen.

### **2.3.2 Merkmalsbasierte Verfahren**

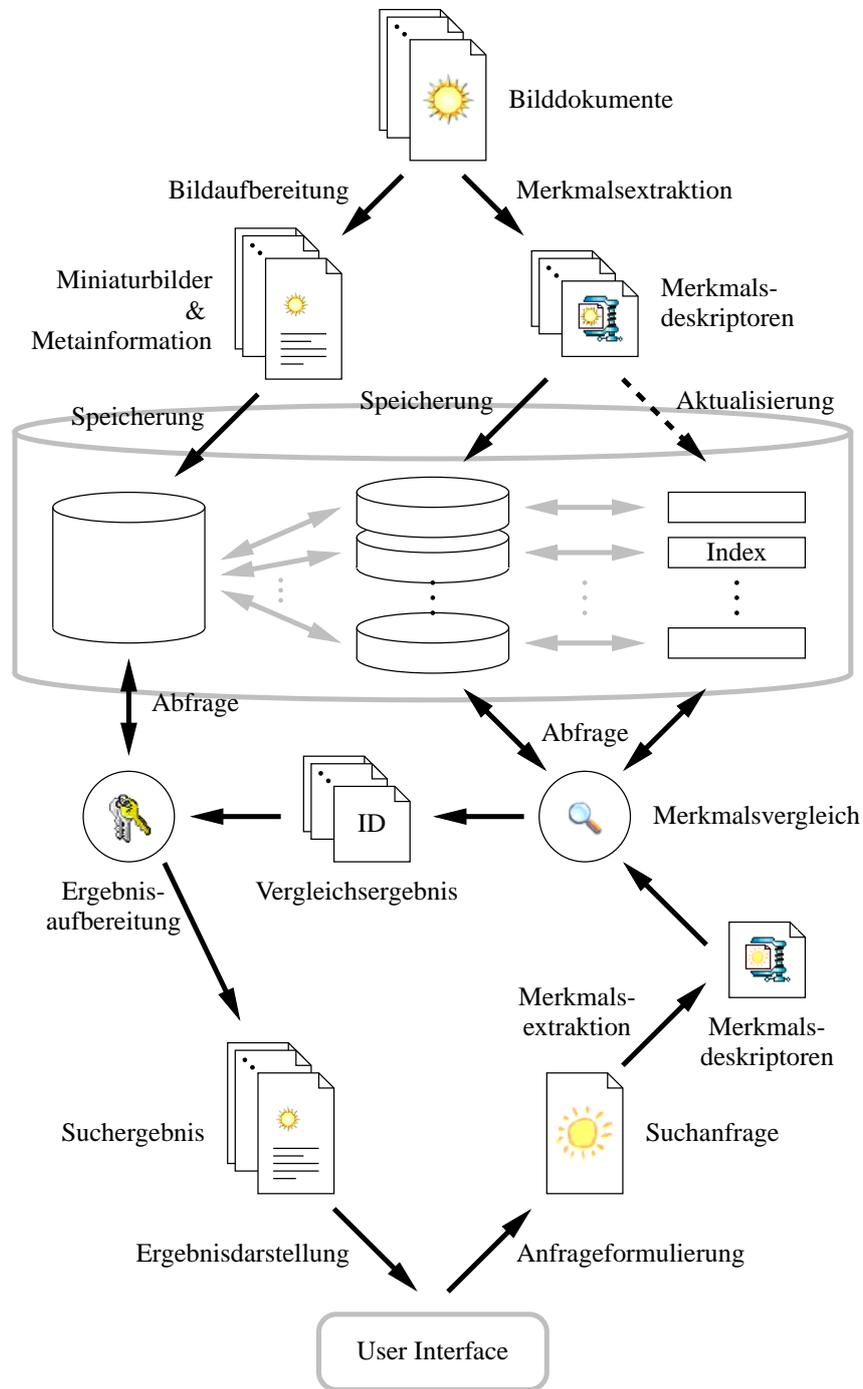
Die genannten Probleme mit textbasierten Verfahren haben zu einem erhöhten Interesse an der Entwicklung von merkmalsbasierten Verfahren (*engl. content-based image retrieval – CBIR*) geführt. Im Gegensatz zu herkömmlichen Verfahren des IR steht bei CBIR die automatische und unbeaufsichtigte Gewinnung von objektiven Beschreibungen des visuellen Inhaltes direkt aus den digitalen Bilddaten im Vordergrund, die zur weiteren Erschließung der Bildinformation herangezogen werden.

Bildarchive sind in erster Linie unstrukturiert, da digitalisierte Bilder aus einer Anordnung diskreter Bildpunkte mit zugeordneten Farbintensitäten bestehen, denen keine unmittelbare Bedeutung zugeordnet werden kann. Daher unterscheiden sich Bildarchive grundsätzlich von Textarchiven, wo die Rohdaten in Form von alphanumerischen Texten bereits logisch strukturiert vorliegen [131].

Die grundlegende Philosophie merkmalsbasierter Verfahren ist die Beschreibung des Bildinhaltes auf der Grundlage visueller Merkmale (siehe 5). Die für ein bestimmtes Bild aus einem Merkmal abgeleitete Beschreibung des Bildinhaltes wird im Allgemeinen als Merkmalsdeskriptor bezeichnet. Die Merkmalsdeskriptoren werden mit Hilfe geeigneter Algorithmen der Digitalen Bildverarbeitung, Mustererkennung, Computer Vision und der Künstlichen Intelligenz durch Analyse der Eigenschaften und Relationen der diskreten Bildpunkte direkt aus den digitalen Daten eines Bildes gewonnen. Aufgrund der subjektiven Wahrnehmung von optischen Eigenschaften in verschiedenen inhaltlichen Kontexten kann jedoch kein einzelnes, gesamtheitliches Merkmal des visuellen Bildinhaltes gefunden werden, das allen Anforderungen gerecht wird. Daher existieren mehrere Möglichkeiten, unterschiedliche Aspekte eines Bildes aus verschiedenen Gesichtspunkten in einem Merkmalsdeskriptor zu repräsentieren.

Die prinzipielle Funktionsweise merkmalsbasierter Verfahren des VIR ist in Abbildung 2.7 schematisch dargestellt. Das klassische IR-Paradigma (siehe 2.2) erfordert es, dass für jedes Bild, das der Datenbasis des CBIR-Systems hinzugefügt wird, eine Menge von Deskriptoren vorberechnet wird. In der Regel werden pro Bild mehrere Merkmalsdeskriptoren erzeugt, die jeweils unabhängig voneinander ein unterschiedliches visuelles Merkmal des Bildinhaltes repräsentieren. Dies ermöglicht es dem Benutzer später, eine differenzierte Auswahl der für seine Suchanfrage wichtigen Merkmale zu treffen. Parallel dazu wird ein Miniaturbild erzeugt und zusammen mit den Merkmalsdeskriptoren und sonstiger Metainformation in der Datenbasis gespeichert.

Der Benutzer übergibt dem CBIR-System seinen Informationsbedarf in Form einer Suchanfrage. Das CBIR-System extrahiert aus der Suchanfrage die ihm bekannten Merkmale in Form von Merkmalsdeskriptoren und berechnet die Ähnlichkeit zwischen den Merkmalen der Suchanfrage und den Merkmalen aller Bildern der Datenbasis durch sukzessiven Vergleich mit den entsprechenden Merkmalsdeskriptoren. Als Ergebnis werden die Bilder der Datenbasis anschließend in einer Rangfolgeliste gemäß der berechneten Ähnlichkeiten geordnet. Zur Ergebnismenge gehören prinzipiell alle Bilder der Datenbasis. Aus praktischen Gründen der Übersichtlichkeit wird dem Benutzer jedoch nur ein kleiner Ausschnitt der Ergebnismenge präsentiert; der Benutzer bekommt die ähnlichsten Bilder zu seiner Suchanfrage in Form einer Liste von Miniaturbildern dargestellt. Er erhält daher immer ein Ergebnis, auch wenn im eigentlichen Sinne kein ähnliches Bild in der Datenbasis existiert.



**Abbildung 2.7:** Prinzipielle Funktionsweise merkmalsbasierter Verfahren des VIR

Mit der Bildarchivierung und -suche mittels merkmalsbasierter Verfahren sind Kosten assoziiert, die jeweils proportional zu der benötigten Zeit und den verwendeten Ressourcen sind. Die mit der Zeit zur Erfassung des Bildmaterials in der Datenbasis verbundenen Kosten können vernachlässigt werden, da es sich bei merkmalsbasierten Verfahren um automatische Methoden handelt, die ohne menschliche Intervention offline verarbeitet werden. Allein der für die Merkmalsdeskriptoren benötigte Speicherplatz in der Datenbasis ist ein Kostenfaktor. Die mit der Zeit für die Suche nach Bildmaterial verbundenen Kosten müssen jedoch hingegen – obwohl ebenfalls automatisiert – berücksichtigt werden. Da es sich bei der Suche um einen interaktiven vom Benutzer zielgerichteten, iterativen Prozess handelt, muss der Benutzer während des gesamten Zeitraumes online sein und verursacht so Kosten. Der Zeitraum, den der Benutzer nach der Formulierung der Suchanfrage bis zur Darstellung des Suchergebnisses durch das CBIR-System passiv wartend verbringt, sollte daher so klein wie möglich gehalten werden (schon aus psychologischen Gründen – siehe 2.2.2). Während dieser Zeit muss für jedes Merkmal genau einmal der Merkmalsdeskriptor berechnet werden. Der sukzessive Vergleich des Merkmalsdeskriptors der Suchanfrage mit allen Merkmalsdeskriptoren der Datenbasis schlägt proportional mit der Anzahl der gespeicherten Bildern zu Buche.

### 2.3.2.1 Merkmale

Da Beschreibungen visueller Merkmale eines Bildes grundsätzlich in großer Anzahl und mannigfaltiger Ausprägung existieren können, wurde in [139] folgende Kategorisierung vorgeschlagen:

**Inhaltsunabhängige Merkmale** (*engl. content-independent features*) – Die Merkmale beziehen sich nicht unmittelbar auf den visuellen Inhalt, stehen aber in Zusammenhang mit dem Bild. Beispiele solcher Merkmale sind:

- Bildformat (Quer-/Hochformat)
- Urheber und Eigentümer
- Ort und Datum der Aufnahme
- Nutzungsbestimmungen

**Inhaltsabhängige Merkmale** (*engl. content-dependent features*) – Die Merkmale beziehen sich unmittelbar auf den visuellen Inhalt des Bildes. Dies kann grundsätzlich auf zwei verschiedenen Ebenen erfolgen:

**Syntaktische Merkmale** (*engl. low-level features*) – Die Merkmale beschreiben visuelle Aspekte des Bildes aus syntaktischer Sicht. Beispiele solcher syntaktischer Merkmale sind:

- Farbe
- Textur
- Form
- räumliche Anordnung

**Semantische Merkmale** (*engl. high-level features*) – Die Merkmale beschreiben konzeptionelle Aspekte des Bildes aus semantischer Sicht. Beispiele solcher semantischer Merkmale sind:

- Abstraktionen (Objekte/Personen/Szenen)
- Bedeutungen
- Emotionen und Eindrücke
- Beziehungen zur Realität

Für eine Suchanfrage nach inhaltlichen Kriterien können offenkundig allein inhaltsabhängige Merkmale herangezogen werden können. Inhaltsunabhängige Merkmale werden in der Regel für die eigentliche Suche nur unterstützend zur Filterung herangezogen (z.B. ein Suchergebnis soll nur Bilder im Querformat enthalten).

Eine – aus Sicht des Benutzers – wünschenswerte vollautomatische und umfassende Identifikation von semantischen Merkmalen wird in absehbarer Zeit technologisch nicht realisierbar sein. Die internationale Forschungsgemeinschaft hat gerade erst damit begonnen, im Bereich der automatischen Extraktion semantischer Primitive (z.B. Erkennung von einfachen Objekten [165], Erkennung von Gesichtern [176]) erste nennenswerte Erfolge zu erzielen. Es ist jedoch noch einen langer Weg, bevor das Ziel einer umfassenden semantischen Analyse von Bildmaterial mit uneingeschränktem Inhalt erreicht sein wird [36]. In der Zwischenzeit wird man sich auf merkmalsbasierte Verfahren auf der Basis von syntaktischen Merkmalen zur Lösung der Probleme im Bereich CBIR konzentrieren müssen.

Aus den theoretischen Voraussetzungen und den Anforderungen eines CBIR-Systems lassen sich allgemeine Anforderungen an die Eigenschaften von Merkmalen und Merkmalsdeskriptoren ableiten:

**Objektivität** – Die in einem Merkmal verkörperten Eigenschaften des visuellen Inhaltes sind im Bild objektiv messbar. Der aus einer solchen Messung resultierende Merkmalsdeskriptor ist jederzeit ohne zusätzliches Wissen deterministisch reproduzierbar.

**Konnexität** – Das Bild und der aus dem Bild gewonnene Merkmalsdeskriptor sind inhärent verbunden.

**Diskriminanz** – Der aus dem Bild gewonnene Merkmalsdeskriptor macht das Bild von allen anderen Bildern unterscheidbar.

**Vergleichbarkeit** – Die Merkmalsdeskriptoren zweier Bilder sind qualitativ und quantitativ objektiv vergleichbar. Das Vergleichsergebnis repräsentiert die Ähnlichkeit der beiden Bilder bezüglich der in den Merkmalsdeskriptoren verkörperten visuellen Eigenschaften der Bilder.

**Kompaktheit** – Die Datenrepräsentationen der Merkmalsdeskriptoren sind kompakt.

Die Forderung nach kompakten Merkmalsdeskriptoren ist im Wesentlichen durch zwei Aspekte motiviert: Erstens sollte der Bedarf an zusätzlichem Speicher in der Datenbasis für die Merkmalsdeskriptoren so gering wie möglich gehalten werden, und zweitens sollte die Komplexität eines Vergleichs zweier Merkmalsdeskriptoren aus zeitlichen Gründen ebenfalls möglichst gering gehalten werden. In der Regel bildet die Merkmalsextraktion dazu die daten-intensive Bilddomäne in einen daten-kompakten Merkmalsraum ab. Diese Abbildung ist zwangsweise mit einem Informationsverlust verbunden. In einem entsprechend kompakten Merkmalsraum können unterschiedliche Bilder prinzipiell den gleichen Merkmalsdeskriptor zugewiesen bekommen. Dieser Sachverhalt steht jedoch im Widerspruch zum postulierten Diskriminanzprinzip.

**Beispiel 2.4 (Diskriminanz vs. Kompaktheit)** *Die mittlere Helligkeit eines Bildes kann als Merkmal aufgefasst werden. Die Bilddomäne wird dabei auf einen äußerst kompakten Merkmalsraum – respektive einen skalaren Wert – abgebildet. Offensichtlich ist dieses Merkmal nicht dazu geeignet, zwischen einer großen Anzahl von Bildern zu diskriminieren.*

Bei der Entwicklung eines Algorithmus zur Merkmalsextraktion ist daher ein Kompromiss zwischen Diskriminanz und Kompaktheit der Merkmalsdeskriptoren zu finden.

### 2.3.2.2 Ähnlichkeit

Die Bewertung von wahrnehmungsorientierter Ähnlichkeit kann unterscheiden werden, ob sie ohne menschliche Fokussierung (*engl. pre-attentative*) oder mit menschlicher Fokussierung (*engl. attentive*) erfolgt [130]. Im ersten Fall erfolgt die Beurteilung der Ähnlichkeit von Bildern ohne, im zweiten Fall mit Miteinbeziehung des menschlichen kognitiven Erkennungsapparates. Die Möglichkeit zum kognitiven Sehen ermöglicht es dem Menschen, im Zusammenhang mit seinem individuellen Wissen den Bildinhalt zu interpretieren. Daher kann der Mensch seine Betrachtung auf wichtige Bereiche des Bildes fokussieren. Geringe Abweichungen reichen in wichtigen Bereichen des Bildes aus, damit solche Bilder als überdurchschnittlich unähnlich

beurteilt werden, während starke Abweichungen in unwichtigen Bereichen zu Folge haben, dass solche Bilder trotzdem als überdurchschnittlich ähnlich betrachtet werden. CBIR-Verfahren basieren auf der Beurteilung von Ähnlichkeit ohne Fokussierung, da die genaue Funktionsweise des kognitiven Sehapparates des Menschen weitestgehend unerforscht ist. Als Konsequenz daraus kann sich die Beurteilung von Ähnlichkeit nur an den im Bild vorhandenen Reizen orientieren, wobei jeder im Bild vorhandene Reiz gleiche Relevanz hat.

Es gibt verschiedene Ansätze, das Ausmaß der Ähnlichkeit bezüglich eines visuellen Merkmals zweier Bilder zu messen. Das wohl bedeutenste und renommierteste Konzept ist das metrische Modell [132]. Das metrische Modell basiert auf der Annahme, dass die menschliche Wahrnehmung von visueller Ähnlichkeit näherungsweise einem psychologischen Modell korrespondiert, das einem metrischen Raum entspricht. Folglich modelliert ein Merkmalsdeskriptor dann die visuellen Eigenschaften eines Bildes in einer Art und Weise, dass er als Punkt im korrespondierenden Merkmalsraum aufgefasst werden kann. Die Ähnlichkeit zweier Bilder bestimmt sich dann aus der messbaren Distanz zwischen den beiden korrespondierenden Punkten im Merkmalsraum.

Das metrische Ähnlichkeitsmodell weist vom Standpunkt der Psychologie einige Unzulänglichkeiten auf [157]. So entspricht z.B. die menschliche Beurteilung der Ähnlichkeit von Formen und Gestalten nachweislich nicht dem metrischen Modell. Trotzdem besitzt das metrische Modell eine Menge Vorteile, die für eine Anwendung in CBIR-Systemen sprechen:

- Das metrische Modell bietet eine adäquate Annäherung zur Beurteilung der Ähnlichkeit von Farbreizen.
- Das metrische Modell hat sich bereits in der Mustererkennung und Computer Vision für viele Anwendungsfälle bewährt.
- Das metrische Modell bietet einen konsistenten Rahmen für die Indexierung von Merkmalsräumen.

### 2.3.2.3 Kombination von Merkmalen

Sollen im Rahmen einer Suchanfrage mehrere unterschiedliche Merkmale kombiniert und gewichtet werden, so wird zunächst für jedes gewählte Merkmal eine unabhängige Anfrage ausgewertet. In einem anschließenden Schritt müssen die verschiedenen Ergebnislisten der einzelnen Merkmale zu einer Gesamtergebnisliste vereinigt werden – diesen Prozess nennt man auch Verschmelzung (*engl. merging*) [138]. Dazu wird in der Regel aus den Ergebnislisten der einzelnen Merkmale ein neuer Ergebniswert be-

rechnet und die Bilder entsprechend dieses Wertes neu geordnet. Je nach Berechnung des Ergebniswertes unterscheidet man folgende Arten der Verschmelzung:

**Merge-by-Score** – Es wird für jedes Bild der Datenbasis ein Ergebniswert

$$\bar{\delta}_i = \frac{1}{M} \sum_{j=1}^M w_j \delta_i^j \quad (2.1)$$

aus einer Linearkombination der einzelnen Ähnlichkeitsmaße berechnet, wobei  $M$  die Anzahl der ausgewählten Merkmale,  $w_j$  die Wichtung des  $j$ -ten Merkmals und  $\delta_i^j$  das Ähnlichkeitsmaß bezüglich des  $j$ -ten Merkmals zwischen der Suchanfrage und dem  $i$ -ten Bilde bezeichnet. Vorteil dieser Methode ist, dass der Ergebniswert  $\bar{\delta}_i$  weiterhin ein quantitatives Maß für die Ähnlichkeit eines Bildes zur Suchanfrage ist. Der Nachteil ist, dass alle Ähnlichkeitsmaße  $\delta^j$  auf denselben Wertebereich standardisiert sein und zudem in linearer Beziehung zueinander stehen müssen. Das trifft in der Regel nur für die wenigsten Merkmals-typen zu.

**Merge-by-Rank** – Es wird für jedes Bild der Datenbasis ein Positionswert

$$\bar{p}_i = \frac{1}{M} \sum_{j=1}^M w_j p_i^j \quad (2.2)$$

aus einer Linearkombination des Ranges berechnet, wobei  $M$  die Anzahl der ausgewählten Merkmale,  $w_j$  die Wichtung des  $j$ -ten Merkmals und  $p_i^j$  der Rang oder die Position des  $i$ -ten Bildes in der Ergebnisliste des  $j$ -ten Merkmals bezeichnet. Der Vorteil dieser Vorgehensweise ist, dass sie sich auf alle Kombinationen von Ähnlichkeitsmaßen  $\delta^j$  – auch nicht-lineare, zueinander nicht-kompatible – anwenden lässt. Nachteilig ist, dass der Positionswert  $\bar{p}_i$  selbst kein Maß der Ähnlichkeit eines Bildes zur Suchanfrage ist, sondern lediglich ein relationales Ordnungskriterium.

#### 2.3.2.4 Semantische Lücke

Das größte Problem merkmalsbasierter Verfahren mittels syntaktischer Merkmale entsteht durch die Existenz einer so genannten semantischen Lücke (*engl. semantic gap*) zwischen System und Benutzer [142][35]. Der Begriff „semantische Lücke“ bezeichnet die fehlende Koinzidenz zwischen der aus den Bilddaten extrahierten visuellen Information und der aus den gleichen Bilddaten gewonnenen subjektiven Interpretation eines Betrachters in einer bestimmten Situation.

Eine linguistische Beschreibung eines Bildes durch einen Betrachter ist immer kontextabhängig und repräsentiert eine semantische Sicht auf den Bildinhalt (siehe 2.3.1).

Der Betrachter tendiert dazu, im Bild vorhandene Objekte zu identifizieren und ihnen eine Bedeutung beizumessen. Merkmalsbasierte Bildbeschreibungen hingegen repräsentieren eine rein syntaktische Sicht auf den Bildinhalt und sind in der Regel unabhängig vom Kontext eines bestimmten Betrachters. Ein Benutzer würde mittels einer sprachorientierten Suchanfrage instinktiv immer nach ähnlichen Bedeutungen in den Bildinhalten suchen, wohingegen merkmalsbasierte IR-Systeme nur eine Suche nach Ähnlichkeit in der visuellen Erscheinung von Bildern ermöglichen. Diese Diskrepanz zwischen System und Benutzer wird semantische Lücke genannt.

Eine sprachorientierte Suchanfrage ist für ein merkmalsbasiertes IR-System nicht geeignet. Um diese semantische Lücke zwischen System und Benutzer zu überbrücken, müssen dem Benutzer geeignete Paradigmen zur Erstellung einer Suchanfrage zur Verfügung gestellt werden.

### 2.3.2.5 Suchparadigmen

Die menschliche Sprache scheidet als Grundlage für eine Suchanfrage an ein merkmalsbasiertes CBIR-System aus oben genannten Gründen aus. Daher muss nach Alternativen gesucht werden. Die Frage, die es zu beantworten gilt, lautet: Wie kann der Benutzer eine prägnante Anfrage nach einer visuellen Information formulieren? Dabei muss genügend syntaktische Information über die vom Benutzer gesuchten visuellen Kriterien in der Suchanfrage vorhanden sein, damit das CBIR-System ein sinnvolles Ergebnis produzieren kann. Die CBIR-Forschungsgemeinschaft hat zwei grundsätzlich verschiedene Suchparadigmen identifiziert [178]:

**Query-by-Example (QBE)** – Dieses Suchparadigma erfordert die Bereitstellung eines Musterbildes seitens des Benutzers als Referenzbeispiel für die von ihm gesuchte visuelle Information. Das CBIR-System sucht dann gemäß den vom Benutzer eingestellten Suchparametern zu der Vorlage im syntaktischen Sinne ähnliche Bilder. Der Hauptvorteil dieser Methode liegt in der Tatsache, dass eine Suchanfrage sowohl vom Benutzer als auch vom CBIR-System naturgemäß verstanden wird. Auf seiten des Benutzers ist zudem kein spezialisiertes Wissen über die interne Funktionsweise des CBIR-Systems notwendig. Der Benutzer kann dabei auf unterschiedliche Weise das Musterbild erzeugen:

**Query-by-Image** – Ein vorhandenes Bild wird als Referenz benutzt, um ähnliche Bilder in der Datenbasis zu finden. Bei dem Bild kann es sich z.B. um ein Bild in schlechter Qualität handeln oder um ein Bild der Ergebnismenge einer früheren Suche. Die Bilder können durch den Benutzer prinzipiell verändert werden (z.B. um ungewünschte Details zu entfernen).

**Query-by-Sketch** – Oftmals ist der Benutzer nicht im Besitz eines geeigneten Bildes für eine visuelle Suchanfrage. Eine Möglichkeit ist das Erstellen einer Musterskizze durch Malen, die die Vorstellungen des Benutzers gesamtheitlich bezüglich der Anordnung von Farben grob charakterisieren. Die Effektivität einer solchen Anfrage hängt sehr stark von den künstlerischen Fähigkeiten des Benutzers ab, seine Vorstellungen graphisch auszudrücken.

**Query-by-Icon** – Ikonische Anfragen eignen sich für Merkmale mit einem gewissen Grad an semantischer Information. In gewisser Weise repräsentiert eine solche Suchanfrage einen Fragesatz in einer visuellen Sprache. Die vorhandene Menge an Ikonen bildet dabei das zur Verfügung stehende Vokabular. Aufgrund des Mangels an semantischen Merkmalen spielen ikonische Anfragen heute eine untergeordnete Rolle.

**Query-by-Feature (QBF)** – Dieses Suchparadigma erfordert, dass der Benutzer explizit die signifikanten Merkmale der von ihm gesuchten visuellen Information in der Suchanfrage spezifiziert. Dieser Typ erfordert für die im CBIR-System vorhandenen Merkmalstypen jeweils eine spezialisierte graphische Benutzerschnittstelle. Auf seiten des Benutzers ist zudem detailliertes Wissen über die Interna des CBIR-Systems notwendig. QBF ist somit nur für Experten praktikabel. QBIC<sup>2</sup> ist ein Beispiel für ein solches existierendes CBIR-System [114][45].

#### 2.3.2.6 Aktuelle Forschungsthemen

Eine Menge von wichtigen Forschungsthemen können aus den Erkenntnissen der Analyse merkmalsbasierter Verfahren abgeleitet werden. In unbestimmter Reihenfolge sind dies:

**Extraktion visueller Merkmale** – Der Entwicklung neuer Algorithmen zur Extraktion von syntaktischer als auch semantischer Merkmale sind der zentrale Fokus der CBIR-Forschung. Obwohl in diesem Bereich bereits signifikante Forschungsergebnisse erbracht wurden, gibt es immer noch einen großen Bedarf an robusten, performanten Methoden, die die semantische Lücke verkleinern bzw. schließen könnten. Ein besonders attraktiver Ansatz bietet hier die Erkennung von impliziten Verbindungen von syntaktischen und semantischen Merkmalen [26]. Des Weiteren sind weiterhin spezialisierte CBIR-Systeme für Applikationen mit sehr spezifischen Anforderungen (z.B. Medizin, Verbrechensbekämpfung, Urheberrechtsverletzung) von Interesse.

---

<sup>2</sup>[www.qbic.almaden.ibm.com](http://www.qbic.almaden.ibm.com)

**Merkmalsbasierte Indexierung** – Die Entwicklung von Indexen ist für Bildsammlungen mit entsprechender Größe von fundamentaler Bedeutung. Der Vergleich der Suchanfrage geschieht im Allgemeinen durch das sequentielle, sukzessive Abarbeiten einer Liste aller der in der Datenbasis vorhandenen Bilder. Diese, einer linearen Suche entsprechenden, Vorgehensweise ist sehr zeitaufwändig und ineffizient: Der Aufwand ist direkt proportional zur Größe der Datensammlung. Können CBIR-Systeme so heute noch Bildsammlungen mit mehreren tausend Bildern in Echtzeit durchsuchen, so wird zukünftig die Geschwindigkeit bei sequentieller Verarbeitung von Bildsammlungen mit zu erwartender Größe von mehreren Millionen Bildern immer mehr zum Flaschenhals werden. Daher wird ein Indexierungsverfahren benötigt, das die Anzahl der in Frage kommenden Bilder a priori einschränkt und den mit dem Vergleich verbundenen Rechenaufwand auf ein vertretbares Minimum reduziert, ohne jedoch die Qualität des Retrievalergebnisses zu beeinträchtigen. Diese Aufgabe wird im Allgemeinen als merkmalsbasierte Indexierung (*engl. similarity indexing*) bezeichnet [173].

**Datenbankmodelle** – Die Entwicklung eines angemessenen Datenbankmodells zur Speicherung visueller Information ist immer noch Gegenstand einer Debatte in der Forschungsgemeinschaft. In fast allen existierenden CBIR-Systemen werden relationale Datenbanksysteme (RDBMS) zur Speicherung von inhaltsbasierter Information herangezogen, allerdings geht die Funktionalität des zugrundeliegenden Datenmodells selten über die eines primitiven Lagers für Deskriptorendaten hinaus. Das CBIR-System als Applikation ist weiterhin für die Verwaltung, Organisation, Zugriff und Auswertung der in der Datenbank gespeicherten Daten verantwortlich – eine Aufgabe, die eigentlich der Datenbank zukommen sollte. Daher ist noch viel Forschungsarbeit nötig, bis die ersten wirklichen „Bild-Datenbanken“ entwickelt sind, die ein integriertes, dem CBIR-Paradigma angepasstes Datenmodell verwirklichen [118].

**Benutzerschnittstellen** – Der Entwicklung graphischer Benutzerschnittstellen kommt eine zentrale Rolle bei der Etablierung von CBIR-Systemen beim Benutzer zu. Ein CBIR-System wird ohne eine einfache und schnell verständliche Bedienung keine Akzeptanz finden. Die Schwierigkeit, eine entsprechende Benutzerschnittstelle auf dieser Basis zu entwickeln, wird häufig unterschätzt. Es muss untersucht werden, welche Paradigmen zur synergistischen Kooperation zwischen Benutzer und CBIR-System geeignet sind. Obwohl in diesem Gebiet viele Ideen existieren, wird beachtlicher Aufwand nötig sein, diese in der Praxis zu evaluieren.

**Relevanzrückkopplung** – Im fundamentalen Gegensatz zu den computer-zentrischen, vollautomatischen CBIR-Verfahren zur Archivierung von visueller Information kann der Mensch bei der merkmalsbasierten Suche als wichtiger Bestandteil interaktiv eingebunden werden: Der Benutzer kann die Suche anhand seiner subjektiven Beurteilung bezüglich der Relevanz vorheriger Anfrageergebnisse zielgerichtet und iterativ verfeinern. Er muss dazu die Bilder eines ihm präsentierten Ergebnisses in Relation zu seiner Suchanfrage als relevant oder irrelevant – ohne eine nähere Begründung zu geben – kennzeichnen. Das CBIR-System kann diese Information nun dazu verwenden, die internen Parameter der Suchauswertung so zu verändern, dass sich die Ergebnisqualität bei gleicher Suchanfrage durch Reduktion der Anzahl der im Ergebnis vorhandenen Bilder mit irrelevanter Information (*engl. false-positives*) verbessert, und ursprünglich nicht im Ergebnis vorhandene Bilder mit relevanter Information (*engl. false-negatives*) aufgespürt werden. Dies kann beispielsweise durch die implizite Veränderung der Gewichte bei der Kombination unterschiedlicher Merkmale bei der Suchanfrage erfolgen (oftmals ist der Benutzer aufgrund mangelnder Systemkenntnis nicht in der Lage, diese selbstständig nach seinen Präferenzen zu verändern). Dieser Prozess wird Relevanzrückkopplung (*engl. relevance feedback*) genannt [27][110][128].

## 2.4 Evaluation

Die Evaluation von Verfahren und Systemen spielt allgemein im IR eine wichtige Rolle. Dabei gilt es, ein IR-System bezüglich der Unterstützung des Benutzers bei der Lösung seines speziellen Problems zu beurteilen. Ein IR-System unterstützt den Benutzer, falls es auf seine Suchanfrage Dokumente mit relevanter Information liefert.

**Definition 2.1 (Relevanz)** *Eine Information gilt als relevant, falls ein Benutzer die Information zur Lösung eines Problems in einer konkreten Situation nutzen und verwenden kann.* ■

**Bemerkung:** *Die Relevanz eines Dokumentes hängt maßgeblich vom Benutzer und seiner jeweiligen Situation ab und ist somit hochgradig subjektiv. Die Beurteilung eines Dokumentes bezüglich seiner Relevanz erfolgt immer binär, d.h. ein Dokument wird entweder als relevant oder irrelevant eingestuft. Ein Dokument wird insgesamt als relevant beurteilt, auch wenn nur ein Teil des Dokumentes relevante Information enthalten sollte (egal wie groß der Teil im Verhältnis zu dem Rest des Dokumentes ist).*

Bei der Beurteilung von Verfahren und Systemen des IR unterscheidet man zwischen Effizienz und Effektivität:

**Effizienz** – Beurteilt den ökonomischen Umgang mit verfügbaren Ressourcen wie z.B. Rechenzeit und Speicherplatz. Die Bewertung erfolgt durch theoretische und empirische Analysen der verwendeten Algorithmen.

**Effektivität** – Beurteilt das Kosten-Nutzen-Verhältnis bei der Anwendung eines Verfahrens bzw. eines Systems. Beim IR wird primär der Nutzen gemessen, d.h. die Qualität der Retrievalergebnisse (Werden alle relevanten Dokumente gefunden?). Hierzu sind nicht-experimentelle, analytische Methoden kaum geeignet. Die Bewertung erfolgt daher in der Regel durch experimentelle Beurteilung.

Zur Evaluation der Leistungsfähigkeit eines IR-Systems werden in erster Linie die Maße Präzision (*engl. precision rate*) und Vollständigkeit (*engl. recall rate*) zur Beurteilung der Effektivität verwendet [28].

**Definition 2.2 (Präzision und Vollständigkeit)** Sei  $A$  die Menge aller Objekte des Retrievalergebnisses und  $B$  die Menge aller relevanten Objekte der Dokumentenbasis. Die Präzision gibt den Anteil der relevanten Dokumente unter den gefundenen Dokumenten an:

$$\text{Präzision} := \frac{|A \cap B|}{|A|}$$

Die Vollständigkeit gibt den Anteil der relevanten Dokumente an, die gefunden wurden:

$$\text{Vollständigkeit} := \frac{|A \cap B|}{|B|} \quad \blacksquare$$

**Bemerkung:** Die Idealwerte für die Präzision und Vollständigkeit sind jeweils 1. Für die Präzision ergibt sich dieser Idealwert, wenn das Suchergebnis z.B. nur ein einziges relevantes Dokument enthält. Der Idealwert für die Vollständigkeit ergibt sich, wenn die Menge aller Objekte im Suchergebnis z.B. gleich der Menge aller Objekte der Dokumentenbasis ist. In der Realität wird die Anzahl der Objekte im Suchergebnis zwischen diesen beiden Extremen liegen. Präzision und Vollständigkeit verhalten sich zueinander invers. Eine Verkleinerung der Antwortmenge durch eine spezifischere Anfrage hat eine höhere Präzision, aber eine kleinere Vollständigkeit zur Folge. Umgekehrt hat eine Vergrößerung der Antwortmenge durch eine allgemeinere Anfrage eine größere Vollständigkeit, aber eine niedrigere Präzision zur Folge.

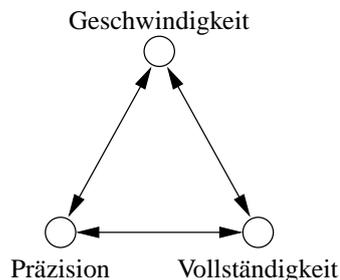
Zu einer umfassenden Evaluation der Leistungsfähigkeit muss im Allgemeinen noch ein dritter Faktor – die Geschwindigkeit (*engl. speed*) – als Maß der Effizienz eines IR-Systems berücksichtigt werden [84].

**Definition 2.3 (Geschwindigkeit)** Sei  $\tau_{query}$  der Zeitpunkt, an dem die Suchanfrage gestellt wurde und  $\tau_{result}$  der Zeitpunkt, an dem das Suchergebnis als Antwort des IR-Systems auf die Suchanfrage dem Benutzer vorliegt. Die Geschwindigkeit gibt die Zeit an, die vom Zeitpunkt des Stellens der Suchanfrage bis zu dem Zeitpunkt des Vorliegens des Suchergebnisses verstreicht:

$$\text{Geschwindigkeit} := \tau_{result} - \tau_{query} \quad \blacksquare$$

**Bemerkung:** Im Gegensatz zur physikalischen Definition repräsentiert die Geschwindigkeit eines IR-Systems eine Zeitspanne. Relativ betrachtet entspricht eine kurze Zeitspanne einer hohen Geschwindigkeit und eine lange Zeitspanne einer niedrigen.

Die Geschwindigkeit ist ein maßgeblicher Faktor für die Benutzbarkeit eines IR-Systems (siehe 2.2.2). Sie steht in konkurrierender Wechselwirkung mit den Faktoren Präzision und Vollständigkeit (siehe Abbildung 2.8). Die Forschung ist mit der Lösung des schwierigen Problems konfrontiert, einen ausgewogenen Kompromiss zwischen den konkurrierenden Faktoren zwecks Optimierung der Leistungsfähigkeit eines IR-Systems bei ständig steigender Anzahl von Dokumenten und Benutzern zu finden.



**Abbildung 2.8:** Konkurrierende Faktoren eines IR-Systems

### 2.4.1 Spezielle Probleme des VIR

Die anfängliche Motivation für Verfahren und Systeme des VIR ist es, die gleichen Evaluationstechniken wie in der verwandten Disziplin IR anzuwenden. Dies ist jedoch nicht ganz unproblematisch; die Identifizierung von relevanten Dokumenten ist für eine bestimmte Suchanfrage im Vergleich zu Textdokumenten aufgrund des größeren Spielraums bei der Interpretation visueller Inhalte wesentlich schwieriger. Dokumenten mit visuellen Inhalten fehlt die starke Beziehung zwischen einer Bedeutung und der damit verbundenen statistischen Häufigkeit der Präsenz bestimmter Elementar-einheiten, wie sie bei Textdokumenten in Form von Wörtern zu finden ist. Eine signifikante Varianz kann zwischen der Relevanzbeurteilung unterschiedlicher Benutzer, ja

sogar zwischen den Relevanzbeurteilungen desselben Benutzers zu unterschiedlichen Zeitpunkten festgestellt werden.

Die Beurteilung von CBIR-Systemen gestaltet sich noch schwieriger; CBIR-Systeme beantworten eine Suchanfrage nicht mit einer undifferenzierten Menge relevanter Dokumente, sondern mit der Darstellung eines festgelegten Bruchteils einer Rangfolge von Dokumenten, die zur Suchanfrage bezüglich einer mehr oder weniger komplexen Konfiguration von visuellen Eigenschaften ähnlich sind. Wenn die Gesamtzahl der relevanten Dokumente in der Basis größer als die Anzahl der im Ergebnis dargestellten Dokumente ist, dann ist z.B. das Maß der Vollständigkeit eine bedeutungslose Größe.

Typischerweise wird daher eine qualitative Evaluation anhand von existierenden Dokument- und Anfragesammlungen durchgeführt, wobei hauptsächlich die Präzision des VIR-Systems beurteilt wird. Die a priori Bestimmung der relevanten bzw. irrelevanten Dokumente für eine bestimmte Suchanfrage gestaltet sich als schwierige Aufgabe, da kein Benutzer über eine exakte Kenntnis der zugrundeliegenden Dokumentenbasis verfügt. Generell erscheint es unmöglich, eine Basis von mehreren Tausend bzw. Millionen Dokumenten umfassend zu beurteilen und die Menge der relevanten Dokumente exakt zu bestimmen.

Erste Veröffentlichungen der frühen neunziger Jahre zum Themengebiet CBIR haben sich damit begnügt, die Ergebnisse ein oder mehrerer Suchanfragen abzdrukken. Diese Methode ist darauf zugeschnitten, einen Eindruck bezüglich der „Güte“ des Verfahrens zu vermitteln. Diese Darstellung ermöglicht jedem Betrachter, selbst zu beurteilen, inwiefern die Antwort des CBIR-Systems mit der persönlichen, mentalen Erwartung übereinstimmt. Im Wesentlichen werden dabei folgende Fragestellungen analysiert:

- Wie groß ist der Anteil von subjektiv „guten“ Bildern?
- Warum wurde ein bestimmtes Bild als ähnlich eingestuft?
- Warum wurde ein bestimmtes Bild im Vergleich zu einem anderen als ähnlicher bzw. unähnlicher eingestuft?
- Warum ist ein bekanntes Bild nicht wie erwartet im Ergebnis?
- Wie kann das Ergebnis mit einer neuen Suchanfrage verbessert werden?

Dies führt dazu, dass der Betrachter eher dazu neigt zu beurteilen, wie „gut“ ihm die Antwort gefällt, anstatt wie relevant sie für ihn ist [39].

Diese Vorgehensweise gibt jedoch weder ein objektives Maß zur Evaluation der Qualität der Retrievalergebnisse noch die Möglichkeit zum Vergleich mit anderen Verfahren. Generell ist der experimentelle Vergleich verschiedener Verfahren und Systeme

ohne gemeinsame Grundlage äußerst kompliziert und oft eine strittige Angelegenheit. Verschiedene Forschungsteams verwenden in der Regel unterschiedliche, nicht-öffentliche Dokument- und Anfragesammlungen, die einen direkten Vergleich ihrer Retrievalergebnisse mit denen anderer Teams unmöglich machen. Zu groß ist dabei der maßgebliche Einfluss, den die interne Zusammensetzung einer Dokumentensammlung hinsichtlich der Kohärenz ihrer Bestandteile auf das Retrievalergebnis eines bestimmten Algorithmus haben kann.

### 2.4.2 Benchmarking

Eine standardisierte Dokument- und Anfragesammlung als verbindlicher Maßstab (*engl. benchmark*) zur Evaluation von IR-Verfahren ist daher notwendig. Der heutige Status auf dem Gebiet des CBIR ist jedoch noch weit entfernt von dem des IR, wo das Ziel einer standardisierten Dokument- und Anfragesammlung seit 1992 im Rahmen der Text Retrieval Conference (TREC) mit Nachdruck verfolgt wird. Die TREC wird vom amerikanischen National Institute of Standards and Technology (NIST) und der Defense Advanced Research Projects Agency (DARPA) von offizieller Seite unterstützt. Forschungsteams, die an der TREC teilnehmen, müssen während der Konferenz mehrere Gigabyte an Textdokumenten indexieren; die Ergebnisse werden nach festgelegten Richtlinien objektiv von TREC-Offiziellen beurteilt und nach der Konferenz öffentlich bekannt gemacht.

Erst kürzlich ist man dazu übergegangen, ähnliche Maßnahmen zur standardisierten Beurteilung von CBIR-Retrievalergebnissen zu ergreifen. Bezüglich der Festlegung von Richtlinien zur Dokumentation von Ergebnissen wurden erste Vorschläge gemacht [106]. Das Benchathlon Projekt<sup>3</sup> ist ein erster Schritt in Richtung der Zusammenstellung einer standardisierten Bildersammlung, dem es jedoch noch an entsprechendem Bekanntheitsgrad in der Forschungsgemeinschaft und an entsprechender Unterstützung von offizieller Seite fehlt. Darüber hinaus müssen noch viele praktische Probleme gelöst werden: Zum einen muss die Bildersammlung nicht durch Urheberrecht eingeschränkt frei verfügbar sein, und zum anderen muss eine ausreichende Vielfalt von Bildern aus den verschiedensten Bilddomänen vorhanden sein. Des Weiteren sollte die Anzahl der Bilder ausreichend groß sein, damit das Verhältnis der konkurrierenden Faktoren Geschwindigkeit und Qualität untersucht werden kann. Im Gegensatz zum IR, in dem Dokumentensammlungen in der Größenordnung von einer Million und mehr üblich sind, wird heutzutage im CBIR Verfahren immer noch mit Bildersammlungen in der Größenordnung von einigen wenigen tausend Bildern evaluiert.

---

<sup>3</sup>[www.benchathlon.net](http://www.benchathlon.net)

## Kapitel 3

# Modellbildung

Die grundlegende Methode zur Nachbildung eines wirklichen Prozesses in einem technischen System ist der Entwurf eines Modells. Primäres Ziel der Modellbildung ist dabei, die wesentlichen Eigenschaften des realen Prozesses vereinfacht abzubilden, indem von seinen Unregelmäßigkeiten abstrahiert wird. Sekundäres aber nicht minder wichtiges Ziel ist die Einigung auf eine gemeinsame Sprache bzw. Symbolik. Dazu soll zunächst im Folgenden der Prozess der visuellen Informationsverarbeitung beim Menschen genauer betrachtet werden.

Die visuelle Informationsverarbeitung des Menschen beginnt mit dem Auge. Durch die Augenlinse wird ein vertikal gespiegeltes Bild auf die Netzhaut (Retina) projiziert. Das Auge wird in seiner Funktion gerne mit einem Kamerasystem verglichen. Jedoch gibt es wohl mehr Unterschiede als Gemeinsamkeiten, die für die optische Wahrnehmung von entscheidender Bedeutung sind. Die Nah- oder Fernakkommodation des Auges erfolgt beispielsweise durch die Verformung der Linse. Die Änderung der Brennweite bewirkt, dass bei unterschiedlichen Entfernungen ein jeweils scharfes Bild auf der Retina entsteht. Bei einem Kamerasystem hingegen, wird der Abstand zwischen Linse und Projektionsebene verändert.

Die Qualität der Abbildung eines Bildes auf der Retina ist verglichen mit der eines technischen Linsensystems wesentlich schlechter. Es kommen optische Fehler wie z.B. sphärische und chromatische Aberration, Astigmatismus, Koma, Bildfeldwölbung und nicht uniforme Auflösung zum Tragen [67]. Im Vergleich zur passiven Sensorik eines technischen Kamerasystems ist die Retina jedoch ein hochspezialisierter Prozessor zur Bildvorverarbeitung, der zeitliche, räumliche und farbliche Information extrahiert, abgebildete Strukturen analysiert und diese an die visuellen Zentren im Gehirn weiterleitet. Anatomisch betrachtet ist die Retina kein eigenständiges Organ, sondern ein ausgelagerter Teil des Gehirns und somit Teil des komplexen kognitiven Wahrnehmungsapparates des Menschen.

Die rechnergestützte Erfassung von visueller Information unterscheidet sich grundsätzlich in der Form, wie sie zur Verfügung gestellt wird. Im Folgenden wird ein Modell zur merkmalsbasierten Informationsverarbeitung definiert, das auf der Grundlage von Bildern arbeitet, die mit einem technischen Kamerasystem digital erfasst wurden.

### 3.1 Bilder

Physikalisch betrachtet stellen Bilder eine flächenhafte Verteilung der Bestrahlungsstärke  $\xi$  von sichtbarem Licht in einer Ebene dar. Mathematisch kann diese als eine kontinuierliche Funktion  $\xi(x, y)$  zweier räumlicher Variablen  $x$  und  $y$  beschrieben werden:

$$\begin{aligned} \xi : \mathbb{R} \times \mathbb{R} &\rightarrow \mathbb{R}_0^+ \\ (x, y) &\mapsto \xi(x, y) \end{aligned} \quad (3.1)$$

Eine fundamentale Voraussetzung der rechnergestützten Bildverarbeitung ist, dass die Bildinformation in digitaler Form vorliegt. Dazu muss die Bildfunktion  $\xi(x, y)$  in eine Darstellung von Werten auf einem zweidimensionalen, diskreten Gitter endlicher Ausbreitung überführt werden. Die Verwendung eines kartesischen Gitters ist dabei die allgegenwärtigste Lösung, die zudem auch am einfachsten technisch realisierbar ist.

**Definition 3.1 (Rasterung)** Die Abbildung der kontinuierlichen, räumlichen Variablen einer Bildfunktion auf eine kartesische  $M \times N$  Bildmatrix mit endlicher Höhe  $M \in \mathbb{N}$  und endlicher Breite  $N \in \mathbb{N}$

$$\begin{aligned} S : \mathbb{R} \times \mathbb{R} &\rightarrow \{0, 1, \dots, M-1\} \times \{0, 1, \dots, N-1\} \\ (x, y) &\mapsto [m, n] \end{aligned} \quad (3.2)$$

wird Rasterung (engl. *sampling*) genannt. ■

Ein Punkt in der zweidimensionalen Bildmatrix wird als Pixel (engl. *picture element*) bezeichnet. Der mit dem Pixel assoziierte Wert  $\xi[m, n]$  gibt dabei die mittlere Bestrahlungsstärke von Licht der von ihm repräsentierten rechteckigen Region innerhalb des Gitters an. Zur digitalen Darstellung muss der Wert eines Pixels auf eine endliche Anzahl diskreter Werte abgebildet werden.

**Definition 3.2 (Quantisierung)** Die Abbildung des kontinuierlichen Wertebereichs eines Bildpunktes auf eine endliche Anzahl  $Q \in \mathbb{N}$  von Intensitätsstufen

$$\begin{aligned} Q : \mathbb{R}_0^+ &\rightarrow \{0, 1, \dots, Q-1\} \\ \xi &\mapsto x \end{aligned} \quad (3.3)$$

wird Quantisierung (engl. *quantization*) genannt. ■

**Definition 3.3 (Digitales Bild)** Die  $M \times N$  Bildmatrix

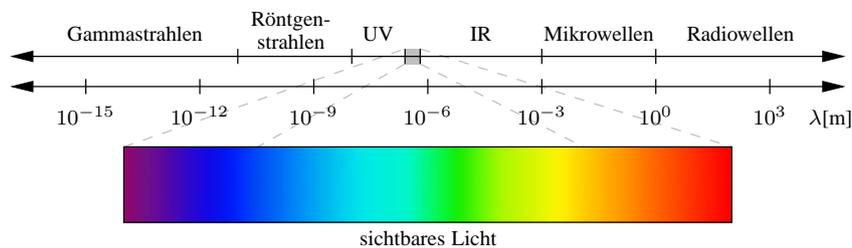
$$X = \left( x[m, n] \right) \quad (3.4)$$

mit endlicher Höhe  $M \in \mathbb{N}$  und endlicher Breite  $N \in \mathbb{N}$  sowie einem diskreten Wertebereich  $x[m, n] \in \{0, 1, \dots, Q - 1\}$  mit einer endlichen Anzahl  $Q \in \mathbb{N}$  unterschiedlicher Intensitätsstufen ist ein digitales Bild. ■

**Bemerkung:** Die Tiefe eines digitalen Bildes bezeichnet die Informationsmenge, mit der die Intensität eines Bildpunktes beschrieben wird. Sie wird in der Einheit bit (engl. binary digit) angegeben.

## 3.2 Farbbilder

Die in einem Bildpunkt dargestellte Bestrahlungsstärke gibt die Energie von elektromagnetischer Strahlung  $\phi(\lambda)$  an, die im Bereich der Wellenlängen 380nm bis 780nm vom Menschen als farbiges Licht wahrgenommen wird (siehe Abbildung 3.1). Farbe ist folglich keine physikalische Eigenschaft, sondern ein Sinneseindruck, der im menschlichen Gehirn erzeugt wird. Neben anderen elementaren Wahrnehmungen wie z.B. Kontrast, Tiefe und Bewegung ist Farbe eine der wichtigsten Quellen visueller Information [99].



**Abbildung 3.1:** Übersicht des Spektrums elektromagnetischer Strahlung

Um das Attribut Farbe in einem Pixel als eine weitere Dimension visueller Information beschreiben zu können, verwendet man das Modell des mehrkanaligen Multispektralbildes. Jeder Bildkanal für sich betrachtet stellt dabei die Intensität von Licht eines bestimmten Ausschnitts des elektromagnetischen Spektrums dar.

**Definition 3.4 (Digitales Multispektralbild)** Ein digitales Bild mit  $K$  Kanälen

$$\mathbf{X} = (X_0, X_1, \dots, X_{K-1}) \quad (3.5)$$

ist ein digitales Multispektralbild, wenn die einzelnen Bildkanäle den identischen diskreten Wertebereich haben, aber jeweils die Intensität des Lichtes von unterschiedlichen Ausschnitten des elektromagnetischen Spektrums darstellen. ■

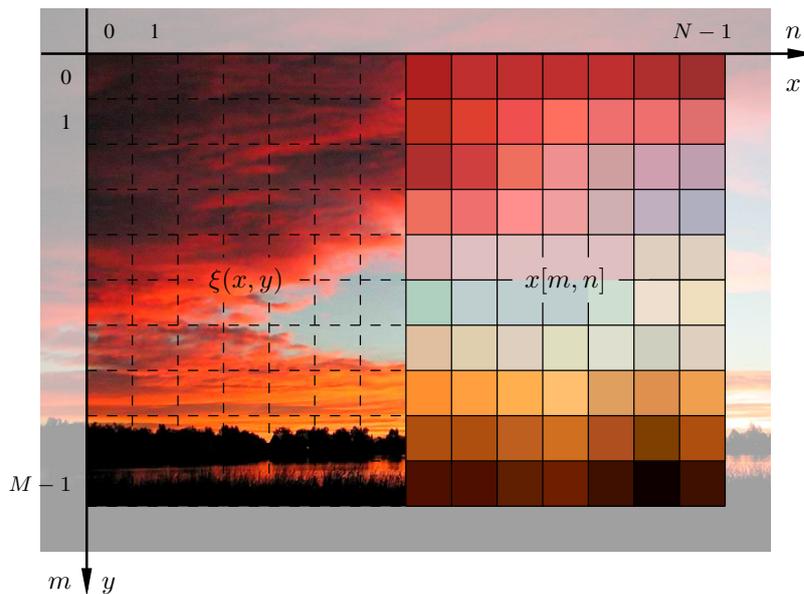
Die Farbe eines Pixels wird in der rechnergestützten Bildverarbeitung in der Regel als Punkt im dreidimensionalen RGB-Farbraum repräsentiert (siehe 4.1.1). Die digitale Repräsentation eines RGB-Farbbildes besteht folglich aus einem 3-kanaligen Multispektralbild.

**Definition 3.5 (Digitales RGB-Farbbild)** Ein 3-kanaliges Multispektralbild mit der Kanalzuordnung Rot-, Grün- und Blauauszug

$$\mathbf{X} = (R, G, B) \quad (3.6)$$

ist ein digitales RGB-Farbbild. ■

**Beispiel 3.1 (Digitales RGB-Farbbild)** In Abbildung 3.2 ist in der linken Hälfte das Originalbild und in der rechten Hälfte ein durch Rasterung und Quantisierung erzeugtes digitales RGB-Farbbild dargestellt. Deutlich ist zu erkennen, dass durch den Prozess der Rasterung und Quantisierung Information verloren geht, d.h. das digitale Abbild ist nur eine Approximation des Originalbildes.



**Abbildung 3.2:** Digitales RGB-Farbbild

**Definition 3.6 (Bilddomäne)** Das Universum  $\Xi$  aller digitalen Bilder  $\mathbf{X}_i$  wird Bild-domäne (engl. image domain) genannt. ■

Die Bilddomäne ist für CBIR von pragmatischer Bedeutung. Sie erlaubt eine Unterscheidung zwischen eingeschränkten und uneingeschränkten Bilddomänen, wobei der

Übergang fließend ist. Bilder einer eingeschränkten Bilddomäne besitzen eine limitierte, berechenbare Variabilität ihres äußeren Erscheinungsbildes in allen relevanten Aspekten des Bildinhaltes. Die semantische Bedeutung von Bildern einer eingeschränkten Bilddomäne ist in aller Regel wohldefiniert und eindeutig. Bilder einer uneingeschränkten Bilddomäne besitzen hingegen eine unendliche, unberechenbare Variabilität ihres äußeren Erscheinungsbildes, und infolgedessen eine unbegrenzte Heterogenität im Hinblick auf deren semantische Bedeutungen.

**Beispiel 3.2 (Bilddomäne)** *Die Menge aller Passfotos von Personen (Frontalansicht, neutraler Hintergrund, freies Gesicht, keine Kopfbedeckung) bildet eine eingeschränkte Bilddomäne. Die semantische Bedeutung jedes Bildes ist eindeutig und klar bestimmt. Trotzdem ist das Bild einer einzelnen Person einzigartig und von allen anderen unterscheidbar. Obwohl die Bilder diesbezüglich eine unendliche Variabilität in visuellen Details besitzen müssen, herrschen offensichtlich einige geometrische, physikalische und farbliche Beschränkungen bezüglich des visuellen Erscheinungsbildes vor. Diese können als vorhandenes Wissen für CBIR-Systeme genutzt werden.*

*Die Menge aller Bilder des WWW bildet eine uneingeschränkte Bilddomäne. Die Bedeutung jedes Bildes ist polysemisch (vieldeutig). Es lassen sich keine für ein CBIR-System nutzbaren Aussagen aus den visuellen Eigenschaften der Bilder ableiten.*

Die Bilddomäne eines CBIR-Systems ist gleichzeitig dessen Definitionsbereich, also die Menge von Bildern, auf die dieses System anwendbar ist.

### 3.3 Merkmalsextraktion

Der Vorgang der visuellen Informationsverarbeitung, der beim Menschen scheinbar mühelos abläuft, ist in der technischen Umsetzung weitaus schwieriger. Die rechnergestützte Verarbeitung erfolgt immer auf der Grundlage digitalisierter Bilder, die eine räumliche Anordnung diskreter Bildpunkte mit zugeordneten Farbeigenschaften repräsentieren. Verglichen mit dem kognitiven Wahrnehmungsapparat des Menschen ist die Erkennungsleistung und -geschwindigkeit eines auf digitalen Daten operierenden Systems sehr gering. Demgegenüber steht eine dem Menschen überlegene Fähigkeit, aus Bildern quantitative Größen präzise und objektiv zu bestimmen. Hier kann das visuelle Erkennungssystem des Menschen nur relative und subjektive Angaben treffen.

Der Vorgang der Merkmalsextraktion eines Menschen unterscheidet sich daher von der eines technischen Systems auf fundamentale Art und Weise. Im Gegensatz zum Menschen, der semantische Merkmale aus Bildern extrahiert, arbeitet ein CBIR-System auf der Basis von syntaktischen Merkmalen, die messbare Eigenschaften visueller Stimuli

des Bildinhaltes verkörpern. Diese werden mittels wohldefinierter Berechnungsvorschriften direkt aus den digitalen Bilddaten gewonnen.

Es existieren prinzipiell viele Möglichkeiten, unterschiedliche Aspekte eines Bildes in Merkmalen zu beschreiben. Es gibt keine allgemeingültige Methode, ein einzelnes Merkmal zu definieren, das allen Anforderungen gerecht wird. Daher werden in der Regel mehrere Merkmale verwendet, die jeweils unabhängig voneinander bestimmte, unterschiedliche Aspekte verkörpern (z.B. statistische geometrische und/oder topologische Eigenschaften). Da die syntaktischen Merkmale auf der Grundlage komplexer mathematischer Berechnungsvorschriften definiert sind, ist eine natursprachliche Interpretation des Merkmals in der Regel schwierig.

Im Folgenden wird ein einheitliches Modell definiert, das einen formalen Rahmen für die Entwicklung geeigneter Berechnungsverfahren zur Darstellung visueller Bildinhalte mittels syntaktischer Merkmale vorgibt. Aus diesem Rahmenwerk lässt sich eine einheitliche Architektur mit standardisierten Schnittstellen für ein CBIR-System ableiten, das durch eine geeignete Auswahl von Merkmalen im Hinblick auf ein bestimmtes Anwendungsgebiet zugeschnitten werden kann.

**Definition 3.7 (Merkmalsextraktion und -typ)** Sei  $\mathcal{F}$  eine Abbildung, die jedem digitalen Bild  $X_i$  der Bilddomäne  $\Xi$  einen Wert  $\omega_i$  zuweist:

$$\begin{aligned} \mathcal{F} : \Xi &\rightarrow \Omega \\ X_i &\mapsto \omega_i \end{aligned} \tag{3.7}$$

Die Abbildung  $\mathcal{F}$  wird Merkmalsextraktion (engl. *feature extraction*) genannt, falls jeder Wert  $\omega_i$  das Ergebnis einer wohldefinierten Berechnungsvorschrift angewandt auf das Bild  $X_i$  repräsentiert. Die Berechnungsvorschrift wird in der umgangssprachlichen Verwendung auch als „Merkmalstyp“ (engl. *feature type*) bezeichnet. ■

**Bemerkung:** Die Abbildung  $\mathcal{F}$  ist in der Regel nicht bijektiv, d.h. es existiert keine inverse Abbildung  $\mathcal{F}^{-1}$ .

**Definition 3.8 (Merkmalsdeskriptoren und -domäne)** Der einem digitalen Bild  $X_i$  durch  $\mathcal{F}$  zugeordnete Wert  $\omega_i$  heißt Merkmalsdeskriptor (engl. *feature descriptor*). Das Universum  $\Omega$  aller Merkmalsdeskriptoren  $\omega_i$  bezüglich eines bestimmten Merkmalstyps wird Merkmalsdomäne (engl. *feature domain*) genannt. ■

**Bemerkung:** In Abhängigkeit von den speziellen Eigenschaften des Merkmalstyps kann die Kardinalität der Merkmalsdomäne a priori bestimmt oder unbestimmt sein.

Die aus einem Bild extrahierten Merkmalsdeskriptoren stellen die jeweiligen verkörperten Eigenschaften in mehr oder weniger komplexen Datenstrukturen (z.B. Zahlen,

Felder, Listen, Bäume und Verbünde) dar. Die Datenmenge aller Merkmalsdeskriptoren eines Bildes ist in der Regel sehr viel kleiner als die Datenmenge des Bildes selbst. Durch die Berechnung der Merkmalsdeskriptoren wird eine signifikante Datenreduktion im Hinblick auf die Beschreibung des Bildinhaltes erreicht.

### 3.4 Ähnlichkeit

Das Paradigma des VIR erfordert es, dass eine qualitative und quantitative Aussage bezüglich der Ähnlichkeit von Bildinhalten getroffen werden muss. Die besondere Schwierigkeit liegt hierbei in der Definition des Begriffes „ähnlich“. Generell kann die Ähnlichkeit immer nur bezüglich eines bestimmten visuellen Kriteriums festgestellt werden. Im CBIR wird die Ähnlichkeit von Bildern daher immer auf der Grundlage eines bestimmten Merkmalstyps beurteilt. Um die Ähnlichkeit zweier Bilder zu bestimmen, wertet man ein Ähnlichkeitsmaß zwischen den zugehörigen Merkmalsdeskriptoren des Merkmalstyps aus.

**Definition 3.9 (Ähnlichkeitsmaß)** *Die Abbildung eines Paares von Merkmalsdeskriptoren auf einen nicht-negativen, reellen Wert*

$$\begin{aligned} \delta : \Omega \times \Omega &\rightarrow \mathbb{R}_0^+ \\ (\omega_i, \omega_j) &\mapsto \delta(\omega_i, \omega_j) \end{aligned} \tag{3.8}$$

heißt Ähnlichkeitsmaß (engl. *similarity measure*). ■

**Bemerkung:** *Ein Wert 0 kennzeichnet das höchste Maß an Ähnlichkeit (Reflexivität), ein unendlicher großer Wert das höchste Maß an Unähnlichkeit. Folglich handelt es sich im eigentlichen Sinne um ein Unähnlichkeitsmaß.*

*Aus pragmatischen Gründen wird häufig ein Ähnlichkeitsmaß gewählt, das ein Paar von Merkmalsdeskriptoren auf das begrenzte Intervall  $[0, 1]$  abbildet. In diesem speziellen Fall entspricht der Wert 1 dem höchsten Maß an Unähnlichkeit.*

Die Merkmalsdomäne ist der Wertebereich eines Merkmalstyps, dessen Elemente als eine Menge von Punkten eines Raumes aufgefasst werden können. Im Zusammenhang mit einem Ähnlichkeitsmaß wird auf dieser Menge nun eine bestimmte Struktur eingeführt. Diese formalisiert die anschauliche Vorstellung, dass es zu jedem Punkt eines Raumes andere Punkte gibt, die seiner Umgebung zugeordnet sind. Ein Raum, auf dem eine derartige Struktur definiert ist, heißt auch topologischer Raum.

**Definition 3.10 (Topologischer Merkmalsraum)** *Das geordnete Paar  $(\Omega, \delta)$  wird topologischer Merkmalsraum (engl. *topological feature space*) genannt.* ■

Soll die Ähnlichkeit zweier Merkmalsdeskriptoren durch „Nähe“ im Merkmalsraum ausgedrückt werden, so muss die Definition des Ähnlichkeitsmaßes um die metrischen Axiome erweitert werden. Der Merkmalsraum reflektiert somit die Ähnlichkeit zweier Bilder als Funktion der Distanz zwischen den zugehörigen Merkmalsdeskriptoren.

**Definition 3.11 (Metrischer Merkmalsraum)** *Der Merkmalsraum  $(\Omega, \delta)$  heißt metrischer Merkmalsraum (engl. metric feature space), falls für alle Elemente  $\omega_i, \omega_j, \omega_k$  des Merkmalsraumes die Eigenschaften*

(i) *Reflexivität*

$$\delta(\omega_i, \omega_i) = 0 \quad (3.9)$$

(ii) *Positivdefinitheit*

$$\delta(\omega_i, \omega_j) \geq 0 \quad (3.10)$$

(iii) *Symmetrie*

$$\delta(\omega_i, \omega_j) = \delta(\omega_j, \omega_i) \quad (3.11)$$

(iv) *Transitivität*

$$\delta(\omega_i, \omega_j) \leq \delta(\omega_i, \omega_k) + \delta(\omega_k, \omega_j) \quad (3.12)$$

gelten. Man nennt  $\delta(\omega_i, \omega_j)$  auch die Distanz zwischen  $\omega_i$  und  $\omega_j$ . ■

**Beispiel 3.3 (Metrischer Merkmalsraum)** *Alle Merkmalsdeskriptoren  $\omega_i, \omega_j \in \mathbb{R}^n$  bilden im Zusammenhang mit der Minkowski-Metrik*

$$\delta(\omega_i, \omega_j) = \left( \sum_{k=1}^n |\omega_{i_k} - \omega_{j_k}|^r \right)^{\frac{1}{r}}, \quad r \in \mathbb{R}^+$$

*einen metrischen Merkmalsraum.*

Die Elemente eines metrischen Merkmalsraumes müssen nicht notwendigerweise aus einem Vektorraum stammen. Die Definition eines metrischen Raumes ist weit reichender als die eines Vektorraumes. Daher bildet das Modell des metrischen Merkmalsraumes die Grundlage für das allgemeinste und umfassendste Konzept zur merkmalsbasierten Suche nach visuellen Inhalten, auch wenn es vom Standpunkt der Psychologie des Menschen einige Unzulänglichkeiten aufweist (siehe 2.3.2.2).

### 3.5 Merkmalsbasierte Suche

Die merkmalsbasierte Suche wird immer auf der Grundlage einer konkreten Datenbasis von Bildern durchgeführt.

**Definition 3.12 (Bildsatz)** Die endliche Menge

$$\mathcal{X} = \{X_1, X_2, \dots, X_K\} \subseteq \Xi \quad (3.13)$$

von  $|\mathcal{X}| = K \in \mathbb{N}$  Bildern heißt Bildsatz (engl. image dataset). ■

Jedem Bild des Datensatzes wird für einen bestimmten Merkmalstyp genau ein Merkmalsdeskriptor zugeordnet. Die endliche Menge dieser Deskriptoren bildet die Grundgesamtheit für die merkmalsbasierte Suche.

**Definition 3.13 (Merkmalsdatensatz)** Die endliche Menge

$$\mathcal{W} = \{\omega_1, \omega_2, \dots, \omega_K\} \subseteq \Omega \quad (3.14)$$

von  $|\mathcal{W}| = K \in \mathbb{N}$  Deskriptoren eines Merkmalstyps, die den Bildern aus  $\mathcal{X}$  durch (3.7) zugeordnet werden, wird Merkmalsdatensatz (engl. feature dataset) genannt. ■

Eine Suchanfrage, der ein entsprechender Merkmalsdeskriptor zugeordnet ist, kann ebenfalls als ein Punkt im metrischen Merkmalsraum interpretiert werden. Die durch die Metrik induzierte Topologie ordnet jedem Merkmalsdeskriptor als Element der Merkmalsdomäne eine bestimmte Nachbarschaft zu. Das Problem der Suche nach ähnlichen Merkmalsdeskriptoren lässt sich dann folgendermaßen formulieren: Bestimme für eine beliebige Anfrage aus  $K$  gegebenen Punkten im metrischen Raum jeweils die nächsten Nachbarn. Dieses Problem wird in der rechnergestützten Verarbeitung auch als Nachbarschaftssuche bezeichnet. Es lassen sich prinzipiell zwei verschiedene Varianten der Nachbarschaftssuche unterscheiden:

**$k$ -Nächste-Nachbar-Suche** – die Ergebnismenge soll eine bestimmte Anzahl  $k$  von nächsten Nachbarn zur Anfrage enthalten (engl. fixed mass approach).

**Bereichssuche** – die Ergebnismenge soll alle nächsten Nachbarn innerhalb eines bestimmten Abstandes  $d_{\max}$  zur Anfrage enthalten (engl. fixed size approach).

Je nach Variante lässt sich die Zusammensetzung der Ergebnismenge durch die Größen  $k$  bzw.  $d_{\max}$  beeinflussen. Auf den ersten Blick erscheint die Bereichssuche für CBIR sinnvoller, da der Parameter „Ähnlichkeit“ direkt durch den Abstand  $d_{\max}$  in der Ergebnismenge kontrolliert werden kann. Aus praktischen Gründen ist die optimale Wahl von  $d_{\max}$  jedoch schwierig, da in Abhängigkeit vom unbekanntem Datensatz in der Ergebnismenge entweder zu wenige oder zu viele ähnliche Kandidaten zu finden sind. Daher kommt in der Praxis häufig die  $k$ -Nächste-Nachbar-Suche zum Einsatz, da die Überschaubarkeit der Ergebnismenge über den Parameter  $k$  steuerbar ist. Das Problem der  $k$ -Nächste-Nachbar-Suche ist, dass in der Ergebnismenge unähnliche Kandidaten vorhanden sein, bzw. ähnliche fehlen können.

**Definition 3.14 ( $k$ -Nächste-Nachbar-Suche)** Sei  $k \in \mathbb{N}$  und  $\omega_Q \in \Omega$  der Merkmalsdeskriptor einer Suchanfrage, der nicht notwendigerweise auch Element von  $\mathcal{W}$  sein muss. Unter einer  $k$ -Nächste-Nachbar-Suche (engl.  $k$ -nearest neighbor search) versteht man die Bestimmung einer Teilmenge

$$\mathcal{Q}_{NN} = \{\omega'_1, \omega'_2, \dots, \omega'_k\} \subseteq \mathcal{W} \quad (3.15)$$

von  $k$  zur Anfrage  $\omega_Q$  benachbarten Deskriptoren aus dem Merkmalsdatensatz unter der Voraussetzung, dass die Eigenschaften

$$\delta(\omega_Q, \omega'_i) \leq \delta(\omega_Q, \omega'_{i+1}), \quad i = 1, 2, \dots, k-1 \quad (3.16)$$

und

$$\delta(\omega_Q, \omega'_k) \leq \delta(\omega_Q, \eta), \quad \forall \eta \in \mathcal{W} \setminus \mathcal{Q} \quad (3.17)$$

gelten. ■

**Bemerkung:** Die Menge  $\mathcal{Q}$  ist eindeutig bis auf Permutationen innerhalb von Gruppen von Merkmalsdeskriptoren  $\omega'_i$  mit gleicher Distanz zur Anfrage  $\omega_Q$ . Speziell kann die Menge  $\mathcal{Q}$  nicht eindeutig sein, falls es mehrere Merkmalsdeskriptoren  $\omega_j \in \mathcal{W}$  gibt, deren Abstand zu  $\omega_Q$  gleich  $\delta(\omega_Q, \omega'_k)$  ist.

**Definition 3.15 (Bereichssuche)** Sei  $\delta_{\max} \geq 0$  eine vorgegebene Suchdistanz und  $\omega_Q \in \Omega$  der Merkmalsdeskriptor einer Suchanfrage, der nicht notwendigerweise auch Element von  $\mathcal{W}$  sein muss. Unter einer Bereichssuche (engl. range search) versteht man die Bestimmung einer Teilmenge

$$\mathcal{Q}_R = \{\omega'_1, \omega'_2, \dots, \omega'_\ell\} \subseteq \mathcal{W} \quad (3.18)$$

von zur Anfrage  $\omega_Q$  benachbarten Deskriptoren aus dem Merkmalsdatensatz unter der Voraussetzung, dass die Eigenschaften

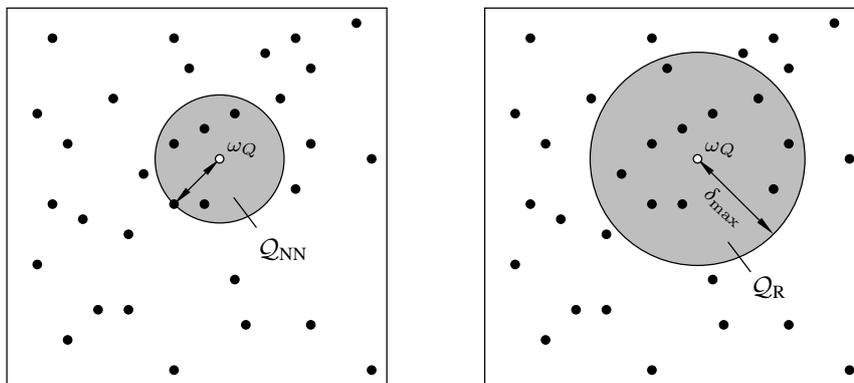
$$\delta(\omega_Q, \omega'_i) \leq \delta(\omega_Q, \omega'_{i+1}), \quad i = 1, 2, \dots, |\mathcal{Q}| \quad (3.19)$$

und

$$\delta(\omega_Q, \omega'_i) \leq \delta_{\max}, \quad \forall \omega'_i \in \mathcal{Q} \quad (3.20)$$

gelten. ■

**Beispiel 3.4 (Nachbarschaftssuche)** Abbildung 3.3 verdeutlicht die zwei prinzipiell unterschiedlichen Varianten von Nachbarschaftssuchen anhand eines zweidimensionalen Merkmalsraumes  $[0, 1] \times [0, 1]$  mit Euklidischem Distanzmaß. Die Deskriptoren des Merkmalsdatensatzes  $\mathcal{W}$  sind mit schwarzen Punkten gekennzeichnet, der Merkmalsdeskriptor der Suchanfrage  $\omega_Q$  mit einem weißen Punkt. 3.3(a) stellt die mittels einer  $k$ -Nächste-Nachbar-Suche ermittelte Ergebnismenge  $\mathcal{Q}_{NN}$  für  $k = 5$  dar, 3.3(b) die durch eine Bereichssuche bestimmte Ergebnismenge  $\mathcal{Q}_R$  für  $\delta_{\max} = \sqrt{0.08}$ .

(a)  $k$ -Nächste-Nachbar-Suche

(b) Bereichssuche

**Abbildung 3.3:** Schematische Darstellung der Nachbarschaftssuche im metrischen Merkmalsraum



# Kapitel 4

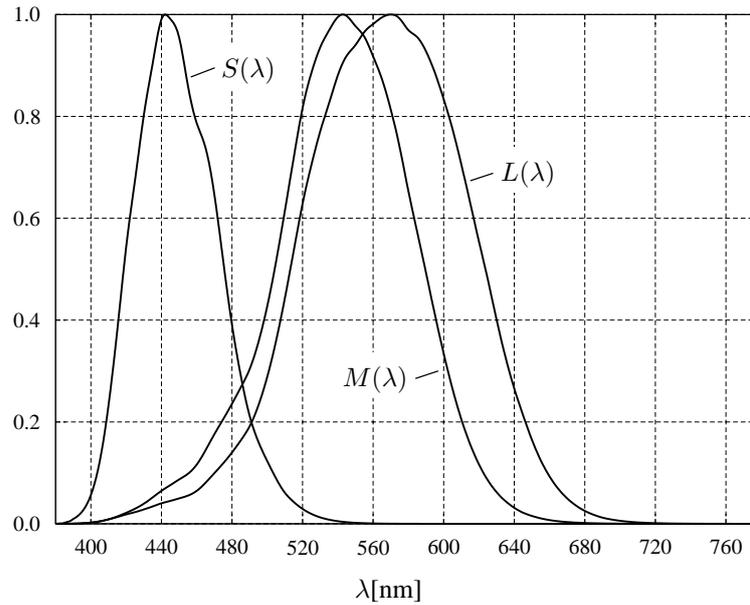
## Farbe

Eine wichtige Quelle visueller Information in digitalen Bildern ist das Medium Farbe. Um Farbe für die Beschreibung von Bildinhalten im CBIR nutzbar zu machen, wird das Phänomen Farbe zunächst näher untersucht.

### 4.1 Stand der Technik

Es wurde experimentell nachgewiesen, dass drei Arten von farbempfindlichen Rezeptoren (*S*-, *M*- und *L*-Zapfen – *engl. short-, middle-, and long-wavelength cones*) in der menschlichen Retina existieren, die für das Farbempfinden des Menschen verantwortlich sind [148]. Diese Rezeptoren haben unterschiedliche spektrale Empfindlichkeiten. Sie umfassen Spektralbänder mit maximalen Sensivitäten bei 442nm, 543nm bzw. 570nm, die sich jedoch beträchtlich gegenseitig überlagern [150][149] (siehe Abbildung 4.1). Die ungleichmäßige Farbempfindlichkeit des menschlichen Auges und die damit verbundenen Nichtlinearitäten machen die technische Darstellung von Farbe äußerst schwierig.

Die Existenz der drei verschiedenen Arten von Farbrezeptoren bedingt, dass genau drei unabhängige numerische Parameter zur Beschreibung einer Farbe benötigt werden. Daher ist die Grundlage der Farbmessung (Colorimetrie) ein trichromatisches Farbsystem, mit dem Farben als Linearkombination von drei Primärfarben (*engl. primary colors*) dargestellt werden können. Ein technisch leicht zu realisierendes Primärfarbsystem bilden die monochromatischen Spektralfarben Rot (700.0nm), Grün (546.1nm) und Blau (435.8nm), die ursprünglich von der Commission Internationale de l'Éclairage (CIE) und später von der International Standardisation Organization (ISO) als Standard festgelegt wurden [77]. Die Primärfarben Rot, Grün und Blau werden im Folgenden mit *R*, *G* und *B* bzw. RGB bezeichnet.



**Abbildung 4.1:** Relative spektrale Empfindlichkeit der  $S$ -,  $M$ - und  $L$ -Zapfen

Ausgangspunkt für die Farbmessung ist der experimentelle Vergleich von Farben (*engl. color matching*). Ein Beobachter wird dabei mittels einer geeigneten Apparatur in die Lage versetzt, die Farbe einer von ihm manipulierbaren Mischung des Lichts aus den drei CIE RGB-Primärfarben so einzustellen, bis es der vorgegebenen Farbe eines monochromatischen Lichts subjektiv entspricht. Auf diese Weise wird bestimmt, welche RGB-Anteile ein Beobachter im Licht mit einer bestimmten Wellenlänge sieht. Die so ermittelten Anteile werden in Form eines Zahlentripels  $(r, g, b)$ , dem so genannten Dreifarbenwert (*engl. tristimulus value*), festgehalten.

Die Resultate dieser Versuche liefern die Grundlage der objektiven Farbmessung. Aus einer großen Anzahl von Messversuchen mit normalsichtigen Personen wurden die Mittelwerte der Dreifarbenwerte bestimmt. Trägt man diese in Abhängigkeit der Wellenlänge auf, so erhält man die Spektralwertkurven (*engl. color matching functions*) der RGB-Primärfarben (siehe Abbildung 4.2). Die Dreifarbenwerte einer beliebigen (nicht zwingenderweise monochromatischen) Farbe können nun aus Sicht eines hypothetischen Standardbeobachters wie folgt modelliert werden:

$$\begin{aligned}
 R &= \int \phi(\lambda) \bar{r}(\lambda) d\lambda \\
 G &= \int \phi(\lambda) \bar{g}(\lambda) d\lambda \\
 B &= \int \phi(\lambda) \bar{b}(\lambda) d\lambda
 \end{aligned} \tag{4.1}$$

Die Spektralwerte der RGB-Spektralwertkurven nehmen für einige bestimmte Wellenlängen negative Werte an. Dies bedeutet, dass die Farben dieser Wellenlängen nicht durch additive Mischung der RGB-Primärfarben reproduziert werden können. Daher hat die CIE 1931 durch mathematische Transformation der RGB-Primärfarben die virtuellen Primärfarben  $X$ ,  $Y$  und  $Z$  eingeführt, deren zugehörige Spektralwertkurven nur positive Werte aufweisen (siehe Abbildung 4.3):

$$\begin{aligned} X &= \int \phi(\lambda) \bar{x}(\lambda) d\lambda \\ Y &= \int \phi(\lambda) \bar{y}(\lambda) d\lambda \\ Z &= \int \phi(\lambda) \bar{z}(\lambda) d\lambda \end{aligned} \quad (4.2)$$

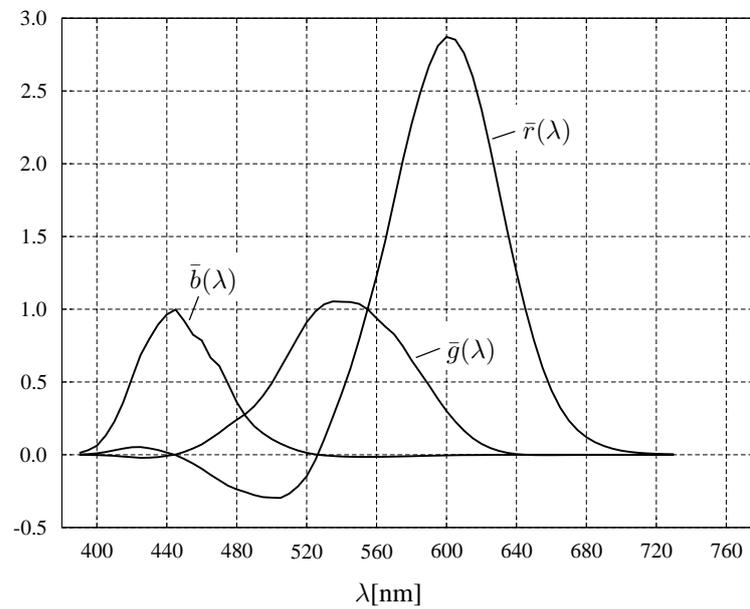
Das so modifizierte CIE XYZ-Farbsystem kann alle Farben einer bestimmten Wellenlänge des sichtbaren Spektrums durch additive Überlagerung der XYZ-Primärfarben erzeugen. Es hat aber den Nachteil, dass die verwendeten Primärfarben  $X$ ,  $Y$  und  $Z$  unanschaulich sind, da sie als theoretische Größen keinen direkten Bezug mehr zu reell existierenden Farben haben (obwohl die Auswahl so getroffen wurde, dass die Farbe Weiß durch gleiche Werte für  $X$ ,  $Y$  und  $Z$  repräsentiert wird).

Aus den XYZ-Primärfarben kann man drei neue Variablen, die Chromazitätskoordinaten  $x$ ,  $y$  und  $z$  ableiten:

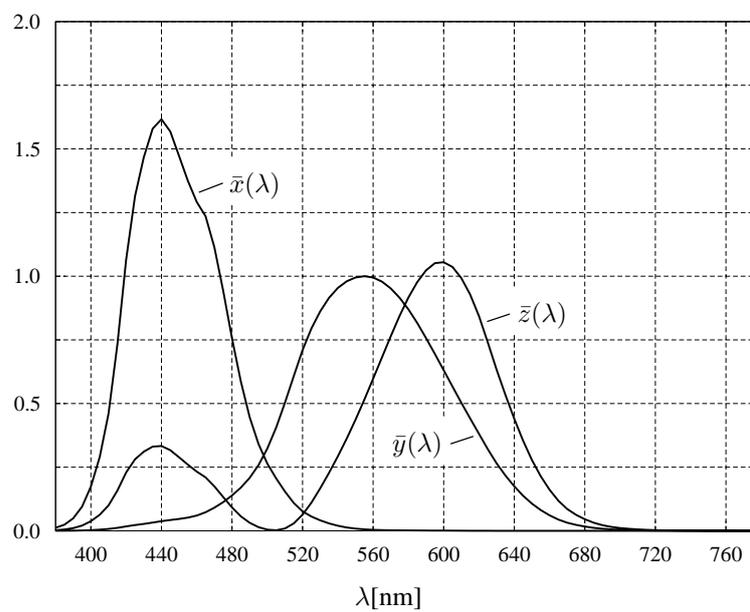
$$\begin{aligned} x &= \frac{X}{X + Y + Z} \\ y &= \frac{Y}{X + Y + Z} \\ z &= \frac{Z}{X + Y + Z} \end{aligned} \quad (4.3)$$

Da  $x + y + z = 1$ , sind nur zwei der neuen Größen unabhängig. Die Chromazitätskoordinaten  $x$  und  $y$  sind für die Spezifikation der Chromatizität einer Farbe ausreichend, und lassen sich in einem zweidimensionalen Schaubild, dem Chromatizitätsdiagramm, darstellen (siehe Abbildung 4.4). Die monochromatischen Spektralfarben liegen auf dem Rand der wie eine Zunge geformten Kurve. Alle Punkte innerhalb dieser Kurve repräsentieren eine physikalisch realisierbare Farbe. Bei  $x = \frac{1}{3}$  und  $y = \frac{1}{3}$  befindet sich in der Mitte des Diagramms die Farbe Weiß.

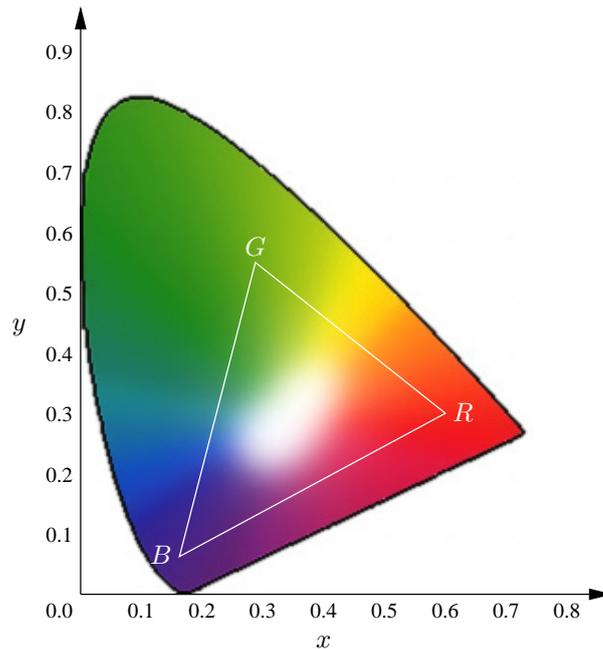
Wenn aus dem Chromatizitätsdiagramm drei beliebige Punkte als Primärfarben eines trichromatischen Farbsystems gewählt werden, so liegen alle durch additive Kombination dieser Primärfarben realisierbaren Farben innerhalb eines dreieckigen Bereiches



**Abbildung 4.2:** RGB-Spektralwertkurven eines CIE-Standardbeobachters



**Abbildung 4.3:** XYZ-Spektralwertkurven eines CIE-Standardbeobachters



**Abbildung 4.4:** CIE  $xy$ -Chromazitätsdiagramm

mit den Primärfarben als Ecken. Dieser Bereich wird Farbgamut (*engl. color gamut*) des entsprechenden Farbsystems genannt.

Das Chromazitätsdiagramm ist ein Hilfsmittel zur Farbdefinition aus physiologischer Sicht, korrespondiert aber weder mit der Farbrepräsentation technischer Geräte noch mit der Art und Weise, wie Farbe vom Menschen wahrgenommen wird. Daher ist die Definition anderer geeigneter Farbsysteme für diese Zwecke notwendig.

#### 4.1.1 RGB-Farbsystem

Das RGB-Farbsystem ist ein – physikalisch, wie auch mathematisch – additives Farbmodell und besitzt ausnahmslos positive Koeffizienten  $R$ ,  $G$  und  $B$  im Wertebereich  $[0, 1]$ . Stellt man den RGB-Farbraum als dreidimensionales, kartesisches Koordinatensystem dar, so liegt die Gesamtheit aller RGB-Farben im Einheitswürfel (siehe Abbildung 4.5). Auf der Diagonalen des Würfels ( $R = G = B$ ) liegen die Intensitätswerte (Grauwerte).

Der RGB-Farbraum eignet sich ideal für die Darstellung von Farben auf selbstleuchtenden Kathodenstrahlröhren von Monitoren, die Farbe auf der Basis von additiver Überlagerung von rotem, grünem und blauem Phosphor (Leuchtstoff) erzeugen. Die Primärfarben  $R$ ,  $G$  und  $B$  sowie der daraus resultierende Weißpunkt sind dabei geräteabhängig. In Abbildung 4.4 ist exemplarisch die Lage der Primärfarben  $R$ ,  $G$  und

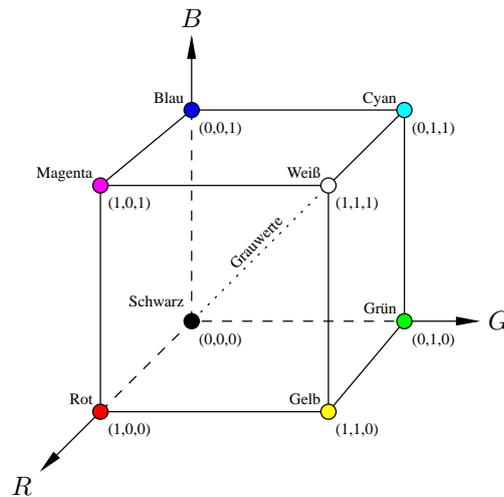


Abbildung 4.5: RGB-Farbraum

$B$  und der zugehörige Farbgamut eines RGB-Farbsystems in das Chromazitätsdiagramm eingezeichnet. Die genaue Lage von  $R$ ,  $G$  und  $B$  kann in Abhängigkeit vom jeweiligen Gerät variieren. Daher ist das RGB-Farbsystem nicht standardisiert.

Obwohl die Wahl der Primärfarben prinzipiell frei ist, sollte die Eigenschaft erhalten bleiben, dass die Mischung der Primärfarben mit maximaler Intensität ( $R = G = B = 1$ ) die Farbe Weiß ergibt. Als Quasi-Standard wird in der Regel ein Weiß definiert, das der Strahlung eines Schwarzen Körpers bei einer Temperatur von 6500 Kelvin entspricht. Diese Festlegung ermöglicht die Transformation von CIE XYZ- nach RGB-Farbwerten:

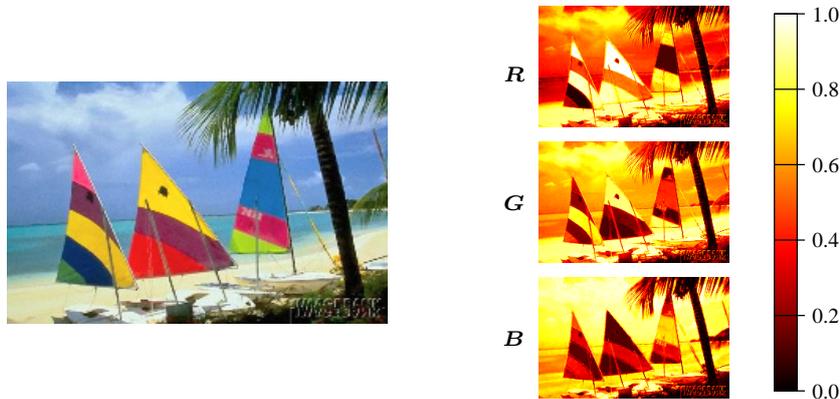
$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 2.741 & -1.147 & -0.426 \\ -1.118 & 2.028 & 0.034 \\ 0.137 & -0.332 & 1.105 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (4.4)$$

Das RGB-Farbsystem kann als geräteunabhängig angesehen werden, sobald der Weißpunkt als Referenz eines bilddarstellenden Gerätes innerhalb spezifizierter Toleranzen liegt. Das RGB-Farbsystem hat sich daher als Grundlage für digitale Farbbilder in der Datenverarbeitung durchgesetzt, da Bildinformation ohne Umwandlung der Farbwerte direkt auf einem Monitor dargestellt werden können.

Eine Farbe wird daher in der rechnergestützten Bildverarbeitung als Vektor  $(R, G, B)$  im RGB-Farbraum repräsentiert. Die digitale Repräsentation eines Farbbildes besteht folglich aus einem 3-kanaligen Multispektralbild (siehe Definition 3.5). Jeder Kanal für sich betrachtet stellt die Intensität des Lichts eines bestimmten Ausschnitts des elektromagnetischen Spektrums dar. Ein Kanal lässt sich unabhängig von den anderen als

Intensitätsbild darstellen – in diesem Zusammenhang spricht man von der Zuordnung Rot- ( $R$ ), Grün- ( $G$ ) und Blaukanal ( $B$ ).

**Beispiel 4.1 (RGB-Farbkanäle)** *Abbildung 4.6 zeigt den Rot-, Grün- und Blaukanal eines digitalen Bildes in Falschfarbendarstellung.*



**Abbildung 4.6:** RGB-Farbkanäle eines digitalen Bildes

*Es ist deutlich zu erkennen, dass eine gesamtheitliche Betrachtung erfolgen muss. Eine Farbinformation ist immer über alle Farbkanäle verteilt. Ein beträchtlicher Rotanteil bedeutet beispielsweise nicht wie man fälschlicherweise annehmen könnte, dass eine Variation der Farbe „Rot“ vorliegt. Je nach Grün- und Blauanteil könnte es sich ebenfalls um die Farben „Weiß“ oder „Gelb“ handeln.*

Die Farbwahrnehmung des Menschen orientiert sich stark an einer natürlichen Beschreibung von Farben mit den Grundeigenschaften Farbton, Sättigung und Intensität. Der Farbton bezeichnet die dominante Spektralfarbe, die Sättigung den Grad der Farbreinheit und die Intensität die Leuchtkraft einer Farbe. Diese Begriffe lassen sich jedoch in dem technisch-orientierten RGB-Farbsystem nicht wiederfinden. Aus psychologischer Sicht des Benutzers ist es wünschenswert, ein Farbsystem als Grundlage zu haben, das die Farben nach einem logischen Plan gemäß der menschlichen Wahrnehmung ordnet. Nach allgemeiner Einschätzung hat der Ansatz des amerikanischen Malers Albert Henry Munsell<sup>1</sup> [109] den größten Einfluss hinsichtlich der Konstruktion eines solchen Systems gehabt.

#### 4.1.2 Munsells Farbordnungssystem

Munsell hatte als Ziel, ein rationales Konzept zur Beschreibung des Phänomens Farbe zu entwerfen, das ohne – aus seiner Sicht – irreführende Benennungen mittels Farbna-

<sup>1</sup>1858–1918

men auskommt und statt dessen eine eindeutige dezimale Notation zur Identifikation der Farben verwendet. Als Startpunkt für die Konstruktion seines Farbordnungssystems wählte Munsell eine Kugel, auf deren Äquator in gleichen Abständen ausgewählte Farben so angeordnet sind, dass sich gegenüberliegende Paare jeweils zu Unbunt mischen (Prinzip der Kompensativität). Auf der Achse der Kugel befindet sich eine Skala von unbunten Grautönen, die sich von Weiß am Nordpol bis Schwarz am Südpol erstreckt. Horizontal zur Achse erstrecken sich Abstufungen der Farben beginnend von Grau an der Achse bis zu voll saturierten Farben an der Hülle. Munsells Bezeichnung der drei definierenden Parameter des Modells heißen „Hue“ (Farbton), „Value“ (Helligkeit) und „Chroma“ (Sättigung).

Grundlage des Farbordnungssystems ist die Anordnung der von Munsell intuitiv gewählten Hauptfarbtöne Rot (R), Gelb (Y), Grün (G), Blau (B) und Purpur (P) und der Zwischenfarbtöne Gelbrot (YR), Grüngelb (GY), Blaugrün (BG), Purpurblau (PB) und Rotpurpur (RP) auf einem zehnteiligen Farbkreis um das neutrale Grau (N) in der Kreismitte herum (siehe Abbildung 4.7). Der Farbkreis unterscheidet insgesamt 100 Farbtöne, indem jedes Farbsegment in weitere zehn Stufen unterteilt ist. Ein Farbton ist durch die jeweilige Farbstufe innerhalb des Farbsegmentes und die dem Farbsegment zugehörige Haupt- bzw. Zwischenfarbe eindeutig gekennzeichnet.

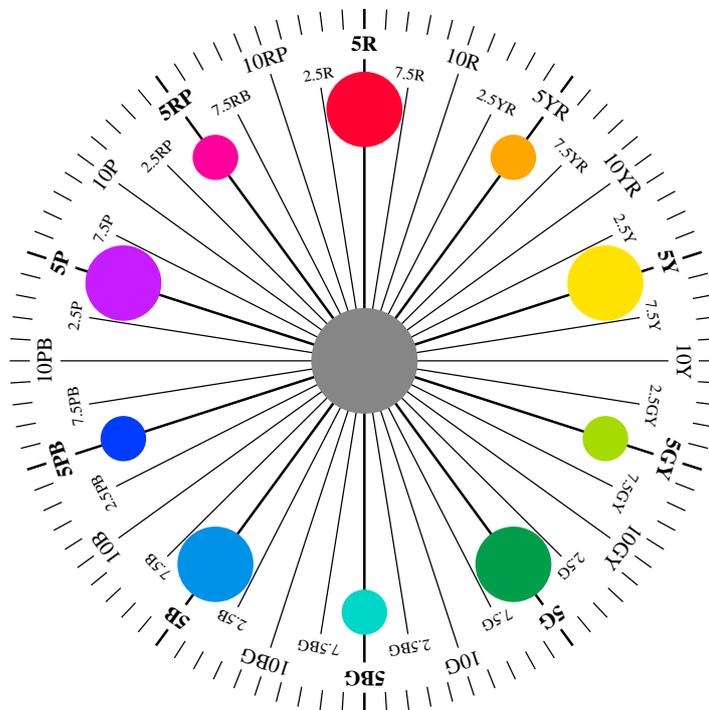


Abbildung 4.7: Munsell-Farbkreis

Der Bereich der Value-Skala ist in zehn Stufen eingeteilt. Die Stufen sind dabei so gewählt, dass sie einer gleichmäßigen Änderung der Quadrat-Wurzel aus der gemessenen reflektierten Intensität entsprechen. Die Chroma-Skala besitzt ein offenes Ende. Jede Farbe wird mittels der Munsell-Notation  $HV/C$  (Hue Value/Chroma) eindeutig bezeichnet (z.B. ist Zinnoberrot 5R 5/26). Die wohl wichtigste Feststellung von Munsell ist, dass Chroma nicht für jeden Wert von Hue und Value gleich ist. Die volle Sättigung einer Farbe kann für verschiedene Farbtöne an unterschiedlichen Stellen des Farbraumes erreicht werden. Danach erstreckt sich die Chroma-Skala, je nachdem wie unterschiedlich intensiv die entsprechenden Farben wahrgenommen werden. Entgegen der ursprünglich von Munsell angestrebten Form einer Kugel, ist der resultierende Farbkörper ein unsymmetrischer Torso (siehe Abbildung 4.8).

Das Munsell-Farbordnungssystem ordnet folglich eine endliche Anzahl von Farben (insgesamt etwa 1200) unter Berücksichtigung der empfindungsgemäßen Gleichabständigkeit der Farben in einem dreidimensionalen Farbkörper an, der hochgradig unsymmetrisch ist. Da Munsell als Maler sein Farbordnungssystem nicht mit technischen Hilfsmitteln sondern mit handbemalten Plättchen konstruierte, ergibt sich aus der Vorgehensweise Munsells keine explizite Umwandlungsvorschrift für RGB-Farbkoordinaten nach Munsells Farbordnungssystem. Obwohl in [105] eine mathematische Transformation von RGB-Farbkoordinaten in den Munsell-Farbraum vorgeschlagen wird, existiert keine technisch einfach zu realisierende Abbildung, da der Munsell-Farbkörper nicht regulär ist. Außerdem bietet das Farbordnungssystem keine Grundlage für den Vergleich von Farben, die nicht benachbart sind.

### 4.1.3 HSL-Farbsystem

Das HSL-Farbsystem wurde in erster Linie entwickelt, um Benutzern von Programmen mit graphischen Oberflächen ein sehr einfaches, intuitives Werkzeug zur Auswahl von Farben zu geben. Das Farbsystem besteht aus drei unabhängigen Parametern  $H$ ,  $S$  und  $L$ , die sinngemäß den Eigenschaften Farbton (*engl. hue*), Sättigung (*engl. saturation*) und Helligkeit (*engl. luminance*) entsprechen.

Der von den Farbkoordinaten  $H$ ,  $S$  und  $L$  aufgespannte Farbraum entspricht einem rotationssymmetrischen Zylinder. Auf dem Umfang des Zylindermantels sind die Farbtöne orthogonal zur Rotationsachse als Farbkreis angeordnet. Die entsprechende Farbkoordinate  $H$  wird als normalisierter Winkel um die Rotationsachse angegeben. Die Werte von  $H$  laufen von 0 (Rot) über  $\frac{1}{6}$  (Gelb),  $\frac{2}{6}$  (Grün),  $\frac{3}{6}$  (Blau),  $\frac{4}{6}$  (Cyan),  $\frac{5}{6}$  (Magenta) zurück zu 1 respektive 0 (Rot). In dieser Beziehung ähnelt der HSL-Farbkreis sehr dem Munsell'schen Farbkreis. Obwohl Munsell eine andere Auswahl der Farbtöne getroffen hat, sind in beiden Systemen die Farbtöne entlang eines kreisfö-

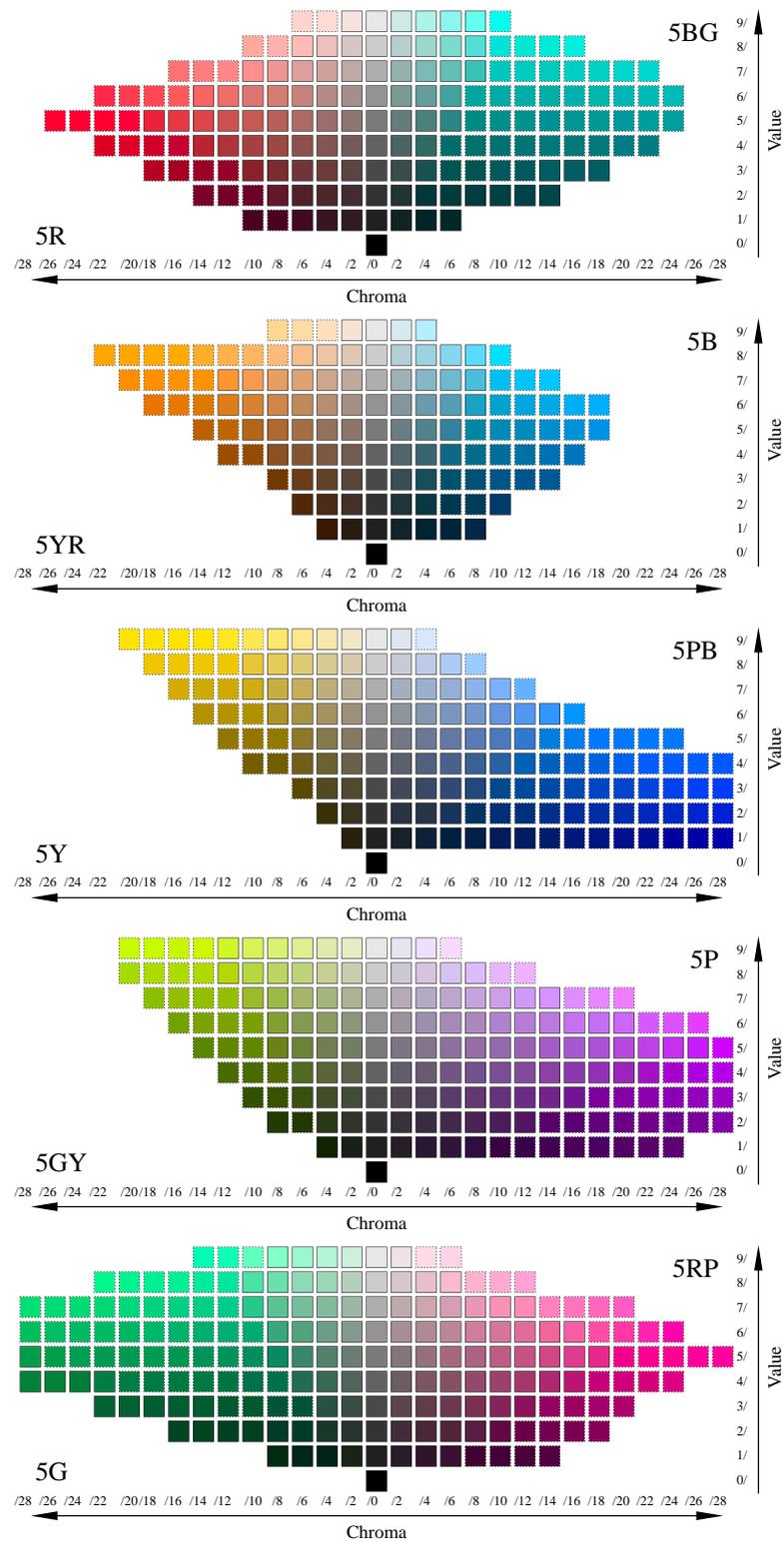


Abbildung 4.8: Value- und Chroma-Skalen der Munsell-Haupt- und Zwischenfarbtöne

migen Musters angeordnet. Die Farbkoordinate  $L$  gibt den Grad der Helligkeit einer Farbe an. Die  $L$ -Achse liegt entlang der Rotationsachse des Zylinders und ist linear von 0.0 (keine Helligkeit – Schwarz) bis 1.0 (maximale Helligkeit – Weiß) unterteilt. Im unteren Teil des Zylinders ( $L < 0.5$ ) befinden sich folglich generell dunklere Farben als im oberen Teil ( $L > 0.5$ ). Die Farbkoordinate  $S$  gibt die Sättigung einer Farbe an. Die  $S$ -Achse liegt orthogonal zur Rotationsachse und ist linear von 0.0 (keine Sättigung) bis 1.0 (volle Sättigung) unterteilt. Im Gegensatz zum Munsell'schen Chroma ist die Unterteilung der Sättigung  $S$  für alle Farbtöne und -helligkeiten einheitlich.

Die HSL-Zylinderkoordinaten lassen sich durch eine nicht-lineare Transformation aus den kartesischen Koordinaten des RGB-Farbsystems berechnen. Mit den Substitutionen  $max = \max\{R, G, B\}$  und  $min = \min\{R, G, B\}$  ergibt sich:

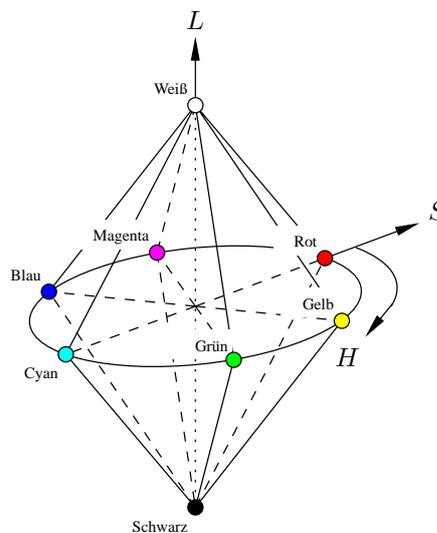
$$L = \frac{max + min}{2} \quad (4.5)$$

$$S = \begin{cases} 0 & \text{für } max = min \\ \frac{max - min}{max + min} & \text{für } max \neq min \wedge L < \frac{1}{2} \\ \frac{max - min}{2 - (max + min)} & \text{für } max \neq min \wedge L \geq \frac{1}{2} \end{cases} \quad (4.6)$$

$$H = \begin{cases} 0 & \text{für } max = min \\ \frac{1}{6} \left( 5 + \frac{R - B}{R - G} \right) & \text{für } R = max \wedge G = min \\ \frac{1}{6} \left( 1 - \frac{R - G}{R - B} \right) & \text{für } R = max \wedge G \neq min \\ \frac{1}{6} \left( 1 + \frac{G - R}{G - B} \right) & \text{für } G = max \wedge B = min \\ \frac{1}{6} \left( 3 - \frac{G - B}{G - R} \right) & \text{für } G = max \wedge B \neq min \\ \frac{1}{6} \left( 3 + \frac{B - G}{B - R} \right) & \text{für } B = max \wedge R = min \\ \frac{1}{6} \left( 5 - \frac{B - R}{B - G} \right) & \text{für } B = max \wedge R \neq min \end{cases} \quad (4.7)$$

Die Transformation ist invertierbar, daher können RGB- und HSL-Farben gegenseitig eindeutig zugeordnet werden; das HSL-Farbsystem ist in sich komplett.

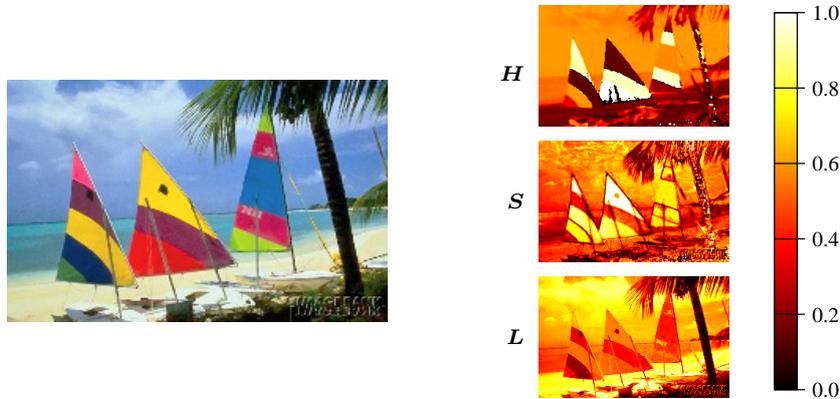
Das HSL-Farbsystem stellt eine technisch einfache Alternative zur natürlichen Beschreibung von Farben dar. In gewisser Hinsicht stellt das HSL-Farbsystem eine kontinuierliche Approximation des Munsell-Farbsystems in einem regulären Koordinatensystem dar, ohne dabei die systematischen Abweichungen der Farbempfindung des Menschen zu kompensieren. Das HSL-Farbsystem besitzt jedoch einen gravierenden Nachteil. Stellt man das HSL-Farbsystem als dreidimensionalen Farbraum dar, so zeigt sich, dass die Grundfläche des Zylinders ( $L = 0$ ) Schwarz und die Deckfläche ( $L = 1$ ) Weiß ist. Folglich ist die Deutung des Abstandes von der Rotationsachse als Sättigung falsch: Die HSL-Farbe  $(1.0, 1.0, 1.0)$  (Weiß) wäre ansonsten eine voll gesättigte Farbe – ein Widerspruch in sich selbst. Daher wird der HSL-Farbraum in der Regel nicht in Form eines Zylinders sondern eines Doppelkegels mit einer schwarzen und einer weißen Spitze interpretiert (siehe Abbildung 4.9). In diesem Fall unterliegt die Farbkoordinate  $S$  jedoch gewissen Einschränkungen – sie ist nicht mehr linear unabhängig von  $H$  und  $L$ .  $S$  repräsentiert nicht mehr die Sättigung einer Farbe, sondern deren „Bunttheit relativ zu ihrer eigenen Helligkeit“. Ein weiterer Nachteil des HSL-Farbsystems ist, dass es im Bereich sehr kleiner und sehr großer Werte von  $L$  instabile Werte von  $H$  liefert.



**Abbildung 4.9:** HSL-Farbraum

Von entscheidendem Vorteil ist hingegen, dass sich die Darstellung von Farben stark an der wahrnehmungsorientierten Beschreibung von Farben mit den Eigenschaften Farbton, Sättigung und Intensität – ähnlich dem Farbordnungssystem von Munsell – orientiert. Der Abstand zweier Farben im HSL-Farbsystem entspricht dabei ungefähr der Ähnlichkeit, wie sie vom Menschen wahrgenommen wird, vorausgesetzt man betrachtet eine hinreichend kleine Umgebung im HSL-Farbraum.

**Beispiel 4.2 (HSL-Farbkanäle)** *Abbildung 4.10 zeigt den H-, S- und L-Kanal eines digitalen Bildes in Falschfarbendarstellung.*



**Abbildung 4.10:** HSL-Farbkanäle eines digitalen Bildes

*Es ist deutlich zu erkennen, dass – wie im RGB-Farbsystem auch (siehe Beispiel 4.1) – eine gesamtheitliche Betrachtung der Farbkanäle erfolgen muss.*

## 4.2 BCR-Farbrepräsentation

Grundsätzliches Ziel des CBIR ist die Befriedigung des Informationsbedarfes eines Benutzers. Daher muss in einem CBIR-System die Art und Weise modelliert werden, wie der Benutzer die Farbinformation eines Bildes wahrnimmt. Das zur Speicherung digitaler Bilder verwendete RGB-Farbsystem (siehe 4.1.1) umfasst eine Bandbreite von mehreren Millionen verschiedener Farbnuancen. Das RGB-Farbmodell orientiert sich im Wesentlichen an dem, was aus technischer Sicht auf einem bilddarstellenden Ausgabegerät realisierbar ist. Die menschliche Wahrnehmung vermag jedoch eine solche Vielfalt von Farben nicht zu unterscheiden.

Der Mensch beurteilt die Welt der Farben anhand ihm bekannter prototypischer Muster [127]. Das menschliche Wahrnehmungsvermögen unterteilt den Farbraum in sich teilweise überlappende Farbkategorien, deren Grenzen nur unscharf definiert sind. Es ist daher nicht genau festgelegt, welche Farben zu welcher Kategorie gehören. Jede Kategorie wird typischerweise durch genau ein prototypisches Farbmuster repräsentiert, die gewissermaßen das beste Beispiel dieser Kategorie darstellt. Je unähnlicher eine bestimmte Farbe zu dem Prototyp einer Farbkategorie ist, desto weniger passt sie zu dieser Kategorie selbst.

In der Studie von Berlin und Kay [9] wurde untersucht, welchen Einfluss die Sprache auf die menschliche Wahrnehmung von Farbe hat. Eine wesentliche Schlussfolgerung

dieser Studie ist, dass dem Mensch – unabhängig von einer bestimmten Sprache – maximal elf grundsätzlich verschiedene Begriffe zur Bezeichnung von Farben zur Verfügung stehen: „Schwarz“, „Weiß“, „Rot“, „Grün“, „Gelb“, „Blau“, „Braun“, „Orange“, „Pink“, „Violett“ und „Grau“. Berlin und Kay stellen die These auf, dass nicht etwa die Sprache, sondern vielmehr das menschliche Wahrnehmungsvermögen der limitierende Faktor für die Evolution einer solch geringen Anzahl von Farben ist, die durch symbolische Farbnamen unterschieden werden können.

Menschen sind sehr konsistent bei der Zuordnung einer beliebigen Farbe zu einer der elf sprachlich unterscheidbaren Farbkategorien. Dies deutet darauf hin, dass die prototypischen Muster der entsprechenden Farbnamen durch die alltägliche Erfahrung sehr gut antrainiert sind. Soll jedoch eine bestimmte Farbe darüber hinaus näher beschrieben werden, kommt der Mensch begrifflich in Schwierigkeiten. In der Regel erweitert er dazu Farbnamen durch Zusätze, die die Abweichung der Farbe relativ zu dem jeweiligen Prototypen der Farbkategorie präzisieren (z.B. „Dunkelblau“, „Sattgrün“ oder „Gelbbraun“). Eine weitere Methode der Beschreibung ist, sich auf Objekte zu beziehen, die gedanklich mit einem bekannten charakteristischen Farbmuster verbunden werden können (z.B. „Zinnoberrot“, „Grasgrün“ oder „Türkisblau“). Der Mensch neigt generell dazu, eine Farbe nicht durch absolute, quantitative Angaben ihrer Eigenschaften, sondern durch relative, qualitative Angaben bezüglich ihm bekannter Farbmuster zu beschreiben.

Das Hauptproblem ist, dass keine allgemeingültige Vorschrift zur lexikalischen Beschreibung von Farben existiert, die eine einheitliche und präzise Benennung von Farben garantiert. Ein Farbname kann bei verschiedenen Personen in deren Vorstellung mit verschiedenen Farbmustern assoziiert sein. Als Grundlage werden daher häufig Farbnamenverzeichnisse (*engl. color name dictionaries*) verwendet, die eine Menge von Farbnamen jeweils im Zusammenhang mit den entsprechenden Mustern der bezeichneten Farben abbilden. Eine beliebige Farbe kann durch Vergleich mit den Mustern eines vorliegenden Farbnamenverzeichnisses begrifflich eindeutig bezeichnet und vermittelt werden. Der Kreativität der menschlichen Vorstellungskraft bei der Erfindung neuer Farbnamen kann so wirkungsvoll entgegengetreten werden.

Die Inter-Society Color Council (ISCC) hat 1955 auf Vorschlag des National Bureau of Standards (NBS) die Vergabe von einheitlichen Farbnamen zur eindeutigen Identifizierung von Farben vorgeschlagen [111]. Die grundlegendere Frage ist jedoch, mit welcher Granularität Farbe unterschieden werden soll (d.h. wieviele verschiedene Farben das jeweilige Farbnamenverzeichnis unterscheidet). Diese Frage lässt sich jedoch nur für einen konkreten Anwendungsfall beantworten, da in dessen Kontext bestimmte Farben bedeutungsvoller als andere sein können. Daher existiert eine Unmenge von Farbnamenverzeichnissen, die für unterschiedliche Zwecke zusammengestellt wurden.

Das wesentliche Merkmal einer für das CBIR geeigneten Farbdarstellung muss folglich sein, die Farben eines Bildes in geeigneter Weise auf eine kleine Menge relevanter Farben abzubilden. Es ergibt daher einen Sinn, eine beliebige Farbe in Relation zu einer festgelegten Menge von bekannten Farben zu beschreiben, anstatt sie durch drei unabhängige Größen von messbaren Farbeigenschaften zu quantifizieren. Die im Rahmen dieser Arbeit zu diesem Zweck entwickelte BCR-Farbrepräsentation (*engl. base color ratio*) stellt eine Farbe durch die verhältnismäßigen Anteile dar, die zur Reproduktion dieser Farbe durch Mischen einer vorher definierten Menge von Basisfarben notwendig sind.

**Definition 4.1 (BCR-Farbrepräsentation)** *Gegeben sei eine geordnete, endliche Menge*

$$C = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K\} \subset \mathbb{R}^3 \quad (4.8)$$

von  $K \in \mathbb{N}$  Basisfarben in allgemeiner Lage, die in einem trichromatischen Bezugssystem definiert sind. Jede beliebige Farbe  $\mathbf{c}$  lässt sich als eine Linearkombination dieser Basisfarben gemäß

$$\mathbf{c} = \chi_1 \mathbf{c}_1 + \chi_2 \mathbf{c}_2 + \dots + \chi_K \mathbf{c}_K \quad (4.9)$$

mit  $\chi_k \in \mathbb{R}$  und der Normierung

$$\chi_1 + \chi_2 + \dots + \chi_K = 1 \quad (4.10)$$

darstellen. Das geordnete  $K$ -Tupel

$$\chi(\mathbf{c}, C) = (\chi_1, \chi_2, \dots, \chi_K) \quad (4.11)$$

wird eine BCR-Repräsentation der Farbe  $\mathbf{c}$  bezüglich der Basisfarbmenge  $C$  genannt. ■

**Bemerkung:**  $\chi(\mathbf{c}, C)$  repräsentiert die Schwerpunktkoordinaten oder die baryzentrischen Koordinaten der Farbe  $\mathbf{c}$  bezüglich des von der Basisfarbmenge  $C$  aufgespannten  $K$ -dimensionalen affinen Raums.

Die Größen  $\chi_k$  geben den relativen Anteil der entsprechenden Basisfarben  $\mathbf{c}_k$  des Mischungsverhältnisses an, das nötig ist, um die Farbe  $\mathbf{c}$  aus der Basisfarbmenge zu erzeugen. Prinzipiell können einzelne  $\chi_k$  auch negative Werte annehmen.

**Beispiel 4.3 (BCR-Farbrepräsentation)** Gegeben sei die geordnete Menge

$$C = \{\mathbf{c}_K, \mathbf{c}_R, \mathbf{c}_G, \mathbf{c}_B, \mathbf{c}_Y, \mathbf{c}_C, \mathbf{c}_M, \mathbf{c}_W\}$$

von acht Basisfarben, die im Bezugsfarbsystem RGB definiert sind:

$$\begin{array}{llll} \mathbf{c}_K = (0, 0, 0) & \mathbf{c}_R = (1, 0, 0) & \mathbf{c}_G = (0, 1, 0) & \mathbf{c}_B = (0, 0, 1) \\ \mathbf{c}_Y = (1, 1, 0) & \mathbf{c}_C = (0, 1, 1) & \mathbf{c}_M = (1, 0, 1) & \mathbf{c}_W = (1, 1, 1) \end{array}$$

Eine BCR-Farbrepräsentation der Farbe  $\mathbf{c} = (0.6, 0.2, 0.8)$  bezüglich der Basisfarbmengemenge  $C$  ist

$$\chi(\mathbf{c}, C) = (0.0, 0.2, 0.0, 0.4, 0.0, 0.0, 0.2, 0.2)$$

Die Farbe kann folglich durch Mischen von 20% der Farbe  $\mathbf{c}_R$ , 40% der Farbe  $\mathbf{c}_B$ , 20% der Farbe  $\mathbf{c}_M$  und 20% der Farbe  $\mathbf{c}_W$  erzeugt werden. Dieser Zusammenhang ist in Abbildung 4.11 veranschaulicht.

$$\begin{array}{l} \text{c} = 0.0 \text{ c}_K + 0.2 \text{ c}_R + 0.0 \text{ c}_G + 0.4 \text{ c}_B \\ + 0.0 \text{ c}_Y + 0.0 \text{ c}_C + 0.2 \text{ c}_M + 0.2 \text{ c}_W \end{array}$$

**Abbildung 4.11:** Beispiel einer BCR-Farbrepräsentation

Diese Idee ist eigentlich nicht neu: Schon im frühen Kindesalter lernen Menschen bereits mit diesem Konzept intuitiv umzugehen, indem sie beispielsweise versuchen, eine bestimmte Farbe mittels verfügbarer Farben eines Wasserfarbkastens zu reproduzieren. Entspricht keine der vorhandenen Wasserfarben der gewünschten Farbe, so kann sie durch Mischen verschiedener Wasserfarben erzeugt werden. Der Mensch beginnt ganz selbstverständlich mit derjenigen Wasserfarbe, die der gewünschten Farbe am ähnlichsten ist. Weitere Wasserfarben werden nach subjektivem Ermessen zugemischt, um die Farbe des Gemisches schrittweise anzupassen. An dieser Vorgehensweise wird deutlich, dass der Mensch eine Farbe durch relativen Vergleich mit vorgegebenen Farbmustern beurteilt.

Die BCR-Farbrepräsentation definiert auf der Grundlage der Basisfarbmengemenge ein lokales  $K$ -dimensionales Koordinatensystem. Das Koordinatensystem steht in einer engen Beziehung zu dem Bezugsfarbsystem, in dem die Basisfarben definiert sind. Im Allgemeinen kann zur Definition der Basisfarben jedes Farbsystem als Bezug verwendet

werden, das genau drei voneinander unabhängige Farbkanäle besitzt (trichromatisches Farbsystem). Die BCR-Farbrepräsentation ist folglich kein eigenständiges Farbsystem, sondern eine alternative Repräsentationsform, die in Symbiose mit einem trichromatischen Bezugsfarbsystem existiert.

Die Definition der BCR-Farbrepräsentation manifestiert einen einheitlichen, aber flexiblen Rahmen zur Darstellung von Farbinformation, der sich weniger an farbtheoretischen, sondern mehr an pragmatischen Überlegungen der menschlichen Farbwahrnehmung orientiert.

### 4.2.1 Grundform

Prinzipielles Ziel der BCR-Farbrepräsentation ist es, jede Farbe des Bezugsfarbraums bezüglich der festgelegten Basisfarbmengens  $C$  eindeutig zu beschreiben. Dazu ist die Kenntnis sowohl einer eindeutigen Vorschrift der Abbildung

$$\begin{aligned} X : \mathbb{R}^3 &\rightarrow \mathbb{R}^K \\ \mathbf{c} &\mapsto \chi(\mathbf{c}, C) \end{aligned} \tag{4.12}$$

von Koordinaten des Bezugsfarbsystems auf BCR-Koordinaten als auch einer eindeutigen Vorschrift der inversen Abbildung  $X^{-1}$  von BCR-Koordinaten auf Koordinaten des Bezugsfarbsystems notwendig. In Definition 4.1 ist jedoch nur eine explizite, eindeutige Vorschrift von  $X^{-1}$  gegeben. Zur Bestimmung der Abbildungsvorschrift von  $X$  lässt sich aus (4.9) und (4.10) das folgende lineare Gleichungssystem mit  $K$  Unbekannten aufstellen:

$$\begin{pmatrix} | \\ | \\ \mathbf{c} \\ | \\ | \\ 1 \end{pmatrix} = \begin{pmatrix} | & | & & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_k & \dots & \mathbf{c}_K \\ | & | & & | & & | \\ 1 & 1 & & 1 & & 1 \end{pmatrix} \cdot \begin{pmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_k \\ \vdots \\ \chi_K \end{pmatrix} \tag{4.13}$$

Da bei Definition der Basisfarbmengens mittels dreidimensionaler Farbkoordinaten in einem trichromatischen Bezugsfarbsystem maximal vier linear unabhängige Gleichungen existieren, ist das Gleichungssystem für  $K > 4$  immer unterbestimmt und somit nicht eindeutig lösbar. Die Abbildung einer Farbe auf das  $K$ -dimensionale Koordinatensystem einer BCR-Farbrepräsentation mit einer Basisfarbmengens von mehr als vier Basisfarben ist daher zwangsläufig mehrdeutig, d.h. es existiert potentiell eine Vielzahl verschiedener BCR-Repräsentationen von ein und derselben Farbe.

**Beispiel 4.4 (Mehrdeutigkeit der BCR-Farbrepräsentation)** Gegeben sei die Menge  $C$  von acht Basisfarben aus Beispiel 4.3. Nach Definition 4.1 ist

$$\chi(\mathbf{c}, C) = (0.2, 0.0, 0.0, 0.0, 0.0, 0.2, 0.6, 0.0)$$

ebenfalls eine BCR-Farbrepräsentation der Farbe  $\mathbf{c} = (0.6, 0.2, 0.8)$ . Die Farbe kann folglich in Alternative zu Beispiel 4.3 durch Mischen von 20% der Farbe  $\mathbf{c}_K$ , 20% der Farbe  $\mathbf{c}_C$  und 60% der Farbe  $\mathbf{c}_M$  erzeugt werden. Dieser Zusammenhang ist in Abbildung 4.12 veranschaulicht.

$$\begin{aligned} \mathbf{c} &= 0.2 \mathbf{c}_R + 0.4 \mathbf{c}_B + 0.2 \mathbf{c}_M + 0.2 \mathbf{c}_W \\ \mathbf{c} &= 0.2 \mathbf{c}_K + 0.2 \mathbf{c}_C + 0.6 \mathbf{c}_M \end{aligned}$$

**Abbildung 4.12:** Beispiel alternativer BCR-Farbrepräsentationen

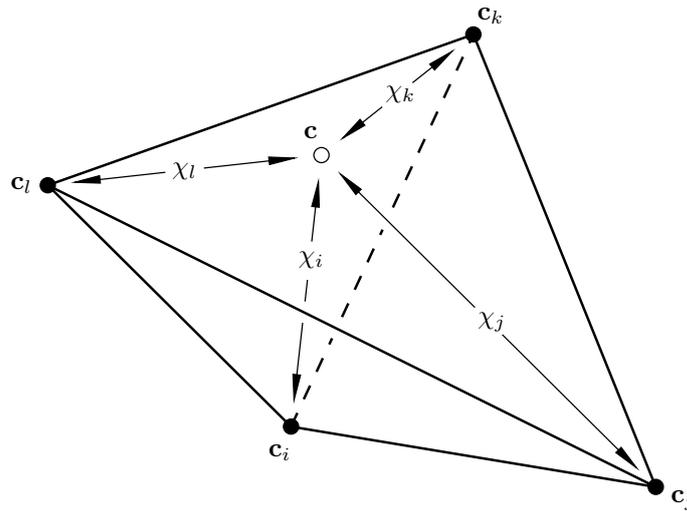
Um eine standardisierte, eindeutige BCR-Farbrepräsentation zu gewährleisten, ist daher die Definition einer Grundform notwendig, die jeder Farbe des Bezugsfarbsystems genau eine BCR-Farbrepräsentation zuordnet. Die Grundform sollte dabei aus intuitiver Sicht die „beste“ Alternative unter allen möglichen BCR-Repräsentationen darstellen. Aus praktischen Erwägungen im Umgang mit Wasserfarben können folgende Prinzipien abgeleitet werden, die die Bestimmung einer Grundform aus der Vielzahl verschiedener BCR-Repräsentationen einer Farbe ermöglichen:

- eine beliebige Farbe sollte als eine Linearkombination möglichst weniger Basisfarben dargestellt werden
- eine beliebige Farbe sollte als eine Linearkombination ausnahmslos nicht-negativer Anteile von Basisfarben dargestellt werden

Damit (4.13) eindeutig lösbar wird, dürfen als Voraussetzung nur genau soviele Unbekannte existieren wie linear unabhängige Gleichungen vorhanden sind. Dies kann mit der Einschränkung erreicht werden, dass maximal vier der  $K$  Größen  $\chi_k$  einen von Null verschiedenen Anteil haben dürfen. In diesem Fall vereinfacht sich (4.13) zu

$$\begin{pmatrix} | \\ \mathbf{c} \\ | \\ 1 \end{pmatrix} = \begin{pmatrix} | & | & | & | \\ \mathbf{c}_i & \mathbf{c}_j & \mathbf{c}_k & \mathbf{c}_l \\ | & | & | & | \\ 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \chi_i \\ \chi_j \\ \chi_k \\ \chi_l \end{pmatrix} \quad (4.14)$$

Das Gleichungssystem (4.14) ist immer genau dann eindeutig lösbar, wenn genau vier nicht-koplanare Basisfarben  $c_i$ ,  $c_j$ ,  $c_k$  und  $c_l$  für die BCR-Repräsentation der Farbe  $c$  ausgewählt werden. Geometrisch beschreiben vier nicht-koplanare Basisfarben im dreidimensionalen Bezugsfarbraum einen Tetraeder. Dieser Sachverhalt ist in Abbildung 4.13 veranschaulicht.



**Abbildung 4.13:** Vier Basisfarben als Eckpunkte eines tetraederförmigen Volumens umhüllen eine Farbe im dreidimensionalen Bezugsfarbraum

Die Lösung von (4.14) besteht darüber hinaus wie gefordert immer dann ausnahmslos aus nicht-negativen Werten, falls die Farbe  $c$  innerhalb des tetraederförmigen Volumens bzw. der konvexen Hülle der vier Basisfarben liegt.

**Definition 4.2 (Punktmenge)** Die nichtleere, endliche Menge

$$P = \{p_1, p_2, \dots, p_K\} \subset \mathbb{R}^n \quad (4.15)$$

von  $K \in \mathbb{N}$  paarweise verschiedenen Punkten im  $n$ -dimensionalen Raum heißt eine Punktmenge (engl. point set) der Größe  $K$ . ■

**Definition 4.3 (Konvexe Hülle)** Gegeben sei eine Punktmenge  $P$  der Größe  $K$  im  $n$ -dimensionalen Raum. Die Menge aller Konvexkombinationen von Punkten  $p_i \in P$

$$\text{conv}(P) = \left\{ x \in \mathbb{R}^n \mid x = \sum_{i=1}^K \lambda_i p_i, \sum_{i=1}^K \lambda_i = 1, \lambda_i \geq 0 \right\} \subset \mathbb{R}^n \quad (4.16)$$

heißt die konvexe Hülle (engl. convex hull) von  $P$ . ■

**Bemerkung:**  $\text{conv}(P)$  ist die kleinste konvexe Menge  $M$ , die  $P$  vollständig enthält, d.h. es gibt keine andere konvexe Menge  $M'$  mit  $M \subset M' \subseteq \text{conv}(P)$  und  $M \neq M'$ .

Prinzipiell existieren  $\binom{K}{4}$  Möglichkeiten, aus einer Menge von  $K$  Basisfarben vier so auszuwählen, dass der entsprechende Tetraeder im Bezugsfarbsystem eine Farbe umhüllt. Daher muss für die Definition einer Grundform der BCR-Farbsystem eine weitere Kriterium gefunden werden, dass jeder Farbe im Bezugsfarbsystem eindeutig vier Basisfarben aus der Basisfarbmenge zuordnet. Aus intuitiver Sicht sollten die Basisfarben so gewählt werden, dass sie zu der jeweiligen Farbe möglichst „nahe“ liegen, d.h. der zugehörige Tetraeder sollte die Farbe möglichst „kompakt“ umhüllen. Ganz allgemein entspricht dieses Problem der Partitionierung eines Raumes in eine Menge von Simplexen bezüglich einer vorgegebenen Menge von in diesem Raum unregelmäßig verteilten Punkten.

**Definition 4.4 (Simplex)** Gegeben sei eine Menge

$$V = \{v_0, v_1, \dots, v_k\} \subset \mathbb{R}^n \quad (4.17)$$

von  $k + 1$  affin unabhängigen Punkten. Die Menge aller Punkte

$$S = \text{conv}(V) = \left\{ x \in \mathbb{R}^n \mid x = \sum_{i=0}^k \lambda_i v_i, \sum_{i=0}^k \lambda_i = 1, \lambda_i \geq 0 \right\} \subset \mathbb{R}^n \quad (4.18)$$

heißt  $k$ -dimensionales Simplex oder kurz  $k$ -Simplex mit den Ecken  $v_0, v_1, \dots, v_k$ . ■

**Bemerkung:** Ein  $k$ -Simplex ist für  $k = 0$  ein Punkt, für  $k = 1$  eine Strecke,  $k = 2$  ein Dreieck und für  $k = 3$  ein Tetraeder.

Eines der bekanntesten Konzepte zur Lösung dieses topologischen Problems ist die Delaunay-Zerlegung. Die Delaunay-Zerlegung ist eng verwandt mit dem Voronoi-Diagramm, einer der grundlegenden Strukturen in der geometrischen Datenverarbeitung [117]. Das Voronoi-Diagramm stellt den Raum als eine verallgemeinerte topologische Karte bezüglich einer Punktmenge dar, die im Folgenden zur Unterscheidung von gewöhnlichen Punkten als *Orte* bezeichnet werden. Das Voronoi-Diagramm setzt sich aus den so genannten Voronoi-Zellen zusammen. Die Voronoi-Zelle bezeichnet die Menge aller Punkte des Raumes, die dem Ort einer Voronoi-Zelle näher liegen als jedem anderen Ort.

**Definition 4.5 (Voronoi-Zelle)** Gegeben sei eine Punktmenge  $P$  im  $n$ -dimensionalen Raum. Die Menge aller Punkte

$$\text{VC}(p_i, P) = \left\{ x \in \mathbb{R}^n \mid |p_i - x| < |p_j - x|, \forall p_j \in P \setminus \{p_i\} \right\} \quad (4.19)$$

heißt Voronoi-Zelle (engl. Voronoi cell) des Ortes  $p_i \in P$  bezüglich  $P$ . ■

**Bemerkung:** Alle Voronoi-Zellen  $\text{VC}(p_i, P)$  sind offene polyedrische Mengen mit  $p_i \in \text{VC}(p_i, P)$ . Weiterhin sind die  $\text{VC}(p_i, P)$  konvex, aber nicht zwangsweise beschränkt.

Die Struktur des Voronoi-Diagramms ist gerade durch die Punkte bestimmt, die in keiner Voronoi-Zelle liegen. Das Voronoi-Diagramm besteht aus den Voronoi-Kanten (Mengen von Punkten, die zu zwei Orten den gleichen Abstand haben) und Voronoi-Knoten (Punkte, die zu mindestens drei Orten den gleichen Abstand haben).

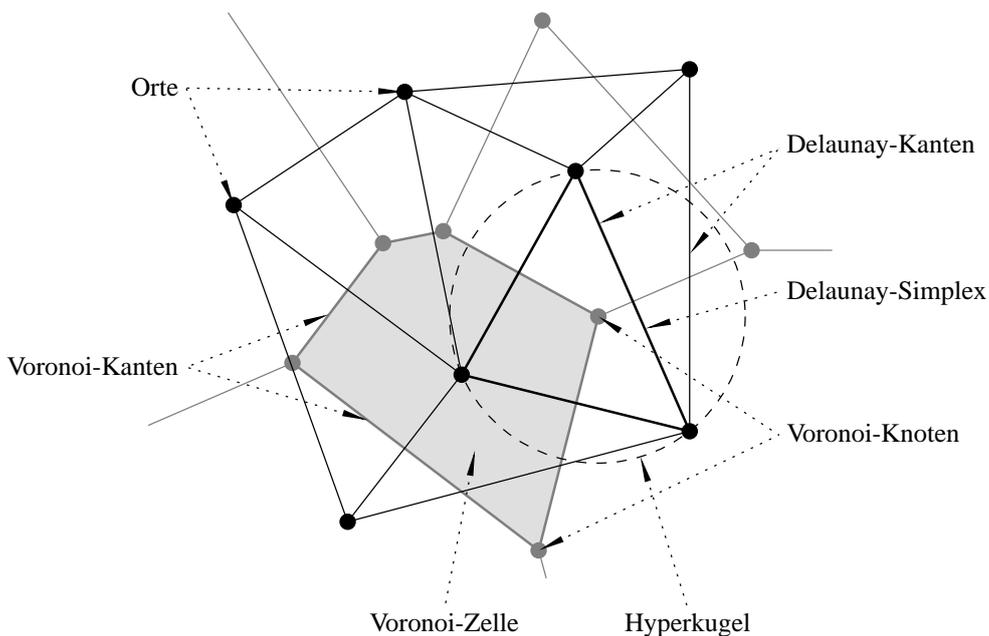
**Definition 4.6 (Voronoi-Diagramm)** Gegeben sei eine Punktmenge  $P$  im  $n$ -dimensionalen Raum. Die Menge aller Punkte

$$\text{VD}(P) = \mathbb{R}^n \setminus \bigcup_{p_i \in P} \text{VC}(p_i, P) \quad (4.20)$$

heißt *Voronoi-Diagramm* (engl. *Voronoi diagram*) bezüglich  $P$ . ■

**Bemerkung:**  $\text{VD}(P)$  hat eine endliche Anzahl von Voronoi-Kanten und -Knoten, deren Anzahl jeweils linear zur Anzahl der Orte zunimmt. Sind die Orte von  $P$  nicht kollinear, so ist  $\text{VD}(P)$  zusammenhängend.

Die Delaunay-Zerlegung ist der duale Graph des Voronoi-Diagramms. Orte, deren Voronoi-Zellen aneinandergrenzen, werden durch eine Delaunay-Kante verbunden. Abbildung 4.14 verdeutlicht diese Dualität an einem zweidimensionalen Beispiel.



**Abbildung 4.14:** Schematische Darstellung der Dualität des Voronoi-Diagramms und der Delaunay-Zerlegung

**Definition 4.7 (Delaunay-Zerlegung)** Gegeben sei eine Punktmenge  $P$  im  $n$ -dimensionalen Raum.  $E$  sei die Menge aller Verbindungsstrecken von Orten  $p_i, p_j \in P$ , die eine gemeinsame Voronoi-Kante besitzen. Der Graph

$$\text{DT}(P) = (P, E) \quad (4.21)$$

heißt die Delaunay-Zerlegung (engl. Delaunay tessellation) bezüglich  $P$ . Die Kanten von  $\text{DT}(P)$  heißen Delaunay-Kanten bezüglich  $P$ . ■

**Bemerkung:**  $\text{DT}(P)$  kann als abstrakter Graph (d.h. ohne geometrische Information) als Liste von Punktpaaren dargestellt werden. Für  $\text{VD}(P)$  existiert keine so offensichtliche Darstellungsweise.

Ursprünglich wurde die Delaunay-Zerlegung zur Triangulierung zweidimensionaler Ebenen entwickelt, d.h. der resultierende Graph setzt sich aus 2-Simplexen (Dreiecken) zusammen. Die Delaunay-Zerlegung lässt sich jedoch auf beliebig viele Dimensionen erweitern. Erweitert man die Delaunay-Zerlegung beispielsweise auf drei Dimensionen, so entspricht das Resultat einer Tetraedrisierung des entsprechenden Raumes, d.h. der resultierende Graph besteht aus 3-Simplexen (Tetraedern). Die Simplexe der Delaunay-Zerlegung besitzen die Eigenschaft der leeren Umkugeln, d.h. die Umkugel eines Delaunay-Simplex enthält keine weiteren Orte.

**Definition 4.8 (Delaunay-Simplexe)** Gegeben sei eine Punktmenge  $P$  im  $n$ -dimensionalen Raum. Geht eine Hyperkugel durch genau  $n + 1$  Orte  $q_1, q_2, \dots, q_{n+1} \in P$ , und enthält die Hyperkugel keine weiteren Orte in seinem Inneren, so heißt der eindeutig bestimmte konvexe Simplex  $S$  mit den Ecken  $q_1$  bis  $q_{n+1}$  ein Delaunay-Simplex bezüglich  $P$ . ■

**Bemerkung:** Die Vereinigungsmenge aller Delaunay-Simplexe ist die Delaunay-Zerlegung. Es ist offensichtlich, dass die Voronoi-Knoten die Hyperkugelmittelpunkte der Delaunay-Simplexe sind. Für den Sonderfall, dass auf der Hyperkugel mehr als  $n + 1$  Orte liegen, ist die Delaunay-Zerlegung nicht eindeutig.

Die aus der Delaunay-Zerlegung hervorgehenden Simplexe sind überschneidungsfrei und überdecken die konvexe Hülle der Punktmenge vollständig. Diese Eigenschaften garantieren die Eindeutigkeit des Resultats einer Delaunay-Zerlegung. Rajan hat in [123] gezeigt, dass der maximale absolute Approximationsfehler bei der Nachbildung von topologischen Relationen durch gebietsweise lineare Interpolation minimal ist, wenn die Partitionierung des Definitionsraumes gemäß der Delaunay-Zerlegung vorgenommen wird. Die Delaunay-Zerlegung stellt somit die günstigste Form der Partitionierung eines Raumes in Simplexe bezüglich einer vorgegebenen Menge von unregelmäßig verteilten Punkten dar. Die Kompaktheit der aus der Delaunay-Zerlegung

hervorgehenden Simplexe ist damit gewährleistet. Die Delaunay-Zerlegung eignet sich folglich in idealer Weise für die Bestimmung der für die Grundform der BCR-Repräsentation einer Farbe günstigsten vier Basisfarben.

Eine Grundform der BCR-Farbrepräsentation existiert jedoch nur für Farben, die innerhalb eines aus der Delaunay-Zerlegung des Bezugsfarbraums bezüglich der Basisfarbmenge hervorgehenden Tetraeder liegen. Daraus folgt unmittelbar, dass der Definitionsbereich der BCR-Grundform durch die Vereinigungsmenge aller Delaunay-Tetraeder bestimmt ist. Aufgrund der Eigenschaft der vollständigen Überdeckung des partitionierten Raumes ist die Vereinigungsmenge der Tetraeder identisch mit der konvexen Hülle der Basisfarbmenge.

**Definition 4.9 (BCR-Grundform)** Die BCR-Farbrepräsentation  $\chi(\mathbf{c}, C)$  einer beliebigen Farbe

$$\mathbf{c} \in \text{conv}(C) \quad (4.22)$$

innerhalb der konvexen Hülle der Basisfarbmenge  $C$  ist in Grundform, falls die folgenden Eigenschaften erfüllt sind:

(i) Alle Koeffizienten von  $\chi(\mathbf{c}, C)$  sind nicht-negativ.

$$\chi_i \geq 0, \quad i \in \{1, 2, \dots, K\} \quad (4.23)$$

(ii) Es existiert ein Simplex

$$S = (\mathbf{c}_i, \mathbf{c}_j, \mathbf{c}_k, \mathbf{c}_l) \in \text{DT}(C) \quad (4.24)$$

mit

$$\mathbf{c} \in \text{conv}(S) \quad (4.25)$$

so dass sich (4.9) zu

$$\mathbf{c} = \chi_i \mathbf{c}_i + \chi_j \mathbf{c}_j + \chi_k \mathbf{c}_k + \chi_l \mathbf{c}_l \quad (4.26)$$

reduziert. ■

**Bemerkung:** Existieren für eine Farbe  $\mathbf{c} \in \text{conv}(C)$  mehrere Tetraeder  $S$ , für die  $\mathbf{c} \in \text{conv}(S)$  gilt, dann handelt es sich um Tetraeder, die eine gemeinsame Seite, Kante oder Ecke besitzen. Für diesen Fall sind die BCR-Farbrepräsentationen von  $\mathbf{c}$ , die sich aus den Baryzentrischen Koordinaten der jeweiligen Tetraeder ergeben, identisch, d.h. es existiert für eine Farbe  $\mathbf{c} \in \text{conv}(C)$  immer eine eindeutige BCR-Grundform.

Die konvexe Hülle der Basisfarbmenge überdeckt den Bezugsfarbraum nicht zwingenderweise vollständig. Daher existiert nicht für jede Farbe des Bezugsfarbraums für die gewählte Basisfarbmenge eine eindeutige BCR-Farbrepräsentation in Grundform.

**Beispiel 4.5 (BCR-Grundform)** Gegeben sei die geordnete Menge

$$C = \{\mathbf{c}_N, \mathbf{c}_K, \mathbf{c}_W, \mathbf{c}_R, \mathbf{c}_Y, \mathbf{c}_G, \mathbf{c}_B, \mathbf{c}_P\}$$

von acht Basisfarben, die im Bezugssystem RGB definiert sind:

$$\begin{array}{ll} \mathbf{c}_N = (0.5, 0.5, 0.5) & \mathbf{c}_K = (0.0, 0.0, 0.0) \\ \mathbf{c}_W = (1.0, 1.0, 1.0) & \mathbf{c}_R = (1.0, 0.0, 0.2) \\ \mathbf{c}_Y = (1.0, 0.9, 0.0) & \mathbf{c}_G = (0.0, 0.6, 0.3) \\ \mathbf{c}_B = (0.0, 0.4, 0.8) & \mathbf{c}_P = (0.6, 0.1, 0.8) \end{array}$$

Die Delaunay-Zerlegung der Basisfarbmenge ergibt

$$\begin{aligned} \text{DT}(C) = \{ & (\mathbf{c}_N, \mathbf{c}_K, \mathbf{c}_R, \mathbf{c}_Y), (\mathbf{c}_N, \mathbf{c}_K, \mathbf{c}_Y, \mathbf{c}_G), \\ & (\mathbf{c}_N, \mathbf{c}_K, \mathbf{c}_G, \mathbf{c}_B), (\mathbf{c}_N, \mathbf{c}_K, \mathbf{c}_B, \mathbf{c}_P), (\mathbf{c}_N, \mathbf{c}_K, \mathbf{c}_P, \mathbf{c}_R), \\ & (\mathbf{c}_N, \mathbf{c}_W, \mathbf{c}_R, \mathbf{c}_Y), (\mathbf{c}_N, \mathbf{c}_W, \mathbf{c}_Y, \mathbf{c}_G), (\mathbf{c}_N, \mathbf{c}_W, \mathbf{c}_G, \mathbf{c}_B), \\ & (\mathbf{c}_N, \mathbf{c}_W, \mathbf{c}_B, \mathbf{c}_P), (\mathbf{c}_N, \mathbf{c}_W, \mathbf{c}_P, \mathbf{c}_R) \} \end{aligned}$$

Die Farbe  $\mathbf{c}_1 = (0.40, 0.68, 0.35)$  liegt innerhalb des Delaunay-Simplex  $S$  mit den Ecken  $\mathbf{c}_N, \mathbf{c}_W, \mathbf{c}_Y$  und  $\mathbf{c}_G$ :

$$\mathbf{c}_1 \in \text{conv}(S) \subset \text{conv}(C)$$

Daher existiert eine eindeutige BCR-Farbrepräsentation der Farbe  $\mathbf{c}$  bezüglich der Basisfarbmenge  $C$  in Grundform:

$$\chi(\mathbf{c}_1, C) = (0.2, 0.0, 0.1, 0.0, 0.2, 0.5, 0.0, 0.0)$$

Dieser Zusammenhang ist in Abbildung 4.15 veranschaulicht.

$$\begin{aligned} \mathbf{c}_1 &= 0.2 \mathbf{c}_N + 0.0 \mathbf{c}_K + 0.1 \mathbf{c}_W + 0.0 \mathbf{c}_R \\ &+ 0.2 \mathbf{c}_Y + 0.5 \mathbf{c}_G + 0.0 \mathbf{c}_B + 0.0 \mathbf{c}_P \end{aligned}$$

**Abbildung 4.15:** Beispiel einer BCR-Farbrepräsentation in Grundform

Im Gegensatz dazu liegt die Farbe  $\mathbf{c}_2 = (0.50, 0.95, 0.25)$  im Bezugssystem außerhalb der konvexen Hülle der Basisfarbmenge  $C$ . Für  $\mathbf{c}_2$  existiert daher keine eindeutige BCR-Farbrepräsentation bezüglich der Basisfarbmenge  $C$  in Grundform.

### 4.2.2 Ähnlichkeit

Um die BCR-Farbrepräsentation zur Beschreibung von Bildinhalten im Rahmen des CBIR einsetzen zu können, muss die Ähnlichkeit zweier BCR-Farbrepräsentationen mittels einer Metrik bestimmbar sein (siehe 2.3.2.2). Am einfachsten lässt sich ein Ähnlichkeitsmaß definieren, das sich aus der elementweisen Summation des Durchschnitts zweier BCR-Farbrepräsentationen ergibt. Für ein eindeutiges Ergebnis ist allerdings zwingend erforderlich, dass beide BCR-Farbrepräsentationen in Grundform vorliegen.

**Definition 4.10 (Ähnlichkeitsmaß)** Gegeben seien zwei Farben  $\mathbf{c}_1$  und  $\mathbf{c}_2$ , für die jeweils eine BCR-Farbrepräsentation bezüglich einer Basisfarbmenge  $C$  in Grundform existiert. Dann bezeichnet

$$\delta(\mathbf{c}_1, \mathbf{c}_2, C) = \delta(\chi(\mathbf{c}_1, C), \chi(\mathbf{c}_2, C)) = \sum_{k=1}^K \min\{\chi_k(\mathbf{c}_1, C), \chi_k(\mathbf{c}_2, C)\} \quad (4.27)$$

das Maß an Ähnlichkeit der Farben  $\mathbf{c}_1$  und  $\mathbf{c}_2$  bezüglich der Basisfarbmenge  $C$ . ■

**Bemerkung:** Das Ähnlichkeitsmaß  $\delta$  ist bezüglich der Delaunay-Zerlegung einer Basisfarbmenge abschnittsweise lokal definiert, d.h. zwei Farben besitzen genau dann ein gewisses Maß an Ähnlichkeit, wenn sie innerhalb benachbarter Delaunay-Tetraeder liegen, die mindestens eine Ecke gemeinsam haben. Analog dazu sind zwei Farben grundsätzlich in gleichem Maße unähnlich, falls sie innerhalb nicht benachbarter Delaunay-Tetraeder liegen, die keine Ecke gemeinsam haben.

Diese Definition der Ähnlichkeit folgt dabei dem Prinzip ausschließlich Gleiches mit Gleichem zu vergleichen. Auf diese Art und Weise gibt das Maß an, zu welchem Bruchteil das Mischungsverhältnis der Basisfarben übereinstimmt. Dies impliziert, dass der Begriff der Ähnlichkeit in der BCR-Farbrepräsentation abhängig von der ausgewählten Basisfarbmenge ist.

**Beispiel 4.6 (Ähnlichkeitsmaß)** Gegeben sei die Menge  $C$  von acht Basisfarben aus Beispiel 4.5. Die BCR-Farbrepräsentationen der Farben  $\mathbf{c}_1 = (0.40, 0.68, 0.35)$  und  $\mathbf{c}_2 = (0.05, 0.41, 0.41)$  in Grundform lauten

$$\begin{aligned} \chi(\mathbf{c}_1, C) &= (0.2, 0.0, 0.1, 0.0, 0.2, 0.5, 0.0, 0.0) \\ \chi(\mathbf{c}_2, C) &= (0.1, 0.2, 0.0, 0.0, 0.0, 0.4, 0.3, 0.0) \end{aligned}$$

Das Ähnlichkeitsmaß der Farben  $\mathbf{c}_1$  und  $\mathbf{c}_2$  bezüglich der Basisfarbmenge  $C$  berechnet sich aus dem Durchschnitt von  $\chi(\mathbf{c}_1, C)$  und  $\chi(\mathbf{c}_2, C)$ :

$$\delta(\mathbf{c}_1, \mathbf{c}_2, C) = 0.5$$

Zur relativen Ähnlichkeit der Farben  $c_1$  und  $c_2$  tragen ausschließlich die Übereinstimmungen der Mischanteile der Basisfarben  $c_N$  und  $c_G$  bei. Dieser Zusammenhang ist in Abbildung 4.16 veranschaulicht.

$$\begin{array}{r}
 c_1 = 0.2 c_N + 0.1 c_W + 0.2 c_Y + 0.5 c_G \\
 c_2 = 0.1 c_N + 0.2 c_K + 0.4 c_G + 0.3 c_B \\
 \delta = 0.1 + 0.0 + 0.0 + 0.0 + 0.4 + 0.0
 \end{array}$$

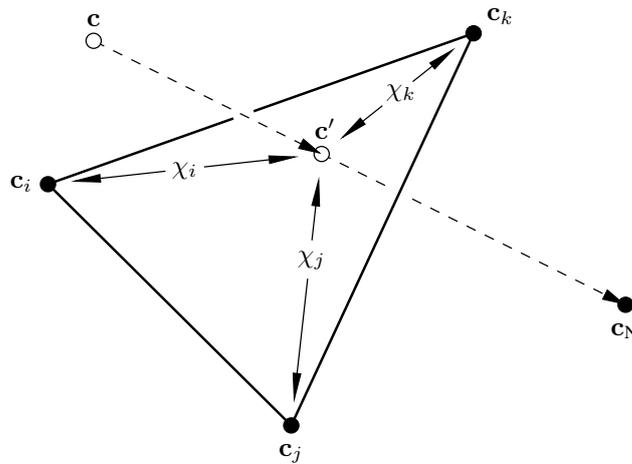
**Abbildung 4.16:** Beispiel einer Berechnung der Ähnlichkeit zweier BCR-Farbrepräsentationen

### 4.2.3 Approximation der Grundform

Ein beliebiges Bild kann grundsätzlich eine unvorhersagbare Verteilung verschiedener Farben des gesamten Bezugsfarbraums beinhalten. Daher besteht die Notwendigkeit, dass für alle Farben des Bezugsfarbraums ein Ähnlichkeitsmaß definiert sein muss. Je nach Auswahl der Basisfarbmenge für eine BCR-Farbrepräsentation kann der Fall auftreten, dass für Farben aus Teilen des Bezugsfarbraums, die außerhalb der konvexen Hülle der Basisfarbmenge liegen, keine BCR-Grundform existiert und somit für diese Farben gemäß Definition 4.10 die Bestimmung der Ähnlichkeit nicht möglich ist.

Jede Farbe  $c$ , die im Bezugsfarbraum außerhalb der konvexen Hülle der Basisfarbmenge liegt, kann immer durch eine andere ähnliche Farbe  $c'$  approximiert werden, für die eine BCR-Grundform existiert. Eine optimale Wahl stellt dabei die Farbe am Schnittpunkt der Strecke  $\overline{cc_N}$  und der Oberfläche der konvexen Hülle der Basisfarbmenge dar, wobei  $c_N$  die neutrale Farbe des Bezugsfarbraums bezeichnet. Dadurch ist gewährleistet, dass die stellvertretende Farbe  $c'$  sowohl den identischen Farbton wie  $c$  besitzt, als auch in der unmittelbaren Umgebung von  $c$  liegt. Diese Wahl impliziert, dass alle Farben entlang der Strecke  $\overline{cc'}$  durch dieselbe Farbe vertreten werden.

Die Oberfläche der konvexen Hülle einer Basisfarbmenge ist aus Dreiecken (2-Simplexen) zusammengesetzt, die ihrerseits Seitenflächen der aus der Delaunay-Zerlegung hervorgehenden Tetraeder (3-Simplexe) sind. Für die stellvertretende Farbe  $c'$  am Schnittpunkt existiert folglich eine BCR-Grundform, bei der maximal drei Basisfarben einen von Null verschiedenen Anteil haben können (siehe Abbildung 4.17). Voraussetzung ist allerdings, dass die neutrale Farbe innerhalb der konvexen Hülle der Basisfarbmenge liegt und dadurch die Existenz der stellvertretenden Farbe  $c'$  garantiert ist.



**Abbildung 4.17:** Projektion einer Farbe in Richtung der neutralen Farbe auf die konvexe Hülle der Basisfarbmenge

Zur Bestimmung der baryzentrischen Koordinaten  $\chi_i$ ,  $\chi_j$  und  $\chi_k$  des Schnittpunktes  $\mathbf{c}'$  bezüglich des von den drei Farben  $\mathbf{c}_i$ ,  $\mathbf{c}_j$  und  $\mathbf{c}_k$  gebildeten Dreiecks lässt sich das folgende lineare Gleichungssystem aufstellen:

$$\begin{pmatrix} | \\ \mathbf{c}' \\ | \\ 1 \end{pmatrix} = \begin{pmatrix} | & | & | \\ \mathbf{c}_i & \mathbf{c}_j & \mathbf{c}_k \\ | & | & | \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \chi_i \\ \chi_j \\ \chi_k \end{pmatrix} \quad (4.28)$$

Es existieren mehr unabhängige Gleichungen als Unbekannte, d.h. (4.28) ist überbestimmt. Das Gleichungssystem ist jedoch grundsätzlich lösbar, falls die Punkte  $\mathbf{c}'$ ,  $\mathbf{c}_i$ ,  $\mathbf{c}_j$  und  $\mathbf{c}_k$  koplanar sind, d.h. in einer Ebene liegen. Durch Ungenauigkeiten bei der expliziten Berechnung des Schnittpunktes kann das System schnell numerisch instabil werden. Daher müssen aufwendige Verfahren, wie z.B. die Singular Value Decomposition, zur Lösung von (4.28) eingesetzt werden.

Es gibt jedoch einen einfacheren Weg zur Bestimmung  $\chi_i$ ,  $\chi_j$  und  $\chi_k$ , der keine explizite numerische Berechnung des Schnittpunktes erfordert und zugleich numerisch stabil ist. Der Schnittpunkt  $\mathbf{c}'$  kann in Abhängigkeit von  $\mathbf{c}$  und  $\mathbf{c}_N$  in parametrisierter Form dargestellt werden:

$$\mathbf{c}' = \mathbf{c} - \alpha(\mathbf{c} - \mathbf{c}_N) \quad 0 < \alpha < 1 \quad (4.29)$$

Setzt man (4.29) in (4.28) ein, so erhält man nach Umformung das bestimmte und eindeutig lösbare Gleichungssystem

$$\begin{pmatrix} | \\ | \\ \mathbf{c} \\ | \\ | \\ 1 \end{pmatrix} = \begin{pmatrix} | & | & | & | \\ \mathbf{c}_i & \mathbf{c}_j & \mathbf{c}_k & (\mathbf{c} - \mathbf{c}_N) \\ | & | & | & | \\ 1 & 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \chi_i \\ \chi_j \\ \chi_k \\ \alpha \end{pmatrix} \quad (4.30)$$

Folglich kann jeder beliebigen Farbe des Bezugsfarbraums – direkt oder indirekt über eine stellvertretende Farbe – eine BCR-Farbrepräsentation in Grundform zugewiesen werden.

**Definition 4.11 (Approximation der BCR-Grundform)** Gegeben sei eine geordnete, endliche Menge  $C = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K\} \subset \mathbb{R}^3$  von  $K \in \mathbb{N}$  Basisfarben in allgemeiner Lage sowie die im Inneren von  $\text{conv}(C)$  liegende neutrale Farbe  $\mathbf{c}_N$ . Für alle Farben  $\mathbf{c} \in \mathbb{R}^3 \setminus \text{conv}(C)$  sei  $\mathbf{c}'$  der  $\mathbf{c}$  am nächsten liegende Schnittpunkt der Geraden durch  $\mathbf{c}$  und  $\mathbf{c}_N$  mit der konvexen Hülle  $\text{conv}(C)$ . Für  $\mathbf{c}'$  existiert ein Simplex

$$S = (\mathbf{c}_i, \mathbf{c}_j, \mathbf{c}_k) \in \text{conv}(C) \quad (4.31)$$

mit  $\mathbf{c}' \in S$ , so dass sich die Linearkombination

$$\mathbf{c} \approx \mathbf{c}' = \sum_{i=1}^K \chi_i \mathbf{c}_i \quad (4.32)$$

auf

$$\mathbf{c} \approx \mathbf{c}' = \chi_i \mathbf{c}_i + \chi_j \mathbf{c}_j + \chi_k \mathbf{c}_k \quad (4.33)$$

reduziert. Dann ist  $\chi(\mathbf{c}', C)$  die Approximation der BCR-Grundform von  $\mathbf{c}$  bezüglich der Basisfarbmengemenge  $C$ . ■

**Beispiel 4.7 (Approximation der BCR-Grundform)** Gegeben sei die Menge  $C$  von acht Basisfarben aus Beispiel 4.5. Gemäß Definition 4.10 existiert für die Farbe  $\mathbf{c} = (0.00, 1.00, 0.00)$  keine BCR-Grundform, da  $\mathbf{c} \notin \text{conv}(C)$ . Die Strecke  $\overline{\mathbf{c}\mathbf{c}_N}$  schneidet die Oberfläche der konvexen Hülle der Basisfarbmengemenge  $C$  innerhalb des Simplex  $S = (\mathbf{c}_W, \mathbf{c}_Y, \mathbf{c}_G)$ . Das Lösen des Gleichungssystems (4.30) ergibt  $\chi_W = 0.0909$ ,  $\chi_Y = 0.2098$  und  $\chi_G = 0.6993$ . Aus (4.33) folgt, dass

$$\mathbf{c}' = \chi_W \mathbf{c}_W + \chi_Y \mathbf{c}_Y + \chi_G \mathbf{c}_G = (0.3007, 0.6994, 0.3007)$$

und somit

$$\chi(\mathbf{c}, C) \approx \chi(\mathbf{c}', C) = (0.0, 0.0, 0.0909, 0.0, 0.2098, 0.6993, 0.0, 0.0).$$

#### 4.2.4 Auswahl der Basisfarben

Der Bezugsfarbraum kann durch eine wohl überlegte Auswahl von Basisfarben so strukturiert werden, dass er das Phänomen Farbe in einem bestimmten Kontext sinnvoll widerspiegelt. In diesem Zusammenhang macht es wenig Sinn, von *der* BCR-Farbrepräsentation zu sprechen. Vielmehr wird mit diesem Begriff eine ganze Klasse verschiedener Repräsentationen mit unterschiedlichen Eigenschaften beschrieben, die im Einzelfall durch die Wahl der Basisfarben konfiguriert werden kann. Daher ist für die BCR-Farbrepräsentation die Wahl der Basisfarbmenge von besonderer Bedeutung.

Das Bezugsfarbsystem, in dem die Basisfarben definiert sind, legt lediglich den Definitionsbereich der BCR-Repräsentation fest. Als mögliche Kandidaten für die Wahl des Bezugsfarbsystems kommen beispielsweise RGB, YIQ,  $U^*V^*W^*$ , CIE XYZ, CIE  $L^*u^*v^*$ , CIE  $L^*a^*b^*$  und andere in Betracht. Es ist jedoch nahe liegend, das RGB-Farbsystem als Bezugsfarbsystem zu wählen, da es für digitale Farbbilder in der rechnergestützten Bildverarbeitung etabliert ist. Daher wird im Folgenden – ohne Einschränkung der Allgemeinheit – das RGB-Farbsystem als Grundlage der BCR-Farbrepräsentation stillschweigend vorausgesetzt.

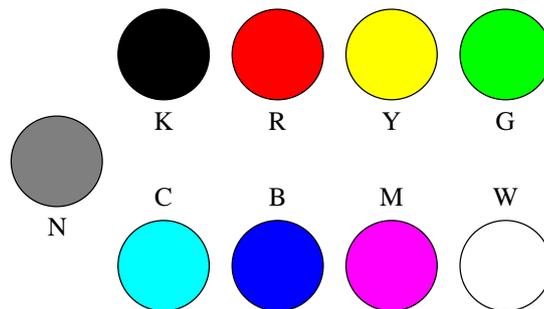
Grundlage für die Auswahl von Basisfarben ist, dass für jede Farbe des Bezugsfarbraums eine BCR-Farbrepräsentation in Grundform bezüglich der gewählten Basisfarbmenge gefunden werden kann. Dies ist der Fall, solange die neutrale Farbe innerhalb der konvexen Hülle der Basisfarbmenge liegt (siehe 4.2.3). Daher bietet es sich an, die neutrale Farbe als notwendige Basisfarbe grundsätzlich vorzuschreiben. Die neutrale Farbe des RGB-Farbsystems ist durch den Mittelpunkt des regulären Einheitswürfels eindeutig bestimmt. Sie wird im Folgenden als „Neutralgrau“ bezeichnet.

##### 4.2.4.1 RGB9-Basisfarben

Ein denkbares Ziel bei der Auswahl weiterer Basisfarben ist, den gesamten RGB-Farbraum abzudecken, um für jede Farbe die Existenz einer eindeutigen BCR-Farbrepräsentation in Grundform zu garantieren. Eine solche Basisfarbmenge muss notwendigerweise die acht Eckfarben des RGB-Einheitswürfels (siehe Abbildung 4.5) enthalten. Da keine weiteren Basisfarben außerhalb des Einheitswürfels definiert werden können, ist die konvexe Hülle aller Basisfarbmengen, die die acht Eckfarben enthalten, dann automatisch der RGB-Einheitswürfel selbst. Die minimale Basisfarbmenge, die den gesamten RGB-Farbraum überdeckt, besteht folglich aus neun Basisfarben. Im Folgenden wird diese Basisfarbmenge auch als RGB9-Basisfarbmenge bezeichnet. Die normalisierten RGB-Werte der RGB9-Basisfarben sind in Tabelle 4.1 aufgelistet, die zugehörigen Farbmuster in Abbildung 4.18 dargestellt.

**Tabelle 4.1:** Normalisierte RGB-Werte der RGB9-Basisfarbmenge

Farbname	Abk.	R	G	B
Neutralgrau	N	0.5000	0.5000	0.5000
RGB Schwarz	K	0.0000	0.0000	0.0000
RGB Rot	R	1.0000	0.0000	0.0000
RGB Gelb	Y	1.0000	1.0000	0.0000
RGB Grün	G	0.0000	1.0000	0.0000
RGB Cyan	C	0.0000	1.0000	1.0000
RGB Blau	B	0.0000	0.0000	1.0000
RGB Magenta	M	1.0000	0.0000	1.0000
RGB Weiß	W	1.0000	1.0000	1.0000

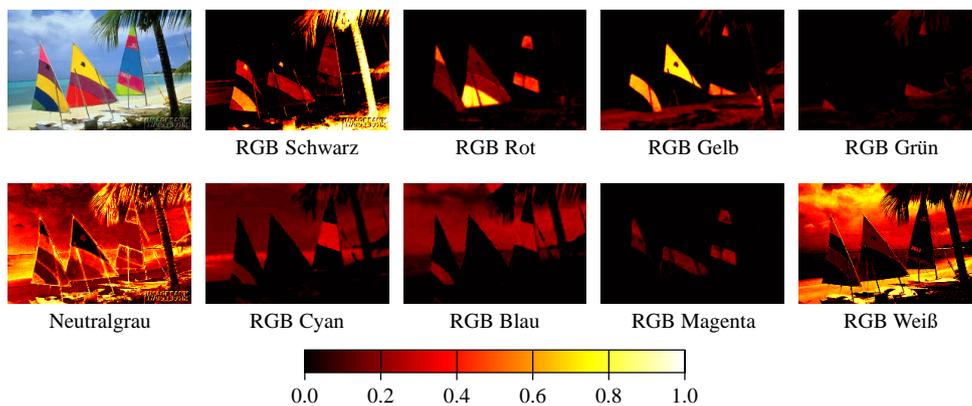
**Abbildung 4.18:** Farbmuster der RGB9-Basisfarbmenge

Die Delaunay-Zerlegung des RGB-Einheitswürfels bezüglich der RGB9-Basisfarbmenge ist nicht eindeutig. Jeweils vier Eckfarben einer Seitenfläche des Würfels bilden zusammen mit der neutralen Farbe eine Pyramide mit quadratischer Grundfläche, deren fünf Eckpunkte gemeinsam auf einer Hyperkugel liegen. Es existieren daher prinzipiell zwei gleichwertige Möglichkeiten, eine solche Pyramide in zwei Tetraeder zu zerlegen. Mittels einer Nebenbedingung kann die Eindeutigkeit jedoch künstlich erzwungen werden. Aus intuitiver Sicht ist es sinnvoll, dass jeder Delaunay-Tetraeder genau zwei „unbunte“ und zwei „bunte“ Farben als Ecken hat, d.h. neben „Neutralgrau“ muss zusätzlich entweder „RGB Schwarz“ oder „RGB Weiß“ eine Ecke des Delaunay-Tetraeders sein. Die resultierende Delaunay-Zerlegung (siehe Tabelle 4.1) stellt somit sicher, dass jeder Farbe des RGB-Farbraums eine eindeutige Grundform

bezüglich der RGB9-Basisfarbmenge zugeordnet werden kann.

Jedes digitale RGB-Farbbild kann pixelweise in die der RGB9-Basisfarbmenge entsprechende BCR-Grundform umgewandelt werden. Das Resultat kann als 9-kanaliges Bild dargestellt werden, wobei jeder Kanal die Farbinformation bezüglich einer bestimmten Basisfarbe repräsentiert. Da für jede RGB-Farbe eine eindeutige BCR-Grundform bezüglich der RGB9-Basisfarbmenge existiert, kann die BCR-Repräsentation ohne Informationsverlust wieder in die ursprüngliche RGB-Repräsentation zurückgewandelt werden.

**Beispiel 4.8 (RGB9-Farbkanäle)** *Abbildung 4.19 stellt die bildliche Information, die in den Farbkanälen der BCR-Repräsentation in Grundform enthalten ist, in Falschfarben dar. Jeder Farbkanal repräsentiert die Farbinformation des Originalbildes bezüglich einer Farbe der RGB9-Basisfarbmenge.*



**Abbildung 4.19:** RGB9-Farbkanäle eines Bildes in BCR-Grundform

*Es ist deutlich zu erkennen, dass im Gegensatz zu Abbildung 4.6 keine gesamtheitliche Betrachtung aller Farbkanäle erfolgen muss; die einzelnen Farbkanäle stellen jeweils die Bildinformation bezüglich einer bestimmten Basisfarbe dar. Jeder Bildpunkt eines Farbkanals stellt die Ähnlichkeit der Farbe im Originalbild zu dem prototypischen Muster der Basisfarbe auf einer Skala von „ähnlich“ (1.0) bis „unähnlich“ (0.0) dar.*

Die Relevanz der RGB9-Basisfarbmenge im Kontext von CBIR-Anwendungen ist jedoch unklar. Bilddatenbanken mit uneingeschränkter Domäne bestehen in der Regel vorwiegend aus Abbildern realer Szenen mit fotografischem Ursprung. Statistisch gesehen enthalten solche Bilder vorwiegend ungesättigte Farben, die in der weiteren Umgebung der unbunten Grauwert-Diagonale des RGB-Würfels liegen. Tabelle 4.2 stellt diesen Sachverhalt am Beispiel einer typischen Bilddatenbank dar. Die bunten Basisfarben (R, Y, G, C, B, M) machen statistisch gesehen lediglich 23% des Anteils an

der Grundgesamtheit aller Farben aus. Anders ausgedrückt: Aus der Sicht einer CBIR-Anwendung ist die Wahl der RGB-Eckfarben als Basisfarben daher wenig sinnvoll.

**Tabelle 4.2:** Relative Anteile der RGB9-Basisfarben an der Grundgesamtheit aller Farben am Beispiel einer typischen Bilddatenbank

Farbname	Abk.	Anteil [%]
Neutralgrau	N	32.26
RGB Schwarz	K	23.23
RGB Rot	R	7.30
RGB Gelb	Y	7.99
RGB Grün	G	0.72
RGB Cyan	C	3.65
RGB Blau	B	3.12
RGB Magenta	M	0.22
RGB Weiß	W	21.51

#### 4.2.4.2 DIN5023-Basisfarben

Die Auswahl der RGB9-Basisfarben war in erster Linie durch technische Kriterien motiviert. Vielmehr muss bei der Auswahl jedoch berücksichtigt werden, dass die Basisfarben auch aus Sicht des Anwenders im Kontext CBIR bedeutungsvoll sind. Eine Möglichkeit stellt die Übertragung des vom Menschen zur Beschreibung von Farbe verwendeten sprachlichen Konzepts auf die BCR-Farbrepräsentation dar. Dazu müssten prinzipiell die prototypischen Muster der elf in der menschlichen Sprache vorhandenen Farbnamen (siehe Seite 62) als Basisfarben gewählt werden.

Das Hauptproblem bei der Bestimmung dieser Basisfarben liegt in der unterschiedlichen Auffassung einzelner Individuen, welches Farbmuster mit einem bestimmten Farbnamen korrespondiert. Abbildung 4.20 zeigt das Ergebnis eines Experiments, das als Aufgabe zu untersuchen hatte, wie groß die Variation der Wahl solcher prototypischer Farben innerhalb einer Gruppe von Menschen mit gleichem kulturellen und sprachlichen Hintergrund ist. Ziel der Probanden war es, aus einer Auswahl von insgesamt 371 zusammengestellten Farbtafeln die aus subjektiver Sicht jeweils repräsentativsten Farbmuster der elf Farbnamen auszuwählen. Es ist deutlich zu erkennen, dass

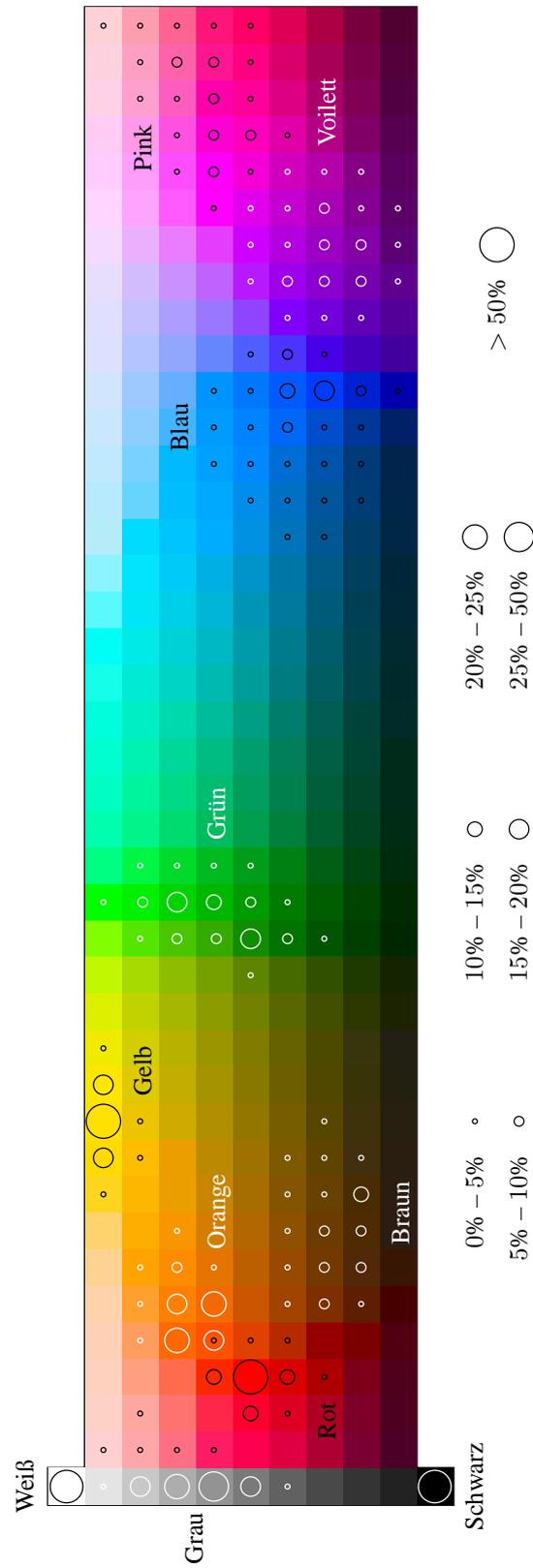


Abbildung 4.20: Unterschiedliche Auffassungen einzelner Individuen bei der Wahl prototypischer Farbmuster

bei der Bestimmung des Musters der Farben Weiß, Schwarz, Rot, Orange und Gelb relative Einigkeit herrscht. Die signifikanten Variationen bezüglich der Farbnamen Braun, Grün, Blau, Violett und Pink zeigen jedoch, dass eine eindeutige, unstrittige Bestimmung dieser Farbmuster unmöglich ist. Diese Erkenntnis ist umso erstaunlicher, wenn man bedenkt, dass der Entscheidungsraum bei diesem Experiment bereits von mehreren Millionen auf 371 unterschiedliche Farbnuancen reduziert wurde.

Das Deutsche Institut für Normung (DIN) hat 1989 eine Norm veröffentlicht, die festlegt, welches Sortiment an Farben in einem kommerziell vertriebenen Wasserfarb-Malkasten enthalten sein muss [30]:

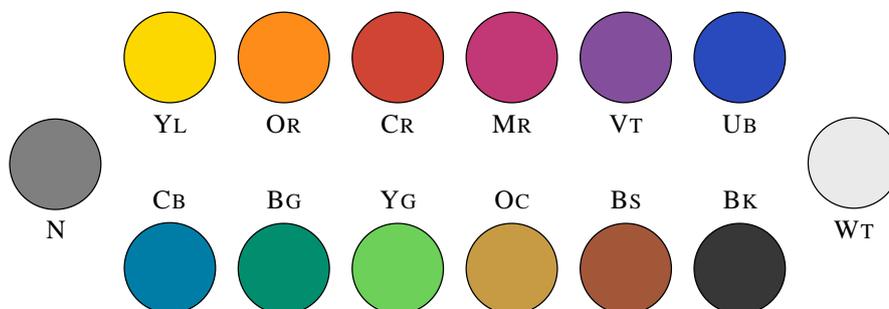
„Zweck dieser Norm ist die Zusammenstellung eines Malkastens mit einem Sortiment von 13 ausgewählten Deckfarben für Schule, Studium und Freizeit. Die Auswahl wurde so getroffen, dass ein großer Umfang an Mischfarben erzeugt werden kann. Neben den unbunten Deckfarben Schwarz und Weiß werden hierzu möglichst hochgesättigte Farben in einem angenähert gleichmäßig nach Bunttonzahl . . . gestuften Farbkreis benötigt. Gleichzeitig enthält der Farbkreis drei Farben, die angenähert dem Buntton der Grundfarben des Mehrfarbendruckes . . . entsprechen: Gelb, Magentarot, Cyanblau. Damit wird sowohl die Unterrichtung über die drucktechnische Bedeutung dieser Farben ermöglicht, als auch die Farbskala für künstlerische Zwecke erweitert. Die beiden maltechnisch bedeutenden Erdfarben (Ockergelb, Gebrannte Siena) wurden dem Farbsortiment hinzugefügt, obwohl deren Farben auch durch Mischen anderer Deckfarben erreicht werden können.“

Die Zusammenstellung dieser Farben wurde offensichtlich einerseits nach wahrnehmungsorientierten Überlegungen aus Sicht der Farblehre getroffen, andererseits wurde aber auch eine ausreichend große Überdeckung des Farbspektrums berücksichtigt. Es ist nicht überraschend, dass die Farben Grün, Blau und Braun mit jeweils zwei Farbtönen vertreten sind, die im vorgestellten Farbexperiment nicht ausreichend eindeutig bestimmt werden konnten. Die um die neutrale Farbe erweiterte Auswahl von Farben kann als ausgewogener Kompromiss zwischen konkurrierenden Erwägungen aufgefasst werden und stellt somit einen attraktiven Kandidaten als repräsentative Basisfarbmenge für CBIR-Anwendungen dar.

Die Norm spezifiziert die 13 Farben durch deren Farbmaßzahlen  $Y_{xy}$ . Unter der Annahme eines Weißpunktes von  $6500^\circ$  Kelvin und einem Gamma-Korrektur-Wert von 2.2 können diese Farbmaßzahlen in normalisierte RGB-Werte umgewandelt werden (siehe Tabelle 4.3), so dass die Farben auf einem Monitor entsprechend dargestellt werden. Die Farbmuster der Basisfarbmenge, die im Folgenden auch als DIN5023-Basisfarbmenge bezeichnet wird, sind in Abbildung 4.21 bildlich dargestellt.

**Tabelle 4.3:** Normalisierte RGB-Werte der DIN5023-Basisfarbmenge

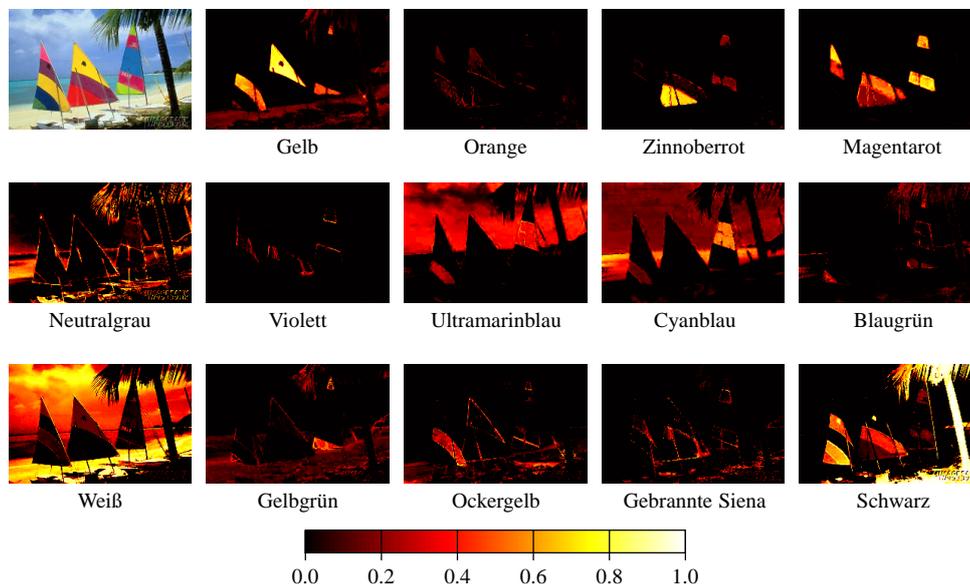
Farbname	Abk.	R	G	B
Neutralgrau	N	0.5000	0.5000	0.5000
Gelb	YL	0.9932	0.8476	0.0000
Orange	OR	0.9971	0.5483	0.0990
Zinnoberrot	CR	0.8176	0.2679	0.2023
Magentarot	MR	0.7617	0.2141	0.4615
Violett	VT	0.5131	0.3083	0.6176
Ultramarinblau	UB	0.1610	0.2887	0.7420
Cyanblau	CB	0.0000	0.4889	0.6712
Blaugrün	BG	0.0000	0.5556	0.4302
Gelbgrün	YG	0.4221	0.8157	0.3496
Ockergelb	OC	0.7963	0.6059	0.2668
Gebrannte Siena	Bs	0.6383	0.3394	0.2234
Schwarz	BK	0.2154	0.2150	0.2150
Weiß	WT	0.9205	0.9189	0.9190

**Abbildung 4.21:** Farbmuster der DIN5023-Basisfarbmenge

Die Delaunay-Zerlegung des RGB-Farbraums bezüglich der DIN5023-Basisfarbmengen ist eindeutig (siehe Tabelle A.2). Die konvexe Hülle der DIN5023-Basisfarbmengen überdeckt den RGB-Farbraum nicht vollständig, da die acht Eckfarben des RGB-Würfels nicht selbst Teil der Basisfarbmengen sind. Tatsächlich umfasst die konvexe Hülle nur etwa 23.7% des gesamten Farbraums, d.h. die Mehrheit aller in einem digitalen Bild darstellbaren Farben muss approximiert werden. Die konvexe Hülle der DIN5023-Basisfarbmengen im RGB-Farbraum ist in Abbildung 4.23 als Farbkörper in verschiedenen Ansichten dargestellt.

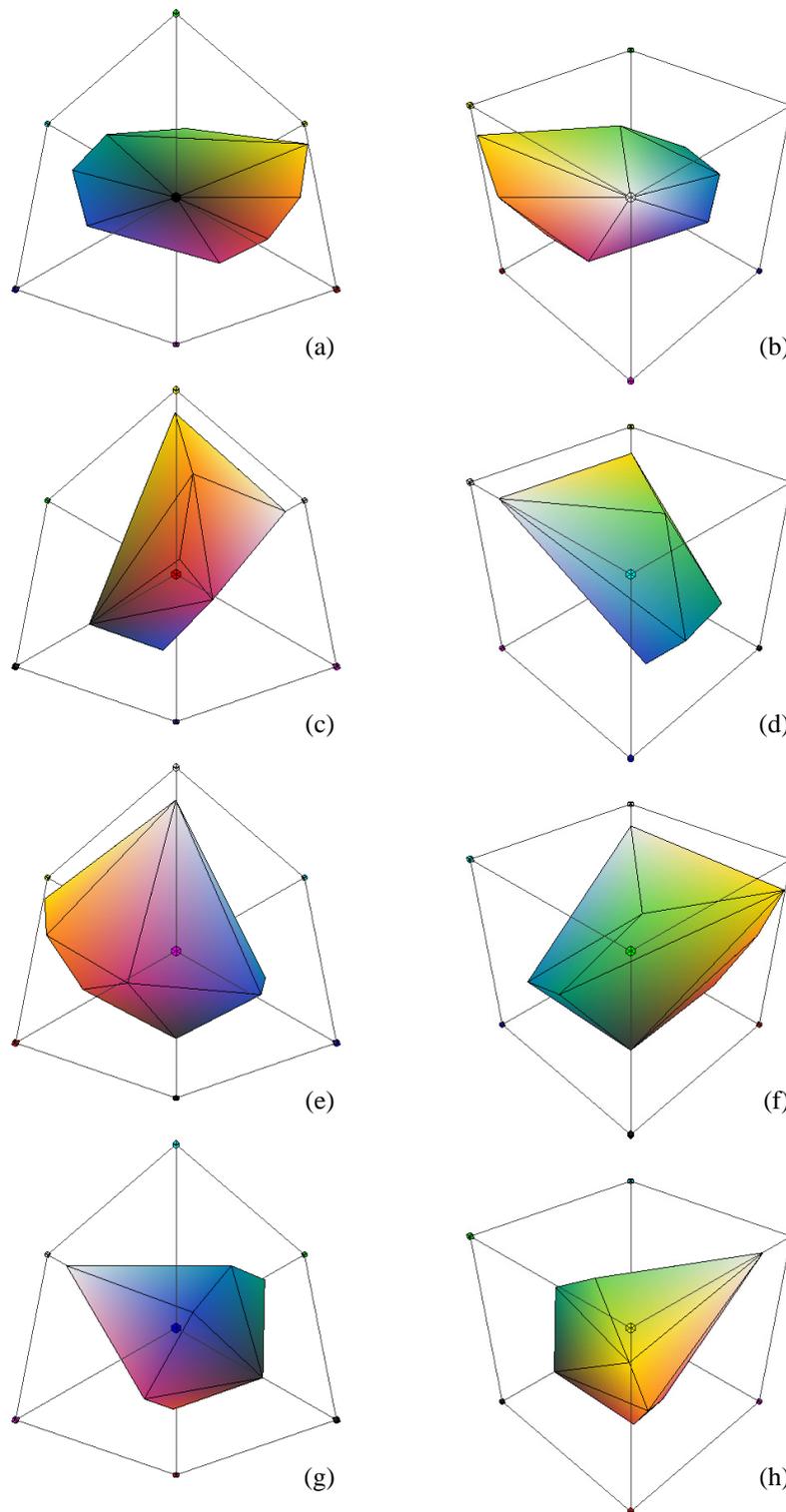
Jedes digitale RGB-Farbbild kann dennoch in eine eindeutige BCR-Grundform umgewandelt werden, wenn Farben außerhalb der konvexen Hülle der DIN5023-Basisfarbmengen jeweils durch eine entsprechende BCR-Grundform approximiert werden (siehe 4.2.3). Die resultierende BCR-Grundform kann als 13-kanaliges Bild dargestellt werden. Da für jede RGB-Farbe keine eineindeutige BCR-Grundform bezüglich der DIN5023-Basisfarbmengen existiert, kann die BCR-Repräsentation nicht wieder in die ursprüngliche RGB-Repräsentation exakt zurückgewandelt werden.

**Beispiel 4.9 (DIN5023-Farbkanäle)** *Abbildung 4.22 stellt die bildliche Information, die in den Farbkanälen der BCR-Repräsentation in Grundform enthalten ist, in Falschfarben dar. Jeder Farbkanal repräsentiert die Farbinformation des Originalbildes bezüglich einer Farbe der DIN5023-Basisfarbmengen.*



**Abbildung 4.22:** DIN5023-Farbkanäle eines Bildes in BCR-Grundform

*Es ist deutlich zu erkennen, dass die einzelnen Farbkanäle jeweils die Bildinformation bezüglich einer bestimmten Farbe der DIN5023-Basisfarbmengen darstellen.*



**Abbildung 4.23:** Konvexe Hülle der DIN5023-Basisfarbmeng im RGB-Farbraum. In (a)–(h) ist jeweils die Ansicht auf die konvexe Hülle von einer Ecke des RGB-Würfels in Blickrichtung zur neutralen Farbe dargestellt.

#### 4.2.4.3 Evaluation

Die BCR-Repräsentation wurde zum Zweck der bedeutungsvollen Darstellung farblicher Information mit Bezug auf einen bestimmten Kontext entworfen. In diesem Sinne wurde die BCR-Repräsentation durch die Wahl der DIN5023-Basisfarbmenge ausdrücklich so konfiguriert, dass die resultierenden Farbkanäle aus wahrnehmungsorientierter Sicht eines Betrachters im Kontext CBIR verwertbare Information enthalten. Da objektive Methoden zur Beurteilung der Qualität von Retrievalergebnissen im CBIR fehlen (siehe 2.4.1), kann die Wahl der DIN5023-Basisfarben nur subjektiv beurteilt werden. Anhand von Beispiel 4.9 kann jeder Betrachter für sich selbst entscheiden, ob in den einzelnen Farbkanälen relevante und aussagekräftige Farbinformation in Bezug auf das Originalbild dargestellt ist.

Objektiv kann lediglich untersucht werden, ob eine Basisfarbmenge die Grundgesamtheit aller in einem bestimmten Bilddatensatz vorhandenen Farben in adäquater Weise repräsentiert. Dazu gilt es, den Verlust an Information einzuschätzen, der durch die Approximation der Farben außerhalb der konvexen Hülle entsteht. Eine BCR-Farbrepräsentation kann mittels (4.9) jederzeit wieder in die ursprüngliche Repräsentation des Bezugsfarbraums zurücktransformiert werden. Auf diese Weise können Bilder rekonstruiert und mit den Originalen verglichen werden. Der Verlust an Farbinformation kann so für ein Bild quantitativ bestimmt werden. Abbildung 4.24 stellt diesen Sachverhalt anhand zweier Beispiele für die DIN5023-Basisfarbmenge dar. Die rekonstruierten Bilder erscheinen gegenüber den Originalen kontrastärmer, die Farben insgesamt weniger gesättigt. Der Unterschied zwischen Original und Rekonstruktion erscheint auf den ersten Blick gering. Zur genaueren Einschätzung der Diskrepanz zwischen Original und Rekonstruktion ist der relative Fehler in einer Disparitätskarte veranschaulicht (der relative Fehler ergibt sich aus der Euklidischen Distanz der RGB-Werte im Verhältnis zur Länge der Diagonalen des RGB-Einheitswürfels).

Anhand eines exemplarischen Bilddatensatzes von 1000 Bildern mit verschiedensten fotografischen Inhalten wurde empirisch ermittelt, dass 53.5% aller Farbwerte innerhalb der konvexen Hülle der DIN5023-Basisfarbmenge lagen. Folglich werden mehr als die Hälfte aller Farben von Bildern fotografischen Ursprungs durch die DIN5023-Basisfarbmenge ohne Informationsverlust in BCR-Grundform repräsentiert. Ein Fehler lässt sich entsprechend den BCR-Werten einzelnen Farbkanälen zuordnen. Auf diese Weise lässt sich abschätzen, wie groß der Beitrag einer einzelnen Basisfarbe zum Gesamtfehler ist (offensichtlich können nur Basisfarben, die auf der Oberfläche des BCR-Farbkörpers liegen, mit einem Fehler assoziiert werden). Tabelle 4.4 stellt den mittleren Fehler der DIN5023-Farbkanäle für den untersuchten Bilddatensatz dar. Die relativ kleinen Fehlerbeträge der „bunten“ Basisfarben deuten darauf hin, dass hochgesättigte Farben in der Tat in Fotografien eine geringe Auftrittswahrscheinlichkeit

**Tabelle 4.4:** Mittlerer Fehler pro DIN5023-Farbkanal

Farbname	Abk.	Fehler
Gelb	YL	0.00131
Orange	OR	0.00174
Zinnoberrot	CR	0.00123
Magentarot	MR	0.00070
Ultramarinblau	UB	0.00248
Cyanblau	CB	0.00333
Blaugrün	BG	0.00062
Gelbgrün	YG	0.00050
Schwarz	BK	0.04204
Weiß	WT	0.01546

**Abbildung 4.24:** Rekonstruktion eines Bildes aus DIN5023-Farbkanälen

haben (siehe auch Tabelle 4.2). Die Basisfarben Schwarz und Weiß tragen im Gegensatz dazu mehr als 82% zum Gesamtfehler bei. Das liegt zum einen an den generell höheren Auftretswahrscheinlichkeiten dieser Farben, zum anderen aber auch an der Tatsache, dass ein sattes Schwarz und ein helles Weiß keine Seltenheit in Fotografien sind. Die Wahl eines noch dunkleren Schwarz und eines noch helleren Weiß hätte den DIN5023-Farbkörper in den entsprechenden Regionen des RGB-Farbraums ausgedehnt: Der Gesamtfehler wäre deutlich geringer ausgefallen. Trotzdem lässt sich die Schlussfolgerung ziehen, dass die DIN5023-Basisfarbmenge – mit der genannten Einschränkung – eine geeignete Wahl für Bilddatensätze mit Inhalten fotografischen Ursprungs ist.

# Kapitel 5

## Merkmalsextraktion

Die Aufgabe von CBIR ist die Gewinnung von Information aus Bildern. Der wichtigste Prozess dabei ist – sowohl beim Menschen als auch in der rechnergestützten Verarbeitung – die Extraktion visueller Merkmale aus dem Bildinhalt. Visuelle Merkmale stellen bestimmte Aspekte eines Bildes dar, die zur Beschreibung der Bildinformation in Form von kompakten Merkmalsdeskriptoren herangezogen werden. Obwohl dieser Vorgang in der menschlichen Wahrnehmung augenblicklich vonstatten zu gehen scheint, ist der damit verbundene Prozess nicht einfach und transparent zu erklären. Dessen Modellierung stellt nach wie vor ein ungelöstes Problem nicht nur in den Ingenieurwissenschaften, sondern auch in den Natur- und Humanwissenschaften dar.

### 5.1 Stand der Technik

Die Forschungsgemeinschaft hat auf dem Gebiet des CBIR eine Unmenge verschiedener teils simpler, teils komplexer Lösungsvorschläge hervorgebracht. Die folgenden Abschnitte geben einen Überblick über die in der Literatur dokumentierten Entwicklungen im Bereich der Merkmalsextraktion. Prinzipiell wäre es ein verlockender Ansatz, alle Merkmale auf ihre Eignung im Bezug auf ein bestimmtes CBIR-Problem zu untersuchen. Diese Brute-Force-Methode ist aufgrund der exponentiellen Komplexität dieses Unterfangens aber praktisch unmöglich. Daher basiert die Gewinnung neuer Erkenntnisse auf dem Gebiet des CBIR hauptsächlich auf konzeptionellen Grundlagen und empirischen Ergebnissen.

#### 5.1.1 Farbe

Das Interesse an Farbe im CBIR kann der Tatsache beigemessen werden, dass Farbe – technisch betrachtet – im Gegensatz zu skalaren Grauwerten durch den größeren Wer-

tebereich mehr Möglichkeiten bietet, Bildinhalte zu unterscheiden. Die Eigenschaft Farbe spielt daher bei der merkmalsbasierten Beschreibung fast immer eine entscheidende Rolle. Im Folgenden werden jedoch zunächst Merkmale untersucht, die zur Beschreibung von Bildinhalten ausschließlich die Eigenschaft Farbe verwenden.

Einer der ersten Ansätze, der in diese Richtung verfolgt wurde, war die Verwendung von Farbhistogrammen, die die Verteilung der statistischen Auftrittswahrscheinlichkeiten einzelner Farben in einem Bild kennzeichnen. Swain und Ballard [154] schlagen vor, den Durchschnitt zweier Farbhistogramme, eine  $L_1$ -Metrik, als Ähnlichkeitsmaß zu verwenden. Um die Ähnlichkeit ähnlicher, aber nicht identischer Farben stärker im Maß zu berücksichtigen, haben Ioka [75] und Niblack et al. [114] eine  $L_2$ -verwandte Metrik für den Vergleich eingeführt. Stricker und Orengo [153] haben darüber hinaus die Verwendung kumulierter Histogramme vorgeschlagen, um die Empfindlichkeit der typischerweise spärlich besetzten Histogramme gegenüber kleinen Abweichungen und Störungen zu vermindern. Farbhistogramme sind seit jeher eine beliebte Wahl, wenn es um eine einfache Beschreibung von Bildern geht.

Neben Histogrammen sind etliche andere Varianten von Farbbeschreibungen im CBIR zum Einsatz gekommen. In [153] wird die Verwendung von Farbmomenten vorgeschlagen. Die mathematische Grundlage dieser Vorgehensweise begründet sich in der Tatsache, dass jede Farbverteilung durch ihre statistischen Momente charakterisiert werden kann. Da die Information in den Momenten niedriger Ordnung konzentriert ist, wurden lediglich die ersten drei Momente (Mittelwert, Varianz und Schiefe) zur Beschreibung verwendet. Als Ähnlichkeitsmaß wurde eine gewichtete euklidische Metrik vorgesehen. Die Repräsentation durch Momente stellt eine Erweiterung von Histogrammen dar, da das Histogramm das erste statistische Moment einer Verteilung ist. Pass et al. [119] haben diese Darstellung dahingehend modifiziert, dass sie zusätzlich jeden Pixel als kohärent oder nicht kohärent klassifiziert haben, je nachdem ob der Pixel zu einer großen homogenen Region gleicher Farbe gehört oder nicht.

Ein Problem jeder Darstellung der Eigenschaft Farbe ist, dass sich das zur Speicherung digitaler Bilder üblicherweise verwendete RGB-Farbmodell grundsätzlich nicht an der Farbwahrnehmung des Menschen orientiert. Eine signifikante Verbesserung aus Sicht des CBIR stellt das HSV-Farbmodell dar, dessen Farbkanäle sich stark an der wahrnehmungsorientierten Farbpfindung des Menschen orientieren und jede Farbe mit den Eigenschaften Farbton (*engl. hue*), Sättigung (*engl. saturation*) und Helligkeit (*engl. value*) darstellen. Smith und Chang [144] haben in ihrer Arbeit die Verwendung von Histogrammen untersucht, die durch Projektion von Farben in einen solchen Farbraum entstehen. Eine Unmenge verschiedener Transformationen, die jeweils die Information eines RGB-Bildes in ein anderes Farbsystem umwandeln, sind in Zusammenhang mit verschiedenen Merkmalsrepräsentationen untersucht worden. Eine um-

fassende Auswahl von Farbmodellen inklusive ihrer Vor- und Nachteile ist in [50] zu finden. Darüber hinaus haben sich Gevers und Smeulders [52][53] damit beschäftigt, Farbrepräsentationen bezüglich kleiner Änderungen der Beleuchtung und der Blickrichtung invariant zu machen.

Neben der Auswahl einer geeigneten Farbdarstellung stellt sich darüber hinaus noch das Problem, dass Menschen die Fähigkeiten besitzen, die Farbe eines Objektes als scheinbar konstant wahrzunehmen, obwohl sie tatsächlich eine große Varianz an Beleuchtungsstärke aufweist. Dieses Problem wurde als erstes von Forsyth [46] betrachtet. Finlayson [44][48] hat dessen Arbeit dahingehend verbessert, dass spekulare Reflektionen von Objekten und variierende Beleuchtungsverhältnisse von der Farbdarstellung im Bild dekoriert werden. Eine andere Methode, mit den Ungleichmäßigkeiten in der Betrachtung von Bildszenen mit Oberflächenreflektionen zurechtzukommen, ist, nach Häufungen im Farbraum zu suchen und diese in Gruppen zusammenzufassen. Pauwels und Frederix [120] stellen so fest, welche Pixel im Bild von welchem farbigen Objekt herkommen.

Das Hauptproblem jeder Darstellung von Farbe jedoch ist, dass das jeweils verwendete Farbmodell die Bandbreite mehrerer Millionen unterscheidbarer Farbnuancen umfasst. Die menschliche Wahrnehmung vermag jedoch konzeptionell nur wenige Farben zu unterscheiden (siehe 4.2). Die Berücksichtigung aller technisch realisierbaren Farbschattierungen führt zu Merkmalsbeschreibungen, die unnötige Redundanzen enthalten. In den meisten Fällen wird der Farbraum daher in eine sinnvolle Anzahl zusammenhängender Farbzellen unterteilt, die jeweils durch eine repräsentative Farbe vertreten werden. Wang et al. [171] haben den Einfluss verschiedener Unterteilungsstrategien auf die Leistungsfähigkeit von CBIR-Systemen untersucht. Sie kommen zu der Schlussfolgerung, dass adaptive Verfahren wie z.B. Clustering-Techniken generell den regulären gleichförmigen Quantisierungsmethoden überlegen sind. Ein unterschiedlicher Ansatz wurde im Rahmen dieser Arbeit verfolgt [170]. Begründet durch die Auffassung, dass die aus der Literatur bekannten Quantisierungsverfahren in den wenigsten Fällen mit der menschlichen Farbempfindung korrespondieren, wird eine Strategie zur Unterteilung von Farbräumen vorgeschlagen, die sich an einer Menge von Farben orientiert, die in der begrifflichen Welt des Menschen eine Bedeutung haben (siehe 4.2).

In den folgenden Abschnitten wird Farbe als Basiseigenschaft jeweils im Kontext mit anderen Eigenschaften und Zusammenhängen verwendet.

### 5.1.2 Textur

Der Begriff „Textur“ bezieht sich ganz allgemein auf bildliche Muster, die sich zwar aus mehreren unterscheidbaren Farben zusammensetzen, aber trotzdem gewisse Homogenitätseigenschaften aufweisen [145]. Texturen sind nicht durch einzelne Pixel, sondern durch die Beziehungen mehrerer Pixel in einer lokalen Nachbarschaft definiert. Texturmerkmale beschreiben visuelle Strukturen, die fast allen Oberflächen eigen sind. Sie können so wichtige Information über Objekte und deren wechselseitigen Beziehungen zur unmittelbaren Umgebung enthalten [65]. Im Bereich der Mustererkennung und Computer Vision hat sich die Texturanalyse von Bildmaterial in der Vergangenheit bereits als äußerst nützlich erwiesen. Daher existiert bereits eine Vielzahl von Forschungsergebnissen, auf die im Kontext CBIR zurückgegriffen werden kann.

Eine der wichtigsten Methoden zur Beschreibung von Texturmerkmalen stellen die von Haralick et al. [65] eingeführten Co-Occurrence Matrizen dar, mit deren Hilfe sie statistische Parameter aus Grauwertbildern bestimmen. Co-Occurrence Matrizen modellieren die Wahrscheinlichkeit des paarweisen Auftretens von Grauwerten bei festem Abstand mit Hilfe einer Statistik zweiter Ordnung. In der Regel bilden sich in der Matrix charakteristische Muster aus, die Rückschlüsse auf die zugrundeliegende Textur erlauben. Co-Occurrence Matrizen eignen sich aufgrund ihres Datenvolumens nicht direkt als Texturmerkmale, weshalb in [65] eine sinnvolle Teilmenge der abgeleiteten Statistik als prototypische Texturmaße vorgeschlagen wurde. Gotlieb et al. [57] haben experimentell herausgefunden, dass sich von diesen die Texturmaße *contrast*, *inverse deference moment* und *entropy* am besten zur Unterscheidung von Texturen eignen. Auf der Grundlage von Co-Occurrence Matrizen wurden später von Huang et al. [73] sowie im Rahmen dieser Arbeit von Kovalev und Volmer [85] texturbeschreibende Merkmale entwickelt, die charakteristische Farbmuster in Bildern in einer für das CBIR geeigneten Weise darstellen.

Die Ergebnisse von Untersuchungen des visuellen Wahrnehmungsapparates des Menschen haben Tamura et al. [155] dazu veranlasst, das Phänomen Textur von einem anderen Blickwinkel aus zu erforschen. Sie haben die Eigenschaften *coarseness*, *contrast*, *directionality*, *linelikeness*, *regularity* und *roughness* mittels mathematischer Modelle approximiert. Im Gegensatz zu Co-Occurrence Matrizen haben Tamura's Modelle den Vorteil, dass sie alle nachweislich in einer psychologischen Studie für den Menschen als bedeutungsvoll erachtet wurden. Diese Tatsache macht sie sehr interessant für CBIR. Das QBIC [114] und das MARS System [118] haben diese Texturmerkmale eingesetzt und weiterentwickelt. Liu und Picard verfolgen in [95] einen ähnlichen Ansatz, indem sie die Eigenschaften *periodicity*, *directionality* und *randomness* modellieren, die ungefähr der Art und Weise entsprechen, wie Menschen natursprachlich eine Textur beschreiben.

In den frühen neunziger Jahren hat sich in der Bildverarbeitung die Wavelet Transformation etabliert. Die Forschungsgemeinschaft hat schnell erkannt, dass die Wavelet-Transformation sich ebenso für die Repräsentation von Texturmerkmalen eignet, insbesondere durch ihre immanente Eigenschaft, Bildinhalte in einer Auflösungshierarchie darstellen zu können. Smith und Chang [143] verwenden die Mittelwerte und Varianzen der einzelnen Wavelet-Subbänder als Texturrepräsentationen. Laine und Fan [89] extrahieren auf ähnliche Weise charakteristische Signaturen aus Texturen mittels einer – der Wavelet Transformation verwandten – Wavelet Packet Zerlegung. Gross et al. [60] haben die Wavelet-Transformation in Verbindung mit der Karhunen-Loeve Methode und Kohonen Netzwerken zur Texturanalyse eingesetzt. In [100] haben Manjunath und Ma verschiedene Wavelet-basierte Repräsentationen im Hinblick auf ihre Eignung zur Beschreibung von Bildtexturen beurteilt. Sie kommen zu der Schlussfolgerung, dass die Gabor Wavelet-Transformation unter den Testkandidaten die besten Ergebnisse lieferte.

Andere Arten von Transformationen wurden ebenfalls untersucht. Vor allem die Beschreibung von Texturen durch fraktale Zerlegungen hat viel Aufmerksamkeit auf sich gezogen [81][160]. Eine vergleichende Studie zur Texturklassifikation mittels transformationsbasierter Eigenschaften kann in [124] gefunden werden. Eine ausführliche Beschreibung der Verfahren ist in [135] zu finden.

### 5.1.3 Form

Der theoretisch beste Weg, um semantische Information aus Bildern zu gewinnen, ist die Identifizierung sichtbarer Objekte. Dazu ist eine Zerlegung der gesamten Bildszenen in einzelne Segmente notwendig, so dass jedes Segment ausschließlich die Pixel der Silhouette eines realen Objektes enthält. Es ist offensichtlich, dass eine rechnergestützte, verlässliche Objektsegmentierung für Bilddomänen mit uneingeschränkten Inhalten nicht realisierbar ist [142]. Die Ausnahme hierzu bildet der Einsatz von technisch hochentwickelten Verfahren zur Segmentierung von Bildern aus einer sehr speziellen Anwendungsdomäne (z.B. Bilddatenbanken von Marken- und Warenzeichen [174][33]). Die Beschreibung von Objektformen setzt immer ein geeignetes Segmentierungsverfahren voraus, auf dessen spezielle Eigenschaften im Folgenden aber nicht eingegangen werden soll.

Bei der Beschreibung eines Objektes aufgrund eindeutiger Charakteristiken seiner Form spielt die Information, wo sich dieses Objekt im Bild befindet, in der Regel nur eine untergeordnete Rolle. Daher stellt sich an die Beschreibung von Objekten die Anforderung, dass sie bezüglich der affinen Transformationen Translation, Rotation und Skalierung invariant sein müssen. Generell können die Merkmale zur Beschreibung

von Form und Gestalt in zwei unterschiedliche Kategorien unterteilt werden: kontur- und regionenbasierte Verfahren.

Ein erfolgreich eingesetzter Vertreter der konturbasierten Klasse ist der Fourier-Deskriptor, der auf den Arbeiten von Zahn und Roskies [179] und Persoon und Fu [121] basiert. Die Kontur eines Objektes wird mittels der Fourier-Transformation in eine Beschreibung überführt, die die Eigenschaften der Objektgestalt reflektiert. Um den Einfluss der digitalen Rasterung auf die Kontur zu reduzieren, haben Rui et al. [129] die Verwendung eines modifizierten Fourier-Deskriptors vorgeschlagen, der gegenüber Rauschen und geometrischen Transformationen robuster ist.

Ein weiteres einfaches Verfahren allgemeiner Verwendbarkeit ist die regionenbasierte Beschreibung von Objekten mittels Moment-Invarianten. Auf der Basis der Arbeit von Hu [72], der sieben solcher Momente identifizierte, sind viele abgeleitete Versionen entwickelt worden. Yang und Algrejtsen [175] haben zunächst eine schnelle Methode zur Berechnung von Moment-Invarianten auf der Basis binärer Bilder entwickelt. Aufgrund der Tatsache, dass die meisten Invarianten im „Trial-and-Error“-Verfahren gefunden wurden, versuchten Kapur et al. [82] entsprechende geometrische Momente systematisch zu finden. In einer Studie [93] haben Li und Ma gezeigt, dass Fourier-Deskriptoren (kontur-basiert) und geometrische Momente (regionen-basiert) generell über eine einfache lineare Transformation miteinander verwandt sind.

Eakins et al. [31] [32] verfolgen einen anderen Ansatz, indem sie eine Anzahl heuristisch bestimmter Merkmale sowohl des äußeren Randes (*aspect ratio*, *circularity*, *transparency* und *relative area*) als auch innenliegender Strukturen (*right-angleness*, *sharpness*, *complexity directedness* und *straightness*) modellieren, die die menschliche Wahrnehmung gestalterischer Elemente widerspiegeln sollen.

Technische kompliziertere Methoden umfassen Ansätze, die Formen mittels kanonischer Deformationen geometrischer Primitive darstellen. In [134] und [133] werden Objekte durch Bestimmung der modalen Deformationen beschreiben, die notwendig sind, um die äußere Form des Objektes aus prototypischen Körpern zu rekonstruieren. Del Bimbo und Pala [29] verwenden elastische Konturen, um solche Übereinstimmungen zu finden. So vielversprechend diese Ansätze auch sind, so sind sie jedoch aufgrund des hohen Rechenaufwands, der zur Berechnung der Konturoptimierung nötig ist, nur eingeschränkt einsatzfähig.

In [104] haben Mehre et al. die Performanz verschiedener formbasierter Merkmale in einer umfassenden Studie untersucht. Die experimentellen Ergebnisse auf Basis einer Bilddatenbank mit 500 Marken- und Warenzeichen haben gezeigt, dass die Kombination mehrerer Verfahren jeweils die einzelnen Verfahren in der Leistung übertroffen haben. Daraus lässt sich ableiten, dass kein einzelner der verfolgten Ansätze das

Merkmal „Form“ gesamtheitlich beschreiben konnte. Das Ergebnis verwundert nicht, da auch Menschen in der Realität Schwierigkeiten haben, die Gestalt eines Körpers treffend zu beschreiben. Es ist allerdings die Unzuverlässigkeit der Segmentierungsalgorithmen, insbesondere im Kontext komplexer Bildszenen (inhomogener Hintergrund, mehrere sich möglicherweise überdeckende Objekte), die ursächlich dafür verantwortlich ist, dass formbasierte Merkmale sich im Einsatz für allgemeine Bildsammlungen nicht eignen. Daher findet man sie nur in CBIR-Systemen, die speziell für ein bestimmtes Anwendungsgebiet zugeschnitten sind. Formbasierte Merkmale sind in [161] im Detail beschrieben.

#### 5.1.4 Räumliche Anordnung

Bildstrukturen werden in der Regel durch die räumliche Anordnung (*engl. spatial layout*) von Objekten charakterisiert. Eine Beschreibung solcher komplexer Bildstrukturen erfordert, dass sowohl die Eigenschaften einzelner lokaler Bildelemente als auch deren räumliche Beziehungen untereinander mittels Merkmalswerten beschrieben werden müssen. Aus diesem Grund handelt es sich, im Vergleich zu den vorgenannten, nicht um grundsätzlich andersartige Merkmale, sondern vielmehr um wohl bekannte farb-, textur- oder formbasierte Merkmale, die bezüglich ihrer wechselseitigen Relationen erweitert wurden. Daher wird im Folgenden nur auf die Beschreibung der neuen, räumlichen Dimension eingegangen.

Der wohl einfachste und natürlichste Ansatz wurde ursprünglich von Niblack et al. [39] und Chua et al. [23] verfolgt. Sie unterteilen das Bild in eine reguläre Anordnung gleichförmiger Subbilder, für die jeweils das jeweilige Merkmal lokal berechnet wird. Eine Variation dieses Ansatzes findet sich in der Arbeit von Lu et al. [97], die ein Bild dazu in eine hierarchische Quadtree-Struktur zerlegt. Obwohl konzeptionell einfach, haben diese Ansätze auf Basis regulärer Zerlegungen in Subbilder das allgemeine Problem, dass sie sehr rechen- und speicherintensiv sind, ohne dabei eine akkurate Darstellung von Objekten und deren Relationen liefern zu können. Um einen guten Kompromiss zwischen Kompaktheit und Präzision zu erreichen, haben Stricker und Dimai [152] die Berechnung der ersten drei Farbmomente aus fünf sich teilweise überlappenden, unscharfer Regionen vorgeschlagen.

Fuh et al. [47] erweitern diesen Ansatz dahingehend, dass sie die topologischen Relationen von beliebig geformten Regionen mit homogener Farbe in Graphen beschreiben. Die Arbeiten von Hsu et al. [71] und Smith und Chang [146] gehen noch einen Schritt weiter. Sie approximieren farblich homogene Regionen durch achsenparallele Rechtecke, deren Farb- und Texturmerkmale zusammen mit Größen- und Ortsangaben der Rechtecke zur Beschreibung der räumlichen Anordnung herangezogen wer-

den. Eine alternative Form der Repräsentation wird in [19] von Chang et al. und – in erweiterter Form – in [61] von Gudivada und Raghavan vorgeschlagen. Die Segmente eines Bildes werden hier durch den Schwerpunkt und einen symbolischen Namen gekennzeichnet. Alle diese Vorgehensweisen setzen jedoch voraus, dass mittels einer Segmentierung eine bedeutungsvolle Bildzerlegung gefunden wird. Die automatische Segmentierung ist aber für Bilder mit einer entsprechend großen visuellen Variabilität nicht verlässlich. Des Weiteren entspricht das Resultat einer Menge von Bildsegmenten, die ein bestimmtes visuelles Homogenitätskriterium erfüllen, in den seltensten Fällen der eigentlich erstrebenswerten Segmentierung des Bildes in einzelne Objekte.

Aus diesem Grund gibt es in der Forschungsgemeinschaft verstärkte Bemühungen, die räumliche Anordnung direkt, ohne Segmentierung des Bildes in Regionen oder Objekte, zu beschreiben. Eine solche Beschreibung ist im Gegensatz zu den vorgenannten Verfahren nicht invariant gegenüber affinen Transformationen. Einen ersten Ansatz in diese Richtung haben Hirata und Kato [68] verfolgt, die eine Anzahl von statistischen Eigenschaften aus einem in der Auflösung stark reduzierten Bild berechnet haben. In [78] schlagen Jacobs et al. vor, die Eigenschaft der Wavelet-Transformation auszunutzen, bildliche Information in hybrider Form in Abhängigkeit sowohl vom Ort als auch von der Frequenz darstellen zu können. Sie verwenden dazu eine binäre Signatur aus den betraglich signifikantesten Koeffizienten. Diese Idee wurde von Wang et al. [172] und im Rahmen dieser Arbeit [167] aufgegriffen und die Signaturen dahingehend erweitert, dass sie eine höhere visuelle Diskriminanz aufweisen.

### 5.1.5 CBIR-Systeme

Es existiert eine Vielzahl von kommerziellen CBIR-Systemen, die von verschiedenen Unternehmen am Markt angeboten werden. Darüber hinaus existiert an Universitäten und Forschungslaboratorien eine noch viel größere Zahl prototypischer CBIR-Systeme, die meist mit dem Ziel entwickelt wurden, spezielle Verfahren experimentell zu testen. Eine vollständige Übersicht aller CBIR-Systeme zusammenzutragen, ist praktisch unmöglich. Aus diesem Grund wird an dieser Stelle auch darauf verzichtet. Stattdessen wird auf zwei exzellente und umfassende Studien von Venters und Cooper [163] sowie von Veltkamp und Tanase [162] verwiesen, die eine große Zahl von CBIR-Systemen inklusive ihrer besonderen Eigenschaften und Fähigkeiten dokumentiert haben.

## 5.2 Merkmale

Prinzipiell erscheint es verlockend, möglichst viele Merkmale für ein CBIR-System zu verwenden. Aus statistischer Sicht sollte die Zahl der Fehlerkennungen sinken, je mehr Unterscheidungskriterien verwendet werden. Das stimmt allerdings nur für den Fall, dass die zur Verfügung stehende Datenbasis unendlich groß ist. In den meisten praktischen Anwendungen stehen nur spärliche Bilddatensätze zur Verfügung, die aus statistischer Sicht nur unzureichende Indikatoren bilden, um repräsentative Aussagen bezüglich der allgemeinen Eignung von Merkmalen machen zu können. Zwar kann man die spezielle Eignung eines Merkmals für einen bestimmten Datensatz ermitteln, diese lässt sich jedoch nicht auf andere Datensätze verallgemeinern. Folglich verkleinert sich die Wahrscheinlichkeit eines guten Retrievalergebnisses, wenn die Anzahl der verwendeten Merkmale größer wird (da auch die Anzahl unbekannter Parameter größer wird). Dies stellt ein klassisches Paradoxon dar. Daher ist die prinzipielle Aufgabe bei der Auswahl von Merkmalen, eine möglichst kleine Menge starker Merkmale zu bilden, die sich für eine spezielle CBIR-Anwendung optimal eignen.

### 5.2.1 Histogramme

Eines der grundlegendsten Verfahren zur Erzeugung von Merkmalsdeskriptoren ist die Verwendung von Histogrammen. Ganz allgemein formuliert repräsentiert ein Histogramm die empirische Verteilung eines kardinalen Merkmals in einem bestimmten Objekt. Dazu muss der Merkmalsraum allerdings in eine endliche Anzahl diskreter Merkmalsklassen eingeteilt sein. Das Histogramm kann dann vereinfachend als ein-dimensionaler Vektor angesehen werden, dessen Werte die Auftrittshäufigkeiten der jeweiligen Merkmalsklassen widerspiegeln.

**Definition 5.1 (Histogramm)** Gegeben sei der Merkmalsraum  $V$  eines kardinalen Merkmals sowie das geordnete  $K$ -Tupel

$$\mathbf{h}(v) = (h_1(v), h_2(v), \dots, h_K(v)) \quad (5.1)$$

von  $K \in \mathbb{N}$  Zuweisungsfunktionen

$$h_k : V \rightarrow [0, 1], v \mapsto h_k(v) \quad (5.2)$$

mit den Eigenschaften

$$\sum_k h_k(v) = 1 \quad \forall v \quad (5.3)$$

und

$$h_k(v) \neq h_l(v) \quad \forall k \neq l \quad (5.4)$$

die einen Merkmalswert  $v \in V$  auf  $K$  diskrete Merkmalsklassen abbilden. Die normierte Verteilungsfunktion

$$\mathbf{H}(V') = \frac{1}{|V'|} \sum_{v \in V'} \mathbf{h}(v) \quad (5.5)$$

der Klassenhäufigkeiten von endlichen vielen Merkmalswerten aus einer Stichprobe  $V' \subset V$  heißt Histogramm (engl. histogram). In diesem Zusammenhang heißt  $\mathbf{h}(v)$  die Klassenzugehörigkeitsfunktion des Merkmals  $v$ . ■

**Bemerkung:** Ist die Stichprobe groß genug, dann ist das Histogramm der Verteilung der statistischen Auftrittswahrscheinlichkeiten der Merkmalsklassen ähnlich.

Um ein Histogramm als Merkmalsdeskriptor im Rahmen des CBIR einsetzen zu können, muss grundsätzlich ein Ähnlichkeitsmaß zwischen den Histogrammen zweier Stichproben definiert sein (siehe Definition 3.11). Die Komplexität der Berechnung dieses Maßes sollte möglichst gering sein, da das Paradigma der merkmalsbasierten Suche den Vergleich des Deskriptors der Suchanfrage mit einer möglicherweise großen Anzahl von Deskriptoren des Merkmalsdatensatzes erfordert. Aus intuitiver Sicht lässt sich die Ähnlichkeit am effizientesten durch die Berechnung des Durchschnitts der entsprechenden Histogramme bestimmen.

**Definition 5.2 (Durchschnitt)** Gegeben sei der Merkmalsraum  $V$  eines kardinalen Merkmals sowie ein Histogramm

$$\mathbf{H}(V_i) = (H_1(V_i), H_2(V_i), \dots, H_K(V_i)) \quad (5.6)$$

das die Auftrittshäufigkeiten von  $K$  Merkmalsklassen in der Stichprobe  $V_i \subset V$  bezeichnet. Dann heißt

$$\cap(\mathbf{H}(V_i), \mathbf{H}(V_j)) = \sum_{k=1}^K \min \{H_k(V_i), H_k(V_j)\} \quad (5.7)$$

der Durchschnitt (engl. intersection) der Histogramme von Stichprobe  $V_i$  und  $V_j$ . ■

**Bemerkung:** Die Berechnung des Durchschnitts zweier Histogramme ist äquivalent zu der Berechnung der  $L_1$ -Metrik [154]:

$$\sum_{k=1}^K \min \{H_k(V_i), H_k(V_j)\} = 1 - \frac{1}{2} \sum_{k=1}^K |H_k(V_i) - H_k(V_j)| \quad (5.8)$$

Daraus folgt, dass das Ähnlichkeitsmaß

$$\delta(\mathbf{H}(V_i), \mathbf{H}(V_j)) = 1 - \cap(\mathbf{H}(V_i), \mathbf{H}(V_j)) \quad (5.9)$$

die metrischen Eigenschaften erfüllt und das Paar  $(\mathbf{H}, \delta)$  ein metrischer Merkmalsraum ist.

Da jeder Histogrammwert  $H_k(V_i)$  die Auftrittshäufigkeit einer bestimmten Merkmalsklasse einer Stichprobe  $V_i$  repräsentiert, erhält man durch die Bildung von (5.9) eine direkte Aussage, wieviele Merkmalswerte zweier Stichproben  $V_i$  und  $V_j$  bezüglich ihrer Klassenzugehörigkeit übereinstimmen. Die Bestimmung der Ähnlichkeit folgt dabei dem Prinzip, nur Gleiches mit Gleichem zu vergleichen, d.h. es werden bei der Berechnung des Durchschnitts zweier Histogramme keinerlei Relationen zwischen unterschiedlichen Merkmalsklassen berücksichtigt. Von signifikanter Bedeutung für die Eignung eines Histogramms zur Beschreibung eines bestimmten Merkmals ist daher die Abbildung des Merkmalsraums auf die einzelnen Merkmalsklassen.

Die Anzahl der im Histogramm vertretenen Merkmalsklassen ist in der Regel von ihrem praktischen Nutzen in der realen Anwendung abhängig. Soll beispielsweise das Merkmal „Helligkeit“ von Bildern mittels eines Histogramms beschrieben werden, so ist die Frage, wieviele unterschiedliche Helligkeitsstufen durch die Klasseneinteilung sinnvollerweise unterschieden werden sollen, nur im Kontext der Anwendung zu beantworten. Neben der sinnvollen Auswahl der diskreten Merkmalsklassen selbst ist auch die Ausprägung bzw. Ausdehnung der einzelnen Klassen im Merkmalsraum von Bedeutung, die durch die Festlegung der entsprechenden Zuweisungsfunktionen bestimmt werden.

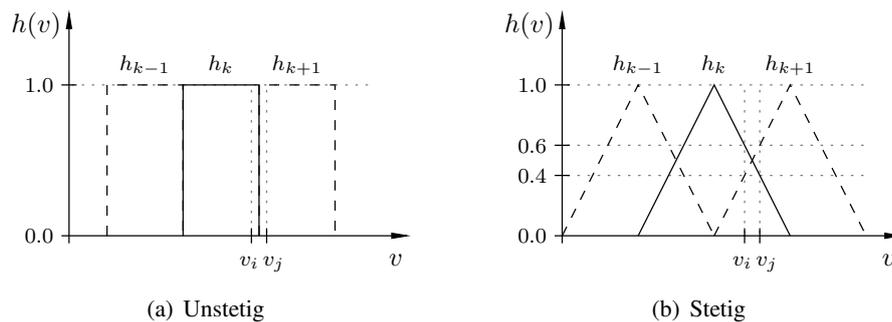
Betrachtet man den Fall der Zuweisung aller Merkmalswerte des Merkmalsraums zu jeweils genau einer Merkmalsklasse, so würden sich die Zuweisungsfunktionen zu

$$h_i(v) = \begin{cases} 1 & \text{falls } v \text{ Element der Merkmalsklasse } i \\ 0 & \text{sonst} \end{cases} \quad (5.10)$$

vereinfachen. Der Merkmalsraum wäre folglich vollständig in disjunkte Merkmalsklassen unterteilt. Als problematisch erweist sich in diesem Fall der Histogrammbildung die Abbildung des Merkmalsraums in den Histogrammraum an den Grenzen der Merkmalsklassen. Es ist offensichtlich, dass diese Abbildung unstetig ist. Zwei nur durch eine beliebig kleine Abweichung unterschiedliche Merkmalswerte könnten prinzipiell auf zwei unterschiedliche – und damit unähnliche – Merkmalsklassen abgebildet werden, obwohl sie unmittelbar benachbart und somit im Sinne des metrischen Modells ähnlich wären. Dieser Widerspruch lässt sich durch die ausschließliche

Verwendung von stetigen Zuweisungsfunktionen  $h_k(v)$  beseitigen. Sind alle Zuweisungsfunktionen eines Histogramms stetig, so spricht man von einem unscharfen Histogramm (*engl. fuzzy histogram*), da ein Merkmalswert  $v$  gewichtet mehreren Merkmalsklassen zugeordnet wird.

**Beispiel 5.1 (Histogrammzuweisungsfunktionen)** *Abbildung 5.1 zeigt die Auswirkung der Verwendung unstetiger und stetiger Histogrammzuweisungsfunktionen auf die Ähnlichkeit zweier benachbarter Merkmalswerte.*



**Abbildung 5.1:** Schematische Darstellung der Abbildung des Merkmalsraums mittels Histogrammzuweisungsfunktionen

In *Abbildung 5.1(a)* wird der Merkmalswert  $v_i$  zu 100% der Merkmalsklasse  $k$  und  $v_j$  zu 100% der Merkmalsklasse  $k + 1$  zugewiesen. Die Berechnung von  $\delta(\mathbf{h}(v_i), \mathbf{h}(v_j))$  ergibt den Wert 1.0, d.h. die Merkmalswerte  $v_i$  und  $v_j$  sind in diesem Fall maximal unähnlich. In *Abbildung 5.1(b)* hingegen wird der Merkmalswert  $v_i$  zu 60% der Merkmalsklasse  $k$  und zu 40% der Merkmalsklasse  $k + 1$  und der Merkmalswert  $v_j$  zu 40% der Merkmalsklasse  $k$  und zu 60% der Merkmalsklasse  $k + 1$  zugewiesen. Die Berechnung des Ähnlichkeitsmaßes ergibt den Wert 0.2, d.h. die Merkmalswerte  $v_i$  und  $v_j$  sind ähnlich. Es ist offensichtlich, dass stetige Zuweisungsfunktionen im Zusammenhang mit dem Histogrammdurchschnitt als Ähnlichkeitsmaß erforderlich sind.

### 5.2.1.1 Verteilungen

Es kommen prinzipiell viele verschiedene visuelle Merkmale eines Bildes in Betracht, die zur merkmalsbasierten Beschreibung von Bildinhalten mittels Histogrammen geeignet sind. Die Farbe eines Bildpunktes ist jedoch das wohl elementarste visuelle Merkmal, das in einem Bild vorhanden ist. Zudem steht das Merkmal Farbe in den Bilddaten unmittelbar zur Verfügung und muss nicht erst mittels komplexer Berechnungsvorschriften als Zwischenprodukt erzeugt werden. Daher bietet sich die globale Verteilung des Merkmals Farbe an, um auf einfache Weise das Aussehen eines Bildes

zu beschreiben. Da jeder diskrete Bildpunkt einen eigenen Farbwert besitzt, kann ein Farbhistogramm auf der Grundlage einer empirischen Stichprobe berechnet werden, dessen Umfang der Anzahl aller im Bild vorhandenen Pixel entspricht.

Der zugrundeliegende Merkmalsraum ist in diesem Fall der RGB-Farbraum, der in der Regel – bei einer Quantisierung von 8-bit pro Farbkanal – 16777216 verschiedene unterscheidbare Farbnuancen umfasst. Die Berücksichtigung aller Farben als Merkmalsklassen resultiert jedoch in Histogrammen mit hoher Dimensionalität, die nur spärlich besetzt sind und damit aus wahrnehmungsorientierter Sicht unnötige Redundanzen enthalten (siehe 4.2). Zudem ist die „Eins-zu-Eins“-Abbildung des Farbraums in den Histogrammraum immanent mit den im Rahmen von unstetigen Zuweisungsfunktionen auftretenden Problemen behaftet.

Daher ist es notwendig, den Farbraum als Merkmalsraum in eine sinnvolle Anzahl zusammenhängender Farbklassen zu unterteilen, die jeweils durch eine repräsentative Farbe vertreten werden. Die Unterteilung des mehrdimensionalen Farbraums in diskrete Farbklassen erfordert als Konsequenz die Erweiterung der Definition von Histogrammen auf mehrdimensionale Merkmalsräume mittels mehrdimensionaler Zuweisungsfunktionen  $h_k(v_1, v_2, \dots, v_n)$ . Die grundsätzliche Forderung nach Massenerhaltung (d.h. die Summe aller Zuweisungsfunktionen ergibt 1) und Stetigkeit der Zuweisungsfunktionen bleibt durch diese Erweiterung unberührt. Da die Handhabung mehrdimensionaler stetiger Zuweisungsfunktionen bei zunehmender Anzahl der Farbklassen schwierig ist, wird im Folgenden ein unterschiedlicher Ansatz zur Berechnung von Farbhistogrammen verfolgt.

Die BCR-Grundform  $\chi(c, C)$  einer Farbe  $c$  bezüglich einer Basisfarbmeng  $C$  vereinigt alle Eigenschaften einer Klassenzugehörigkeitsfunktion (siehe Definition 4.1 und 5.1) und kann daher äquivalent zur Berechnung eines Farbhistogramms eingesetzt werden.

**Definition 5.3 (Farbhistogramm)** *Gegeben sei ein digitales RGB-Farbbild  $\mathbf{X}$  mit endlicher Höhe  $M \in \mathbb{N}$  und endlicher Breite  $N \in \mathbb{N}$  sowie eine Basisfarbmeng  $C$*

$$C = \{c_1, c_2, \dots, c_K\} \quad (5.11)$$

*von  $K \in \mathbb{N}$  Basisfarben, die im RGB-Farbraum definiert sind. Dann ist der Mittelwert der Farben aller Bildpunkte  $\mathbf{x}_{m,n}$  in BCR-Grundform bezüglich  $C$*

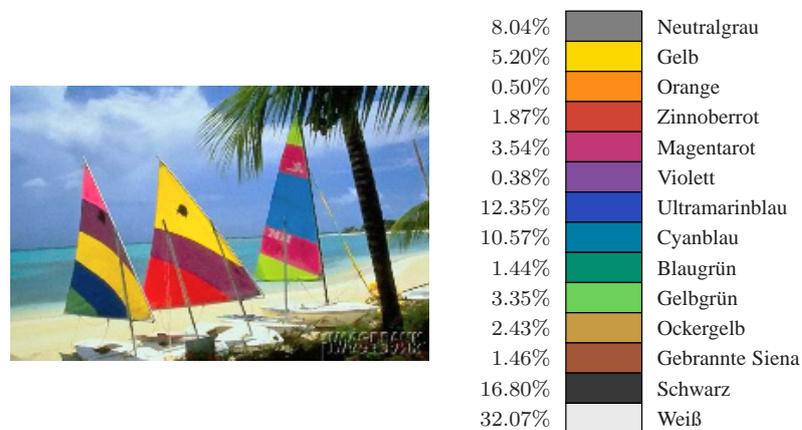
$$H(\mathbf{X}, C) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \chi(\mathbf{x}_{m,n}, C) \quad (5.12)$$

*ein unscharfes Farbhistogramm des Bildes  $\mathbf{X}$ . Die Histogrammwerte  $H_k$  bezeichnen die normierten Auftrittshäufigkeiten der Farbklassen, die durch die jeweiligen Basisfarben  $c_k$  repräsentiert werden. ■*

Jeder Klasse des Farbhistogramms ist folglich genau eine Basisfarbe als prototypisches Muster zugeordnet. Die Werte der BCR-Grundform können dann als die unscharfe Klassenzugehörigkeit einer Farbe zu den entsprechenden Farbklassen interpretiert werden. Im Unterschied zu einem herkömmlichen Histogramm wird die Einteilung des Farbraums in diskrete Klassen in diesem Fall nicht durch die Vorgabe von Zuweisungsfunktionen vorgenommen, sondern vielmehr direkt durch die Wahl der Basisfarben. Die Verwendung der BCR-Grundform bedingt, dass der Farbraum mittels der Basisfarben als Stützstellen gebietsweise linear interpoliert wird. Aus diesem Grund sind die mit einer BCR-Grundform assoziierten Zuweisungsfunktionen immer stetig, auch wenn sie nicht in expliziter Form vorliegen. Das resultierende Farbhistogramm eignet sich folglich zur Berechnung der Ähnlichkeit mit anderen Farbhistogrammen mittels des Histogrammdurchschnitts.

Die Frage der Einteilung des Farbraums in Farbklassen entspricht dadurch einer Frage der Auswahl der Basisfarben. Eine sinnvolle Wahl ist in diesem Zusammenhang beispielsweise die DIN5023-Basisfarbmenge, deren Basisfarben aus Gründen der bedeutungsvollen Darstellung von Farbe im Kontext CBIR gewählt wurden (siehe 4.2.4.2).

**Beispiel 5.2 (Farbhistogramm)** *Abbildung 5.2 zeigt das Farbhistogramm eines digitalen Farbbildes, dessen Histogrammwerte  $H_k$  den relativen Auftretishäufigkeiten der 14 Farben der DIN5023-Basisfarbmenge entsprechen.*



**Abbildung 5.2:** Farbhistogramm eines digitalen Bildes

Die Ähnlichkeit zweier Bilder bezüglich ihrer Farbhistogramme ergibt sich aus Definition 5.2 zu

$$\delta(\mathbf{H}(\mathbf{X}_1, C), \mathbf{H}(\mathbf{X}_2, C)) = \frac{1}{2} \sum_{k=1}^K |H_{1,k} - H_{2,k}| \quad (5.13)$$

**Beispiel 5.3 (Ähnlichkeit)** *Abbildung 5.3 zeigt zwei digitale Farbbilder  $X_1$  und  $X_2$ . Die Ähnlichkeit  $\delta(\mathbf{H}(X_1), \mathbf{H}(X_2))$  der beiden Bilder bezüglich ihrer Farbhistogramme der DIN5023-Basisfarbmenge ist 0.1409. Den größten Anteil zur Ähnlichkeit im Sinne von (5.9) liefern dabei die Farbkanäle Schwarz (0.3053), Gelbgrün (0.2087), Gelb (0.0958), Gebrannte Siena (0.0818) und Ockergelb (0.0717).*



**Abbildung 5.3:** Merkmalsbasierter Vergleich zweier digitaler Farbbilder

Farbhistogramme sind seit jeher eine beliebte Wahl, wenn es um eine einfache Beschreibung von Bildern geht, die relativ robust gegenüber Komplikationen der Bildszene ist. Farbhistogramme sind invariant gegenüber affinen Transformationen der räumlichen Bildkoordinaten, d.h. bei der Beurteilung der Bildähnlichkeit ist die räumliche Anordnung der Bildszene grundsätzlich irrelevant. Histogramme bieten den Vorteil, dass sie unabhängig von Bildgröße und -format sind.

Darüber hinaus lässt sich ein Farbhistogramm in einer sehr kompakten Form repräsentieren. Geht man davon aus, dass eine Fließkommazahl mit einfacher Genauigkeit (15 Nachkommastellen) in einem Maschinenwort (4 Byte) gespeichert werden kann, dann benötigt das Farbhistogramm eines Bildes zur persistenten Speicherung einen Merkmalsdeskriptor der Größe  $4 \cdot K$  Byte, wobei  $K$  die Anzahl der unterschiedlichen Farbklassen des Farbhistogramms bezeichnet. Für die DIN5023-Basisfarbmenge ergibt das eine Datenmenge von 56 Byte pro Deskriptor und Bild.

Eine allzu kompakte Datenrepräsentation eines Merkmalsdeskriptors steht jedoch im Widerspruch zu der allgemeinen Anforderung eines CBIR-Systems, eine sehr große Menge von Bildern ausreichend genau beschreiben und vor allem auch unterscheiden zu können (siehe Seite 25). Die Abbildung von Bildinhalten in einen daten-kompakten Merkmalsraum ist zwangsweise mit einem hohen Verlust an Information verbunden – unterschiedliche Bilder können möglicherweise aufgrund ähnlicher Merkmalsdeskriptoren nicht mehr ausreichend unterschieden werden. Daher ist zu erwarten, dass die Qualität des Retrievalergebnisses mit zunehmender Größe des zugrundeliegenden Bilddatensatzes degradiert. Im Folgenden wird untersucht, wie das Farbhistogramm in diesem Sinne als Merkmal verbessert werden kann.

### 5.2.1.2 Verteilungsmomente

Die Idee, die Wahrscheinlichkeitsverteilung von diskreten Farbklassen zur Beschreibung von Bildinhalten zu verwenden, ist attraktiv. Sie kann jedoch noch erweitert werden, wenn man die Zuweisung einer Farbe zu einer bestimmten Farbkategorie als Zufallsgröße auffasst. Die Wahrscheinlichkeitsverteilung einer Zufallsgröße lässt sich eindeutig durch statistische Maßzahlen charakterisieren. Von besonderer Bedeutung sind in diesem Zusammenhang das erste Moment sowie das zweite und dritte zentrale Moment. Im Normalfall lassen sich die Maßzahlen der Grundgesamtheit mittels einer ausreichend großen Stichprobe schätzen. Daraus folgt, dass sich die empirische Farbverteilung einer Stichprobe aller Bildpunkte ebenfalls durch diese Maßzahlen beschreiben lässt.

Das erste Moment einer Verteilung ist der Erwartungswert (*engl. expectation*) der Zufallsgröße. Der Erwartungswert ist eine statistische Maßzahl zur Kennzeichnung der zentralen Tendenz einer Wahrscheinlichkeitsverteilung. Der Erwartungswert lässt sich aus dem Mittelwert (siehe (5.12)) der Stichprobe schätzen:

$$E_k = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \chi_k(\mathbf{x}_{m,n}, C) \quad (5.14)$$

Das zweite zentrale Moment einer Wahrscheinlichkeitsverteilung ist ihre Varianz (*engl. variance*). Die Varianz ist ein Maß für die Streuung einer Zufallsgröße und bezeichnet die mittleren quadratischen Abweichungen der Werte einer Verteilung ihres Mittelwerts:

$$V_k = \sqrt{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (\chi_k(\mathbf{x}_{m,n}, C) - E_k)^2} \quad (5.15)$$

Das dritte zentrale Moment einer Wahrscheinlichkeitsverteilung ist ihre Schiefe (*engl. skewness*). Die Schiefe ist ein Maß für den Grad der Asymmetrie einer Verteilung:

$$S_k = \sqrt[3]{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (\chi_k(\mathbf{x}_{m,n}, C) - E_k)^3} \quad (5.16)$$

Die Beschreibung des Bildinhaltes durch statistische Maßzahlen der Farbverteilungen stellt eine Erweiterung der Beschreibung durch Farbhistogramme dar, da das Histogramm mit dem ersten Moment einer Verteilung identisch ist. Da jede Zufallsgröße durch drei Maßzahlen charakterisiert ist, müssen für jede durch eine Basisfarbe bestimmte Farbkategorie auch drei Werte im entsprechenden Merkmalsdeskriptor gespeichert werden:

$$\mathbf{M}(\mathbf{X}, C) = ((E_1, V_1, S_1), (E_2, V_2, S_2), \dots, (E_K, V_K, S_K)) \quad (5.17)$$

**Beispiel 5.4 (Farbmomente)** *Abbildung 5.4 zeigt die Farbmomente eines digitalen Farbbildes.  $E_k$  gibt den Erwartungswert,  $V_k$  die Varianz und  $S_k$  die Schiefe der Verteilung der 14 DIN5023-Basisfarben im Bild an.*



	$E_k$	$V_k$	$S_k$
N	0.0804	0.1765	0.2361
YL	0.0520	0.1476	0.2305
OR	0.0050	0.0268	0.0534
CR	0.0187	0.1005	0.1852
MR	0.0354	0.1304	0.2064
VT	0.0038	0.0323	0.0727
UB	0.1235	0.1592	0.1469
CB	0.1057	0.1362	0.1624
BG	0.0144	0.0493	0.0807
YG	0.0335	0.0730	0.1093
Oc	0.0243	0.0836	0.1371
BS	0.0146	0.0672	0.1191
BK	0.1680	0.3158	0.3719
WT	0.3207	0.2944	0.1885

**Abbildung 5.4:** Farbmomente eines digitalen Bildes

Die Ähnlichkeit zweier Bilder bezüglich der Momente ihrer Farbverteilungen lässt sich durch

$$\delta(M(\mathbf{X}_1, C), M(\mathbf{X}_2, C)) = \frac{1}{2} \sum_{k=1}^K |E_{1,k} - E_{2,k}| + \frac{2}{K} \left( \sum_{k=1}^K |V_{1,k} - V_{2,k}| + \sum_{k=1}^K |S_{1,k} - S_{2,k}| \right) \quad (5.18)$$

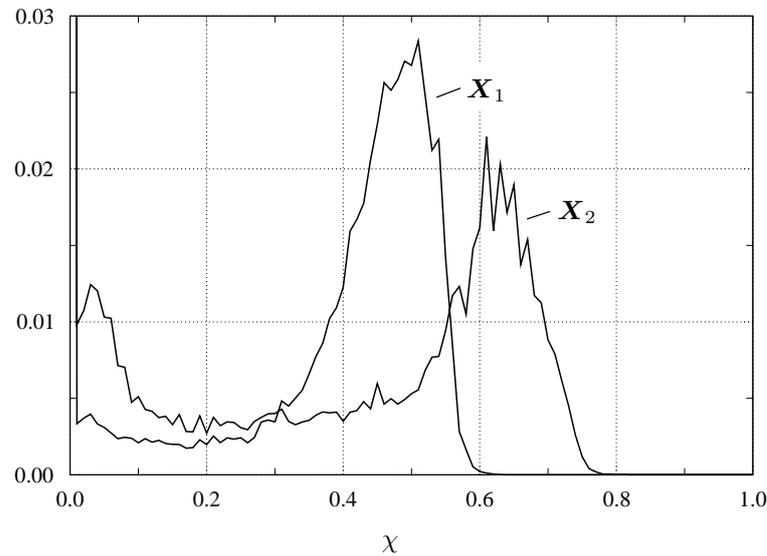
ermitteln. Über Substitution von (5.13) erhält man

$$\delta(M(\mathbf{X}_1, C), M(\mathbf{X}_2, C)) = \delta(H(\mathbf{X}_1, C), H(\mathbf{X}_2, C)) + \frac{2}{K} \left( \sum_{k=1}^K |V_{1,k} - V_{2,k}| + \sum_{k=1}^K |S_{1,k} - S_{2,k}| \right) \quad (5.19)$$

d.h. die Ähnlichkeit von Bildern bezüglich ihrer Farbmomente ergibt sich aus der Ähnlichkeit der Farbhistogramme plus einem Term mit nicht-negativem Wertebereich. Verglichen mit (5.13) ist der Wert von (5.19) folglich für gleiche Bilderpaare immer gleich oder größer. Der zusätzliche Term in (5.19) kann gewissermaßen als Korrekturwert interpretiert werden, der unter Berücksichtigung der Maßzahlen des zweiten und dritten zentralen Moments die Unterscheidungskraft des resultierenden Merkmalsdeskriptors gegenüber dem Farbhistogramm erhöht.

**Beispiel 5.5 (Ähnlichkeit)** *Abbildung 5.5 zeigt die empirische Verteilung der Werte des Gelbgrün-Kanals der Bilder  $\mathbf{X}_1$  und  $\mathbf{X}_2$  aus Abbildung 5.3. Obwohl beide Verteilungen unterschiedlich sind, besitzen sie nahezu identische Mittelwerte ( $E_{1,YG} =$*

0.2087 und  $E_{2,Y_G} = 0.2293$ ). Das erste Moment ist offensichtlich nicht geeignet zwischen den beiden Verteilungen ausreichend zu unterscheiden. Der Unterschied wird jedoch durch das zweite und dritte zentrale Moment erfasst ( $V_{1,Y_G} = 0.2278$ ,  $V_{2,Y_G} = 0.2756$ ,  $S_{1,Y_G} = 0.1528$ , und  $S_{2,Y_G} = 0.2333$ ).



**Abbildung 5.5:** Empirische Verteilung der Werte eines BCR-Farbkanals

Die Ähnlichkeit von  $X_1$  und  $X_2$  bezüglich aller Momente der DIN5023-Farbkanäle errechnet sich damit zu 0.2601.

**Bemerkung:** Der Wert eines Ähnlichkeitsmaßes ist ausschließlich im Kontext des betrachteten Merkmals zu interpretieren. Ähnlichkeitsmaße verschiedener Merkmale besitzen in der Regel weder denselben Wertebereich noch stehen sie in einer linearer Beziehung zueinander. Daher ist es nicht möglich die Ergebnisse aus Beispiel 5.3 und 5.5 direkt zu vergleichen. Ein Vergleich von Ähnlichkeitswerten verschiedener Merkmale ist nur im Zusammenhang einer konkreten Suchanfrage bezüglich einer bestimmten Datenbasis möglich.

Um die Farbmomente eines Bildes persistent zu speichern, benötigt man einen Merkmalsdeskriptor der Größe  $12 \cdot K$  Byte, wobei  $K$  die Anzahl der unterschiedenen Farbklassen bezeichnet. Für die DIN5023-Basisfarbmenge ergibt sich eine Datenmenge von 168 Byte pro Deskriptor und Bild.

### 5.2.1.3 Co-Occurrence Matrizen

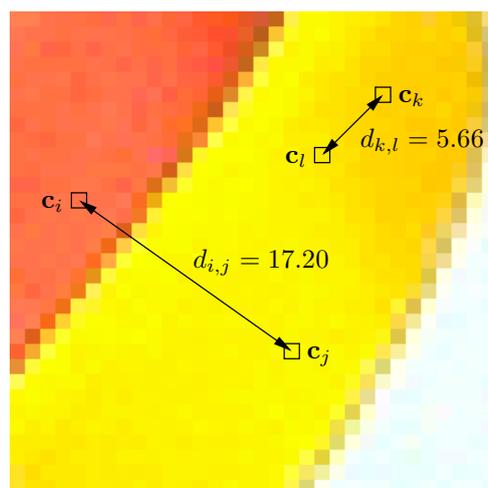
Ganz allgemein betrachtet besteht ein digitales Bild aus der Komposition einer Menge von elementaren Einheiten, den so genannten Pixeln (siehe 3.1). Jedem solchen

Element ist ein Attribut in Form einer Farbe zugeordnet. Darüber hinaus steht jedes einzelne Element über die Art und Weise seiner Anordnung im Bild in einer wechselseitigen Beziehung zu allen anderen Elementen des Bildes. Die Beziehung zweier Elemente ist durch deren räumliche Distanz eindeutig gekennzeichnet.

Statistiken erster Ordnung berücksichtigen ausschließlich Attribute einzelner Pixel (siehe 5.2.1.2) und ignorieren die räumlichen Beziehungen der Pixel untereinander. Es ist offensichtlich, dass Statistiken erster Ordnung, obwohl sie aufgrund der Einfachheit ihrer Berechnung häufig verwendet werden, zur Beschreibung der räumlichen Struktur eines Bildes unzureichend sind. Um Korrelationen von Pixelfarben in einem Merkmalsdeskriptor berücksichtigen zu können, bedarf es aber mindestens einer Statistik zweiter oder höherer Ordnung.

Co-Occurrence Matrizen modellieren die Wahrscheinlichkeit des paarweisen Auftretens eines kardinalen Merkmals in festen geometrischen Abständen mit Hilfe einer Statistik zweiter Ordnung. Ursprünglich wurden Co-Occurrence Matrizen zur Analyse von Texturen in Grauwertbildern eingesetzt [65]. In den letzten Jahren wurde das grundlegende Konzept von Co-Occurrence Matrizen jedoch dahingehend erweitert, dass sie für eine Reihe von Aufgaben im Kontext digitaler Farbbilder anwendbar sind [63].

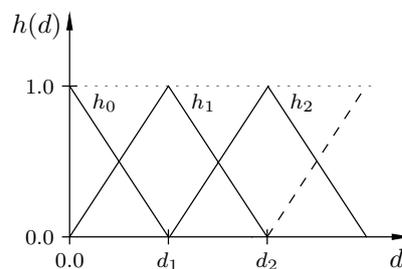
Die für jedes Bild charakteristische räumliche Korrelation von Farben kann mittels einer dreidimensionalen Co-Occurrence Matrix  $\mathbf{W}$  repräsentiert werden, wobei das Matrixelement  $w[c_i, c_j, d_{i,j}]$  die empirische Wahrscheinlichkeit angibt, in dem Bild ein Pixel  $i$  der Farbe  $c_i$  in einem Abstand  $d_{i,j}$  zu einem Pixel  $j$  der Farbe  $c_j$  zu finden. Diesen Sachverhalt stellt Abbildung 5.6 anhand von zwei Pixelpaaren bildlich dar.



**Abbildung 5.6:** Bildliche Darstellung der Deutung der Elemente einer Co-Occurrence Matrix

Ähnlich wie beim Farbhistogramm ist es notwendig, den Farbraum in eine sinnvolle Anzahl von Farbklassen zu unterteilen, um den Nutzen einer Co-Occurrence Matrix zu gewährleisten. Darüber hinaus müssen die Distanzwerte ebenfalls mittels geeigneter stetiger Zuweisungsfunktionen auf eine diskrete Anzahl von Distanzklassen abgebildet werden. Die Co-Occurrence Matrix kann dann als ein  $K \times K \times D$  Histogramm angesehen werden, wobei jedes Element die Auftrittshäufigkeit einer Kombination von  $K$  diskreten Farbklassen in  $D$  diskreten Abständen widerspiegelt.

Zur Abbildung der Pixelfarben auf diskrete Farbklassen bietet sich die BCR-Farbrepräsentation an. Die Abbildung des geometrischen Abstands  $d$  zweier Pixel ( $L_2$ -Norm) auf diskrete Abstandsklassen gestaltet sich schwieriger, da das Abstandsmaß grundsätzlich einen nach oben unbeschränkten Wertebereich  $\mathbb{R}_0^+$  hat. Zwar bietet sich die Bildung eines unscharfen Histogramms mittels der in Abbildung 5.7 dargestellten stetigen Histogrammzuweisungsfunktionen an, jedoch existieren aufgrund der Unbeschränktheit des Wertebereichs unendlich viele Klassen. Für die Beschreibung von Bildinhalten spielen lokale Korrelationen zwischen Farben generell eine wichtigere Rolle als globale Korrelationen [73]. Daher ist es sinnvoll, nur kleine Werte von  $d$  zu berücksichtigen. In der Praxis hat sich gezeigt, dass sich die Beschränkung auf lokale Korrelationen positiv auf die Qualität der Co-Occurrence Matrix im Kontext CBIR auswirkt. Durch diese Maßnahme reduziert sich der Aufwand zur Berechnung einer Co-Occurrence Matrix zudem beträchtlich.



**Abbildung 5.7:** Unschärfe Abbildung der Distanzwerte auf diskrete Klassen

Der Begriff des räumlichen Abstands zweier Pixel ist immanent mit der Größe und Auflösung des betrachteten Bildes verknüpft. Ein Merkmal sollte jedoch idealerweise ausschließlich vom Bildinhalt und nicht von den technischen Parametern der Digitalisierung abhängig sein (siehe 3.1). Dieser unerwünschte Nebeneffekt verschwindet, wenn das Abstandsmaß relativ zu den äußeren Abmessungen des Bildes skaliert wird. Die einfachste Vorgehensweise hierzu ist, das Bild vor der Berechnung des Merkmals auf eine Standardgröße zu normieren. Dazu bietet sich eine Verkleinerung bzw. Vergrößerung des Bildes auf ein Bild mit einer vorgegebenen maximalen Pixelzahl unter Erhaltung des Breiten-Höhen-Verhältnisses des Originals an.

**Beispiel 5.6 (Co-Occurrence Matrix)** Abbildung 5.8 zeigt die normalisierte Co-Occurrence Matrix eines in der Größe auf 8192 Pixel normierten digitalen Farbbildes, dessen Werte die paarweisen Auftretshäufigkeiten aller Farbkombinationen der DIN5023-Basisfarbmengung im Abstand von  $d_1 = 3$  entsprechen.



	N	YL	OR	CR	MR	VT	UB	CB	BG	YG	OC	BS	BK	WT
N	1.86	0.25	0.02	0.05	0.16	0.03	0.59	1.09	0.23	0.46	0.22	0.16	1.48	2.08
YL	0.25	1.49	0.11	0.05	0.12	0.01	0.12	0.11	0.06	0.30	0.42	0.16	0.71	1.12
OR	0.02	0.11	0.01	0.03	0.04	0.00	0.01	0.01	0.00	0.01	0.04	0.01	0.07	0.06
CR	0.05	0.05	0.03	0.73	0.46	0.02	0.01	0.03	0.00	0.02	0.07	0.06	0.21	0.12
MR	0.16	0.12	0.04	0.46	1.15	0.09	0.14	0.12	0.01	0.06	0.09	0.15	0.56	0.31
VT	0.03	0.01	0.00	0.02	0.09	0.02	0.04	0.03	0.00	0.01	0.01	0.03	0.07	0.06
UB	0.59	0.12	0.01	0.01	0.14	0.04	3.50	2.19	0.10	0.05	0.05	0.02	0.66	4.83
CB	1.09	0.11	0.01	0.03	0.12	0.03	2.19	2.39	0.13	0.20	0.07	0.03	0.54	3.60
BG	0.23	0.06	0.00	0.00	0.01	0.00	0.10	0.13	0.10	0.09	0.01	0.01	0.53	0.15
YG	0.46	0.30	0.01	0.02	0.06	0.01	0.05	0.20	0.09	0.39	0.19	0.08	0.50	1.17
OC	0.22	0.42	0.04	0.07	0.09	0.01	0.05	0.07	0.01	0.19	0.34	0.10	0.28	0.79
BS	0.16	0.16	0.01	0.06	0.15	0.03	0.02	0.03	0.01	0.08	0.10	0.11	0.39	0.25
BK	1.48	0.71	0.07	0.21	0.56	0.07	0.66	0.54	0.53	0.50	0.28	0.39	9.16	1.34
WT	2.08	1.12	0.06	0.12	0.31	0.06	4.83	3.60	0.15	1.17	0.79	0.25	1.34	15.7

$\times 10^{-2}$

**Abbildung 5.8:** Co-Occurrence Matrix eines digitalen Bildes

**Bemerkung:** Die Co-Occurrence Matrix für eine bestimmte Abstandsklasse ist quadratisch. Da eine räumliche Korrelation per Definition symmetrisch ist, ist auch die Co-Occurrence Matrix selbst symmetrisch.

Anhand Beispiel 5.6 wird der Bezug zur Texturanalyse erkennbar. Für eine im Verhältnis zur geometrischen Ausdehnung der Korrelation grobe Textur haben Pixelpaare mit großer Häufigkeit identische Farbwerte, daher dominieren die Diagonalelemente von  $\mathbf{W}$ . Bei einer detailreichen Textur haben die Pixelpaare im Mittel unterschiedliche Farbwerte, die Co-Occurrence Matrix ist daher gleichmäßig besetzt.

Im praktischen Umgang stellt sich das Problem, dass Co-Occurrence Matrizen sehr groß sind. So hat  $\mathbf{W}$  für 14 unterschiedliche Farbklassen und drei Abstandsklassen bereits 588 Elemente. Dies hat zur Folge, dass zur Speicherung der Matrix ebenfalls eine entsprechend große Datenmenge benötigt wird. In der Regel bildet sich in der Co-Occurrence Matrix eines Bildes ein charakteristisches Muster aus wenigen Elementen mit betragslich relevanten Werten und vielen Elementen mit vernachlässigbar kleinen Werten. Diese Tatsache kann zur Speicherung der Matrix in einem kompakten Deskriptor ausgenutzt werden, indem alle Elemente mit vernachlässigbar kleinen Werten zu Null gesetzt werden. Die Anzahl der für ein Bild „repräsentativen“ Elemente hängt maßgeblich vom Bildinhalt ab und variiert von Bild zu Bild. Ein triviales Bild mit Pixeln einer identischen Farbe ist durch ein einzelnes Matrixelement ausreichend beschrieben, wohingegen ein Bild mit unkorreliertem Farbrauschen alle Matrixelemente

zur Beschreibung benötigt. Da der Fehler proportional zu der Summe der gekürzten Elemente ist, kann durch eine Auswahl einer geeigneten Anzahl von Matrixelementen sowohl die Datenmenge als auch die Genauigkeit eines Deskriptors kontrolliert werden.

Die Auswahl bestimmter Elemente erfordert aber, dass neben dem Betrag auch der Positionsindex im Deskriptor gespeichert werden muss. Um einen spürbaren Gewinn bezüglich der Größe des Deskriptors erreichen zu können, müssen die Indexe und der Betrag in einem einzelnen Maschinenwort gespeichert werden. Eine mögliche Belegung eines Maschinenworts für eine Matrix mit  $K = 14$  Farbklassen und  $D = 3$  Abstandsklassen sieht beispielsweise wie folgt aus: jeweils 4 Bit für den Index der Farbkategorie, 2 Bit für den Index der Abstandskategorie und 22 Bit für die Mantisse der Fließkommazahl des Betrages (das Vorzeichen sowie der Exponent sind bekannt und müssen nicht gespeichert werden).

Die Ähnlichkeit zweier Co-Occurrence Matrizen lässt sich durch die Berechnung der normalisierten Differenz der einzelnen Elemente bestimmen:

$$\delta(\mathbf{W}_1, \mathbf{W}_2) = 1 - \frac{\sum_{k=1}^K \sum_{l=1}^K \sum_{d=1}^D |w_{1k,l,d} - w_{2k,l,d}|}{\sum_{k=1}^K \sum_{l=1}^K \sum_{d=1}^D w_{1k,l,d} + \sum_{k=1}^K \sum_{l=1}^K \sum_{d=1}^D w_{2k,l,d}} \quad (5.20)$$

Die Normalisierung der Differenz ist notwendig, da die einzelnen Co-Occurrence Matrizen durch die Löschung nicht relevanter Elemente selbst nicht mehr zwangsweise normalisiert sind.

Die Co-Occurrence Matrix speichert die globale Verteilung von lokalen räumlichen Farbkorrelationen eines Bildes. Sie ist unabhängig von der Bildgröße und invariant gegenüber den affinen Transformationen Translation, Rotation und Skalierung, was sie zu einem attraktiven Merkmal für das CBIR macht. Zudem lässt sie sich für eine kleine Anzahl von Farb- und Abstandsklassen leicht und schnell berechnen und in einem kompakten Deskriptor speichern.

## 5.2.2 Signaturen

Eine häufig auftretende Form der Suchanfrage sucht nach einem Bild an sich, d.h. der Bildinhalt wird nicht als Komposition einzelner Bildelemente mit einer globalen räumlichen Anordnung interpretiert, sondern als Gesamtheit (z.B. die Suche nach Fotografien eines Gemäldes). Bilder sind in diesem Fall nur dann ähnlich, wenn das globale Erscheinungsbild übereinstimmt. Aus diesem Grund ist es notwendig, die für jedes

Bild charakteristische räumliche Verteilung von Farben mittels eines Merkmals zu beschreiben. Eine solches Merkmal darf im Gegensatz zu den vorgenannten Verfahren offensichtlich nicht invariant gegenüber affinen Bildtransformationen sein.

Die einfachste Art und Weise zwei Bilder zu vergleichen ist der direkte pixelweise Vergleich. Diese Vorgehensweise ist jedoch aus mehreren Gründen nicht realisierbar. Das Bild selbst wäre das Merkmal. Von einem kompakten Merkmalsdeskriptor könnte folglich keine Rede mehr sein. Ein solcher Vergleich wäre sehr ineffektiv, wenn eine ungenaue Suchanfrage an ein System mit einer großen Anzahl von Bildern gestellt werden würde [78]. Schon kleine Abweichungen und Verzeichnungen haben direkten Einfluss auf die Bestimmung der Ähnlichkeit, da einzelne Pixel explizit miteinander verglichen werden. Zudem stellt sich das praktische Problem, dass die zu vergleichenden Bilder dieselbe Größe haben müssen. Auch die Verkleinerung der Bilder in ein in der Auflösung stark reduziertes Standardformat bringt keine grundsätzliche Lösung der genannten Probleme mit sich (ein Farbbild der Größe  $32 \times 32$  benötigt zur Speicherung bereits 3072 Bytes).

Die im Rahmen dieser Arbeit an dieser Stelle verfolgte Idee basiert auf denselben Prinzipien, die auch bei der Kompression von digitalen Bildern in verschiedenen Dateiformaten angewendet werden. Aus jedem Bild wird dazu als Merkmal eine sehr stark komprimierte Signatur extrahiert, die aus wahrnehmungsorientierter Sicht ausschließlich „wichtige“ bildliche Information enthält und von „unwichtigen“ Details abstrahiert. Die Signatur dient gewissermaßen als eine grobe Skizze für das entsprechende Bild. Mittels dieser Skizzen können Bilder dann bezüglich ihres globalen Erscheinungsbildes verglichen werden.

Das Problem der Extraktion einer entsprechenden Signatur lässt sich wie folgt umformulieren: Wie kann die Datenmenge zur Darstellung einer gewissen Menge an visueller Information in geeigneter Weise reduziert werden? Ähnlich wie bei der Bildkompression ist dazu die Elimination von redundanten Daten notwendig. Dazu muss die Bildmatrix in einen statistisch unkorrelierten Datensatz transformiert werden. Im Gegensatz zur Bildkompression liegt der Fokus bei der Extraktion einer Signatur als Merkmal eher auf der kompakten Darstellungsweise als auf dem Erhalt der visuellen Information aus wahrnehmungsorientierter Sicht.

Für die Dekorrelation der Ortsinformation in Bilddaten kommen prinzipiell verschiedene Transformationen in Betracht. Für die Verwendung der Wavelet-Transformation spricht die Tatsache, dass die resultierenden Transformationkoeffizienten eine fast optimale Varianzverteilung aufweisen [80]. Die Varianz kann als ein Maß des visuellen Informationsgehaltes eines bestimmten Koeffizienten aufgefasst werden. Die Beschränkung der Signatur auf eine Darstellung mit einigen wenigen betraglich signifikanten Koeffizienten resultiert daher in einem insgesamt minimalen mittleren qua-

dratischen Fehler. Des Weiteren lässt sich die Wavelet-Transformation im Vergleich zu anderen Transformationen mit einer deutlich geringeren Komplexität berechnen. Daher bietet sich die Wavelet-Transformation für die Berechnung einer äußerst kompakten Signatur eines Bild an.

### 5.2.2.1 Wavelet-Transformation

Ganz allgemein formuliert überführt die kontinuierliche Wavelet-Transformation eine Funktion  $f(x)$  in eine Darstellung aus einer linearen Superposition von Basisfunktionen  $\psi_{a,b}(x)$ , den so genannten Wavelets

$$WT(a, b) = \int f(x)\psi_{a,b}(x)dx \quad (5.21)$$

Wavelets sind Funktionen, die durch Skalierung, Dehnung und Translation aus einem Mutter-Wavelet  $\psi_0(x)$  gemäß

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi_0\left(\frac{x-b}{a}\right) \quad (5.22)$$

erzeugt werden können, wobei  $a$  der Dehnungsfaktor und  $b$  der Translationsfaktor ist. Das Mutter-Wavelet  $\psi_0(x)$  muss die folgenden Bedingungen erfüllen: Die Funktion besitzt einen kompakten Träger, und die Integration über den Träger ergibt 0. Der Träger (*engl. support*) einer Funktion ist als die kleinste abgeschlossene Menge des Definitionsbereichs, außerhalb derer  $f(x) = 0$  ist.

Um die Wavelet-Transformation auf diskrete Daten anwenden zu können, müssen die Parameter der Wavelet-Transformation ebenfalls diskretisiert werden. Aus Gründen der einfacheren Handhabung und schnelleren Berechnung werden in der Regel dyadische Dehnungs- bzw. Translationsfaktoren verwendet

$$a = 2^v, \quad b = u \cdot 2^v, \quad u, v \in \mathbb{Z} \quad (5.23)$$

Die daraus resultierende Generalisierung von (5.21) ergibt durch Substitution von (5.23) die Definition der diskreten Wavelet-Transformation

$$WT[u, v] = 2^{-\frac{v}{2}} \sum_n f[n]\psi[u \cdot 2^{-v} - n] \quad (5.24)$$

Die Wavelet-Transformation umfasst eine ganze Klasse von Transformationen, die je nach verwendetem Mutter-Wavelet unterschiedliche Eigenschaften und Besonderheiten aufweisen. Das Problem ist, ein Mutter-Wavelet zu finden, das eine Transformation konstituiert, die den gestellten Anforderungen genügt. Das einfachste – und auch im vorliegenden Kontext brauchbarste – Mutter-Wavelet ist das Daubechies-Wavelet erster Ordnung, das auch unter dem Namen Haar-Wavelet bekannt ist.

Das Haar-Wavelet

$$\psi[n] = \begin{cases} 1 & n = 0 \\ -1 & n = 1 \\ 0 & \text{sonst} \end{cases} \quad (5.25)$$

bildet mit der Skalierungsfunktion

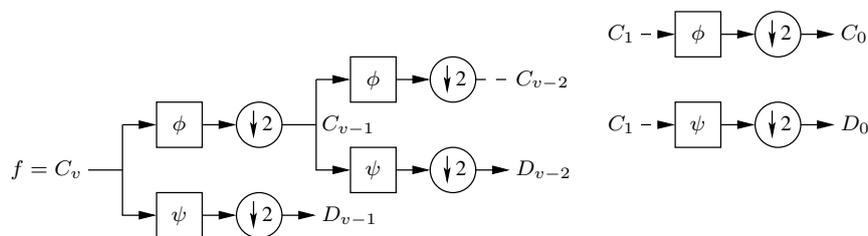
$$\phi[n] = \begin{cases} 1 & n = 0, 1 \\ 0 & \text{sonst} \end{cases} \quad (5.26)$$

eine orthonormale Basis. In diesem Fall ist die Energie der Koeffizienten  $f[n]$  gleich der Energie der Koeffizienten  $WT[u, v]$ , d.h. die Transformation ist energieerhaltend

$$\sqrt{\sum_n f[n]^2} = \sqrt{\sum_u \sum_v WT[u, v]^2} \quad (5.27)$$

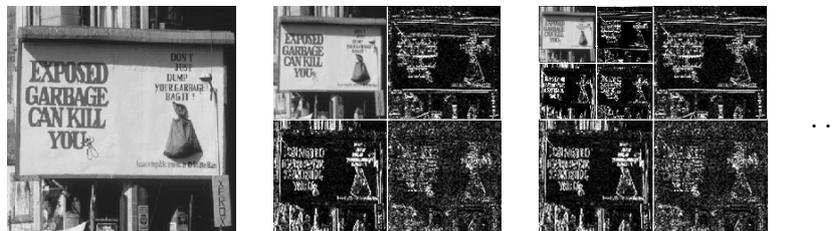
Dies ist ein wichtiges Kriterium zur Einschätzung von Fehlern. Darüber hinaus besitzt das Haar-Wavelet den kleinsten möglichen Support und erfordert keinerlei gesonderte Betrachtung des Signals an den Rändern von  $f[n]$ . Das Haar-Wavelet eignet sich folglich hervorragend zur Transformation von digitalisierten Bildern.

Der effektivste Weg, die diskrete Wavelet-Transformation zu implementieren, ist der Einsatz einer mehrstufigen Quadraturspiegelfilterbank [98]. Ein Signal wird dabei mittels eines aus einem orthogonalen Hochpass- und Tiefpassfilter bestehenden Quadraturspiegelfilters in ein Grobsignal  $C$  und ein Detailsignal  $D$  zerlegt. Ein den Filtern nachgeschalteter Downsampler reduziert jeweils die Redundanz der Signale um den Faktor 2. Das Grobsignal wird sukzessive in mehreren Stufen durch identische Quadraturspiegelfilter in jeweils ein Grob- und ein Detailsignal weiterzerlegt. Als Resultat erhält man Detailsignale, die jeweils unterschiedliche Frequenzbänder des Originalsignals repräsentieren. Die vollständige Filterbank ist in Abbildung 5.9 dargestellt, wobei  $C_v$  das Grob- und  $D_v$  das Detailsignal der Auflösungsstufe  $v$  bezeichnen. Durch diese Vorgehensweise reduziert sich der Aufwand der Wavelet-Transformation auf eine lineare Komplexität.



**Abbildung 5.9:** Mehrstufige Quadraturspiegelfilterbank einer Wavelet-Transformation

Die Wavelet-Transformation kann durch eine einfache Bildung eines Tensorproduktes auf zweidimensionale Signale erweitert werden. Ein Grauwertbild kann daher durch eine sukzessive Transformation aller Bildzeilen und -spalten des Grobbildes in seine zweidimensionale Wavelet-Repräsentation überführt werden. Jede Stufe der zweidimensionalen Wavelet-Transformation erzeugt insgesamt drei Detailbilder; eins mit horizontaler, eins mit vertikaler und eins mit diagonaler Vorzugsrichtung. Voraussetzung für die Anwendung der Wavelet-Transformation ist jedoch, dass das Bild in quadratischer Form mit einer Kantenlänge  $2^n$  vorliegt. Abbildung 5.10 zeigt die sukzessive Vorgehensweise einer zweidimensionalen Haar-Transformation am Beispiel eines Grauwertbildes. Auf jeder Auflösungsstufe bleibt die quadratische Grundform erhalten – es bildet sich inhärent eine Auflösungspyramide.



**Abbildung 5.10:** Zweidimensionale Wavelet-Transformation eines Grauwertbildes

Um ein digitales Farbbild mittels der Wavelet-Transformation in eine Repräsentation überführen zu können, aus der eine Signatur extrahiert werden kann, muss das Farbbild zunächst auf eine Standardgröße mit quadratischem Format normiert werden (in der Praxis hat sich das Format der Größe  $64 \times 64$  bewährt). Mit der Normierung ist immanent eine Verzerrung des Bildinhaltes verbunden, da Bilder in der Regel im Portrait- oder Landschaftsformat mit einem Breiten-Höhen-Verhältnis von 3:4 bzw. 4:3 vorliegen. Eine Signatur kann daher nicht dazu verwendet werden, um nach einem Bild in einem bestimmten Format zu suchen. Diese Information muss zusätzlich zur Signatur in Form von Metainformation in der Datenbasis gespeichert werden.

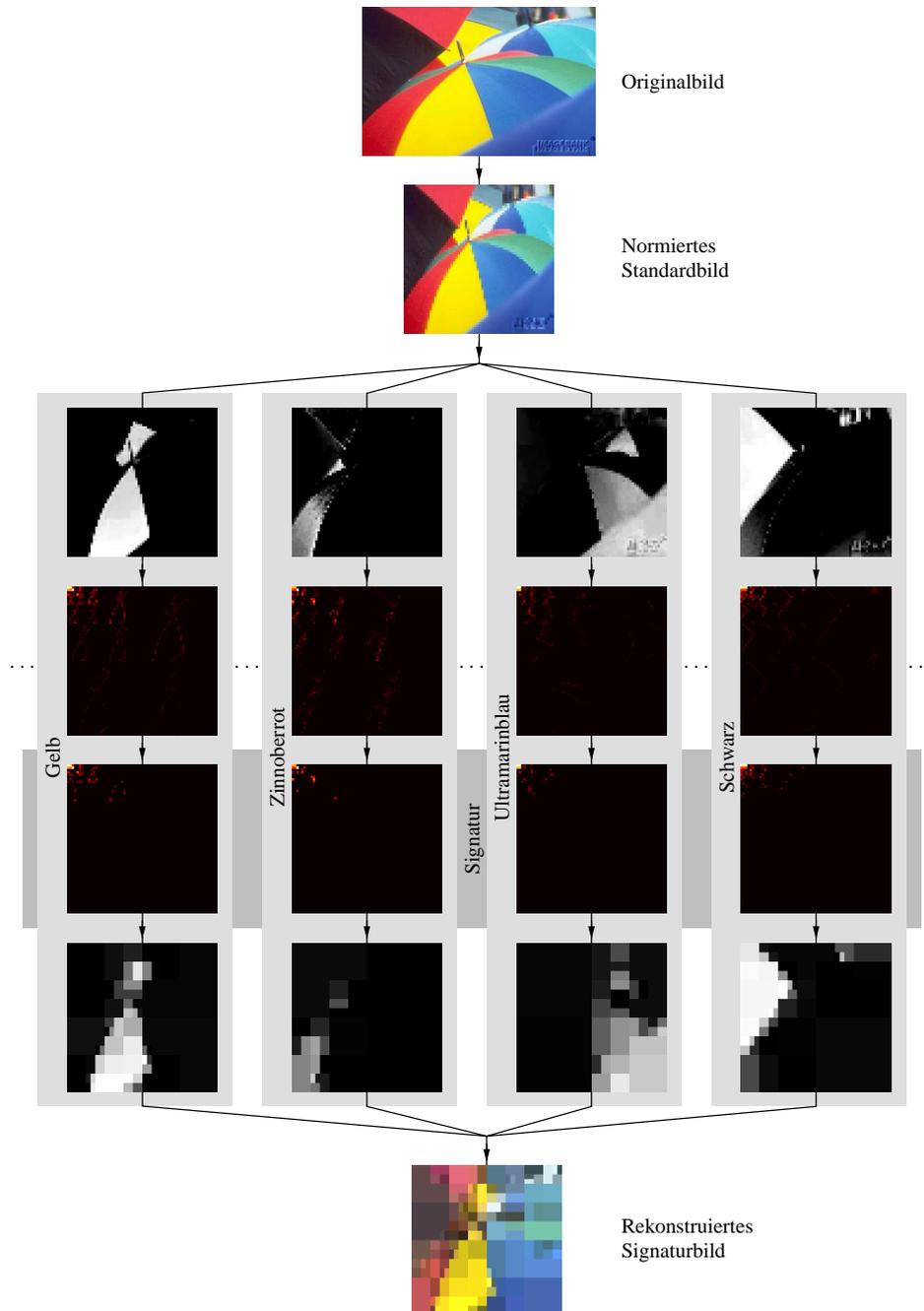
Grundsätzlich ist die Wavelet-Transformation eines Farbbildes unabhängig von dessen Farbrepräsentation, da jeder Farbkanal einzeln (gewissermaßen als Grauwertbild) transformiert wird. Daher besitzt die Wavelet-Repräsentation eines Farbbildes die gleiche Anzahl von Farbkanälen wie das Original. Im Folgenden wird daher ohne Einschränkung der Allgemeinheit vorausgesetzt, dass alle Farbbilder in BCR-Farbrepräsentation vorliegen.

Das Ziel der Extraktion einer kompakten Signatur ist die Auswahl einer kleinen Anzahl repräsentativer Kandidaten aus der Gesamtmenge der Wavelet-Koeffizienten aller

Farbkanäle. Die Signatur sollte dabei das Original aus Sicht des visuellen Informationsgehaltes bestmöglich approximieren, um später eine eindeutige Identifikation des Bildes zu ermöglichen. Da der Approximationsfehler aufgrund der energieerhaltenden Eigenschaften der Wavelet-Transformation proportional zu der Energie der nicht ausgewählten Koeffizienten ist, lässt sich schließen, dass der beste Ansatz die Auswahl der betragsmäßig größten Koeffizienten ist. Prinzipiell kann bei der Auswahl entweder die Anzahl oder der Auswahlsschwellwert fest vorgegeben werden. In der Praxis hat sich gezeigt, dass ein hybrider Ansatz optimale Ergebnisse liefert, d.h. es wird sowohl eine maximale Anzahl als auch ein minimaler Schwellwert für die Auswahl der Koeffizienten festgelegt. Dadurch ist gewährleistet, dass Deskriptoren einerseits eine maximale Größe haben können, andererseits Bilder mit geringer visueller Variabilität durch entsprechend kleine Deskriptoren resultiert werden.

**Beispiel 5.7 (Signaturextraktion)** *Abbildung 5.11 stellt die Verarbeitungskette der Extraktion einer Wavelet-Signatur am Beispiel eines Farbbildes dar. Das Bild wird zunächst auf ein quadratisches Standardformat ( $64 \times 64$ ) normiert und anschließend in seine BCR-Farbrepräsentation überführt. In Abbildung 5.11 sind aus Gründen der Übersichtlichkeit nur vier der insgesamt 14 DIN5023-Farbkanäle dargestellt. Mittels der Haar-Transformation werden die Bilddaten jedes einzelnen Farbkanals dekorreliert. Die Wavelet-Transformierten sind in Abbildung 5.11 vorzeichenlos in Falschfarben dargestellt. An dieser Stelle umfasst die Repräsentation des Bildes  $64 \cdot 64 \cdot 14 = 57344$  vorzeichenbehaftete reelle Koeffizienten. Aus den Transformierten aller 14 Farbkanäle werden im nächsten Schritt die 192 ( $\approx 0.3\%$ ) betragsmäßig größten Koeffizienten in einer kompakten Signatur zusammengefasst. Um dem Betrachter einen Eindruck über den visuellen Informationsgehalt der Signatur zu geben, wird aus der Signatur ein Signaturbild rekonstruiert. Es ist deutlich zu sehen, dass das Signaturbild lediglich das grobe Layout des Originals ohne jegliches Detail wiedergibt. Ein subjektiver Vergleich des Originals und des Signaturbildes aus ausreichend großer Entfernung offenbart, dass das Signaturbild offensichtlich den Teil der visuellen Information enthält, der dem Bild seine charakteristische Identität verleiht.*

Die Auswahl einer kleinen repräsentativen Menge von Koeffizienten macht es ähnlich wie bei der Co-Occurrence Matrix erforderlich, dass sowohl Index als auch der Betrag der Koeffizienten im Deskriptor gespeichert werden müssen. Es ist daher sinnvoll, jeweils einen Koeffizienten in einem Maschinenwort zu speichern, um eine möglichst kompakte Datenrepräsentation erreichen zu können. Für den Index sind 16 Bit notwendig, d.h. der Betrag muss ebenfalls in maximal 16 bit codiert werden. Da die Koeffizienten einen quasi unbeschränkten Wertebereich im reellen Zahlenraum haben (de facto wird der Wertebereich durch die Größe des normalisierten Standardbildes beschränkt), muss neben der Mantisse auch das Vorzeichen und der Exponent berücksicht-



**Abbildung 5.11:** Extraktion einer Signatur aus Wavelet-Koeffizienten aus einem digitalen Farbbild

sichtigt werden. Der Betrag kann daher nur mittels einer Fließkommazahl mit einer sehr geringen Genauigkeit gespeichert werden, was allerdings kein großes Problem darstellt, da der zusätzliche Fehler die Signatur als Approximation des Originals nur geringfügig verfälscht.

Die Ähnlichkeit zweier Wavelet-Signaturen lässt sich durch die Berechnung der Quadratsumme aus den Differenzen der jeweiligen Koeffizienten bestimmen:

$$\delta(\mathbf{WT}_1, \mathbf{WT}_2) = \sqrt{\sum_c \sum_u \sum_v (WT_{1_{c,u,v}} - WT_{2_{c,u,v}})^2} \quad (5.28)$$

Wavelet-Signaturen repräsentieren die visuelle Information eines Bildes in Form einer Skizze in einem kompakten Merkmalsdeskriptor. Sie konservieren die hervorstechendsten Farbcharakteristika eines Bildes unter Berücksichtigung der globalen räumlichen Anordnung. Aufgrund der Tatsache, dass sie auf sehr hohem Niveau von Details abstrahieren, sind Signaturen sehr robust gegenüber Veränderungen des Bildes in jeder Form. Sie eignen sich insbesondere für Suchanfragen ohne detailreiche Vorlagen (z.B. Query-by-Sketch).

### 5.2.3 Evaluation

An dieser Stelle sollte evaluiert werden, ob die vorgestellten Merkmale dazu geeignet sind, an ein CBIR-System gestellte Suchanfragen mit einer Menge relevanter Bilder zu beantworten. Wie schon in 2.4.1 besprochen, ist dies ein schwieriges, wenn nicht gar unmögliches Unterfangen. Im Gegensatz zum Menschen, der ein Bild nach subjektiver Einschätzung entweder als relevant oder irrelevant einstuft, beantworten CBIR-Systeme eine Suchanfrage mit einer Rangfolge von Bildern, die bezüglich ihrer Ähnlichkeit im Hinblick auf ein bestimmtes Merkmal erstellt wurde. Ein relevantes Bild kann zufälligerweise ähnlich zur Suchanfrage sein, muss es aber nicht.

Visuelle Inhalte lassen generell einen größeren Spielraum bei der Interpretation zu, d.h. unterschiedliche Benutzer kommen immer zu unterschiedlichen Beurteilungen. Zudem ist das Ergebnis einer Suchanfrage direkt abhängig von der zugrundeliegenden Datenbasis. Zu groß ist der maßgebliche Einfluss, den die Zusammensetzung der Datenbasis auf das Ergebnis einer Beurteilung eines bestimmten Merkmals haben kann. Jede sinnvolle Beurteilung eines Merkmals würde folglich eine exakte Dokumentation der Bildersammlung erfordern – bei mehreren Tausend Bildern ein aussichtsloses Unterfangen.

Nichtsdestotrotz sind im Anhang B die experimentellen Ergebnisse mehrerer Suchanfragen für die in diesem Kapitel vorgestellten Merkmale dokumentiert. Diese dienen

nicht als Beweis der Leistungsfähigkeit eines Merkmals, in einer bekannterweise unbekanntes Bildmenge syntaktisch oder sogar semantisch ähnliche Bilder auffinden zu können. Vielmehr soll der individuelle Betrachter die Möglichkeit bekommen, die unterschiedlichen Eigenschaften der verschiedenen Merkmale visuell zu erforschen und zu erfassen. Jedes Merkmal verkörpert einen bestimmten, unterschiedlichen Aspekt eines Bildes, dessen Erläuterung mit Worten generell schwierig ist, da die Merkmale auf der Grundlage abstrakter mathematischer Berechnungsvorschriften definiert sind. Anhand der Beispiele kann jeder Betrachter für sich die ihm auffälligen Aspekte entdecken.

## Kapitel 6

# Merkmalsindexierung

Eine Suchanfrage in einem CBIR-System ist grundsätzlich so formuliert, dass in einem konkreten Bilddatensatz nach – bezüglich eines bestimmten Merkmals – ähnlichen Bildern gesucht wird. Auf der Grundlage des metrischen Modells entspricht dieses Problem einer Nachbarschaftssuche im Merkmalsraum (siehe 3). Das Suchen nach einem oder mehreren Nachbarn innerhalb eines endlichen Merkmalsdatensatzes der Größe  $N$  kann trivialerweise durch ein sequentielles, sukzessives Vergleichen aller in dem Datensatz vorhandenen Merkmalsdeskriptoren mit einer Komplexität von  $\mathcal{O}(N)$  durchgeführt werden. Eine solche so genannte lineare Suche ist sehr ineffizient, da der Aufwand direkt proportional mit der Datensatzgröße steigt. Die Entwicklung einer Strategie, die den Aufwand einer Nachbarschaftssuche reduziert, ist daher von fundamentaler Bedeutung für die Benutzbarkeit eines CBIR-Systems.

Um eine effizientere Suche zu ermöglichen, muss die Anzahl der im Rahmen der Suche notwendigen Merkmalsvergleiche deutlich kleiner sein als die Gesamtzahl der im Datensatz vorhandenen Merkmalsdeskriptoren. Daher wird ein geeignetes Suchverfahren benötigt, das nur jene Teile des Suchraums in Betracht zieht, in denen relevante Nachbarn vorkommen können. Das kann prinzipiell nur gelingen, wenn bei der Suche die Topologie des Merkmalsdatensatzes berücksichtigt wird. Diese Information muss zunächst aus einem vorliegenden Datensatz erschlossen und in einer geeigneten Indexstruktur gespeichert werden. Ziel ist die Verwendung dieser Indexstruktur zur Reduktion der Komplexität einer Nachbarschaftssuche, ohne jedoch die Qualität des Ergebnisses zu beeinträchtigen. Diese Aufgabe wird im Allgemeinen als Merkmalsindexierung (*engl. similarity indexing*) bezeichnet.

## 6.1 Stand der Technik

Die folgenden Abschnitte geben einen Überblick über die in der Literatur dokumentierten Entwicklungen im Bereich der Merkmalsindexierung. Alle Methoden bauen auf dem grundsätzlichen Prinzip auf, den Merkmalsraum in kleinere Teilräume zu unterteilen. Da eine Nachbarschaftssuche naturgemäß nur einen verhältnismäßig begrenzten Teil des Merkmalsraums betrifft, ist es eine logische Konsequenz, die Suche auf einige wenige Teilräume zu beschränken. Die Aufgabe der Indexstruktur ist dabei, die topologischen Beziehungen einer solchen Unterteilung abzubilden, um den optimalen Zugriffspfad auf die für eine Suchanfrage relevanten Teilräume zu gewährleisten. Das Hauptproblem des Entwurfs einer solchen Indexstruktur ist, dass sich Objekte eines metrischen Raums nicht einfach bezüglich ihrer räumlichen Nähe ordnen lassen. In anderen Worten, es existiert keine Abbildung des metrischen Raums auf einen eindimensionalen Raum, der die topologischen Beziehungen aller Objekte korrekt wiedergeben würde. Daher kann nicht auf bereits existierende Lösungen eindimensionaler Zugriffsmethoden (z.B. Hashing [38], *B*-Baum [4]) aus dem Bereich herkömmlicher Datenbanken zurückgegriffen werden.

Nievergelt hat in [115] zusammengefasst, welche Anforderungen allgemein an eine Struktur zur Indexierung räumlicher Datensätze gestellt werden:

**Dynamik** – der Index sollte sich fortlaufend den Veränderungen des Datensatzes anpassen. Das gilt insbesondere für das Einfügen, Modifizieren und Löschen von Objekten.

**Unabhängigkeit** – der Index sollte unabhängig von den speziellen Eigenschaften des zugrundeliegenden Datensatzes sein. Dieser Punkt ist besonders wichtig in Hinblick auf Datensätze, die nicht gleichmäßig im Raum verteilt sind.

**Robustheit** – ein Index sollte ausreichend robust gegenüber Fehlbedienung sein.

**Skalierbarkeit** – ein Index sollte bezüglich des Wachstums des zugrundeliegenden Datensatzes skalierbar sein.

**Effizienz** – ein Index sollte im Vergleich zu dem indexierten Datensatz klein sein, damit er vollständig im Hauptspeicher vorgehalten werden kann. In Hinblick auf die zur Verfügung stehenden Speicherressourcen sollte der Index sekundäre und tertiäre Speichermedien nahtlos und effizient einbinden.

**Schnelligkeit** – ein Index sollte eine schnelle Suche im Datensatz ermöglichen. Im ungünstigsten Fall sollte der Index für alle denkbaren Datensätze der Größe  $N$  eine Komplexität von  $\mathcal{O}(\log(N))$  aufweisen.

### 6.1.1 Multidimensionale Zugriffsmethoden

Die Verfahren der Merkmalsindexierung haben ihre Wurzeln in den so genannten multidimensionalen Zugriffsmethoden (*engl. multidimensional access methods*), deren Entwicklung Mitte der 70er Jahre begann. Ursprünglich wurden diese Zugriffsmethoden vor dem Hintergrund geographischer Informationssysteme mit dem Ziel entwickelt, die Suche in Vektorräumen mit geringer Dimensionalität zu optimieren.

#### 6.1.1.1 Vektorräume

Ein Ansatz zur Entwicklung multidimensionaler Zugriffsmethoden stellt naturgemäß die Erweiterung eindimensionaler Zugriffsmethoden auf mehrere Dimensionen dar. Ein typischer Vertreter dieser Kategorie ist das so genannte Grid-File [116]. Der  $n$ -dimensionale Vektorraum wird dabei mittels einer orthogonalen Gitterstruktur, die nicht notwendigerweise gleichförmig ist, unterteilt (siehe Abbildung 6.1(a)). Jedes Objekt im Vektorraum kann nun aufgrund seiner Lage genau einer quaderförmigen Zelle zugeordnet werden. Ein Verzeichnis assoziiert die Objekte eines oder mehrerer solcher Zellen als eine zusammenhängende Einheit, die auf einem sekundären Medium gespeichert wird. Das Verzeichnis selbst wird im Hauptspeicher vorgehalten. Ein bestimmtes Objekt kann so durch maximal zwei Zugriffe garantiert gefunden werden. Eine solche Zugriffsmethode eignet sich jedoch nur bedingt für Nachbarschaftssuchen, da die Anzahl der von der Suche betroffenen Zellen bei höherdimensionalen Räumen schnell exponentiell wächst. Darüber hinaus ist auch eine dynamische Anpassung an Veränderungen des Datensatzes problematisch, ohne dass der Index vollständig überarbeitet werden muss.

Eine der bekannteren Strukturen zur Indexierung zweidimensionaler Vektorräume ist der Quad-Tree [43]. Ein Teilraum wird jeweils rekursiv durch zwei orthogonale Grenzen in vier gleich große Teilräume aufgeteilt, falls die Anzahl der Objekte im Teilraum einen a priori festgelegten Grenzwert übersteigt (siehe Abbildung 6.1(b)). Der Quad-Tree offenbart eine gute Performanz in zweidimensionalen Vektorräumen, falls die hierarchische Struktur des Quad-Tree balanciert ist. Daher eignet sich der Quad-Tree ausschließlich für Datensätze, die sich durch eine gleichmäßige Verteilung der Objektpunkte im Vektorraum auszeichnen. Der Quad-Tree kann grundsätzlich auf beliebig viele Dimensionen erweitert werden, jedoch ist Erweiterung wenig sinnvoll, da im  $n$ -dimensionalen Raum  $2^n$  Teilräume entstehen.

Bentley [6] hat einen speziellen binären Suchbaum, den  $k$ - $d$ -Baum, entwickelt, der mit dem Quad-Tree eng verwandt ist. Jeder Knoten des  $k$ - $d$ -Baum enthält einen Punkt, der den Teilraum des Knotens orthogonal zu einer der Achsen in unterschiedlicher Größe

unterteilt. Die Wahl der Achsen ist abhängig von der Höhe des Knotens im Baum, d.h. die Achsen werden alternierend gewählt (siehe Abbildung 6.1(c)). Im Gegensatz zum Quad-Tree bildet der  $k$ - $d$ -Baum einen Datensatz mit beliebiger Verteilung in einer balancierten hierarchischen Struktur ab. Der  $k$ - $d$ -Baum ist jedoch nur zur Indexierung statischer Datensätze geeignet, da dynamische Veränderungen zur Folge haben, dass die Struktur nicht mehr balanciert ist. Daher eignet sich der  $k$ - $d$ -Baum allenfalls zur Indexierung statischer niederdimensionaler Vektorräume.

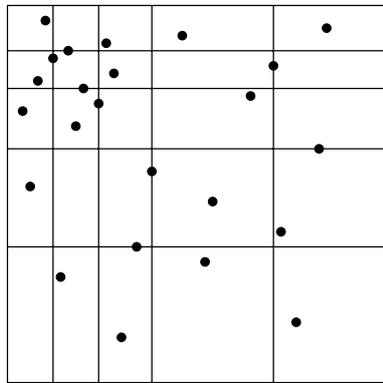
Guttman [62] hat 1984 den  $R$ -Baum mit dem Ziel entwickelt, eine hierarchische Struktur zu entwerfen, die im Gegensatz zu dem Quad- und  $k$ - $d$ -Baum immer perfekt balanciert ist. Im  $R$ -Baum existieren zwei grundsätzlich verschiedene Knotentypen: Blattknoten und innere Knoten. Blattknoten enthalten sowohl die Punktdaten selbst als auch die Information bezüglich der Ausmaße der kleinsten begrenzenden quaderförmigen Zelle (MBB – *engl. minimum bounding box*), die alle Punkte des Blattes enthält. Die Blattknoten werden so gewählt, dass sie nur benachbarte Punkte enthalten und somit möglichst kompakte MBBs haben. Die MBBs einzelner Blattknoten können sich dabei gegenseitig überlappen. Innere Knoten umfassen entweder eine Menge von Blattknoten oder eine Menge von anderen inneren Knoten. Ein innerer Knoten bekommt jeweils eine MBB zugewiesen, die alle MBBs seiner Kindknoten vollständig enthält (siehe Abbildung 6.1(d)). Ein spezieller Einfügealgorithmus garantiert, dass die Baumstruktur stets balanciert ist, d.h. die Blattknoten liegen alle auf derselben Hierarchiestufe des Baums. Der  $R$ -Baum ist eine sehr leistungsstarke Indexstruktur für niederdimensionale Vektorräume, da die Höhe des Baumes logarithmisch zur Anzahl der eingefügten Objekte ist. Eines der großen Probleme des  $R$ -Baum sind die Überlappungen der MBBs: Je mehr sich MBBs überlappen, desto geringer ist ihr Beitrag zur Eingrenzung des Suchraums. Da mit zunehmender Dimensionalität des zugrundeliegenden Vektorraums die Verteilung der begrenzten Anzahl von Punkten im Raum spärlicher wird, führt das automatisch zu MBBs mit größeren Ausmaßen, die sich wiederum mehr überlappen. Daher nimmt die Performanz des  $R$ -Baum mit steigender Dimensionalität des Vektorraums rapide ab.

In den letzten Jahren sind unzählige Erweiterungen bzw. Verfeinerungen des  $R$ -Baums publiziert worden. Das Grundkonzept ist jedoch stets dasselbe. Einen umfassenden und detaillierten Überblick aller in der Literatur dokumentierten Entwicklungen auf dem Gebiet der multidimensionalen Zugriffsmethoden erhält der interessierte Leser in [49]. An dieser Stelle sollen nur die wegweisendsten Entwicklungen kurz vorgestellt werden. Der von Sellis et al. [137] vorgeschlagene  $R^+$ -Baum wurde beispielsweise dahingehend verbessert, dass auf derselben Hierarchiestufe nur noch nicht-überlappende MBBs existieren. Zu diesem Zweck werden Objekte auf mehrere Knoten verteilt, d.h. Objekte werden in mehreren Blattknoten redundant gespeichert. Beckmann et al. [5]

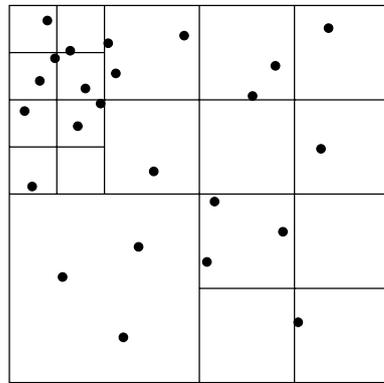
haben den  $R^*$ -Baum entwickelt, der sich vom  $R$ - und  $R^+$ -Baum dahingehend unterscheidet, dass die Knoten der Baumstruktur mittels einer besseren Strategie beim Einfügen von Objekten in einer Art und Weise restrukturiert werden, dass sich die Überlappungen der MBBs verringern. In [79] wird eine Baumstruktur vorgeschlagen, deren MMBs nicht mehr notwendigerweise Seitenflächen haben, die orthogonal zu den Achsen des Vektorraums liegen. Die MBBs des so genannten  $P$ -Baums werden durch Polyeder beschrieben, deren Seitenflächen entlang einer festgelegten Anzahl von Richtungen liegen können. Berchtold et al. [7] verwenden im Rahmen des  $X$ -Baum so genannte Superknoten, die ein oder mehrere Knoten zusammenfassen, falls deren MBBs eine zu starke Überlappung aufweisen. Dadurch entsteht eine Baumstruktur, dessen Knoten fast keine Überlappungen der MBBs mehr aufweisen. Das problematische für die Performanz des  $X$ -Baums ist jedoch, dass Superknoten sehr groß sind und mit linearer Komplexität durchsucht werden müssen.

Eine Weiterentwicklung des  $R^*$ -Baums ist der  $SS$ -Baum [173]. Das wesentlichste Unterscheidungsmerkmal ist, dass die MBBs des  $SS$ -Baum die Form von Hypersphären haben (siehe Abbildung 6.1(e)). Der Algorithmus zum Aufbau eines  $SS$ -Baums unterteilt den Suchraum folglich in isotropische Nachbarschaften und verbessert somit die Performanz von Nachbarschaftssuchen. Der Speicherbedarf der MBBs eines  $SS$ -Baums ist zudem deutlich geringer, da zur Repräsentation einer Hypersphäre lediglich Mittelpunkt und Radius benötigt werden. Eine MBB des  $SS$ -Baums lässt sich im  $n$ -dimensionalen Vektorraum mit maximal  $n + 1$  Parametern darstellen. Im Vergleich dazu sind zur Darstellung einer rechteckigen MBB mindestens  $2n$  Parameter notwendig. Der geringere Speicherbedarf bedingt, dass der Fan-Out der Baumknoten annähernd verdoppelt werden kann. Aufgrund der geringeren zu erwartenden Höhe des  $SS$ -Baums reduziert sich automatisch auch der Suchaufwand.

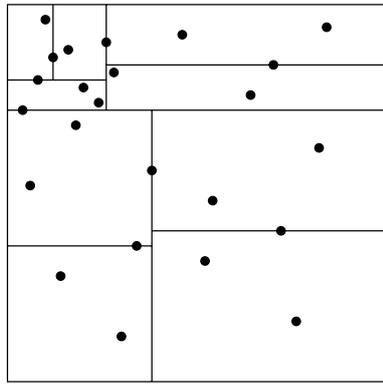
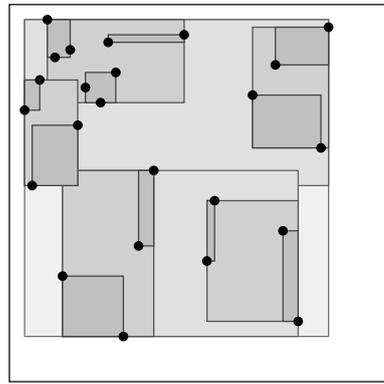
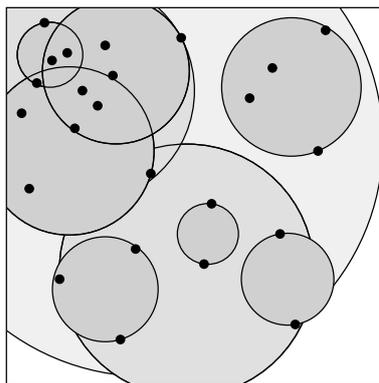
Obwohl der von Kurniawati et al. [87] vorgeschlagene  $SS+$ -Baum in der Struktur dem  $SS$ -Baum gleicht, wird jedoch ein vollständig neuartiger Weg bei der Konstruktion eingeschlagen. Im Gegensatz zum  $SS$ -Baum, der mittels einer herkömmlichen Top-Down-Strategie durch sukzessives Einfügen von Objekten in eine anfängliche leere Baumstruktur aufgebaut wird, wird bei dem Aufbau des  $SS+$ -Baums eine Bottom-Up-Strategie angewendet. Mittels eines Clustering-Algorithmus wird ein Datensatz zunächst in isotropische Nachbarschaften zerlegt, die dann a posteriori aufgrund ihrer topologischen Relationen in einer Baumstruktur zusammengefügt werden. Im Mittel werden dadurch kompaktere MBBs erzeugt. Eine Mischung des  $R^*$ - und des  $SS$ -Baums stellt der von Katayama und Satoh [83] entwickelte  $SR$ -Baum dar, der sowohl mit Hypersphären (in Gebieten mit hoher Datendichte) als auch Hyperquadern (in Gebieten mit geringer Datendichte) arbeitet.



(a) Grid-File



(b) Quad-Tree

(c) *k-d*-Baum(d) *R*-Baum(e) *SS*-Baum

**Abbildung 6.1:** Schematische Darstellung der Unterteilungsstrategien verschiedener multidimensionaler Zugriffsmethoden

Die im Rahmen dieses Abschnitts vorgestellten multidimensionalen Zugriffsmethoden setzen voraus, dass der zu indexierende Raum ein normierter Vektorraum ist. Ein normierter Vektorraum ist ein spezieller Fall eines metrischen Raums – er besitzt über die Existenz einer Norm (Metrik) hinaus zusätzlich eine Geometrie. Die Definition eines metrischen Raums ist daher allgemeiner gefasst als die eines Vektorraums – jeder normierte Vektorraum ist immer auch ein metrischer Raum, umgekehrt ist aber nicht jeder metrische Raum auch ein Vektorraum. Alle vorgestellten multidimensionalen Zugriffsmethoden nutzen die geometrische Struktur eines Vektorraums als Zusatzinformation zur Indexierung aus. Die Kenntniss der Dimensionalität des zugrundeliegenden Raums ist für multidimensionale Zugriffsmethoden beispielsweise eine notwendige Voraussetzung. Metrische Räume, die nicht gleichzeitig auch Vektorräume sind, besitzen jedoch keine explizite Dimensionalität. Daher eignen sich die multidimensionalen Zugriffsmethoden nicht zur Indexierung allgemeiner Merkmale auf der Basis des metrischen Modells.

#### 6.1.1.2 Metrische Räume

Uhlmann [158] hat als erster einen Ansatz zur Konstruktion einer Indexstruktur verfolgt, der nicht nur speziell für Vektorräume, sondern generell für alle metrischen Räume mit kontinuierlichen Distanzfunktionen anwendbar ist. Die Idee wurde später von Yianilos [177] und Chiueh [22] aufgegriffen und zum Konzept des *vp*-Baums vervollständigt. Der *vp*-Baum ist ein binärer Suchbaum, der einen metrischen Raum bezüglich einer Anzahl aus dem Datensatz zufällig ausgewählten so genannten „Aussichtspunkten“ (*engl. vantage points*) aufteilt. Der *vp*-Baum wird konstruiert, indem irgendein beliebiger Punkt als Aussichtspunkt ausgewählt wird und die übrigen Punkte bezüglich dieses Aussichtspunktes in zwei gleich große Partitionen aufgeteilt werden. Als Kriterium wird dabei der Median aller Entfernungen der Punkte zum entsprechenden Aussichtspunkt verwendet. Alle Punkte, die näher als der Median entfernt liegen, werden in den linken, alle Punkte, die weiter als Median entfernt liegen, in den rechten Subbaum einsortiert. Der linke und rechte Subbaum besitzen daher vollständig disjunkte MBBs, die aber jeweils in Vereinigung den gesamten Raum überdecken. Die Prozedur wird rekursiv fortgesetzt, bis ein Knoten nur noch ein Element enthält (der *vp*-Baum ist daher immer perfekt balanciert). Eine Nachbarschaftssuche entlang eines Zweigs des *vp*-Baums kann abgebrochen werden, falls aufgrund topologischer Relationen sicher ausgeschlossen werden kann, dass ein Punkt des Zweigs der Ergebnismenge zugehört. Wie in [177] festgestellt wird, ist die Wahrscheinlichkeit, dass die Suche entlang eines Zweigs nicht fortgesetzt werden muss, jedoch sehr gering. In der Regel müssen daher beide Subbäume eines Knotens besucht werden. Aus diesem Grund ist der *vp*-Baum als Indexstruktur in dieser Form nicht allgemein einsetzbar.

Bozkaya und Ozsoyoglu [12] schlagen zur Verbesserung der Performanz von  $vp$ -Bäumen die Verwendung von Baumstrukturen mit einem höheren Grad vor (der Grad eines Baumes bezeichnet die maximale Anzahl von Kindknoten eines inneren Knotens). Anstelle des Medians werden gleichförmige Percentile der Entfernungen zu einem Aussichtspunkt herangezogen, um den Raum in mehrere disjunkte Subräume aufzuteilen. In [12] wird experimentell gezeigt, dass die so genannten  $mvp$ -Bäume (engl. *multi vantage points*) geringfügig bessere Ergebnisse liefern als binäre  $vp$ -Bäume.

Die Verwendung von Baumstrukturen höheren Grades wird auch von Brin in [13] verfolgt. Im Gegensatz zum  $mvp$ -Baum wird der Raum an einem inneren Knoten jedoch bezüglich mehrerer Aussichtspunkte aufgeteilt. Jeder Punkt des Raumes wird dabei jeweils demjenigen Aussichtspunkt zugeordnet, zu dem er die geringste Entfernung ausweist. Der zugrundeliegende Raum wird dabei in MBBs partitioniert, die wie die Voronoi-Zellen einer Delaunay-Zerlegung aussehen (siehe Abbildung 4.14). Der so genannte GNAT (engl. *geometric near-neighbor access tree*) zeigt eine dem  $mvp$ -Baum vergleichbare Performanz.

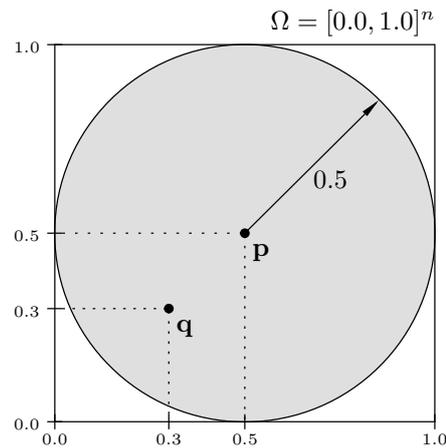
Ciaccia et al. [25] haben das Konzept des  $SS$ -Baum dahingehend erweitert, dass es nicht nur speziell auf Vektorräume, sondern generell für alle metrischen Räume anwendbar ist. Die von ihnen vorgeschlagene hierarchische Indexstruktur, der so genannte  $M$ -Baum, weist zwar eine gewisse Ähnlichkeit zu dem GNAT auf, unterscheidet sich jedoch grundsätzlich bezüglich der Strategie, wie bei einer Nachbarschaftssuche der Baum traversiert wird. Im Gegensatz zum GNAT weist der  $M$ -Baum daher eine gute I/O Performanz mit besseren dynamischen Fähigkeiten auf. In [25] wird experimentell gezeigt, dass der  $M$ -Baum gegenüber dem  $R^*$ -Baum konkurrenzfähig ist.

Die in diesem Abschnitt vorgestellten Indexstrukturen ( $vp$ -Baum,  $mvp$ -Baum, GNAT und  $M$ -Baum) lassen sich auf Problemstellungen einsetzen, die über die herkömmliche Darstellung von Objekten als Punktdaten im Vektorraum hinausgehen. So können beispielsweise auch Zeichenketten indiziert werden, die mit der Levenshtein-Distanz [91] verglichen werden. Die Indexstrukturen eignen sich somit grundsätzlich zur Verwaltung beliebiger Merkmale auf der Basis des metrischen Modells.

### 6.1.1.3 „Fluch der Dimensionen“

Ein Grund für die Flut von veröffentlichten Indexierungsverfahren ist, dass die Indexstrukturen bei einer hohen (auch impliziten) Dimensionalität des zugrundeliegenden Merkmalsraums generell versagen. Verantwortlich für die schlechte Performanz sind eine Reihe von Effekten, die üblicherweise mit dem Schlagwort „Fluch der Dimensionen“ (engl. *curse of dimensionality*) zusammengefasst werden [122]. Die Auswirkungen dieser Effekte sind im folgenden Beispiel anschaulich dargestellt.

**Beispiel 6.1 („Fluch der Dimensionen“)** Gegeben seien ein  $n$ -dimensionaler Raum  $\Omega = [0.0, 1.0]^n$ , zwei Punkte  $\mathbf{p} = (0.5, \dots, 0.5) \in \Omega$  und  $\mathbf{q} = (0.3, \dots, 0.3) \in \Omega$  sowie eine Hypersphäre um  $\mathbf{p}$  mit dem Radius 0.5. Im zweidimensionalen Fall (siehe Abbildung 6.2) hat  $\mathbf{q}$  zu  $\mathbf{p}$  die Distanz  $\delta(\mathbf{p}, \mathbf{q}) = \sqrt{0.2^2 + 0.2^2} = 0.283$  und liegt damit innerhalb des Kreises um  $\mathbf{p}$ .



**Abbildung 6.2:** Veranschaulichung des Phänomens „Fluch der Dimensionen“

In höherdimensionalen Fällen gilt dies jedoch nicht mehr. Die Distanz zwischen  $\mathbf{p}$  und  $\mathbf{q}$  ist allgemein gegeben als  $\delta(\mathbf{p}, \mathbf{q}) = 0.2 \cdot \sqrt{n}$ .  $\mathbf{q}$  liegt innerhalb der Hypersphäre, falls  $\delta < 0.5$  ist. Dies ist aber nur für  $n < (\frac{0.5}{0.2})^2 = 6.25$  der Fall. Das heißt bereits in Räumen mit sieben oder mehr Dimensionen liegt  $\mathbf{q}$  nicht mehr innerhalb der Hypersphäre um  $\mathbf{p}$ . Betrachtet man weiterhin den Kreis im zweidimensionalen Fall, so ist ersichtlich, dass der Kreis 78.54% des Raums überdeckt. Bereits im zehndimensionalen Fall überdeckt dieselbe Hypersphäre nur noch 0.31% des Raums.

Ein hochdimensionaler Raum verhält sich offensichtlich völlig anders, als man aus dem Verhalten eines zweidimensionalen Raums intuitiv schließen würde. Diese Eigenheiten sind dafür verantwortlich, dass hierarchische Strukturen, die zur Indexierung niederdimensionaler Räume konzipiert wurden, zwangsläufig im höherdimensionalen Fall versagen.

### 6.1.2 Dimensionsreduktion

Die Forschungsgemeinschaft hat in der Vergangenheit viele Versuche unternommen, das Problem des Fluchs der Dimensionen in den Griff zu bekommen. Ein oft verfolgter Ansatz ist die explizite Reduktion der Dimensionalität des Merkmalsraums durch eine

Hauptachsentransformation (*engl. principal component analysis*) oder deren Varianten. In [40] schlagen Faloutsos und Lin die Verwendung der Karhunen-Loeve Transformation vor, um einen hochdimensionalen Merkmalsraum mittels eines Abbildraums mit einer geringeren Anzahl von Dimensionen anzunähern. Ng und Sedighian [112] konnten experimentell zeigen, dass, obwohl die Dimensionalität einiger reeler Merkmalsräume signifikant reduziert werden konnte, die Qualität der Suchergebnisse generell proportional mit dem Verlust der Dimensionen degenerierte. Um den Einfluss herkömmlicher Indexierungsverfahren im Rahmen dieses Ansatzes spürbar zu machen, muss die Dimensionalität jedoch soweit reduziert werden, dass die Suchergebnisse in der Regel qualitativ wenig akkurat und somit unbrauchbar sind.

Einen unterschiedlichen Ansatz zur Reduktion der Dimensionalität von allgemeinen metrischen Merkmalsräumen verfolgen Vleugels und Veltkamp in [166]. Ihr Ziel ist es, einen beliebigen metrischen Raum auf einen Vektorraum mit frei wählbarer Dimensionalität abzubilden. Dazu muss a priori eine Anzahl von  $N$  zufällig ausgewählten „Aussichtspunkten“ im metrischen Raum festgelegt werden. Jeder Punkt des Merkmalsraums kann dann durch einen  $N$ -dimensionalen Vektor dargestellt werden, dessen Werte die Distanzen zu den jeweiligen Aussichtspunkten bezeichnen. Offensichtlich ist die Auswahl der Anzahl und Lage der Aussichtspunkte im metrischen Raum der kritische Punkt für dieses Verfahren. In [166] wird dazu ein Auswahlkriterium vorgeschlagen, dass eine Menge von Aussichtspunkten so auswählt, dass der metrische Raum von ihnen „gut genug“ überdeckt wird. Die zentralen Fragen, wie adäquat eine Abbildung den ursprünglichen Raum repräsentiert und wieviele Dimensionen für eine solche Abbildung notwendig sind, werden jedoch nicht beantwortet.

### 6.1.3 Ungefähre Suche

In den letzten Jahren ist in der Literatur ein verstärktes Interesse an Verfahren festzustellen, die das Problem der hohen Dimensionalität zu umgehen versuchen, indem sie Indexierungsstrukturen bereitstellen, die lediglich eine ungefähre Suche in dem jeweiligen Merkmalsraum ermöglichen [1] [74] [180]. Diese Verfahren gehen davon aus, dass der Vergleich visueller Inhalte auf Grundlage des metrischen Modells an sich kein mathematischer, sondern ein heuristischer Ansatz ist, der den Begriff der Ähnlichkeit nur ungenau repräsentiert. Aus diesem Blickwinkel erscheint es unlogisch, auf einer exakten Nachbarschaftssuche zu beharren. Der Ansatz der ungefähren Suche basiert daher auf einer Abschwächung der Anforderungen an die Suche im Merkmalsraum – das Suchergebnis einer ungefähren Suche kann bis zu einem gewissen Grad von dem einer exakten Suche abweichen. Das Resultat sollte aus Sicht der meisten praktischen Anwendungen trotzdem nutzbar sein. Der Ansatz der ungefähren Suche hat das Problem des „Fluchs der Dimensionen“ bisher zwar nicht grundsätzlich gelöst, jedoch

zeigen neuere Untersuchungen ermutigende Resultate [88] [24]. Eine umfassende und eingehende Behandlung der ungefähren Suche ist in der Studie von White und Jain in [173] zu finden, an die der interessierte Leser an dieser Stelle verwiesen wird.

Im Rahmen dieser Arbeit wurde ein Ansatz verfolgt, der verschiedene Aspekte des  $M$ -Baums, der Dimensionsreduktion mittels „Aussichtspunkten“ und der ungefähren Suche aufgreift und zu einer leistungsfähigen Indexierungsstruktur für beliebige metrische Merkmalsräume kombiniert [168] [169]. Die Indexierungsstruktur wird in den folgenden Abschnitten näher vorgestellt.

## 6.2 Bojen-Indexierung

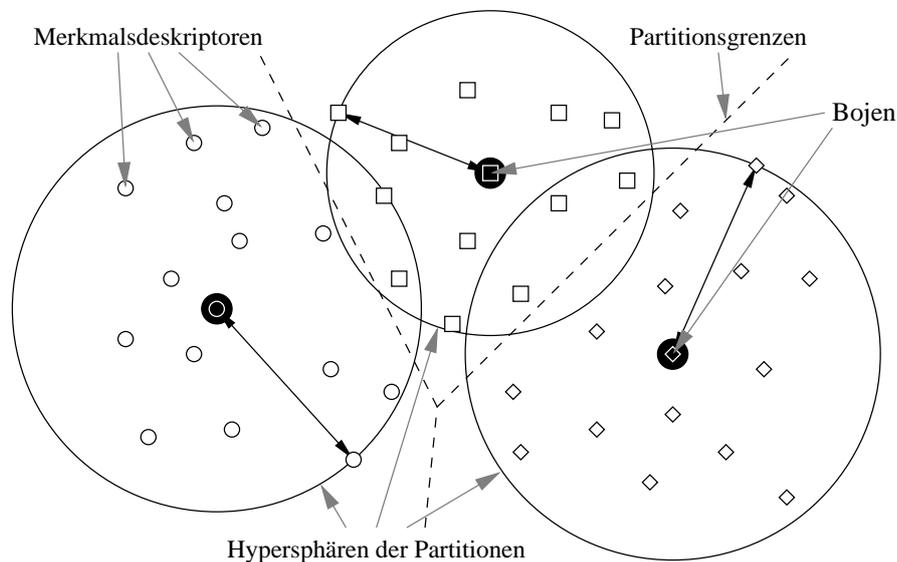
Das Ziel bei der Entwicklung der hier vorgestellten Strategie zur Merkmalsindexierung ist, eine pragmatische Lösung für das Problem der inhaltsbasierten Suche zu finden. Im Unterschied zu Vorgehensweisen, die das Problem der Nachbarschaftssuche in allgemeinen metrischen Räumen auf der Grundlage theoretischer Überlegungen zu lösen versuchen, ermöglicht dieser Ansatz das Ausnutzen von Besonderheiten, die im praktischen Umgang mit realem Bildmaterial auftreten. Insbesondere ist die Tatsache nutzbar, dass die Merkmalsdeskriptoren eines realen Bilddatensatzes nicht gleichmäßig im Merkmalsraum verteilt sind und sich in bestimmten Bereichen des Merkmalsraums stärker konzentrieren als in anderen. Es ist aus dieser Sicht daher wenig sinnvoll, eine bestimmte Verteilung der Merkmalsdeskriptoren im Merkmalsraum vorauszusetzen – jeder Bilddatensatz weist in der Praxis eine eigene charakteristische Verteilung auf.

Die Grundidee des Verfahrens ist daher, die Indexierung jeweils an den vorliegenden Bilddatensatz anzupassen. Dazu wird der Merkmalsraum anhand der in einem Datensatz vorhandenen Verteilung von Merkmalsdeskriptoren in disjunkte, nicht-leere Partitionen unterteilt, die idealerweise jeweils die gleiche Anzahl von Merkmalsdeskriptoren enthalten. Die so entstehenden Partitionen besitzen unterschiedliche Ausdehnungen im Merkmalsraum – in Bereichen mit einer hohen Verteilungsdichte entstehen so kompaktere Partitionen als in Bereichen mit einer geringen Verteilungsdichte.

In allgemeinen metrischen Räumen steht als einziges Mittel zur Repräsentation von topologischer Information die Abstandsmetrik zur Verfügung. Diese Tatsache schließt die Beschreibung der Partition durch explizite Angabe der Grenzen im Merkmalsraum kategorisch aus. Aus diesem Grund wird aus einer Partition ein einzelner Merkmalsdeskriptor ausgewählt, der die Topologie der Partition prototypisch repräsentiert. Bildlich gesprochen vertritt dieser Merkmalsdeskriptor seine assoziierte Partition gewissermaßen als Boje im Merkmalsraum. Diese Metapher gibt dem Verfahren auch den Namen „Bojen-Indexierung“ (*engl. buoy indexing*).

Die Zugehörigkeit eines einzelnen Merkmalsdeskriptors zu einer bestimmten Partition ist eindeutig über seinen Abstand zu den Bojen im Merkmalsraum festgelegt: Jeder Merkmalsdeskriptor wird exklusiv derjenigen Partition zugeordnet, zu dessen Boje er am nächsten liegt. Die Ausdehnung einer Partition im Merkmalsraum wird allein durch die Topologie der im Merkmalsraum platzierten Bojen bestimmt. Die Partitionsgrenzen sind implizit Hyperebenen, die jeweils den identischen Abstand zu einem Bojenpaar im Merkmalsraum haben. Partitionen haben daher immer eine konvexe Form. Die Partitionen sind vergleichbar mit den Voronoi-Zellen, die bei einer Delaunay-Zerlegung einer Menge von Punkten im  $n$ -dimensionalen Raum entstehen (siehe Abbildung 4.14).

Obwohl die Ausdehnung jeder Partition nicht in expliziter Form bestimmbar ist, kann jedoch für jede Partition deren Größe geschätzt werden. Jede Partition ist in ihrer Ausdehnung begrenzt, d.h. es existiert eine Hypersphäre, die die Partition vollständig umhüllt. Nimmt man als Mittelpunkt die jeweilige Boje der Partition an, dann hat die kleinste Hypersphäre, die die Partition vollständig umhüllt, einen Radius der mit dem Abstand des am weitesten von der Boje entfernt liegenden Merkmalsdeskriptors der Partition identisch ist. Die Größe der Hypersphäre gibt keinerlei Auskunft über die tatsächliche Form und Ausdehnung der Partition. Sie legt lediglich eine obere Grenze der Partitionsgröße fest.



**Abbildung 6.3:** Schematische Darstellung der Partitionierung des Merkmalsraums mittels Bojen

Abbildung 6.3 verdeutlicht das Prinzip der Partitionierung eines metrischen Merkmalsraums durch Bojen an einem zweidimensionalen Beispiel. Ein Datensatz mit insgesamt 44 Merkmalsdeskriptoren wird anhand von drei Bojen (die selbst Merkmalsdeskripto-

ren sind) in etwa drei gleich große Partitionen unterteilt. Die Partitionszugehörigkeit der Merkmalsdeskriptoren bezüglich der gewählten Bojentopologie ist durch unterschiedliche Symbole gekennzeichnet. Die gestrichelten Linien repräsentieren die äquidistanten Punkte zwischen benachbarten Partitionen und markieren damit die impliziten Partitions Grenzen. Konzentrische Kreise um die Bojen stellen die jeweils kleinste umhüllende Hypersphäre einer Partition dar. Es ist offensichtlich, dass ein innerhalb einer Hypersphäre einer Partition liegender Merkmalsdeskriptor nicht automatisch auch dieser Partition zugeordnet ist.

Die grundsätzliche Frage ist: Wie wählt man aus einem Datensatz eine geeignete Menge von Merkmalsdeskriptoren als Bojen aus? Im Folgenden wird zur Lösung dieses Problems eine Strategie angewendet, die mittels eines Clustering-Algorithmus einen Datensatz in isotropische Partitionen zerlegt, die den formulierten Anforderungen der Partitionierung eines allgemeinen metrischen Raumes mittels Bojen genügen.

### 6.2.1 Indexerzeugung

Algorithmen, die in einem Datensatz mit nicht ausgezeichneten Daten eine Struktur erkennen und eine entsprechende Einteilung des Datenraumes vornehmen, gehören zu der Gruppe der unüberwachten Clustering-Algorithmen. Unter Clustering versteht man ganz allgemein die „sinnvolle“ Einteilung einer Menge von Objekten in Clustern (*dt. Gruppen*), so dass sich ähnliche Objekte in einem Cluster befinden. Unüberwachte Clustering-Algorithmen haben in der Regel die Anforderung, dass die Anzahl der zu bildenden Cluster a priori festgelegt sein muss.

**Definition 6.1 (K-Clustering)** Sei  $\mathcal{W}$  ein Datensatz und  $k \in \mathbb{N}$  eine positive natürliche Zahl. Unter *K-Clustering* versteht man eine Partitionierung von  $\mathcal{W}$  in  $K$  Untermengen  $\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_k$  unter der Voraussetzung, dass die Eigenschaften

(i) *Vollständigkeit*

$$\bigcup_{i=1}^k \mathcal{W}_i = \mathcal{W} \quad (6.1)$$

(ii) *Nicht-Leerheit*

$$\mathcal{W}_i \neq \emptyset, \quad \forall 1 \leq i \leq k \quad (6.2)$$

(iii) *Paarweise Disjunktheit*

$$\mathcal{W}_i \cap \mathcal{W}_j = \emptyset, \quad \forall 1 \leq i, j \leq k, i \neq j \quad (6.3)$$

gelten. ■

**Bemerkung:** Jedes Element des Datensatzes  $\mathcal{W}$  wird genau einer Untermenge  $\mathcal{W}_i$  zugeordnet.

Das  $K$ -Clustering ist ein NP-vollständiges Problem, d.h. es existiert kein deterministischer Algorithmus, der das Problem mit polynomialer Komplexität lösen kann. Das  $K$ -Clustering formuliert man daher gewöhnlich als Optimierungsproblem, zu dem es aus einer Menge möglicher Lösungen eine bestmögliche Lösung zu finden gilt. Die Qualität jeder potentiellen Lösung wird dabei durch eine lokale Kostenfunktion

$$c : \{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_k\} \rightarrow \mathbb{R}_0^+ \quad (6.4)$$

bewertet, die jedem Cluster  $\mathcal{W}_i$  einen nicht-negativen Kostenwert zuweist. Das Ziel des Optimierungsalgorithmus ist nun die Minimierung der globalen Kostenfunktion

$$c(\mathcal{W}) = \sum_{i=1}^k c(\mathcal{W}_i) \quad (6.5)$$

für den gesamten Merkmalsdatensatz. Im vorliegenden Fall kann die lokale Kostenfunktion eines Cluster sehr einfach als die Summe der Abstände aller Merkmalsdeskriptoren zu der jeweiligen Boje definiert werden

$$c(\mathcal{W}_i) = \sum_{j=1}^{|\mathcal{W}_i|} \delta(\dot{\omega}_i, \omega_{i,j}) \quad (6.6)$$

wobei  $\dot{\omega}_i$  die Boje,  $\omega_{i,j}$  den  $j$ -ten Merkmalsdeskriptor und  $|\mathcal{W}_i|$  die Anzahl der assoziierten Merkmalsdeskriptoren des Clusters  $\mathcal{W}_i$  bezeichnen.

Die lokale Kostenfunktion (6.6) ist automatisch minimal, falls der Mittelwert (Schwerpunkt) aller Merkmalsdeskriptoren als Boje gewählt würde. Das entsprechende Verfahren ist in der Literatur unter dem Namen  $K$ -Means-Clustering wohlbekannt [11]. Das Problem an dieser Stelle ist, dass im metrischen Merkmalsraum kein Mittelwert definiert ist. Daher muss als Boje ein Element des Clusters selbst ausgewählt werden, da nur dann garantiert ist, dass die Boje auch Element des metrischen Merkmalsraum ist

$$\dot{\omega}_i \stackrel{!}{\in} \mathcal{W}_i \subset \mathcal{W} \subset \Omega \quad (6.7)$$

Mit dieser Einschränkung wird (6.6) minimal, falls der Zentralwert (Median) aller Merkmalsdeskriptoren als Boje gewählt wird. Das Verfahren wird daher auch als  $K$ -Medians-Clustering bezeichnet.

Der klassische Ansatz zur Lösung des Optimierungsproblems ist ein Algorithmus, der die globale Kostenfunktion iterativ minimiert. Der Optimierungsalgorithmus konvergiert, falls (6.5) in zwei aufeinanderfolgenden Iterationen konstant bleibt. Die so gefundene stabile Lösung stellt ein lokales Minimum der Kostenfunktion dar.

### 6.2.1.1 Äußere Zwänge

Der  $K$ -Medians-Clustering Algorithmus in seiner Grundform produziert Cluster mit einer nicht vorhersagbaren Größe im Bereich von

$$1 \leq |\mathcal{W}_i| \leq |\mathcal{W}| - k + 1 \quad (6.8)$$

Idealerweise sollte das Verfahren jedoch in einer Partitionierung resultieren, die aus Clustern gleicher Größe besteht. Dies kann näherungsweise erreicht werden, wenn dem Optimierungsalgorithmus äußere Zwänge auferlegt werden. Dazu kommen prinzipiell zwei verschiedene Ansätze in Frage: Einerseits kann die Größe eines Clusters mit einem logarithmischen Faktor in der Kostenfunktion bestraft werden, andererseits besteht die Möglichkeit, mittels einer Split-And-Merge-Strategie die Entwicklung der Clustergrößen direkt zu beeinflussen. Da die Bestimmung der Bojen selbst über die Minimierung der lokalen Kosten erfolgt, sollte diese Funktion nicht verändert werden. Daher bietet sich im vorliegenden Fall die Verwendung einer Split-And-Merge-Strategie an.

Die Ausdehnung der Cluster im Merkmalsraum kann während der Optimierung kontrolliert werden, indem durch einen empirisch gewählten Grenzwert erzwungen wird, dass die Cluster eine maximale Größe nicht überschreiten

$$|\mathcal{W}_i| \stackrel{!}{\leq} W_{\max} \quad (6.9)$$

Falls ein Cluster die Größe  $W_{\max}$  überschreitet, wird das Cluster in zwei Cluster mit ungefähr gleicher Größe aufgeteilt: Die mit dem entsprechenden Cluster assoziierten Merkmalsdeskriptoren werden zufällig auf die beiden Cluster verteilt. Da die Anzahl  $K$  aller Cluster eine konstante Größe ist, muss das kleinste Cluster im Gegenzug gelöscht werden (die entsprechenden Merkmalsdeskriptoren werden gemäß ihres Abstandes zu den Bojen auf andere Cluster umverteilt). Abbildung 6.4 erläutert den so modifizierten Optimierungsalgorithmus in einer hochsprachlichen Form.

Die Auswahl des Grenzwerts  $W_{\max}$  hat direkten Einfluss auf das Konvergenzverhalten des Clustering-Algorithmus. Der Erwartungswert der Größe eines Clusters ist

$$E(|\mathcal{W}_i|) = \frac{N}{k} \quad (6.10)$$

wobei  $N$  die Größe des gesamten Merkmalsdatensatzes bezeichnet. Offensichtlich muss  $W_{\max} > \frac{N}{k}$  gelten, damit der modifizierte Optimierungsalgorithmus konvergieren kann. Dieses Kriterium ist notwendig, aber nicht hinreichend, um die Konvergenz des Verfahrens zu gewährleisten. Eine allzu stringente Auswahl des Grenzwerts resultiert in der Regel in unstabiler Verhalten, da dieser äußere Zwang zu einer ständigen Umverteilung der Cluster führt. Der Algorithmus benötigt folglich eine sehr viel höhere Anzahl von Iterationen bis zur Konvergenz, bzw. er konvergiert möglicherweise gar nicht.

---

```

Ordne alle Deskriptoren zufällig einem Cluster zu
Bestimme die Bojen der Cluster durch Minimierung der lokalen Kosten
Berechne die globale Kosten
repeat
  for all Cluster, deren Größe den Grenzwert überschreitet do
    Finde das kleinste Cluster
    Ordne alle Deskriptoren des kleinsten Clusters anderen Clustern zu
    Verteile Deskriptoren zufällig auf die beiden Cluster
    Bestimme die Bojen der beiden Cluster durch Minimierung der lokalen Kosten
  endfor
  Ordne alle Deskriptoren den Clustern erneut zu
  Bestimme die Bojen der Cluster durch Minimierung der lokalen Kosten
  Berechne die globalen Kosten
until Globale Kosten bleiben in zwei aufeinanderfolgenden Iterationen unverändert

```

---

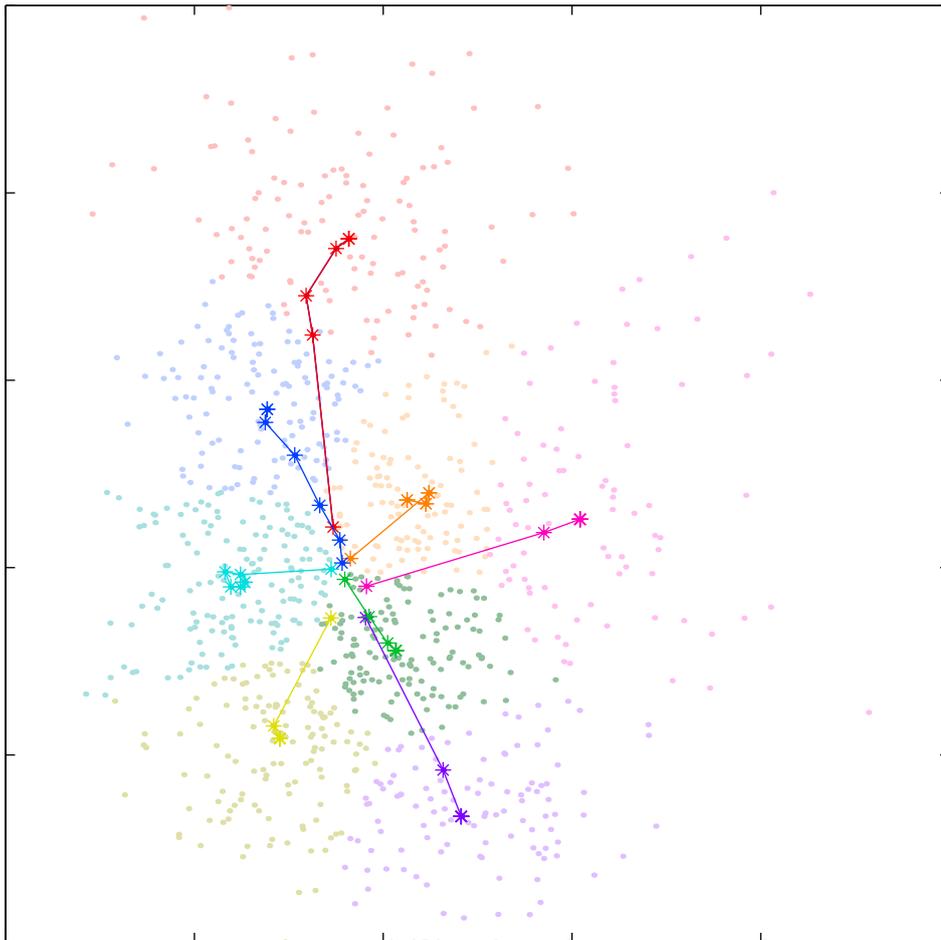
**Abbildung 6.4:** Iterative Optimierung mit äußerem Zwang

### 6.2.1.2 Dynamische Aktualisierung

Ein Index muss fortlaufend den Veränderungen des Datensatzes angepasst werden, da neue Bilder eingefügt, bzw. existierende Bilder verändert oder gelöscht werden können. Dies sollte in der Regel durch eine dynamische Aktualisierung geschehen, ohne dass der Index vollständig überarbeitet werden muss. Beim Hinzufügen eines Bildes wird der entsprechende Merkmalsdeskriptor in den Index eingepflegt, indem er intuitiv dem Cluster zugeordnet wird, zu dessen Boje er den geringsten Abstand hat. Das Löschen eines Bildes erfolgt durch das einfache Entfernen des korrespondierenden Merkmalsdeskriptors aus dem Cluster. Die Modifikation eines Bildes kann durch eine Lösch- gefolgt von einer Hinzufüge-Operation modelliert werden. In jedem Fall erfolgt keine Anpassung der Clusterbojen, da dies für jede Operation die Überarbeitung des gesamten Index zur Folge hätte.

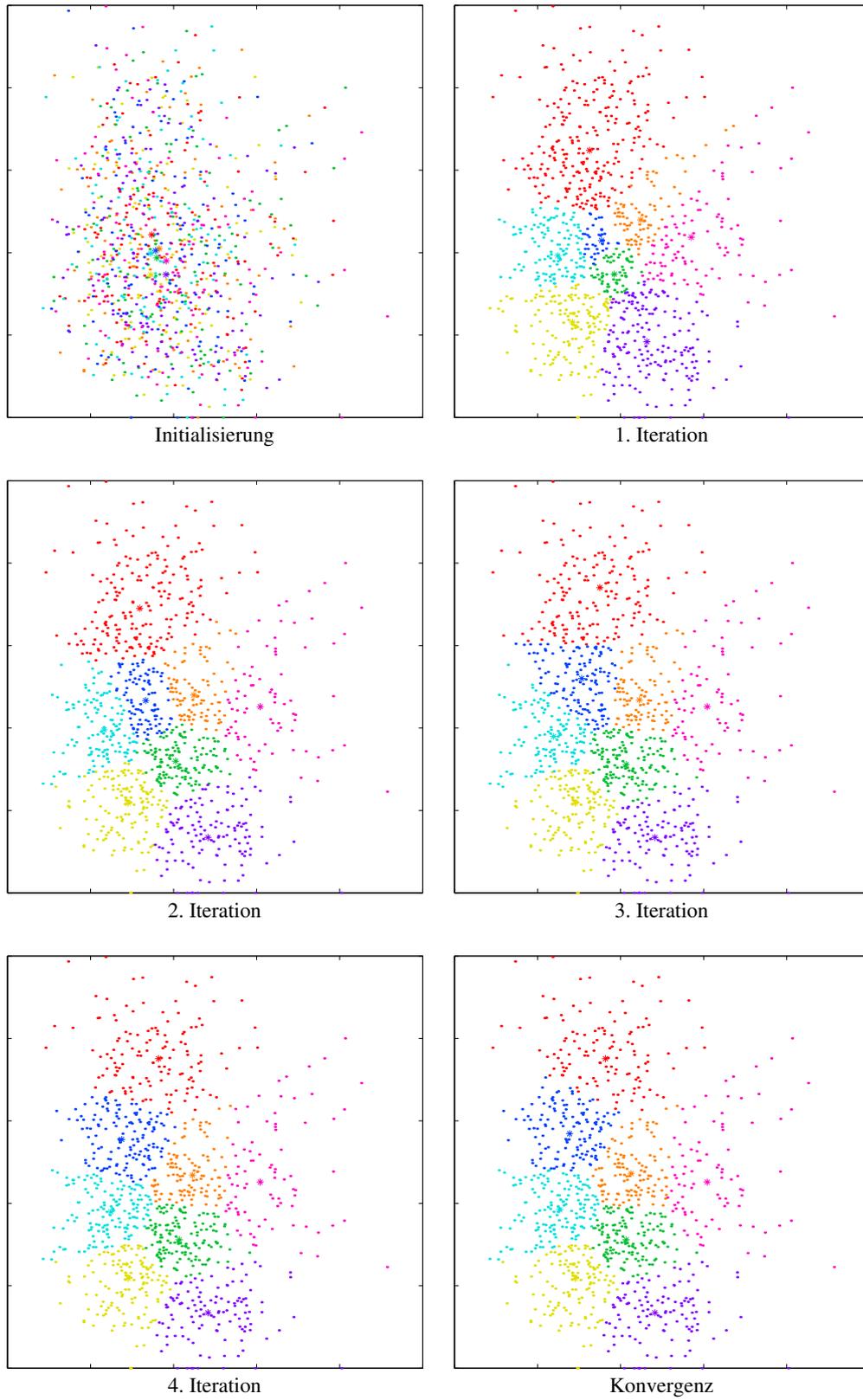
Die Konsequenz einer Vielzahl von Veränderungen ist, dass viele Cluster möglicherweise über die Zeit nicht mehr adäquat durch ihre Boje repräsentiert werden. Dies macht eine Aktualisierung des Index notwendig, die jedoch nur selten durchgeführt werden muss, da sie die prinzipielle Funktionsweise des Index nicht beeinträchtigt. Eine Aktualisierung des Index wird erreicht, indem der Algorithmus aus Abbildung 6.4 mit dem alten Index initialisiert und die iterative Optimierung erneut gestartet wird. Analog dazu kann der Index bei Bedarf durch Hinzufügen einer Anzahl zusätzlicher Cluster erweitert werden, falls einige Cluster die gesetzte Maximalgröße überschreiten. In jedem Fall ist eine Aktualisierung des Index eine Operation, die offline erfolgen kann und den laufenden Betrieb eines Bildarchivs nicht beeinträchtigt.

**Beispiel 6.2 ( $K$ -Medians-Clustering)** *Abbildung 6.6 zeigt die iterative Optimierung des  $K$ -Medians-Clustering über vier Iterationen bishin zur Konvergenz an einem Beispiel im zweidimensionalen Vektorraum. Zu Beginn der Optimierung wird jedes der 1000 Elemente des Datensatzes zufällig einem der acht Cluster zugeordnet. Die jeweiligen Cluster-Zugehörigkeiten der Elemente sind farblich gekennzeichnet, die Bojen zusätzlich durch ein sternförmiges Symbol.*



**Abbildung 6.5:** Trajektorien der Bojen während der iterativen Optimierung

*Die größten Veränderungen der Topologie der Bojen erfolgen innerhalb der ersten Iterationen. Die Trajektorien der Bojen während der Optimierung sind zur Verdeutlichung in *Abbildung 6.5* separat dargestellt. Es ist deutlich zu erkennen, dass sich die Ausdehnung der Cluster den topologischen Gegebenheiten des Datensatzes anpasst: In Bereichen mit einer hohen Verteilungsdichte entstehen kompaktere Partitionen als in Bereichen mit einer geringen Verteilungsdichte.*



**Abbildung 6.6:** Iterative Optimierung des  $K$ -Medians-Clustering

### 6.2.2 Bojen-indexierte Nächste-Nachbar-Suche

Beim Problem der  $k$ -Nächste-Nachbar-Suche (siehe Definition 3.14) lässt sich der Zeitaufwand verringern, indem man eine triviale topologische Eigenschaft des Merkmalsdatensatzes ausnutzt, die das naive Verfahren einer linearen Suche ganz übersieht: Die  $k$  nächsten Nachbarn des Merkmalsdeskriptors einer Suchanfrage müssen in dessen näherer Umgebung zu finden sein. Es sollte folglich nicht notwendig sein, die Entfernungen zu allen – insbesondere den weit entfernten – Merkmalsdeskriptoren zu prüfen. Daher ist offensichtlich die beste Strategie, die Suche auf die unmittelbare Umgebung des Merkmalsdeskriptors der Suchanfrage zu beschränken.

Die Idee der indexierten Nächste-Nachbar-Suche ist daher denkbar einfach: Die Cluster werden sequentiell in der Reihenfolge der Entfernung ihrer Boje zu dem Merkmalsdeskriptor der Suchanfrage durchsucht. Die Suche wird abgebrochen, wenn eine „ausreichende“ Anzahl von in der näheren Umgebung liegenden Clustern durchsucht wurde.

Das Problem dieser Vorgehensweise ist, dass es analytisch unmöglich ist, die genaue Anzahl der Cluster in der näheren Umgebung zu bestimmen, die für die Zusammenstellung eines akkuraten Ergebnisses benötigt werden. Der Grund dafür liegt in der Tatsache, dass die genaue Topologie des Merkmalsdatensatzes in der unmittelbaren Umgebung der Suchanfrage unbekannt ist. Der Merkmalsdeskriptor der Suchanfrage könnte beispielsweise in einem Gebiet mit einer geringen Verteilungsdichte liegen. Die Varianz der Abstandsverteilung dieses Deskriptors zu allen anderen des Merkmalsdatensatzes ist daher ebenfalls gering. Als Konsequenz daraus existieren viele potentielle Nächste-Nachbar-Kandidaten mit annähernd der gleichen Entfernung zur Suchanfrage. Würde die Suche nach der Durchsuchung zu weniger Cluster abgebrochen, könnten einzelne für das Ergebnis relevante Merkmalsdeskriptoren unberücksichtigt bleiben: Die indexierte Suche würde nicht dasselbe Ergebnis liefern, wie die lineare Suche.

Da der Begriff der Ähnlichkeit auf der Grundlage des metrischen Modells an sich ein heuristischer Ansatz ist, erscheint es unlogisch, im Kontext CBIR die Forderung einer exakten Nächste-Nachbar-Suche aufrecht zu erhalten. Ein ungefähres Suchergebnis, das bis zu einem gewissen Grad vom exakten Ergebnis abweichen kann, ist in der Praxis ausreichend. Daher wird im Folgenden der pragmatische Ansatz verfolgt, dass die Suche nach der Durchsuchung eines Perzentils  $q \in (0, 1]$  von Clustern in der näheren Umgebung der Suchanfrage abgebrochen wird. Dabei wird bewusst in Kauf genommen, dass das Ergebnis möglicherweise nicht hundertprozentig akkurat ist.

In der Praxis ist daher die Vorgehensweise bei der Durchführung einer indizierten  $k$ -Nächste-Nachbar-Suche wie folgt: Zunächst werden die Entfernungen der Bojen aller Cluster zu dem Merkmalsdeskriptor der Suchanfrage berechnet. Mit dieser Information

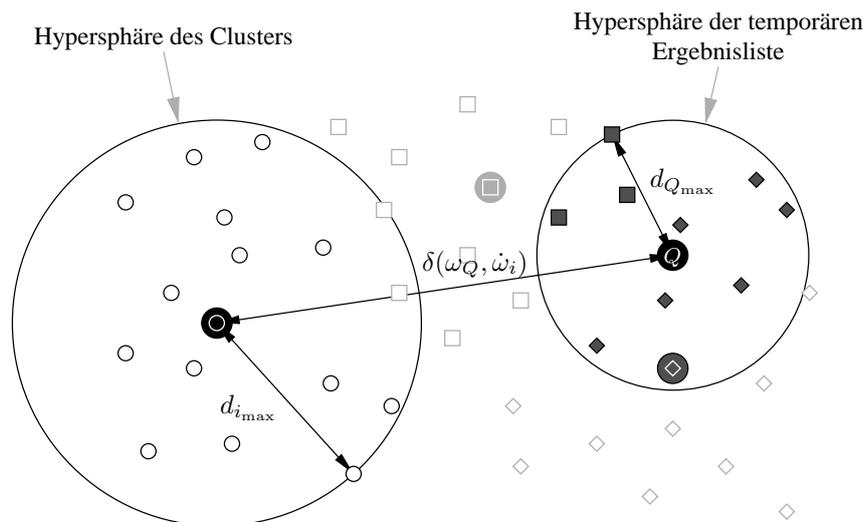
werden alle Cluster gemäß ihrer Nähe zur Suchanfrage sortiert. Die eigentliche Suche wird dann mittels einer geordneten Sequenz von lokalen  $k$ -Nächste-Nachbar-Suchen durchgeführt, die jeweils auf ein Cluster begrenzt sind. Die Suche in einem Cluster selbst erfolgt mit linearer Komplexität unter Prüfung aller dem Cluster assoziierten Merkmalsdeskriptoren. Während der sukzessiven Durchsuchung der Cluster wird eine Liste der vorläufigen  $k$  nächsten Nachbarn geführt, die ständig aktualisiert wird.

Bevor eine Teilsuche in einem Cluster durchgeführt wird, kann über ein Ausschlusskriterium a priori geprüft werden, ob die Suche tatsächlich durchgeführt werden muss oder nicht. Ein Cluster kann übergangen werden, falls

$$\delta(\omega_Q, \dot{\omega}_i) > d_{i_{\max}} + d_{Q_{\max}} \quad (6.11)$$

wobei  $\delta(\omega_Q, \dot{\omega}_i)$  die Distanz zwischen Suchanfrage und Boje des Clusters,  $d_{i_{\max}}$  der Radius der Hypersphäre des Clusters und  $d_{Q_{\max}}$  der Radius der Hypersphäre des vorläufigen Ergebnisses um die Suchanfrage ist. Aufgrund der vorliegenden Topologie kann in diesem Fall die Durchsuchung des Clusters das vorläufige Ergebnis nicht beeinflussen.

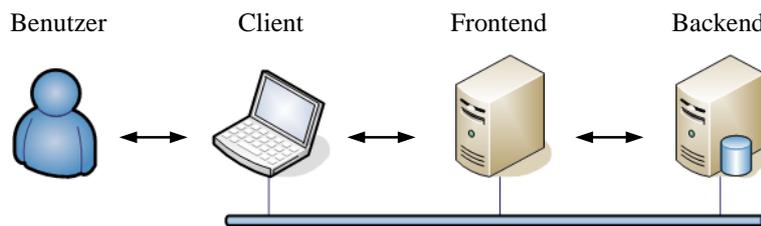
**Beispiel 6.3 (Ausschlusskriterium)** *Abbildung 6.7 zeigt die schematische Darstellung der Anwendung des Ausschlusskriteriums im Verlauf einer indexierten 12-Nächste-Nachbar-Suche. Die Abbildung zeigt die Situation, dass Teilsuchen in den beiden der Suchanfrage am nächsten liegenden Clustern ( $\square, \diamond$ ) bereits eine vorläufige Ergebnisliste ergeben haben. Da sich die Hypersphäre des Clusters ( $\circ$ ) und die der vorläufigen Ergebnisliste nicht schneiden, kann auf eine Durchsuchung des entsprechenden Clusters verzichtet werden.*



**Abbildung 6.7:** Schematische Darstellung des Ausschlusskriteriums

### 6.2.3 Evaluation

Im Folgenden werden experimentelle Ergebnisse, die im Rahmen einer prototypischen Realisierung erzielt wurden, detailliert vorgestellt und hinsichtlich ihrer Eignung für die merkmalsbasierte Bildsuche untersucht. Abbildung 6.8 zeigt den für das Sammeln der Ergebnisse verwendeten Systemaufbau.



**Abbildung 6.8:** Experimenteller Aufbau

Kernstück des Prototyps ist ein Frontend- und ein Backend-Server. Auf dem Frontend-Server läuft die eigentliche Suchmaschine, die den Index (Clusterbojen und -radien) vollständig im Hauptspeicher vorhält. Der Backend-Server dient als Datenbank und ist für die persistente Speicherung aller Merkmalsdeskriptoren des indexierten Datensatzes verantwortlich. Auf Anweisung des Frontend-Servers werden die Merkmalsdeskriptoren eines bestimmten Clusters für die Berechnung der Entfernungen vorübergehend in den Frontend-Server transferiert. Der Client fungiert als Schnittstelle zum Benutzer: Er ist damit für die Formulierung der Suchanfrage und die Präsentation der Ergebnisse verantwortlich. Sowohl Client als auch Frontend- und Backend-Server kommunizieren über ein TCP/IP-basiertes Netzwerk mit der Bandbreite 100 MBit/s.

Die verwendete Bilddatenbank bestand aus 25000 Farbbildern im JPEG-Format, deren Auflösung etwa  $300 \times 200$  (bzw.  $200 \times 300$ ) entsprach. Die Bilder wurden verschiedenen kommerziellen Anbietern von digitalen Bildarchiven zur Verfügung gestellt. Sie umfassen Fotos und Illustrationen aus den verschiedensten thematischen Kollektionen, wie z.B. Menschen, Sport, Kunst, Reisen, Tiere, Natur und Industrie.

Als Merkmal wurde exemplarisch eine Co-Occurrence-Matrix mit 14 Farbklassen und einer Abstandklasse gewählt. Das Merkmal spannt aufgrund der Symmetrie der Co-Occurrence-Matrix einen 105-dimensionalen Vektorraum auf. Der der Untersuchung zugrundeliegende hat folglich eine ausreichende Dimensionalität, um die Eignung des Indexierungsverfahrens zu untersuchen.

### 6.2.3.1 Indexerzeugung

Die Partitionierung eines metrischen Merkmalsraums mittels Bojen ist in erster Linie durch die Größe  $N$  und die Topologie des Merkmalsdatensatzes bestimmt. Sie kann jedoch zusätzlich durch eine empirische Wahl der Parameter  $K$  und  $W_{\max}$  gesteuert werden. Die Anzahl der Cluster  $K$  beeinflusst im Wesentlichen die Präzision einer indexierten  $k$ -Nächste-Nachbar-Suche (siehe 6.2.3.2). Im Folgenden wird der Einfluss der Wahl des oberen Grenzwertes  $W_{\max}$  auf die Ausbildung von Clustern untersucht.

Abbildung 6.9 zeigt die Histogramme der Verteilungen von Clustergrößen im Index für unterschiedliche Grenzwerte  $W_{\max}$ . Es ist deutlich zu sehen, dass ein  $K$ -Median-Clustering ohne äußeren Zwang (siehe 6.9(a)) einen unvorteilhaften Index mit wenigen großen und vielen kleinen Clustern produziert. Die Wahl  $W_{\max} = 75$  hingegen resultiert in einer gaußförmigen Verteilung um den Erwartungswert  $E(|\mathcal{W}|) = 50$  (siehe 6.9(c)). Prinzipiell wäre ein Grenzwert, dessen Wert mit dem Erwartungswert identisch ist, die optimale Wahl. Jedoch verschlechtert sich das Konvergenzverhalten der Optimierung bei abnehmender Differenz von Erwartungs- und Grenzwert erheblich.

Einen ausgewogenen Kompromiss zwischen schnellem Konvergenzverhalten und vorteilhafter Größenverteilung stellt das experimentell bestimmte Optimum von

$$W_{\max} = \frac{3}{2}E(|\mathcal{W}|) \quad (6.12)$$

dar. Tabelle 6.1 zeigt die Laufzeiten des Optimierungsalgorithmus für diesen Grenzwert in Abhängigkeit von der Anzahl der Cluster. Er ist ersichtlich, dass die iterative Partitionierung des Datensatzes bereits für eine kleine Anzahl  $N$  von Merkmalsdeskriptoren eine beachtliche Verarbeitungszeit benötigt, die in etwa proportional zur Anzahl  $K$  der Cluster ist.

**Tabelle 6.1:** Einfluss der Parameter  $N$ ,  $K$  und  $W_{\max}$  auf die Indexerzeugung

$N$	$K$	$W_{\max}$	$c(\mathcal{W})$	$i$	$t$
25000	125	300	6947.59	6	2440 s
25000	250	150	6418.61	7	3439 s
25000	500	75	5883.10	10	7383 s
25000	1000	37	5337.43	11	15481 s

Die Partitionierung der Merkmalsdatensatzes durch das  $K$ -Median-Clustering benötigt zudem eine beträchtliche Menge an Speicherressourcen, da während der Optimierung permanent auf alle Merkmalsdeskriptoren zugegriffen wird. Falls aufgrund eines

zu kleinen Hauptspeichers ein Teil des Datensatzes in sekundären und tertiären Speichermedien vorgehalten werden muss, müssen zusätzliche erhebliche Verzögerungen in Kauf genommen werden. Obwohl die Indexerzeugung in der Regel offline durchgeführt wird, kann es bei entsprechend großen Datensätzen zu nicht mehr tolerierbaren Laufzeiten kommen. Abhilfe schafft in diesem Fall die Verteilung der Rechenleistung und der Speicherressourcen auf eine skalierbare Anzahl von Rechnern. Eine entsprechende Methode zur Parallelisierung des iterativen Optimierungsalgorithmus für das  $K$ -Median-Clustering kann beispielsweise in [151] gefunden werden.

### 6.2.3.2 Bojen-indexierte Nächste-Nachbar-Suche

Das für einen Benutzer unmittelbarste Leistungsmerkmal eines CBIR-Systems ist – neben der Relevanz des Ergebnisses selbst – die Dauer, die zur Beantwortung einer Suchanfrage benötigt wird. Im Folgenden wird daher die zeitliche Effizienz von indexierten im Vergleich zu nicht-indexierten Suchen experimentell untersucht und bewertet.

Die Genauigkeit und die Dauer einer bojen-indexierten Nächste-Nachbar-Suche ist von einer Vielzahl von Einflüssen abhängig:

- den Eigenschaften des Merkmalsraums  $(\Omega, \delta)$
- der Größe und Verteilung des zugrundeliegenden Merkmalsdatensatzes  $\mathcal{W}$  im Merkmalsraum
- dem Merkmalsdeskriptor  $\omega_Q$  der Suchanfrage
- der Größe  $k$  der Ergebnismenge
- der Anzahl  $K$  der Cluster des Indexes
- des Perzentils  $q$  des durchsuchten Merkmalsraums

Die Eigenschaften eines Merkmalsraums werden maßgeblich zur Entwurfszeit durch die Festlegung auf einen bestimmten Algorithmus zur Merkmalsextraktion vorgegeben. Die Auswirkung dieser Eigenschaften auf die Leistung indexierter Suchen kann daher quantitativ nicht greifbar gemacht werden, da sie zur Zeit der Suchanfrage nicht beeinflussbar ist. Gleiches gilt für den Einfluss des zugrundeliegenden Merkmalsdatensatzes: Eine Analyse muss daher immer im Kontext eines bestimmten Merkmalsraums und eines bestimmten Merkmalsdatensatzes durchgeführt werden. Im direkten

Vergleich von indexierter und nicht-indexierter Suche verschwinden diese Einflüsse jedoch weitgehend, so dass eine generelle Aussage über die Leistungsfähigkeit eines bestimmten Indexierungsverfahren unabhängig vom Merkmalsraum und -datensatz gegeben werden kann.

Den maßgeblichsten Einfluss auf das Ergebnis einer Suchfrage hat offensichtlicherweise die Suchanfrage selbst. Diese unerwünschte Abhängigkeit kann vollständig eliminiert werden, falls eine ausreichend große Stichprobe von Suchanfragen mit statistisch unkorreliertem Inhalt für die experimentelle Analyse verwendet wird. Im vorliegenden Fall wurde daher eine Stichprobe von 1000 zufällig ausgewählten Bildern verwendet, um allgemeingültige Aussagen bezüglich der Leistungsfähigkeit des Verfahrens zu erhalten.

Die Parameter  $k$ ,  $K$  und  $q$  ermöglichen eine echte Steuerung des Verhaltens einer bojen-indexierten Suche. Die Anzahl  $K$  der Cluster muss bereits bei der Indexerzeugung festgelegt werden, während das Perzentil  $q$  und die Größe  $k$  der Ergebnismenge noch zur Suchzeit verändert werden können. Eine geschickte Auswahl von  $K$  und  $q$  in Relation zur Größe  $N$  des Merkmalsdatensatzes können in einer signifikanten Beschleunigung der Verarbeitungszeit resultieren, ohne dass die Suche prinzipiell an Genauigkeit verliert. Daher werden im Folgenden die Auswirkungen der Wahl dieser Parameter auf die Suche näher untersucht.

Das Maß zur Beurteilung der Genauigkeit einer indexierten Suche ist definiert als

$$G(\omega_Q, k, K, q) = \frac{|\mathcal{Q}_{\text{NN}} \cap \mathcal{Q}_{\text{INN}}|}{k} \quad (6.13)$$

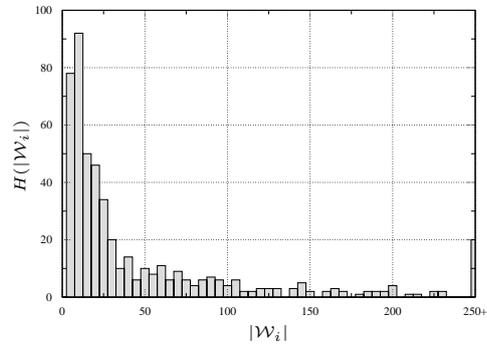
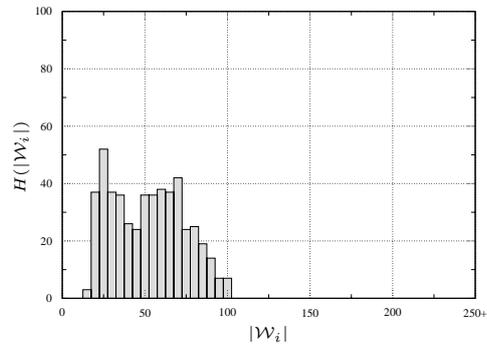
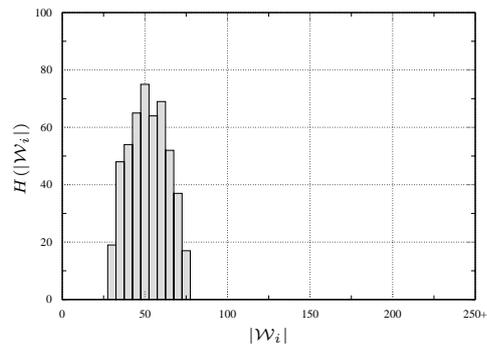
wobei  $\mathcal{Q}_{\text{NN}}$  die „vollständige“ Ergebnismenge der nicht-indexierten, linearen Suche und  $\mathcal{Q}_{\text{INN}}$  die Ergebnismenge der ungefähren indexierten Suche bezeichnet.  $G(\cdot)$  ist mit dem Maß der Vollständigkeit (siehe Definition 2.2) unter der Annahme identisch, dass die lineare Suche alle bezüglich der Suchanfrage relevanten Bilder enthält. Da diese Voraussetzung jedoch praktisch nie zutrifft, wird im Folgenden zur besseren Unterscheidung der Begriff der Genauigkeit im Sinne des Grads der Übereinstimmung von indexierter und linearer Suche verwendet.

Abbildung 6.10 stellt die Genauigkeit und Dauer von bojen-indexierten 20-Nächste-Nachbar-Suchen in Abhängigkeit des Perzentils  $q$  für drei verschiedene Werte von  $K$  gegenüber dar. Die Graphen geben jeweils die Zentralwerte der Ergebnisse über die Stichprobe wieder. Der Zentralwert kennzeichnet das 50%-Quantil der empirisch ermittelten Beobachtungswertverteilung, d.h. die Hälfte aller Beobachtungswerte liegen oberhalb des Zentralwerts, die Hälfte unterhalb. Zusätzlich zum Zentralwert ist in den Graphen das 95%-Quantil der Verteilung jedes Beobachtungswerts gekennzeichnet. 95% aller Beobachtungswerte der Stichproben lagen folglich innerhalb des mit jeweils

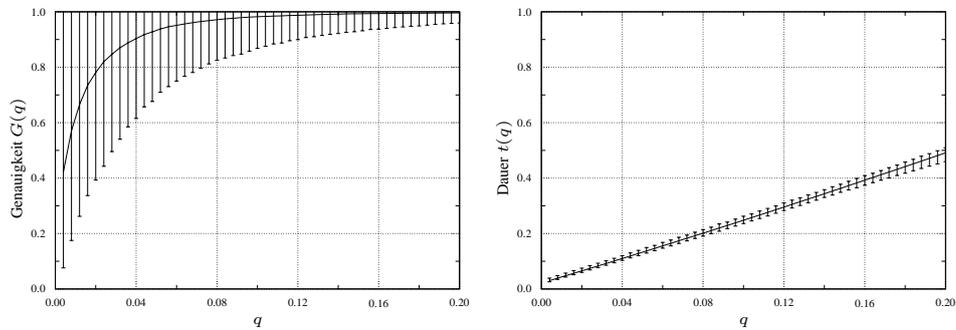
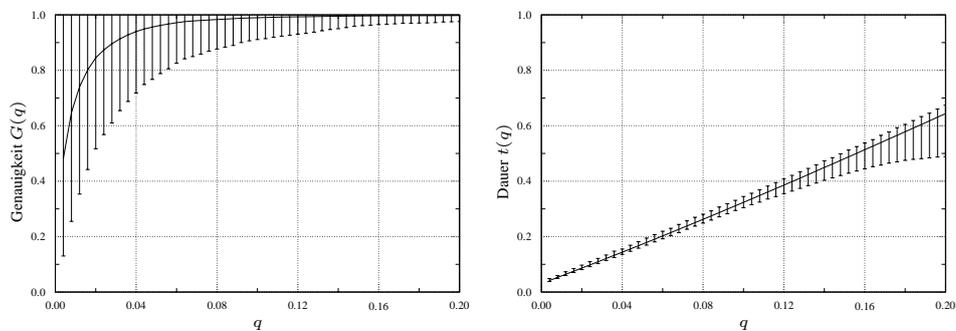
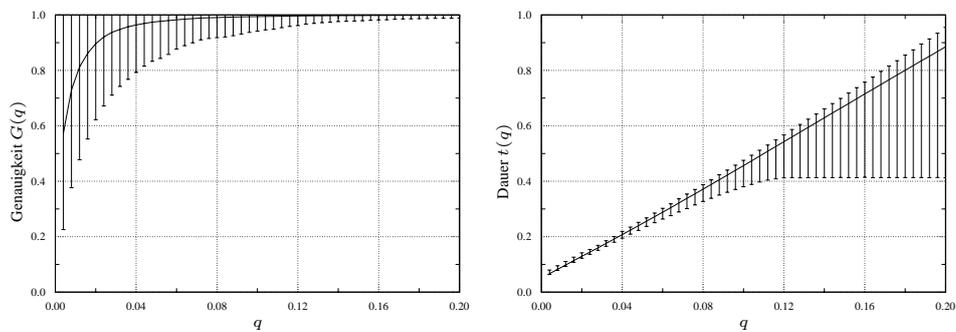
einem Balken markierten Bereichs. Diese Darstellungsform ermöglicht dem Betrachter, sich einen Eindruck über die Verteilung der Beobachtungswerte über die gesamte Stichprobe zu verschaffen.

Es ist offensichtlich, dass sich die Genauigkeit mit zunehmendem Perzentil  $q$  der durchsuchten Cluster schnell 100% nähert. Gleichzeitig nimmt die Dauer der Suche linear mit  $q$  zu. In der Mehrheit aller Fälle konnte eine 95-prozentige Genauigkeit der Suche mit relativ geringen Werten von  $q$  erreicht werden. Für  $K = 500$  und  $q = 0.08$  (durchschnittlich 2000 von 25000 Merkmalsdeskriptoren werden das Suchergebnis berücksichtigt) resultieren beispielsweise 81.3% aller Suchen der Stichprobe in Ergebnissen mit 100-prozentiger Genauigkeit. Gleichzeitig verkürzt sich die Antwortzeit der indexierten Suche im Vergleich zur linearen um den Faktor 3.78. Die relativ geringe Genauigkeit einiger weniger Suchen der Stichprobe ist durch die Tatsache zu erklären, dass die entsprechenden Merkmalsdeskriptoren der Suchanfrage jeweils in Gebieten mit einer geringen Verteilungsdichte des Datensatzes im Merkmalsraum liegen (siehe Seite 137). Die Genauigkeit nimmt folglich mit der Verteilungsdichte des Merkmalsdatensatzes in der näheren Umgebung der Suchanfrage zu. Allgemein ist daher zu erwarten, dass die Genauigkeit einer Suche zunimmt (und nicht abnimmt, wie man intuitiv vermuten könnte), je größer der zugrundeliegende Datensatz ist.

Die breite Verteilung der Beobachtungswerte der Dauer für den Index mit  $K = 1000$  Clustern deutet darauf hin, dass das Ausschlusskriterium (siehe Seite 138) für Suchen in einer entsprechend großen Umgebung ( $q > 0.1$ ) erfolgreich angewendet wird. Der Anteil der betroffenen Suche ist allerdings so gering, dass der Zentralwert von diesem Effekt nicht betroffen ist. Ein Cluster kann offensichtlich nur sehr selten aufgrund seiner Lage im Merkmalsraum a priori von einer Suche ausgeschlossen werden. Dies ist ein Beweis mehr, dass hierarchische Baumstrukturen, die auf eine hohe Wahrscheinlichkeit angewiesen sind, den Merkmalsraum bereits auf hohen Hierarchieebenen eingrenzen zu können, zur Indexierung spärlich besetzter hochdimensionaler Merkmalsräume nicht geeignet sind. Der Ansatz der Bojen-Indexierung, die Cluster in einer nicht-hierarchischen Struktur vorzuhalten, ist daher der bessere Weg.

(a)  $W_{\max} = \infty$ (b)  $W_{\max} = 100$ (c)  $W_{\max} = 75$ 

**Abbildung 6.9:** Histogramme der Verteilungen von Clustergrößen in Abhängigkeit des oberen Grenzwertes  $W_{\max}$

(a)  $K = 250$ (b)  $K = 500$ (c)  $K = 1000$ 

**Abbildung 6.10:** Gegenüberstellung von Genauigkeit und Dauer von bojen-indexierten 20-Nächste-Nachbar-Suchen in Abhängigkeit der Parameter  $K$  und  $q$



## Kapitel 7

# Zusammenfassung

Im Rahmen dieser Arbeit wurde ein Ansatz auf der Grundlage eines allgemeingültigen Modells für die merkmalsbasierte Suche nach visuellen Inhalten in digitalen Bildern erarbeitet. Dieses Modell stellt einen formalen Rahmen für die Entwicklung von Extraktionsalgorithmen zur Beschreibung der unterschiedlichsten Aspekte eines Bildes mittels Merkmalen dar. Das Modell ist aber auch zugleich Bindeglied zwischen Merkmalsextraktion und -indexierung. Damit ist sichergestellt, dass ein zukünftig entwickeltes Extraktionsverfahren, das modellkonform ist, ohne Modifikation mit einem existierenden Indexierungsverfahren zusammenarbeiten kann (das gleiche gilt auch im umgekehrten Sinn für die Entwicklung neuartiger modellkonformer Indexierungsverfahren). Aus diesem allgemeinen Modell lässt sich eine einheitliche Architektur mit standardisierten Schnittstellen für ein CBIR-System ableiten. Ein solches System kann durch die Entwicklung einzelner anwendungsspezifischer Bausteine (in der Regel Merkmalsbeschreibungen) für eine bestimmte Problemstellung zugeschnitten werden.

Grundlage für jeden Algorithmus zur Extraktion eines Merkmals ist die Interpretation der Farben einzelner Bildpunkte. Farbe ist jedoch ein Sinneseindruck, der für das menschliche Individuum einzigartig ist. Menschen desselben Kulturkreises kommunizieren diesen Sinneseindruck durch Verwendung von Namen der von ihnen bekannten Farbmuster (dieser Vorgang ist unabhängig von der individuellen Wahrnehmung des tatsächlichen Sinnesreizes). Die im Rahmen dieser Arbeit entwickelte BCR-Farbrepräsentation ermöglicht die Darstellung der Farbinformation bezüglich einer sorgfältig ausgesuchten Menge solcher Farbmuster. Dadurch ist es möglich, das Konzept von bedeutungsvollen Farbnamen auf ein technisches System zu übertragen. Das der BCR-Repräsentation zugrundeliegende mathematische Gerüst ermöglicht zudem einen einfachen und schnellen Vergleich von Farben. Eine solche Darstellung von Farbinformation kann nahezu für jede Aufgabe im Bereich der digitalen Bildverarbeitung nutzbringend eingesetzt werden. Die BCR-Farbrepräsentation ermöglicht eine

neuartige Gestaltung von intuitiven Bedienelementen zur Spezifikation einer Suchanfrage. Das Konzept, eine Farbe aufgrund ihrer Relationen zu einer Menge Farbmuster zu spezifizieren, birgt klare Vorteile gegenüber konventionellen Bedienelementen, die eine dreidimensionale Orientierung im RGB- oder HSL-Farbraum erfordern.

Auf der BCR-Farbrepräsentation aufbauend wurden exemplarisch einige universelle Extraktionsalgorithmen entwickelt, die nicht-anwendungsspezifische Aspekte eines digitalen Bildes mittels kompakter Merkmalsdeskriptoren beschreiben. Diese können einzeln oder in Kombination verwendet werden, um große Bilddatensätze für die Informationsgewinnung zu erschließen. Die Liste der vorgestellten Merkmale erhebt nicht den Anspruch auf Vollständigkeit, sondern dient eher als exemplarische Realisierung der in dieser Arbeit vorgestellten Konzepte.

Große Bildarchive erfordern große Merkmalsdatensätze. Große Merkmalsdatensätze bedingen wiederum eine entsprechende lange Verarbeitungszeit bei der Suche nach Information. In der vorliegenden Arbeit wird ein Indexierungsverfahren vorgestellt, das die kausale Verbindung zwischen einer großen Menge von Bildern und einer entsprechend langen Suchzeit aufbricht. Das Verfahren basiert auf dem metrischen Modell und ist daher universell für alle Merkmalsbeschreibungen einsetzbar. Das dabei verfolgte Prinzip ist die Lokalisierung der Suche auf die unmittelbare Umgebung der Suchanfrage im Merkmalsraum. Durch die Einsparung von Ressourcen für den Vergleich von per se unähnlichen Bildern kann eine signifikante Beschleunigung der Suche erreicht werden. Da das Verfahren immanent mit einer gewissen Ungenauigkeit behaftet ist, wurde experimentell nachgewiesen, dass das Verfahren bei einer sinnvollen Parameterwahl im praktischen Einsatz eine ausreichend hohe Genauigkeit besitzt. Die wohl innovativste Eigenschaft des Verfahrens ist, dass der Benutzer die Antwortzeit des CBIR-Systems für eine Suche direkt kontrollieren kann. Tatsächlich kann eine Suchanfrage so konfiguriert werden, dass sie innerhalb eines vorgegebenen Zeitintervalls garantiert ein – möglicherweise ungenaues – Suchergebnis liefert. Dies ist eine wichtige Eigenschaft für CBIR-Systeme, da in der Regel mehrere Merkmale gleichzeitig bei einer Suche zur Anwendung kommen. Auf diese Weise kann vermieden werden, dass das in der Berechnung aufwendigste Merkmal zum Flaschenhals des Systems wird.

## 7.1 Ausblick

Die BCR-Farbrepräsentation kann mittels einer geeigneten Basisfarbmenge für einen bestimmten Anwendungsfall konfiguriert werden. Dabei kann es durchaus vorkommen, dass eine Basisfarbmenge mit einer größeren Menge von Basisfarben ausgewählt wird. Durch die entsprechend hohe Dichte von Basisfarben im Bezugsfarbraum ist

der Begriff der Ähnlichkeit zweier Farben automatisch auf kleinere Umgebungen begrenzt. Daher kann es hilfreich sein, die Ähnlichkeitsmetrik dahingehend zu erweitern, dass nicht nur die Ähnlichkeiten einzelner Basisfarben berücksichtigt werden, sondern auch die Ähnlichkeiten zwischen mehreren verschiedenen Basisfarben. Als Resultat würde das Ähnlichkeitsmaß nicht mehr zwingend an die Auswahl der Basisfarben gebunden sein.

Das Hauptproblem der Bojenindexierung ist die Verwaltung von Speicherressourcen während der Indexerzeugung. Falls aufgrund eines zu kleinen Hauptspeichers ein Teil des Datensatzes in sekundären und tertiären Speichermedien vorgehalten werden muss, kann es bei entsprechend großen Datensätzen zu nicht mehr tolerierbaren Berechnungszeiten kommen. Daher ist die Parallelisierung der Indexerzeugung auf eine beliebig skalierbare Anzahl von Rechnern ein zentraler Punkt der Weiterentwicklung dieses Verfahrens.





Tabelle A.2: Delaunay-Zerlegung der DIN5023-Basisfarbmenge

	N	YL	OR	CR	MR	VT	UB	CB	BG	YG	OC	BS	BK	WT	Vol.
1	•							•		•				•	2.09%
2		•								•	•			•	2.03%
3	•				•						•			•	1.63%
4	•									•	•			•	1.55%
5	•						•	•						•	1.39%
6	•						•		•				•		1.27%
7		•								•		•	•		1.25%
8			•		•						•			•	1.21%
9	•								•	•			•		1.00%
10	•					•	•							•	0.92%
11	•									•		•	•		0.88%
12		•	•								•			•	0.87%
13	•				•	•								•	0.83%
14		•								•	•	•			0.67%
15	•					•	•						•		0.63%
16	•					•						•	•		0.55%
17	•							•	•	•					0.54%
18	•									•	•	•			0.50%
19	•				•						•	•			0.49%
20					•	•						•	•		0.46%
21		•							•	•			•		0.41%
22	•						•	•	•						0.34%
23					•	•	•						•		0.29%
24		•	•								•	•			0.28%
25				•	•						•	•			0.26%
26			•	•	•						•				0.25%
27				•	•							•	•		0.23%
28		•	•									•	•		0.21%
29	•				•	•						•			0.20%
30			•	•							•	•			0.15%
31			•	•								•	•		0.13%
32							•	•	•				•		0.10%
33					•	•	•							•	0.05%
	16	7	7	5	11	8	8	5	6	10	12	14	13	10	23.66%

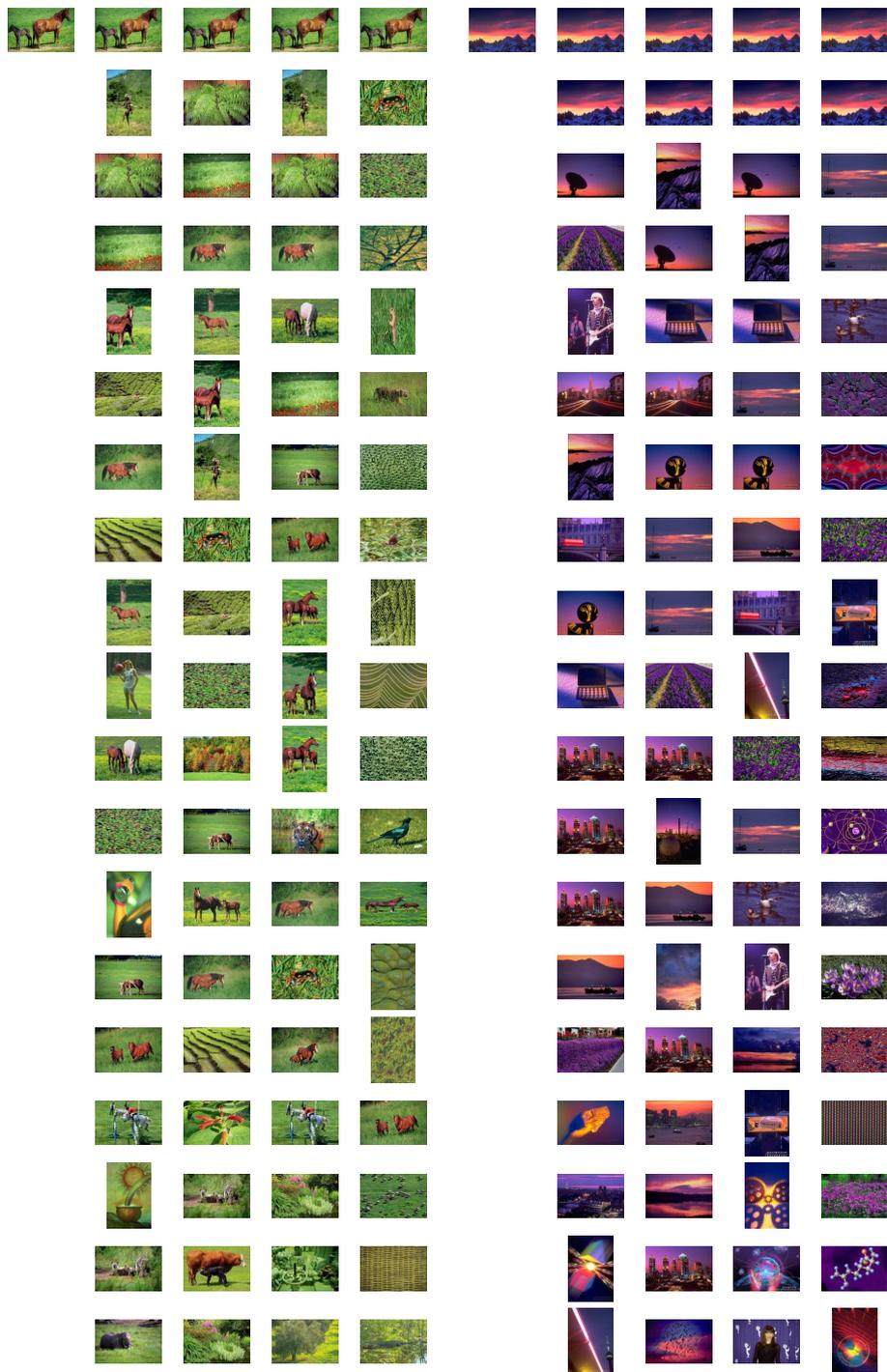
## Anhang B

# Ergebnisse von Suchanfragen

In Abbildung B.1 sind die Ergebnisse mehrerer Suchanfragen dargestellt. Die Suchanfrage selbst ist jeweils das Bild links oben. Eine Spalte von Bildern daneben gibt die Reihenfolge (von oben nach unten) der jeweils 18 ähnlichsten Bilder der Ergebnismenge auf der Basis eines unterschiedlichen Merkmals wieder:

1. Spalte  $\Rightarrow$  Merkmal beschrieben in Abschnitt 5.2.1.1
2. Spalte  $\Rightarrow$  Merkmal beschrieben in Abschnitt 5.2.1.2
3. Spalte  $\Rightarrow$  Merkmal beschrieben in Abschnitt 5.2.1.3
4. Spalte  $\Rightarrow$  Merkmal beschrieben in Abschnitt 5.2.2

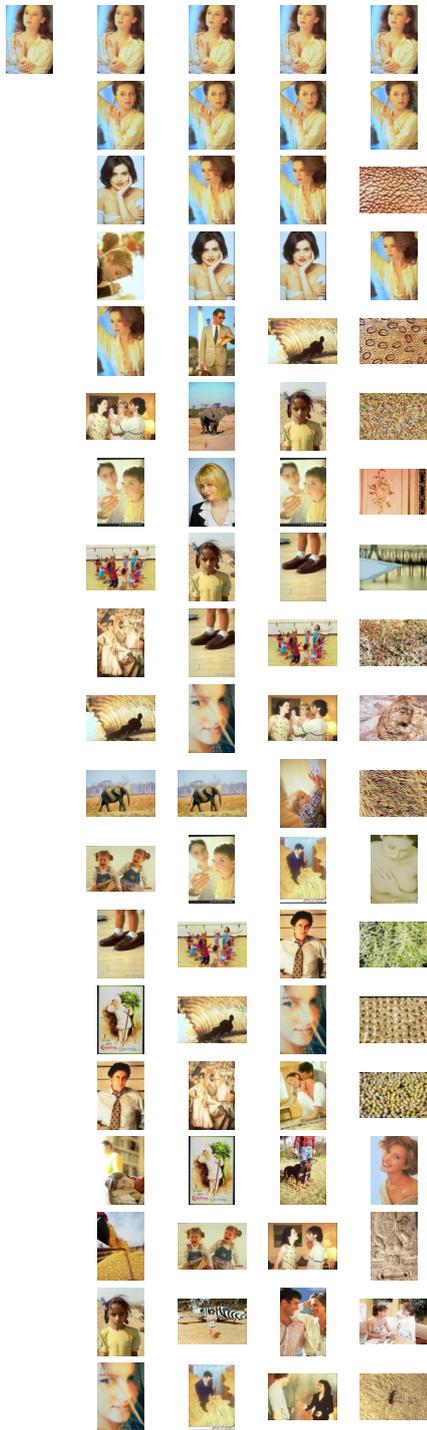
Mittels dieser Darstellung soll dem individuellen Betrachter ermöglicht werden, die unterschiedlichen Eigenschaften verschiedener Merkmale, deren natursprachliche Beschreibung schwierig ist, visuell zu erfassen.



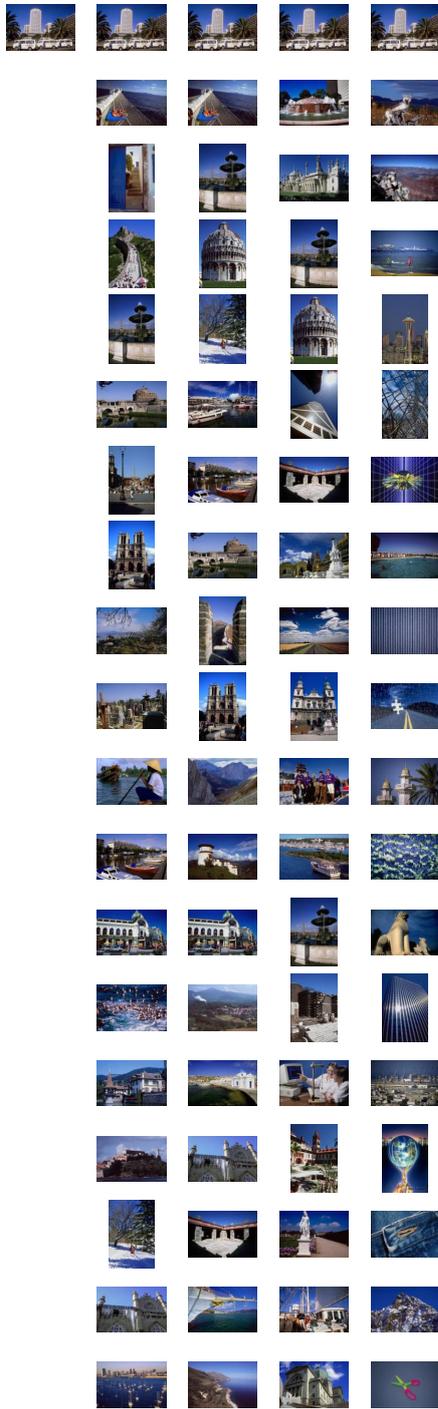
(a) Bild ID 290583110

(b) Bild ID 1551187610

**Abbildung B.1:** Resultate verschiedener Suchanfragen für unterschiedliche Merkmale



(c) Bild ID 23130248



(d) Bild ID 1551368958



# Literaturverzeichnis

- [1] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu. An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions. *Journal of the ACM*, 45(6):891–923, 1998.
- [2] K. Barnard, P. Duygulu, N. De Freitas, D. Forsyth, D. Blei, and M.I. Jordan. Matching Words and Pictures. *Journal of Machine Learning Research, Special Issue on Machine Learning Methods for Text and Images*, 3:1107–1135, 2003.
- [3] K. Barnard and D.A. Forsyth. Learning the Semantics of Words and Pictures. In *Proceedings of the IEEE Int'l Conference on Computer Vision*, pages 408–415, Vancouver, Canada, July 2001.
- [4] R. Bayer and E.M. McCreight. Organization of Large Ordered Indexes. *Acta Informatica*, 1(3):173–189, 1972.
- [5] N. Beckmann, H.P. Kriegel, R. Schneider, and B. Seeger. The  $R^*$ -Tree: An Efficient and Robust Access Method for Points and Rectangles. In *Proceedings of the ACM SIGMOD Int'l Conference on Management of Data*, pages 322–331, Atlantic City, NJ, USA, May 1990.
- [6] J.L. Bentley. Multidimensional Binary Search Trees Used for Associative Searching. *Communications of ACM*, 18(9):509–517, 1975.
- [7] S. Berchtold, D. Keim, and H.P. Kriegel. The  $X$ -Tree: An Index Structure for High-Dimensional Data. In *Proceedings of the 22nd Int'l Conference on Very Large Data Bases*, pages 28–39, Mumbai, India, September 1996.
- [8] J. Berens and G.D. Finlayson. An Efficient Coding of Three-Dimensional Color Distributions for Image Retrieval. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 245–252, London, UK, July 2002.
- [9] B. Berlin and P. Kay. *Basic Colour Terms: Their Universality and Evolution*. University of California Press, 1969.

- [10] S.A. Berrani, L. Amsaleg, and P. Gros. Approximate Searches:  $k$ -Neighbors + Precision. In *Proceedings of the 12th Int'l Conference on Information and Knowledge Management*, pages 24–31, New Orleans, LA, USA, November 2003.
- [11] S.T. Bow. *Pattern Recognition and Image Preprocessing*. Marcel Dekker, 2nd edition, 2002.
- [12] T. Bozkaya and M. Ozsoyoglu. Distance-Based Indexing for High-Dimensional Metric Spaces. In *Proceedings of the ACM SIGMOD Int'l Conference on Management of Data, Symposium on Principles of Database Systems*, pages 357–368, Tuscon, AZ, USA, May 1997.
- [13] S. Brin. Near Neighbor Search in Large Metric Spaces. In *Proceedings of the 21st Int'l Conference on Very Large Data Bases*, pages 574–584, Zürich, Switzerland, September 1995.
- [14] D. Cantone, A. Ferro, A. Pulvirenti, D.R. Recupero, and D. Shasha. Antipole Tree Indexing to Support Range Search and  $k$ -Nearest Neighbor Search in Metric Spaces. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):535–550, 2005.
- [15] L.D. Castelli. V. Bergman, editor. *Image Databases: Search and Retrieval of Digital Imagery*. John Wiley & Sons, 2002.
- [16] R. Cavet, S. Volmer, E. Leopold, J. Kindermann, and G. Paaß. Revealing the Connoted Visual Code: A New Approach to Video Classification. *Computers & Graphics*, 28(3):361–369, 2004.
- [17] G.H. Cha, X. Zhu, D. Petkovic, and C.W. Chung. An Efficient Indexing Method for Nearest Neighbor Searches in High-Dimensional Image Databases. *IEEE Transactions on Multimedia*, 4(1):76–87, 2002.
- [18] S.K. Chang and A. Hsu. Image Information Systems: Where Do We Go from Here? *IEEE Transactions on Knowledge and Data Engineering*, 4(5):431–442, 1992.
- [19] S.K. Chang, Q.Y. Shi, and C.W. Yan. Iconic Indexing by 2D Strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3):413–427, 1987.
- [20] E. Chavez, G. Navarro, R. Baeza-Yates, and J. Marroquin. Searching in Metric Spaces. *ACM Computing Surveys*, 33(3):273–321, 2001.
- [21] Y. Chen and J.Z. Wang. Image Categorization by Learning and Reasoning with Regions. *Journal of Machine Learning Research*, 5:913–939, 2004.

- [22] T. Chiueh. Content-Based Image Indexing. In *Proceedings of the 20th Int'l Conference on Very Large Data Bases*, pages 582–593, Santiago, Chile, September 1994.
- [23] T.S. Chua, K.L. Tan, and B.C. Ooi. Fast Signature-Based Color-Spatial Image Retrieval. In *Proceedings of the IEEE Int'l Conference on Multimedia Computing and Systems*, pages 362–369, Ottawa, Canada, June 1997.
- [24] P. Ciaccia and M. Patella. Using the Distance for Approximate Similarity Queries in High-dimensional Metric Space. In *Proceedings of the 10th Int'l Workshop on Database & Expert Systems Applications*, pages 200–205, Florence, Italy, September 1999.
- [25] P. Ciaccia, M. Patella, and P. Zezula. *M*-Tree: An Efficient Access Method for Similarity Search in Metric Spaces. In *Proceedings of the 23rd Int'l Conference on Very Large Data Bases*, pages 426–435, Athens, Greece, September 1997.
- [26] C. Colombo, A. Del Bimbo, and P. Pala. Semantics in Visual Information Retrieval. *IEEE Multimedia*, 6(3):38–53, 1999.
- [27] I.J. Cox, M.L. Miller, T.P. Minka, and P.N. Yianilos. An Optimized Interaction Strategy for Bayesian Relevance Feedback. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 553–558, Santa Barbara, CA, USA, June 1998.
- [28] A. Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann, 1999.
- [29] A. Del Bimbo and P. Pala. Visual Image Retrieval by Elastic Matching of User Sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):121–132, 1997.
- [30] Deckfarben-Malkasten. DIN Standard 5023, 1989.
- [31] J.P. Eakins, J.M. Boardman, and M.E. Graham. Similarity Retrieval of Trademark Images. *IEEE Multimedia*, 5(2):53–63, 1998.
- [32] J.P. Eakins, K.J. Riley, and J.D. Edwards. Shape Feature Matching for Trademark Image Retrieval. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 28–38, Urbana-Champaign, IL, USA, July 2003.
- [33] J.P. Eakins, K. Shields, and J.M. Boardman. ARTISAN – a Shape Retrieval System Based on Boundary Family Indexing. In *Storage and Retrieval for Image and Video Databases IV*, volume 2670, pages 17–28, La Jolla, CA, USA, February 1996. SPIE.

- [34] P.G.B. Enser. Pictorial Information Retrieval. *Documentation*, 51(2):126–170, 1995.
- [35] P.G.B. Enser and C.J. Sandom. Towards a Comprehensive Survey of the Semantic Gap in Visual Image Retrieval. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 291–299, Urbana-Champaign, IL, USA, July 2003.
- [36] P.G.B. Enser, C.J. Sandom, and P.H. Lewis. Surveying the Reality of Semantic Image Retrieval. In *Proceedings of the 8th Int'l Conference on Visual Retrieval Systems*, Amsterdam, The Netherlands, July 2005.
- [37] R. Fagin, R. Kumar, and D. Sivakumar. Efficient Similarity Search and Classification via Rank Aggregation. In *Proceedings of the ACM SIGMOD Int'l Conference on Management of Data*, pages 301–312, San Diego, CA, USA, June 2003.
- [38] R. Fagin, J. Nievergelt, N. Pippenger, and H.R. Strong. Extensible Hashing – A Fast Access Method for Dynamic Files. *ACM Transactions on Database Systems*, 4(3):315–344, 1979.
- [39] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and Effective Querying by Image Content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, 1994.
- [40] C. Faloutsos and K.I. Lin. FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. In *Proceedings of the ACM SIGMOD Int'l Conference on Management of Data*, pages 163–174, San Jose, CA, USA, May 1995.
- [41] J. Fan, Y. Gao, H. Luo, and G. Xu. Automatic Image Annotation by Using Concept-Sensitive Salient Objects for Image Content Representation. In *Proceedings of the 27th Annual Int'l Conference on Research and Development in Information Retrieval*, pages 361–368, Sheffield, UK, July 2004.
- [42] V. Ferrari, T. Tuytelaars, and L.J. van Gool. Simultaneous Object Recognition and Segmentation by Image Exploration. In *Proceedings of the 8th European Conference on Computer Vision*, pages 40–54, Prague, Czech Republic, May 2004.
- [43] R.A. Finkel and J.L. Bentley. Quad Trees: A Data Structure for Retrieval on Composite Keys. *Acta Informatica*, 4(1):1–9, 1974.
- [44] G.D. Finlayson. Color in Perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):1034–1038, 1996.

- [45] M. Flickner, H.S. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by Image Content: The QBIC System. *IEEE Computer*, 28(9):23–32, 1995.
- [46] D.A. Forsyth. A Novel Algorithm for Color Constancy. *Int'l Journal of Computer Vision*, 5(1):5–36, 1990.
- [47] C.S. Fuh, S.W. Cho, and K. Essig. Hierarchical Color Image Region Segmentation for Content-Based Image Retrieval System. *IEEE Transactions on Image Processing*, 9(1):156–163, 2000.
- [48] B.V. Funt and Finlayson G.D. Color Constant Color Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
- [49] V. Gaede and O. Günther. Multidimensional Access Methods. *ACM Computing Surveys*, 30(2):170–231, 1998.
- [50] T. Gevers. Color-Based Retrieval. In *Principles of Visual Information Retrieval*, chapter 2, pages 11–49. Springer-Verlag, 2001.
- [51] T. Gevers and A.W.M. Smeulders. Color-Based Object Recognition. *Pattern Recognition*, 32(3):453–464, 1999.
- [52] T. Gevers and A.W.M. Smeulders. Content-Based Image Retrieval by Viewpoint-Invariant Image Indexing. *Image and Vision Computing*, 17(7):475–488, 1999.
- [53] T. Gevers and A.W.M. Smeulders. PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval. *IEEE Transactions on Image Processing*, 9(1):102–119, 2000.
- [54] T. Gevers and H.M.G. Stokman. Robust Photometric Invariant Segmentation of Multispectral Images. *Int'l Journal of Computer Vision*, 53(2):135–151, 2003.
- [55] E.H. Gombrich. *The Image & the Eye: Further Studies in the Psychology of Pictorial Representation*. Phaidon Press, reprint, 1994.
- [56] A.A. Goodrum. Image Information Retrieval: An Overview of Current Research. *Informing Science – Special Issue on Information Science Research*, 3(2):63–66, 2000.
- [57] C.C. Gotlieb and H.E. Kreyzig. Texture Descriptors Based on Co-Occurrence Matrices. *Computer Vision, Graphics, and Image Processing*, 51(1):70–86, 1990.
- [58] P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000.

- [59] A. Grigorova and F.G.B. de Natale. Semi-Automatic Feature-Adaptive Relevance Feedback for Content-Based Image Retrieval. In *Proceedings of the 8th Int'l Conference on Visual Retrieval Systems*, Amsterdam, The Netherlands, July 2005.
- [60] M.H. Gross, R. Koch, L. Lippert, and A. Dreger. Multiscale Image Texture Analysis in Wavelet Space. In *Proceedings of the IEEE Int'l Conference on Image Processing*, pages 412–416, Austin, TX, USA, November 1994.
- [61] V.N. Gudivada and V.V. Raghavan. Content-Based Image Retrieval Systems. *IEEE Computer*, 28(9):18–22, 1995.
- [62] A. Guttman. *R*-Trees: A Dynamic Index Structure for Spatial Searching. In *Proceedings of the ACM SIGMOD Int'l Conference on Management of Data*, pages 47–57, Boston, MA, USA, June 1984.
- [63] J.F. Haddon and J.F. Boyce. Co-Occurrence Matrices for Image Analysis. *Electronics & Communication Engineering Journal*, 5(2):71–83, 1993.
- [64] A. Hanjalic, J. Nesvadba, and J. Benois-Pineau. Moving Away from Narrow-Scope Solutions in Multimedia Content Analysis. In *Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, London, UK, November 2005.
- [65] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, 1973.
- [66] J.S. Hare and P.H. Lewis. Salient Regions for Query by Image Content. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 317–325, Dublin, Ireland, July 2004.
- [67] E. Hecht. *Optics*. Addison-Wesley, 4th edition, 2001.
- [68] K. Hirata and T. Kato. Query by Visual Example – Content Based Image Retrieval. In *Proceedings of the 3rd Int'l Conference on Extending Database Technology*, pages 56–71, Vienna, Austria, March 1992.
- [69] P. Howarth and S.M. Rüger. Evaluation of Texture Features for Content-Based Image Retrieval. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 326–334, Dublin, Ireland, July 2004.
- [70] P. Howarth and S.M. Rüger. Trading Precision for Speed: Localised Similarity Functions. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 415–424, Singapore, July 2005.

- [71] C.C. Hsu, W.W. Chu, and R.K. Taira. A Knowledge-Based Approach for Retrieving Images By Content. *IEEE Transactions on Knowledge and Data Engineering*, 8(4):522–532, 1996.
- [72] M.K. Hu. Visual Pattern Recognition by Moment Invariants. *IRE Transactions on Information Theory*, 8(2):179–187, 1962.
- [73] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih. Image Indexing Using Color Correlograms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 762–768, San Juan, Puerto Rico, June 1997.
- [74] P. Indyk and R. Motwani. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 604–613, Dallas, TX, USA, May 1998.
- [75] M. Ioka. A Method of Defining the Similarity of Images on the Basis of Color Information. Technical Report RT-0030, IBM Research, Tokyo Research Laboratory, November 1989.
- [76] Q. Iqbal and J.K. Aggarwal. Feature Integration, Multi-Image Queries and Relevance Feedback in Image Retrieval. In *Proceedings of the 9th Int'l Conference on Distributed Multimedia Systems*, pages 467–474, Miami, FL, USA, September 2003.
- [77] Colorimetric Observers. ISO/CIE Standard 10527, 1986.
- [78] C.E. Jacobs, A. Finkelstein, and D.H. Salesin. Fast Multiresolution Image Querying. In *Proceedings of the ACM SIGGRAPH*, pages 277–286, Los Angeles, CA, USA, August 1995.
- [79] H.V. Jagadish. Spatial Search with Polyhedra. In *Proceedings of the 6th IEEE Int'l Conference on Data Engineering*, pages 311–319, Los Angeles, CA, USA, February 1990.
- [80] A.K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 1988.
- [81] L.M. Kaplan, R. Murenzi, and K.R. Namuduri. Fast Texture Database Retrieval Using Extended Fractal Features. In *Storage and Retrieval for Image and Video Databases VI*, volume 3312, pages 162–175, San Jose, CA, USA, January 1998. SPIE.
- [82] D. Kapur, Y.N. Lakshman, and T. Saxena. Computing Invariants Using Elimination Methods. In *Proceedings of the IEEE Int'l Symposium on Computer Vision*, pages 97–102, Coral Gables, FL, USA, November 1995.

- [83] N. Katayama and S. Satoh. The *SR*-Tree: An Index Structure for High-Dimensional Nearest Neighbor Queries. In *Proceedings of the ACM SIGMOD Int'l Conference on Management of Data*, pages 369–380, Tucson, AZ, USA, May 1997.
- [84] M. Kobayashi and K. Takeda. Information Retrieval on the Web: Selected Topics. Technical Report RT-0264, Tokyo Research Laboratory, IBM Japan Ltd., December 1999.
- [85] V. Kovalev and S. Volmer. Color Co-occurrence Descriptors for Querying-by-Example. In *Proceedings of the 5th Int'l Conference on Multimedia Modeling*, pages 32–38, Lausanne, Switzerland, October 1998.
- [86] R. Kuhlen. Zum Stand pragmatischer Forschung in der Informationswissenschaft. In *Proceedings des 1. Internationalen Symposiums für Informationswissenschaft*, pages 13–18, Konstanz, Germany, 1990.
- [87] R. Kurniawati, J.S. Jin, and J.A. Shepherd. The *SS+*-Tree: An Improved Index Structure for Similarity Searches in a High-Dimensional Feature Space. In *Storage and Retrieval for Image and Video Databases V*, volume 3022, pages 110–120, San Jose, CA, USA, February 1997. SPIE.
- [88] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient Search for Approximate Nearest Neighbor in High Dimensional Spaces. In *Proceedings of the 13th Annual ACM Symposium on Theory of Computing*, pages 614–623, Dallas, TX, USA, May 1998.
- [89] A. Laine and J. Fan. Texture Classification by Wavelet Packet Signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1186–1191, 1993.
- [90] E. Leopold, J. Kindermann, G. Paaß, S. Volmer, R. Cavet, M. Larson, S. Eickeler, and T. Kastner. Integrated Classification of Audio, Video and Speech Using Partitions of Low-Level Features. In *Proceedings of the ECML Workshop on Multimedia Discovery and Mining*, Dubrovnic, Croatia, September 2003.
- [91] V.I. Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, 10(8):707–710, 1966.
- [92] M.S. Lew. Next-Generation Web Searches for Visual Content. *IEEE Computer*, 33(11):46–53, 2000.
- [93] B. Li and S.D. Ma. On the Relation Between Region and Contour Representation. In *Proceedings of the 12th Int'l Conference on Pattern Recognition*, pages 352–355, Jerusalem, Israel, October 1994.

- [94] W. Li and T. Bretschneider. An Effective Multidimensional Index Strategy for Cluster Architectures. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 164–173, Singapore, July 2005.
- [95] F. Liu and Picard R.W. Periodicity, Directionality, and Randomness: Wold Features for Image Modeling and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):517–549, 1996.
- [96] G. Loy and J.O. Eklundh. A Review of Benchmarking Content-Based Image Retrieval Systems. In *Proceedings of the MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation*, Vienna, Austria, September 2005.
- [97] H. Lu, B.C. Ooi, and K.L. Tan. Efficient Image Retrieval by Color Contents. In *Proceedings of the 1st Int'l Conference on Applications of Databases*, pages 95–108, Vadstena, Sweden, June 1994.
- [98] S. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.
- [99] H.A. Mallot. *Computational Vision: Information Processing in Perception and Visual Behavior*. MIT Press, 2nd edition, 2000.
- [100] B.S. Manjunath and W.Y. Ma. Texture Features for Browsing and Retrieval of Image Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–841, 1996.
- [101] B.S. Manjunath, J.R. Ohm, V. Vasudevan, and A. Yamada. Color and Texture Description. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.
- [102] S. Marchand-Maillet. Construction of a Formal Multimedia Benchmark. In *Proceedings of the European Signal Processing Conference*, Toulouse, France, September 2002.
- [103] D.R. Martin, C. Fowlkes, and J. Malik. Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, 2004.
- [104] B.M. Mehre, M.S. Kankanhalli, and W.H. Lee. Shape Measures for Content Based Image Retrieval: A Comparison. *Information Processing Management*, 33(3):319–337, 1997.

- [105] M. Miyahara and Y. Yoshida. Mathematical Transform of (r, g, b) Color Data to Munsell (h, v, c) Color Data. In *Visual Communications and Image Processing*, volume 1001, pages 650–657, Cambridge, MA, USA, November 1988. SPIE.
- [106] H. Müller, W. Müller, D. Squire, S. Marchand-Maillet, and T. Pun. Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals. *Pattern Recognition Letters*, 22(5):593–601, 2001.
- [107] N. Moënné-Loccoz. High-Dimensional Access Methods for Efficient Similarity Queries. Technical Report 05.05, University of Geneva, Computing Science Center, Computer Vision Group, May 2005.
- [108] F. Monay and D. Gatica-Perez. On Image Auto-Annotation with Latent Space Models. In *Proceedings of the 11th ACM Int'l Conference on Multimedia*, pages 275–278, Berkeley, CA, USA, November 2003.
- [109] A.H. Munsell. *The Atlas of the Munsell Color System*. Munsell Color Company, 1915.
- [110] C. Nastar, M. Mitschke, and C. Meihac. Efficient Query Refinement for Image Retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 547–552, Santa Barbara, CA, USA, June 1998.
- [111] The ISCC-NBS Method of Designating Colors and a Dictionary of Color Names. NBS Standard 553, 1955.
- [112] R. Ng and A. Sedighian. Evaluating Multi-Dimensional Indexing Structures for Images Transformed by Principal Component Analysis. In *Storage and Retrieval for Image and Video Databases IV*, volume 2670, pages 50–61, La Jolla, CA, USA, January 1996. SPIE.
- [113] R.T. Ng and J. Han. CLARANS: A Method for Clustering Objects for Spatial Data Mining. *IEEE Transactions on Knowledge and Data Engineering*, 14(5):1003–1016, 2002.
- [114] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanke, C. Faloutsos, and G. Taubin. The QBIC Project: Querying Images by Content Using Color, Texture, and Shape. In *Storage and Retrieval for Image and Video Databases*, volume 1908, pages 173–187, San Jose, CA, USA, January 1993. SPIE.
- [115] J. Nievergelt.  $7 \pm 2$  Criteria for Assessing and Comparing Spatial Data Structures. In *Proceedings of the 1st Symposium on Design and Implementation of Large Spatial Databases*, pages 3–27, Santa Barbara, CA, USA, July 1989.

- [116] J. Nievergelt and H. Hinterberger. The Grid File: An Adaptable Symmetric Multikey File Structure. *ACM Transactions on Database Systems*, 9(1):38–71, 1984.
- [117] A. Okabe, B. Boots, K. Sugihara, and Chiu S.N. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons, 2nd edition, 2000.
- [118] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T.S. Huang. Supporting Similarity Queries in MARS. In *Proceedings of the ACM Conference on Multimedia*, pages 403–413, Seattle, WA, USA, June 1997.
- [119] G. Pass, R. Zabih, and J. Miller. Comparing Images Using Color Coherence Vectors. In *Proceedings of the 4th ACM Conference on Multimedia*, pages 65–73, Boston, MA, USA, November 1996.
- [120] E.J. Pauwels and G. Frederix. Finding Salient Regions in Images: Nonparametric Clustering for Image Segmentation and Grouping. *Computer Vision and Image Understanding*, 75(1-2):73–85, 2000.
- [121] E. Persoon and K.S. Fu. Shape Discrimination Using Fourier Descriptors. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(3):170–179, 1977.
- [122] V. Pestov. On the Geometry of Similarity Search: Dimensionality Curse and Concentration of Measure. *Information Processing Letters*, 73(1–2):47–51, 2000.
- [123] V.T. Rajan. Optimality of the Delaunay Triangulation in  $R^d$ . *Discrete & Computational Geometry*, 12(2):189–202, 1994.
- [124] T. Randen and J.H. Husoy. Filtering for Texture Classification: A Comparative Study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, 1999.
- [125] R. Rautakorpi and J. Iivarinen. A Novel Shape Feature for Image Classification and Retrieval. In *Proceedings of Int'l Conference on Image Analysis and Recognition*, pages 753–760, Porto, Portugal, September 2004.
- [126] I. Rock. *Perception*. Scientific American Library, 1985.
- [127] E. Rosch. Natural Categories. *Cognitive Psychology*, 4(3):328–350, 1974.
- [128] Y. Rui, T.S. Huang, M. Ortega, and S. Mehrotra. Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

- [129] Y. Rui, A.C. She, and T.S. Huang. Modified Fourier Descriptors for Shape Representation – A Practical Approach. In *Proceedings of the 1st Int'l Workshop on Image Databases and Multi Media Search*, Amsterdam, The Netherlands, August 1996.
- [130] S. Santini and R.C. Jain. Gabor Space and the Development of Preattentive Similarity. In *Proceedings of the 13th Int'l Conference on Pattern Recognition*, Vienna, Austria, August 1996.
- [131] S. Santini and R.C. Jain. The Graphical Specification of Similarity Queries. *Journal of Visual Languages and Computing*, 7(4):403–421, 1996.
- [132] S. Santini and R.C. Jain. Similarity Measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):871–883, 1999.
- [133] S. Sclaroff and L. Liu. Deformable Shape Detection and Description via Model-Based Region Grouping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(5):475–489, 2001.
- [134] S. Sclaroff and A. Pentland. Modal Matching for Correspondence and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):545–561, 1995.
- [135] N. Sebe and M.S. Lew. Texture Features for Content-Based Image Retrieval. In *Principles of Visual Information Retrieval*, chapter 3, pages 51–86. Springer-Verlag, 2001.
- [136] N. Sebe and M.S. Lew. Robust Shape Matching. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 17–28, London, UK, July 2002.
- [137] T. Sellis, N. Roussopoulos, and C. Faloutsos. The  $R^+$ -Tree: A Dynamic Index for Multidimensional Objects. In *Proceedings of the 13th Int'l Conference on Very Large Data Bases*, pages 507–518, Brighton, UK, September 1987.
- [138] G. Sheikholeslami, W. Chang, and A. Zhang. Semantic Clustering and Querying on Heterogeneous Features for Visual Data. In *Proceedings of the 6th ACM Int'l Conference on Multimedia*, pages 3–12, Bristol, UK, September 1998.
- [139] A.P. Sheth and W. Klas, editors. *Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media*. McGraw-Hill, 1998.
- [140] S. Siggelkow and H. Burkhardt. Image Retrieval Based on Local Invariant Features. In *Proceedings of the IASTED 3rd Int'l Conference on Signal and Image Processing*, pages 369–373, Las Vegas, NV, USA, October 1998.

- [141] S. Siggelkow and H. Burkhardt. Improvement of Histogram-Based Image Retrieval and Classification. In *Proceedings of the Int'l Conference on Pattern Recognition*, pages 367–370, Quebec, Canada, September 2002.
- [142] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R.C. Jain. Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.
- [143] J.R. Smith and S.F. Chang. Transform Features for Texture Classification and Discrimination in Large Image Databases. In *Proceedings of the IEEE Int'l Conference on Image Processing*, pages 407–411, Austin, TX, USA, November 1994.
- [144] J.R. Smith and S.F. Chang. Single Color Extraction and Image Query. In *Proceedings of the IEEE Int'l Conference on Image Processing*, pages 528–531, Washington, DC, USA, October 1995.
- [145] J.R. Smith and S.F. Chang. Local Color and Texture Extraction and Spatial Query. In *Proceedings of the IEEE Int'l Conference on Image Processing*, pages 1011–1014, Lausanne, Switzerland, September 1996.
- [146] J.R. Smith and S.F. Chang. Integrated Spatial and Feature Image Query. *Multimedia Systems*, 7(2):129–140, 1999.
- [147] R. Stenvert. Bridging the Gap between Pixel and Picture? *Computers and the History of Art*, 2(2):19–34, 1992.
- [148] W.S. Stiles and J.M. Burch. Npl Colour-Matching Investigation: Final Report. *Optica Acta*, 6:1–26, 1959.
- [149] A. Stockman and L.T. Sharpe. Spectral Sensitivities of the Middle- and Long-Wavelength Sensitive Cones Derived from Measurements in Observers of Known Genotype. *Vision Research*, 40(13):1711–1737, 2000.
- [150] A. Stockman, L.T. Sharpe, and C.C. Fach. The Spectral Sensitivity of the Human Short-Wavelength Cones. *Vision Research*, 39(17):2901–2927, 1999.
- [151] K. Stoffel and A. Belkoniene. Parallel  $k/h$ -Means Clustering for Large Data Sets. In *Proceedings of the 5th European Conference on Parallel Processing*, pages 1451–1454, Toulouse, France, August 1999.
- [152] M.A. Stricker and A. Dimai. Color Indexing with Weak Spatial Constraints. In *Storage and Retrieval for Image and Video Databases IV*, volume 2670, pages 29–40, La Jolla, CA, USA, February 1996. SPIE.

- [153] M.A. Stricker and M. Orengo. Similarity of Color Images. In *Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 381–392, San Jose, CA, USA, February 1995. SPIE.
- [154] M.J. Swain and D. Ballard. Color Indexing. *Int'l Journal of Computer Vision*, 7(1):11–32, 1991.
- [155] H. Tamura, S. Mori, and T. Yamawaki. Texture Features Corresponding to Visual Perception. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC–8(6):460–473, 1978.
- [156] H. Tamura and N. Yokoya. Image Database Systems: A Survey. *Pattern Recognition*, 17(1):29–43, 1984.
- [157] A. Tversky. Features of Similarity. *Psychological Review*, 84(4):327–352, 1977.
- [158] J.K. Uhlmann. Satisfying General Proximity/Similarity Queries with Metric Trees. *Information Processing Letters*, 40(4):175–179, 1991.
- [159] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, and H.J. Zhang. Image Classification for Content-Based Indexing. *IEEE Transactions on Image Processing*, 10(1):117–130, 2001.
- [160] J.J. Valdés, L.C. Molina, and S. Espinosa. Behavior Analysis of Fractal Features for Texture Description in Digital Images: An Experimental Study. In *Proceedings of the IEEE Int'l Conference on Pattern Recognition*, pages 3917–3920, Barcelona, Spain, September 2000.
- [161] R.C. Veltkamp and M. Hagedoorn. State of the Art in Shape Matching. In *Principles of Visual Information Retrieval*, chapter 4, pages 87–119. Springer-Verlag, 2001.
- [162] R.C. Veltkamp and M. Tanase. Content-Based Image Retrieval Systems: A Survey. Technical Report UU–CS–2000–34, Utrecht University, Department of Computer Science, October 2000.
- [163] C.C. Venters and M. Cooper. A Review of Content-Based Image Retrieval Systems. Technical Report JTAP–054, JISC Technology Applications Programme, July 2000.
- [164] V.M. Viitaniemi and J.T. Laaksonen. Keyword-Detection Approach to Automatic Image Annotation. In *Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, London, UK, November 2005.

- [165] P.A. Viola and M.J. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, Kauai Island, HI, USA, December 2001.
- [166] J. Vleugels and R.C. Veltkamp. Efficient Image Retrieval through Vantage Objects. In *Proceedings of the 3rd Int'l Conference on Visual Informations Systems*, pages 575–584, Amsterdam, The Netherlands, June 1999.
- [167] S. Volmer. Tracing Images in Large Databases by Comparison of Wavelet Fingerprints. In *Proceedings of the 2nd Int'l Conference on Visual Information Systems*, pages 163–172, La Jolla, CA, USA, December 1997.
- [168] S. Volmer. Buoy Indexing of Metric Feature Spaces for Fast Approximate Image Queries. In *Proceedings of the 6th Eurographics Workshop on Multimedia*, pages 131–140, Manchester, UK, September 2001.
- [169] S. Volmer. Fast Approximate Nearest-Neighbor Queries in Metric Feature Spaces by Buoy Indexing. In *Proceedings of the 5th Int'l Conference on Visual Informations Systems*, pages 36–49, Hsinchu, Taiwan, March 2002.
- [170] S. Volmer and N. Diakopoulos. Contextually Meaningful Representation of Color for the Description of Visual Content. In *Proceedings of the 9th Int'l Conference on Distributed Multimedia Systems*, pages 403–409, Miami, FL, USA, September 2003.
- [171] J. Wang, W.J. Yang, and R. Acharya. Color Clustering Techniques for Color-Content-Based Image Retrieval from Image Databases. In *Proceedings of the 4th IEEE Int'l Conference on Multimedia Computing and Systems*, pages 442–449, Ottawa, Canada, June 1997.
- [172] J.Z. Wang, G. Wiederhold, O. Firschein, and S.X. Wei. Wavelet-Based Image Indexing Techniques with Partial Sketch Retrieval Capability. In *Proceedings of the 4th Int'l Forum on Research and Technology Advances in Digital Libraries*, pages 13–24, Washington, DC, USA, May 1997.
- [173] D.A. White and R.C. Jain. Similarity Indexing: Algorithms and Performance. In *Storage and Retrieval for Image and Video Databases IV*, volume 2670, pages 65–72, La Jolla, CA, USA, February 1996. SPIE.
- [174] J.K. Wu, C.P. Lam, B.M. Mehtre, Y.J. Gao, and A. Narasimhalu. Content-Based Retrieval for Trademark Registration. *Multimedia Tools and Applications*, 3(3):245–267, 1996.

- [175] L. Yang and F. Albrechtsen. Fast Computation of Invariant Geometric Moments: A New Method Giving Correct Results. In *Proceedings of the 12th IAPR Int'l Conference on Pattern Recognition*, pages 201–204, Jerusalem, Israel, October 1994.
- [176] M.H. Yang, D. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
- [177] P.N. Yianilos. Data Structures and Algorithms for Nearest Neighbor Search in General Metric Spaces. In *Proceedings of the 4th Annual ACM SIAM Symposium on Discrete Algorithms*, pages 311–321, Austin, TX, USA, January 1993.
- [178] J.M. Zachary and S.S. Iyengar. Content Based Image Retrieval Systems. In *Proceedings of 2nd IEEE Symposium on Application-Specific Systems and Software Engineering and Technology*, pages 136–143, Richardson, TX, USA, March 1999.
- [179] C.T. Zahn and R.Z. Roskies. Fourier Descriptors for Plane Closed Curves. *IEEE Transactions on Computers*, C-21(3):269–281, 1972.
- [180] P. Zezula, P. Savino, G. Amato, and F. Rabitti. Approximate Similarity Retrieval with  $M$ -Trees. *The VLDB Journal*, 7(4):275–293, 1998.
- [181] X. Zhou, L. Chen, J. Ye, Q. Zhang, and B. Shi. Automatic Image Semantic Annotation Based on Image-Keyword Document Model. In *Proceedings of the Int'l Conference on Image and Video Retrieval*, pages 164–173, Singapore, July 2005.

# Veröffentlichungen

- [1] S. Walter, S. Volmer, and P. Efraimidis. Parallel Rendering of 3D Ultrasound at Interactive Rates. In *Proceedings of the 1st Eurographics Workshop for Parallel Graphics and Visualisation*, pages 189–202, Bristol, UK, September 1996.
- [2] S. Walter, M. Grimm, G. Sakas, and S. Volmer. Akquisition, Rekonstruktion und Visualisierung von 3D Ultraschall. In *Bildverarbeitung für die Medizin: Algorithmen, Systeme, Anwendungen*, Aachen, Germany, November 1996.
- [3] S. Volmer. Tracing Images in Large Databases by Comparison of Wavelet Fingerprints. In *Proceedings of the 2nd Int'l Conference on Visual Information Systems*, pages 163–172, La Jolla, CA, USA, December 1997.
- [4] V. Kovalev and S. Volmer. Color Co-occurrence Descriptors for Querying-by-Example. In *Proceedings of the 5th Int'l Conference on Multimedia Modeling*, pages 32–38, Lausanne, Switzerland, October 1998.
- [5] S. Volmer. Buoy Indexing of Metric Feature Spaces for Fast Approximate Image Queries. In *Proceedings of the 6th Eurographics Workshop on Multimedia*, pages 131–140, Manchester, UK, September 2001.
- [6] F. Morsdorf and S. Volmer. LIVE@WEB.COM – Using CBIR Technology in Interactive Web-TV. In *Proceedings of the 6th Eurographics Workshop on Multimedia*, pages 189–197, Manchester, UK, September 2001.
- [7] S. Volmer. Fast Approximate Nearest-Neighbor Queries in Metric Feature Spaces by Buoy Indexing. In *Proceedings of the 5th Int'l Conference on Visual Informations Systems*, pages 36–49, Hsinchu, Taiwan, March 2002.
- [8] N. Diakopoulos and S. Volmer. Temporally Tolerant Video Matching. In *Proceedings of the ACM SIGIR Workshop on Multimedia Information Retrieval*, Toronto, Canada, August 2003.
- [9] E. Leopold, J. Kindermann, G. Paaß, S. Volmer, R. Cavet, M. Larson, S. Eickeler, and T. Kastner. Integrated Classification of Audio, Video and Speech Using

- Partitions of Low-Level Features. In *Proceedings of the ECML Workshop on Multimedia Discovery and Mining*, Dubrovnic, Croatia, September 2003.
- [10] S. Volmer and N. Diakopoulos. Contextually Meaningful Representation of Color for the Description of Visual Content. In *Proceedings of the 9th Int'l Conference on Distributed Multimedia Systems*, pages 403–409, Miami, FL, USA, September 2003.
- [11] R. Cavet, S. Volmer, E. Leopold, J. Kindermann, and G. Paaß. Revealing the Connoted Visual Code: A New Approach to Video Classification. *Computers & Graphics*, 28(3):361–369, 2004.

# Lebenslauf

## Persönliche Daten

Name	Stephan Volmer
Anschrift	Zwingenberger Straße 32 64342 Seeheim-Jugenheim
Geburtsdatum	02.07.1968
Geburtsort	Lich, Hessen
Staatsangehörigkeit	deutsch
Familienstand	verheiratet

## Schulbildung

08/1974 – 07/1978	Grundschule, Goetheschule Gießen
08/1978 – 06/1987	Gymnasium, Herderschule Gießen, Abschluss Abitur

## Studium

10/1988 – 02/1995	Studium Elektrotechnik, Vertiefungsrichtung Datentechnik, Technische Universität Darmstadt, Abschluss Diplom
08/1992 – 07/1994	Graduate Program Electrical Engineering, State University of New York at Buffalo, Degree Master of Science

## Berufliche Tätigkeiten

03/1995 – 12/2003	Wissenschaftlicher Mitarbeiter, Fraunhofer Institut für Graphische Datenverarbeitung, Darmstadt
seit 02/2004	Softwareingenieur, Zühlke Engineering GmbH, Eschborn

